BIOGEOCHEMISTRY (LIBRETEXTS)



Josh Halpern
LibreTexts



BioGeoChemistry

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

This text was compiled on 05/19/2026

TABLE OF CONTENTS

Licensing

1: Introduction

- 1.1: What is biogeochemistry?
- 1.2: Chemistry and Energetics of the Life Process
- 1.3: Systems, Cycles, Reservoirs, and Fluxes
- 1.4: Gaia - Bioregulation of the Environment

2: Origins

- 2.1: Introduction
- 2.2: Origin of the Elements
- 2.3: Formation and evolution of the Earth
- 2.4: The Earth's crust
- 2.5: Origin of the atmosphere and the oceans
  - 2.5.1: Origin of the Oceans
  - 2.5.2: Origin and Evolution of the Atmosphere
- 2.6: Earthlike Planets - Venus and Mars
  - 2.6.1: The Nearest Planets - An Overview
  - 2.6.2: The Geology of Venus
  - 2.6.3: The Massive Atmosphere of Venus
  - 2.6.4: The Geology of Mars
  - 2.6.5: Water and Life on Mars
  - 2.6.6: Divergent Planetary Evolution
  - 2.6.E: Earthlike Planets - Venus and Mars (Exercises)
- 2.7: Summary

3: The Atmosphere

- 3.1: Introduction
- 3.2: Structure and Composition of the Atmosphere
- 3.3: Origin and Evolution of the Atmosphere
- 3.4: General Circulation
  - 3.4.1: A Simplified Description of the Global Circulation
  - 3.4.2: Radiative Differential Heating
  - 3.4.3: Pressure Profiles
  - 3.4.4: Geostrophic Wind and Geostrophic Adjustment
  - 3.4.5: Thermal Wind Effect
  - 3.4.6: Horizontal Circulation
  - 3.4.7: Extratropical Ridges and Troughs (Rossby Waves)
  - 3.4.8: Three-band Global Circulation
  - 3.4.9: Ekman Spiral of Ocean Currents
  - 3.4.10: Explaining the General Circulation
  - 3.4.11: Jet Streams
  - 3.4.12: Types of Vorticity
  - 3.4.13: Key Terms

- 3.4.14: Review
- 3.4.15: Homework Exercises
- 3.5: Atmospheric Composition
  - 3.5.1: Atmospheric Composition
  - 3.5.2: Changes in Atmospheric Composition
  - 3.5.3: Other Trace Gases
  - 3.5.4: Stratospheric Ozone Formation
  - 3.5.5: The Story of the Atmosphere's PAC-MAN
  - 3.5.6: Where do Cloud Condensation Nuclei (CCN) come from?
  - 3.5.7: Summary and Final Tasks
- 3.6: Biogeochemical reactions in the troposphere
- 3.7: Atmospheric deposition
  - 3.7.1: Acid Deposition
- 3.8: Biogeochemical reactions in the stratosphere
- 3.9: Models
- 3.10: Impacts
- 3.11: Summary

4: The Lithosphere

- 4.1: The Earth - The Ground Beneath Your Feet
- 4.2: Minerals
- 4.3: Rocks
- 4.4: Bedrock
- 4.5: Weathering
- 4.6: The Mineral Particles of the Earth's Surface Materials
- 4.7: Regolith
- 4.8: Sediment
- 4.9: Soils
- 4.10: Soil Chemistry
- 4.11: Soil Biodiversity and Ecology
- 4.12: Soil Nutrient Cycling
- 4.13: Weathering, Erosion, and Mass Movement
  - 4.13.1: Weathering
  - 4.13.2: Mass Movement
  - 4.13.3: Water Erosion
  - 4.13.4: Review and Additional Resources
- 4.14: Summary

5: The Carbon Cycle of Terrestrial Ecosystems

- 5.1: Introduction
- 5.2: Photosynthesis
- 5.3: Respiration
- 5.4: Net Primary Production
  - 5.4.1: Marine Primary Production
    - 5.4.1.1: Primary production
    - 5.4.1.2: The Producers
    - 5.4.1.3: Factors Influencing Production

        - 5.4.1.4: Patterns of Primary Production
    - 5.4.2: Surface Primary Production
- 5.5: Net ecosystem production and Eddy-covariance studies
- 5.6: Remote sensing of primary production and biomass
- 5.7: Global estimates of net primary production and biomass
- 5.8: The fate of net primary production
    - 5.8.1: Introduction to and Components of Food Webs
    - 5.8.2: Quantifying Food Webs
    - 5.8.3: Trophic Cascades
    - 5.8.4: Sources and Attributions
- 5.9: Net primary production and global change
- 5.10: Detritus
- 5.11: Soil organic matter and global change
- 5.12: Summary

6: Biogeochemical Cycling on Land

- 6.1: Biogeochemical cycling in land plants
- 6.2: Nutrient allocations and cycling in land vegetation
- 6.3: Biogeochemical cycling in the soil
    - 6.3.1: Soils and the carbon cycle
    - 6.3.2: Soils and the Water Cycle
- 6.4: Calculating landscape mass balance
- 6.5: Ecosystem models and remote sensing
- 6.6: Ecosystem Monitoring
- 6.7: Human impacts on terrestrial biogeochemistry
    - 6.7.1: Introduction
    - 6.7.2: Human-Induced Changes
    - 6.7.3: Sinks and Cycles
    - 6.7.4: Impacts and Options
    - 6.7.5: References
- 6.8: Summary

7: Wetland Ecosystems

- 7.1: Introduction
- 7.2: Types of wetlands
- 7.3: Productivity in wetland ecosystems
- 7.4: Organic matter storage in wetlands
- 7.5: Microbial metabolism in saturated sediments
- 7.6: Anaerobic and aerobic metabolic pathways
- 7.7: The redox ladder
- 7.8: Wetlands and water quality
- 7.9: Wetlands and global change
    - 7.9.1: Introduction
    - 7.9.2: Carbon Accumulation and GHG Emissions from Freshwater Wetlands (Including Permafrost) In a Changing Climate
    - 7.9.3: Ecological Consequences for Freshwater Wetlands in a Changing Climate
    - 7.9.4: Salt Marsh and Mangrove Response to a Changing Climate and Associated Sea Level Rise

- 7.10: Summary

8: Inland waters

- 8.1: Introduction
- 8.2: The structure of aquatic ecosystems
- 8.3: The special properties of water
- 8.4: Inland waters and their watersheds
  - 8.4.1: The Physical Template
  - 8.4.2: Watershed Structure
  - 8.4.3: Watershed Functions
  - 8.4.4: Summary
- 8.5: Lakes
  - 8.5.1: Freshwater Lakes
  - 8.5.2: Saline Lakes
- 8.6: Streams and Rivers
- 8.7: Estuaries- Where the turf meets the surf
- 8.8: Waves and Beaches
- 8.9: Human impacts on inland waters
  - 8.9.1: Fresh Water Supply and the Water Cycle
  - 8.9.2: Water Usage
  - 8.9.3: Water Scarcity and Solutions
  - 8.9.4: Data Dive- Aqueducts of Rome
  - 8.9.5: Review
- 8.10: Summary

9: The Oceans

- 9.1: Introduction
- 9.2: Properties of Seawater
  - 9.2.1: Properties of Seawater
  - 9.2.2: Specific Heat and Latent Heat Capacity of Water
  - 9.2.3: Energy Associated With Evaporation and Condensation of Water In the Air
  - 9.2.4: Salinity
  - 9.2.5: Formation of Sea Ice
  - 9.2.6: Relationship of Salinity, Density and Temperature
  - 9.2.7: Salinity and Latitude
  - 9.2.8: "Cline Curves" and Mixing (Surface) Zone
  - 9.2.9: Gases Dissolved In Seawater
  - 9.2.10: Local Conditions in the San Diego Region
  - 9.2.11: Quiz Questions Properties of Seawater
- 9.3: Biogeochemistry of the surface ocean
  - 9.3.1: Introduction
  - 9.3.2: Biological Pump
  - 9.3.3: Export from the Upper Ocean to the Depths and Sediments
  - 9.3.4: Summary
- 9.4: 12.4 Biogeochemistry of the Deep Ocean
- 9.5: Biogeochemistry in ocean sediments
  - 9.5.1: Introduction

- 9.5.2: Classifying Sediments
- 9.5.3: Lithogenous Sediments
- 9.5.4: Biogenous Sediments
- 9.5.5: Hydrogenous Sediments
- 9.5.6: Cosmogenous Sediments
- 9.5.7: Sediment Distribution
- 9.6: Marine element cycles
- 9.7: The Oceans and Global Climate Change
- 9.8: Summary

10: The Global Water Cycle

- 10.1: Introduction
- 10.2: Models of the hydrologic cycle
- 10.3: The history of the water cycle
- 10.4: The water cycle and climate change
- 10.5: Summary

11: The Global Carbon and Oxygen Cycles

- 11.1: Introduction
- 11.2: What might stimulate net primary production?
- 11.3: Temporal perspectives of the carbon cycle
- 11.4: Atmospheric methane
- 11.5: Carbon monoxide
- 11.6: Synthesis- Linking the carbon and oxygen cycles
- 11.7: Summary

12: The Global Cycles of Nitrogen, Phosphorus and Potassium

- 12.1: Introduction
- 12.2: The global nitrogen cycle
  - 12.2.1: Chemistry of the nitrogen cycle
- 12.3: Temporal variations in the global nitrogen cycle
- 12.4: Nitrous Oxide (N2O)
  - 12.4.1: Terrestrial
  - 12.4.2: Marine
  - 12.4.3: Human Sources
- 12.5: The global phosphorus cycle
- 12.6: Potassium- An overlooked limiting element?
- 12.7: Linking global biogeochemical cycles
- 12.8: Summary

13: The Global Cycles of Sulfur and Mercury

- 13.1: Introduction
- 13.2: The global sulfur cycle
- 13.3: Interactions of the sulfur and oxygen cycles
- 13.4: Temporal perspectives on the global sulfur cycle
- 13.5: The global mercury cycle
- 13.6: Summary

Index

Glossary

Detailed Licensing

Detailed Licensing

Licensing

*A detailed breakdown of this resource's licensing can be found in **Back Matter/Detailed Licensing**.*

CHAPTER OVERVIEW

1: Introduction

1.1: What is biogeochemistry?

1.2: Chemistry and Energetics of the Life Process

1.3: Systems, Cycles, Reservoirs, and Fluxes

1.4: Gaia - Bioregulation of the Environment

1: Introduction is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

1.1: What is biogeochemistry?

Biogeochemistry is the scientific discipline that explores the interactions between living organisms and the physical and chemical aspects of the environment. It is a field that combines principles from biology, geology, chemistry, and environmental science to study the processes that govern the cycling of elements and compounds in ecosystems. Biogeochemists study how living organisms, including plants, animals, and microorganisms, interact with the solid Earth (geosphere).

In biogeochemical cycles, elements and compounds move through the atmosphere, hydrosphere (water bodies), lithosphere (Earth's crust), and biosphere (living organisms) in a series of complex processes. These processes include photosynthesis, respiration, decomposition, weathering, erosion, and sedimentation. Human activities, such as industrial processes and agriculture, can also significantly impact biogeochemical cycles.

Rather than flowing through an ecosystem, the matter that makes up living organisms is conserved and recycled. The six most common elements associated with organic molecules—carbon, nitrogen, hydrogen, oxygen, phosphorus, and sulfur—take a variety of chemical forms and may exist for long periods in the atmosphere, on land, in water, or beneath Earth’s surface or may rapidly move between reservoirs. Geologic processes, such as weathering, erosion, water drainage, and the subduction of the continental plates, all play a role in the cycling of elements on Earth. Because geology and chemistry have major roles in the study of this process, the recycling of inorganic matter between living organisms and their nonliving environment is called a biogeochemical cycle.

Human impact is an important part of biogeochemistry. Our activities can change the proportion of nutrients that are in reservoirs and in circulation. Deforestation disrupts nutrient cycling by removing trees that contribute organic matter to the soil. Pollution introduces chemicals that can alter soil and water chemistry, affecting plant and animal life.Coal is a reservoir of carbon, but the human use of fossil fuels has released carbon into the atmosphere, increasing the amount of carbon in circulation. Likewise, phosphorous and nitrogen are extracted from geological reservoirs and used and excesses of these elements have caused the overgrowth of plant matter and the disruption of many ecosystems. Excess fertilizer use can lead to nutrient runoff, causing algal blooms and oxygen-deprived "dead zones" in water bodies.

Water, which contains hydrogen and oxygen, is essential to all living processes. The hydrosphere is the area of Earth where water movement and storage occurs: as liquid water on the surface (rivers, lakes, oceans) and beneath the surface (groundwater) or ice, (polar ice caps and glaciers), and as water vapor in the atmosphere. Carbon is found in all organic macromolecules and is an important constituent of fossil fuels. Nitrogen is a major component of our nucleic acids and proteins and is critical to human agriculture. Phosphorus, a major component of nucleic acids, is one of the main ingredients (along with nitrogen) in artificial fertilizers used in agriculture, which has environmental impacts on our surface water. Sulfur, critical to the three-dimensional folding of proteins (as in disulfide binding), is released into the atmosphere by the burning of fossil fuels.

The cycling of these elements is interconnected and occurs on all scales, from microscopic to global and over minutes and millenia. The movement of water is critical for the leaching of nitrogen and phosphate into rivers, lakes, and oceans. The ocean is also a major reservoir for carbon. Thus, mineral nutrients are cycled, either rapidly or slowly, through the entire biosphere between the biotic and abiotic world and from one living organism to another.

Energy flows directionally through ecosystems, entering as sunlight (or inorganic molecules for chemoautotrophs) and leaving as heat during energy transformation between trophic levels. Rather than flowing through an ecosystem, the matter that makes up organisms is conserved and recycled. The law of conservation of mass states that matter is neither created nor destroyed. For example, after a chemical reaction, the mass of the products (ending molecules) will be the same as the mass of the reactants (starting molecules). The same is true in an ecosystem. Matter moves through different media, and atoms may react to form new molecules, but the amount of matter remains constant. Although the Earth receives energy from the Sun, the chemical composition of the planet is more or less fixed. Matter is occasionally added by meteorites, but supplies of essential elements generally do not change.

Understanding biogeochemistry is crucial for addressing environmental challenges, managing natural resources, and mitigating the impacts of human activities on ecosystems and the planet's overall health. Scientists in this field study how nutrients and pollutants move through ecosystems, impacting both natural habitats and human societies. Biogeochemistry is critical for understanding climate change feedback loops. Thawing permafrost releases stored carbon, amplifying global warming. Changes in ocean

chemistry due to increased carbon dioxide levels (ocean acidification) affect marine ecosystems, disrupting biogeochemical processes. In summary, biogeochemistry is a multidisciplinary field that integrates biology, geology, chemistry, and environmental science to unravel the complex web of interactions between living organisms and their environment, with a focus on the cycling of elements and compounds crucial for life.

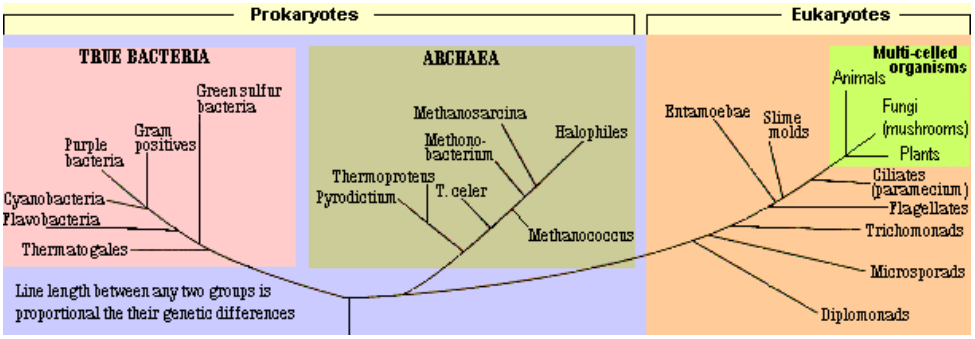
References

1. **Books:**
   - Schlesinger, W. H. (2013). **Biogeochemistry: An Analysis of Global Change.** Academic Press.
   - Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., & Stemple, J. (2000). **The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System.** Science, 290(5490), 291-296.
2. **Scientific Journals:**
   - **Biogeochemistry:** A scientific journal dedicated to studies of chemical, physical, geological, and biological processes influencing the composition and behavior of the Earth's systems. (Published by Springer)
   - **Global Biogeochemical Cycles:** A journal focusing on biogeochemical processes in the Earth's systems and their responses to global change. (Published by Wiley)
   - **Journal of Geophysical Research - Biogeosciences:** A journal covering research articles on the biogeosciences, including the interface between biological, chemical, and physical processes in terrestrial and aquatic environments. (Published by Wiley)
3. **Research Articles:**
   - Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). **The microbial engines that drive Earth's biogeochemical cycles.** Science, 320(5879), 1034-1039.
   - Canfield, D. E., Glazer, A. N., & Falkowski, P. G. (2010). **The evolution and future of Earth's nitrogen cycle.** Science, 330(6001), 192-196.
   - Schimel, J. (2001). **Biogeochemical models: implicit vs. explicit microbiology.** Trends in Ecology & Evolution, 16(8), 365-369.
4. **Review Articles:**
   - Ward, B. B., Eveillard, D., Kirshtein, J. D., & Nelson, J. D. (2007). **Importance of Extant and Extinct Forms of Dissimilatory Nitrite Reductase in Biogeochemical Cycles.** Limnology and Oceanography, 52(6), 2276-2283.
   - Amundson, R., Austin, A. T., Schuur, E. A. G., Yoo, K., Matzek, V., Kendall, C., ... & Biogeochemical Research Community. (2003). **Global patterns of the isotopic composition of soil and plant nitrogen.** Global Biogeochemical Cycles, 17(1), 1031.

1.1: What is biogeochemistry? is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

1.2: Chemistry and Energetics of the Life Process

The biosphere comprises the various regions near the earth's surface that contain and are dynamically affected by the metabolic activity of the approximately 1.5 million animal species and 0.5 million plant species that are presently known and are still being discovered at a rate of about 10,000 per year. The biosphere is the youngest of the dynamical systems of the earth, having had its genesis about 2 billion years ago. It is also the one that has most profoundly affected the other major environmental systems, particularly the atmosphere and the hydrosphere.



"The tree of life"

About a third of the chemical elements cycle through living organisms, which are responsible for massive deposits of silicon, iron, manganese, sulfur, and carbon. Large quantities of methane and nitrous oxide are introduced into the atmosphere by bacterial action, and plants alone inject about 400,000 tons of volatile substances (including some metals) into the atmosphere annually.

> It has been suggested that biological activity might be responsible for the deficiency of hydrogen on Earth, compared to its very high relative abundance in the solar system as a whole. Bacteria capable of reducing hydrogen compounds into H_2 transform this element into a form in which it can escape from the earth; such bacteria may have been especially active in the reducing atmosphere of the early planet.
>
> A second mechamism might be the microbial production of methane, which presently injects about 10^9 T of CH_4 into the atmosphere each year. Some of this reaches the stratosphere, where it is oxidized to CO_2 and H_2O. The water vapor is photolyzed to H_2, which escapes into space. This may be the major mechanism by which water vapor (and thus hydrogen) is transported to the upper atmosphere; the low temperature of the upper atmosphere causes most of the water originating at lower levels to condense before it can migrate to the top of the atmosphere.

The increase in the abundance of atmospheric oxygen from its initial value of essentially zero has without question been the most important single effect of life on earth, and the time scale of this increase parallels the development of life forms from their most primitive stages up to the appearance of the first land animals about 0.5 billion years ago.

Bioenergetics

All life processes involve the uptake and storage of energy, and its subsequent orderly release in small steps during the metabolic process. This energy is taken up in the combination of ADP with inorganic phosphate to form ATP, in which form the energy is stored and eventually delivered to sites where it can provide the free energy needed for driving non-spontaneous reactions such as protein and carbohydrate synthesis.

$ADP + PO_4^{3-} \rightarrow ATP\ \Delta G° = +30\ kJ$

The three main metabolic processes are glycolysis, respiration, and photosynthesis. The first two of these extract free energy from glucose by breaking it up into smaller, more thermodynamically stable fragments. Photosynthesis reverses this process by capturing the energy of sunlight into ATP which then drives the buildup of glucose from CO_2.

Glycolysis and fermentation.

As its name implies, this most primitive (and least efficient) of all metabolic processes is based on the breakdown of a sugar into fragments having a smaller total free energy. Thus the 6-carbon sugar glucose can be broken down into two 3-carbon lactic acid units, or into three 2-carbon ethanol units.

$C_6H_{12}O_6 \rightarrow 2\ CH_3CHOHCOOH\ \Delta G° = –197\ kJ$

In this process, two molecules of ATP are produced, thus capturing 61 kJ of free energy. Since the standard free energy of glucose (with respect to its elements) is –2870 kJ, this represents an overall efficiency of about 2 percent.

The net reaction of glycolysis is essentially a rearrangement of the atoms initially present in the energy source. This is a form of fermentation, which is defined as the enzymatic breakdown of organic molecules in which other organic compounds serve as electron acceptors. Since there is no need for an external oxidizing or reducing agent, there is no change in the oxidation state of the environment.

Respiration

When the enzymatic degradation of organic molecules is accompanied by transfer of electrons to an external (and usually better) electron acceptor, the process is known as respiration. The overall reaction of respiration is the oxidation of glucose to carbon dioxide:

$C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ CO_2 + 6\ H_2O\ \Delta G° = –2380\ kJ$

In this process, 36 molecules of ADP are converted into ATP, thus capturing 1100 kJ of free energy: an efficiency of 38 percent.

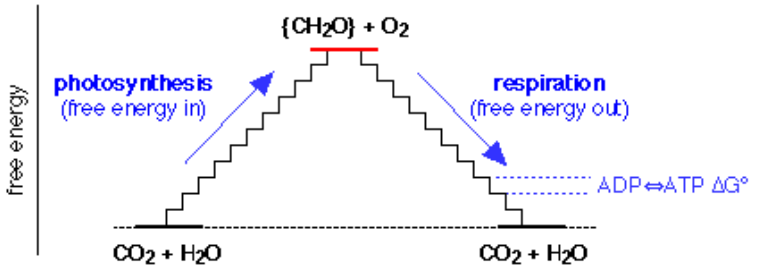
The oxidizing agent (electron sink) need not be oxygen; some bacteria reduce nitrates to NO or to N_2, and sulfates or sulfur to H_2S. These metabolic products can have far-reaching localized environmental effects, particularly if hydrogen ions are involved. Falling down the respiratory ladder presents a succinct picture of oxidation-reduction and the role of non-O_2 electron sinks in biological energy capture.

Photosynthesis

The energy of sunlight is trapped in the form of an intermediate which is able to deliver electrons to successively lower free-energy levels through the mediation of various molecules (mainly cytochromes) comprising an electron-transport chain. The free energy thus gained is utilized in part to reduce CO_2 to glucose, which is then available to supply metabolic energy by glycolysis or respiration. In green plants and eukaryotic algae, the source of hydrogen is water, the net reaction being

$6\ CO_2 + 6\ H_2O \rightarrow C_6H_{12}O_6 + 6\ O_2\ \Delta G° = +2830\ kJ$

For every CO_2 molecule fixed in this way, 469 kJ of free energy must be supplied. Red light of 680 nm wavelength has an energy of 168 kJ/mol; this implies that about three photons must be absorbed for every carbon atom taken up, but experiment indicates that about ten seem to be required.



The photosynthetic - respiration cycle.

Photosynthesis utilizes the energy of red light to add hydrogen (from H_2O) and electrons (from the O in H_2O) to CO_2, reducing it to **carbohydrate**. **Respiration** is the reverse of this process; electrons are removed from the carbohydrate (food is "burned") in small steps, each one releasing a small amount of energy. Some of this energy is liberated as heat, but part of it is used to add a phosphate group to ADP, converting it into ATP. **ATP** (adenosine triphosphate) is to an organism's energy needs as money is to our

material needs; it circulates to wherever it is required in order to bring about energy-requiring reactions or to make muscle cells contract. Each increment of energy given up by ATP converts it back to ADP and phosphate, ready to repeat the cycle.

Green plants are able to operate in both modes during the daylight hours, reverting to respiration-only at night. **Animals** carry out only the right side of the cycle and thus require as source of carbohydrate ("**food**") either directly (by eating plants), or indirectly by eating other animals that eat plants.

There are many kinds of photosynthetic bacteria, but with one exception (cyanobacteria) they are incapable of using water as a source of hydrogen for reducing carbon dioxide. Instead, they consume hydrogen sulfide or other reduced sulfur compounds, organic molecules, or elemental hydrogen itself, excreting the reducing agent in an oxidized state. Green plants, cyanobacteria, green filamentous bacteria and the purple nonsulfur bacteria utilize glucose by respiration during periods of darkness, while the green sulfur bacteria and the purple sulfur bacteria are strictly anaerobic.

This page titled 1.2: Chemistry and Energetics of the Life Process is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

1.3: Systems, Cycles, Reservoirs, and Fluxes

Thinking about the Earth as a system has become, in recent years, much in vogue. The basic idea of a ***system*** is that it is *something that consists of a number of distinctive and diverse parts that function together by a variety of interactions and exchanges of energy and matter*. Your motor vehicle is a good example of a fairly small and simple system: it is a machine with a number of moving or nonmoving parts, which consumes motor fuel and emits various exhaust substances as it travels from place to place. The Earth is, of course, a far larger and also far more complicated system, whose component parts operate on time scales from seconds to many millions of years. This is the basis for what is nowadays called ***Earth system science***. Here I'm only touching upon some important aspects of Earth systems science that are relevant to this course: the hydrologic cycle and the carbon cycle. A whole course could be devoted to those and certain other cycles that are important in the context of the Earth's surface environment.

The concept of cycles plays a fundamental role in systems thinking— although it's not easy to frame a suitable definition of such cycles. A ***cycle*** might best be described as a characteristic succession of events and processes , involving certain kinds of Earth materials, by which the materials reside in certain kinds of places and move among such places in certain ways. Cycles operate through an indefinitely long span of time.

Much of the material that has been covered so far in this course involve, in one way or another, water and sediment. Earlier chapters have dealt in much detail with the Earth's water and sediment: where they are, and how the move. Now it's time for us to be more integrative in our thinking about the Earth's water and sediment. To that end, the next two sections deal with what are called the *hydrologic cycle* and the *sediment cycle*.

Geochemists, in particular, are fond of thinking about certain substances from the standpoint of cycles. Such cycles are called ***geochemical cycles***. The most important, in the context of the Earth's surface environment, is the *carbon cycle*. The carbon cycle is especially important because carbon is the fundamental basis for life. It is also of great importance for the Earth's climate, because carbon dioxide (CO2) is the second important of the so-called greenhouse gases, which play a key role in climate. (Water vapor is the most important, but we humans have virtually no control over the content of water vapor in the atmosphere, whereas we have great influence on the content of carbon dioxide in the atmosphere, owing to the burning of fossil fuels.) The concluding section of this chapter is devoted to some aspects of the Earth's carbon cycle that are most relevant to the Earth's surface environment. Geochemical cycling of certain other elements, most notably nitrogen and phosphorus, have received study because of their role as nutrients.

What pops into your mind when you hear the word *reservoir*? Probably a body of water, small or large, that is impounded behind a dam. In Earth systems science, the term ***reservoir*** is used for *a distinctive kind of place where a certain kind of material is stored, or resides, for some period of time*. That definition probably doesn't enlighten you much, so here are some examples of near-surface reservoirs you have encountered earlier in the course: the atmosphere; glaciers; the soil layer; the aggregate of bodies of fresh water on the continents (rivers and lakes).

Material moves into and out of reservoirs. *The rate at which a given material moves between reservoirs* is called a ***flux***. If the flux of material into and out of a given reservoir is the same for some period of time, that reservoir is said to be in a ***steady state***. Commonly, however, the flux in and the flux out are not equal.

Another concept that is useful to know about when thinking about systems is ***residence time***. In an earlier chapter you learned about the residence time of water in lakes. The same concept can be applied to any substance that resides in some reservoir for some period of time.

This page titled 1.3: Systems, Cycles, Reservoirs, and Fluxes is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) via source content that was edited to the style and standards of the LibreTexts platform.

- **12.1: Introduction - Systems, Cycles, Reservoirs, and Fluxes** by John Southard is licensed CC BY-NC-SA 4.0. Original source: https://ocw.mit.edu/courses/12-090-the-environment-of-the-earths-surface-spring-2007/.

1.4: Gaia - Bioregulation of the Environment

The Gaia Hypothesis

The physical conditions under which life as we know it can exist encompass a relatively narrow range of temperature, pH, osmotic pressure, and ultraviolet radiation intensity. It seems remarkable enough that life was able to get started at all; it is even more remarkable that it has continued to thrive in the face of all the perils that have, or could have occurred, during the past 3 billion years or so.

During the time that life has been evolving, the sun has also been going through the process of evolution characteristic of a typical star; one consequence of this is an increase in its energy output by about 30 percent during this time. If the sun's output should suddenly drop to what it was 3 billion years ago, the oceans would freeze. How is it that the earth was not in a frozen state for the first 1.5 billion years of life's existence? Alternatively, if conditions were somehow suitable 3 billion years ago, why have the oceans not long since boiled away?

A rather non-traditional answer to this kind of question is that the biosphere is far from playing a passive role in which it is continually at the mercy of environmental conditions. Instead, the earth's atmosphere, and to a lesser extent the hydrosphere, may be actively maintained and regulated by the biosphere. This view has been championed by the British geochemist J.E. Lovelock, and is known as the *Gaia hypothesis*.

> *Gaia is another name for the Greek earth-goddess Ge, from which root the sciences of geography, geometry, and geology derive their names. Lovelock's book Gaia: a new look at life on Earth (Oxford, 1979) is a short and highly readable discussion of the hypothesis.*

Evidence in support of this hypothesis is entirely circumstantial, but nevertheless points to important questions that must be answered: how have the climatic and chemical conditions on the earth remained optimal for life during all this time; how can the chemical composition of the atmosphere remain in a state that is tens of orders of magnitude from equilibrium?

Note

Although the Gaia hypothesis has received considerable publicity in the popular press, it has never been very well received by the scientific community, many of whom feel that there is no justification for proposing a special hypothesis to describe a set of connections which can be quite adequately explained by conventional geochemical processes. More recently, even Lovelock has backed away from the teleological interpretation of these relations, so that the Gaia hypothesis should now be more properly described as a set of loosely connected effects, rather than as a hypothesis. Nevertheless, these effects and the mechanisms that might act to connect them are sufficiently interesting that it seems worthwhile to provide an overview of the major observations that led to the development of the hypothesis.

Teleology is the doctrine that natural processes operate with a purpose. See *No longer willful, Gaia becomes respectable.* 1988: Science 240 393-395.

Bioregulation of the Atmosphere

The increase in the oxygen content of the atmosphere as a result of the development of the eucaryotic cell was discussed above. Why has the oxygen content leveled off at 21 percent? It is interesting to note that if the oxygen concentration in the atmosphere were only four percent higher, even damp vegetation, once ignited by lightning, would continue to burn, enveloping vast areas of the earth in a firestorm. Evidence for such a worldwide firestorm that may be related to the extinction of the dinosaurs has recently been discovered. The charcoal layers found in widely distributed sediments laid down about 65 million years ago are coincident with the iridium anomaly believed to be due to the collision of a large meteor with the earth.

- **Oxygen**: Regulation of the oxygen partial pressure is probably achieved by a balance between its production through photosynthesis and its consumption during oxidation of organic matter; the present steady state requires the burial of about 0.1% the carbon that is fixed annually, leaving one O_2 molecule in the air for each atom of carbon removed from the photosynthetic cycle. The large quantities of microbially-produced methane and N_2O also constitute important oxygen sinks; if methanogenic bacteria should suddenly cease to exist, the O_2 concentration would rise by 1% in about 12,000 years. This type

of regulation implies a negative feedback mechanism, in which an increase in atmospheric oxygen would increase the activity of organisms capable of generating metabolic products that react with it.

- **Nitrous oxide**: Nitrous oxide, in addition to serving as an oxygen sink, might also be a factor in the regulation of the intensity of the ultraviolet component of sunlight. N_2O acts as a catalytic intermediate in the decomposition of stratospheric ozone, which shields the earth from excessive ultraviolet radiation.
- **Ammonia**: Ammonia, another atmospheric gas, is produced by the biosphere in approximately the same quantities as methane, 10^9 tons per year, and at the expense of a considerable amount of metabolic energy. The function of NH_3 could well be to regulate the pH of the environment; in the absence of ammonia, the large amounts of SO_2 and HCl produced by volcanic action would reduce the pH of rain to about 3. The fact that the atmospheric concentration of ammonia is only 10^{-8} times that of N_2 should not imply that this "trace" component plays a less significant role in the overall nitrogen cycle than does than N_2. In fact, the annual rates of production of the two gases are roughly the same; the much lower steady-state concentration of NH_3 is due to its faster turnover time.
- **Nitrogen**: As stable as the triply-bonded N_2 molecule is, there is a still more stable form of nitrogen: the hydrated nitrate ion. How is this stability consistent with the predominance of nitrogen in the atmosphere? The answer is that it is not: if it were not for nitrogen-fixing bacteria (powered directly or indirectly by the free energy of ATP captured from sunlight), the nitrogen content of the atmosphere would disappear to almost zero. This would raise the oxygen fraction to disastrously high levels, and the additional NO_3^- concentration would increase the ionic strength and osmotic pressure of seawater to levels inconsistent with most forms of life.

Bioregulation of the Oceans

The input of salts into the sea from streams and rivers is about 5.4×10^8 tons per year, into a total volume of about 1.2×10^9 km^3 yr^{-1} . Upwelling of juvenile water and hydrothermal action at oceanic ridges provide additional inputs of salts. With a few bizarre exceptions such as the brine shrimp and halophilic bacteria, 6 percent is about the maximum salinity level that organisms can tolerate. The internal salinities of cells must be maintained at much lower levels (around 1%) to prevent denaturation of proteins and other macromolecules whose conformations are dependent on electrostatic forces. At higher levels than this, the electrostatic interaction between the salt ions and the cell membrane destroys the integrity of the latter so that it can no longer pump out salt ions that leak in along the osmotic gradient. At the present rate of salt input, the oceans would have reached their present levels of salinity millions of years ago, and would by now have an ionic strength far to high to support life, as is presently the case in the landlocked Dead Sea.

> *The present average salinity of seawater is 3.4 percent. The salinity of blood, and of many other intra- and intercellular fluids in animals, is about 0.8 percent. If we assume that the first organisms were approximately in osmotic equilibrium with seawater, then our body fluids might represent "fossilized" seawater as it existed at the time our predecessors moved out of the sea and onto the land.*

By what processes is salt removed from the oceans in order to maintain a steady-state salinity? This remains one of the major open questions of chemical oceanography. There are a number of answers, mostly based on strictly inorganic processes, but none is adequately supported by available evidence. For example, Na^+ and Mg^{2+} ions could adsorb to particulate debris as it drops to the seafloor, and become incorporated into sediments. The requirement for charge conservation might be met by the involvement of negatively charged silicate and hydroxyaluminum ions. Another possible mechanism might be the burial of salt beds formed by evaporation in shallow, isolated arms of the sea, such as the Persian Gulf. Extensive underground salt deposits are certainly found on most continents, but it is difficult to see how this very slow mechanism could have led to an unfluctuating salinity over shorter periods of highly variable climatic conditions.

The possibility of biological control of oceanic salinity starts with the observation that about half of the earth's biomass resides in the sea, and that a significant fraction of this consists of diatoms and other organisms that build skeletons of silica. When these organism die, they sink to the bottom of the sea and add about 300 million tons of silica to sedimentary rocks annually. It is for this reason that the upper levels of the sea are undersaturated in silica, and that the ratio of silica to salt in dead salt lakes is much higher than in the ocean.

These facts could constitute a basis for a biological control of the silica content of seawater; any link between silica and salt could lead to the control of the latter substance as well. For example, the salt ions might adsorb onto the silica skeletons, and be carried down with them; if the growth of these silica-containing organisms is itself dependent on salinity, we would have our negative feedback mechanism.

The continual buildup of biogenic sedimentary deposits on the ocean floor might possibly deform the thin oceanic crust by its weight, and cause local heating by its insulating properties. This could conceivably lead to volcanic action and the formation of new land mass, thus linking the lithosphere into Gaia.

Contributors and Attributions

- Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook

This page titled 1.4: Gaia - Bioregulation of the Environment is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

2: Origins

2.1: Introduction
2.2: Origin of the Elements
2.3: Formation and evolution of the Earth
2.4: The Earth's crust
2.5: Origin of the atmosphere and the oceans
2.5.1: Origin of the Oceans
2.5.2: Origin and Evolution of the Atmosphere
2.6: Earthlike Planets - Venus and Mars
2.6.1: The Nearest Planets - An Overview
2.6.2: The Geology of Venus
2.6.3: The Massive Atmosphere of Venus
2.6.4: The Geology of Mars
2.6.5: Water and Life on Mars
2.6.6: Divergent Planetary Evolution
2.6.E: Earthlike Planets - Venus and Mars (Exercises)
2.7: Summary

2: Origins is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.1: Introduction

According to the **Big Bang theory**, the universe blinked violently into existence 13.77 billion years ago (Figure 2.1.1). The Big Bang is often described as an explosion, but imagining it as an enormous fireball isn't accurate. The Big Bang involved a sudden expansion of matter, energy, and space from a single point. The kind of Hollywood explosion that might come to mind involves expansion of matter and energy *within* space, but during the big bang, space *itself* was created.

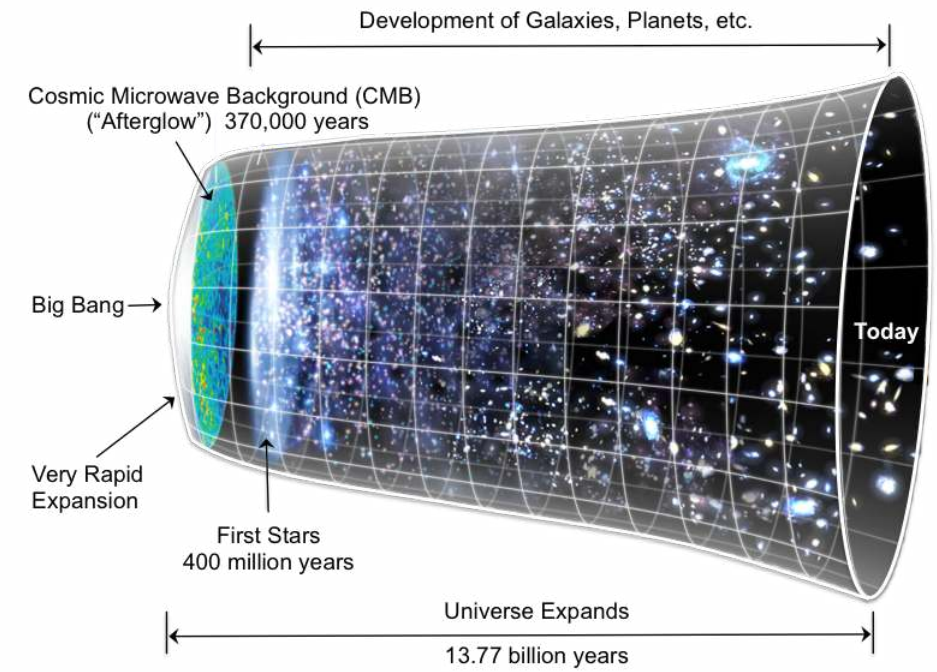


Figure 2.1.1 The Big Bang and development of the universe (Steven Earle, "Physical Geology").

At the start of the Big Bang, the universe was too hot and dense to be anything but a sizzle of particles smaller than atoms, but as it expanded, it also cooled. Eventually some of the particles collided and stuck together. Those collisions produced hydrogen and helium, the most common elements in the universe, along with a small amount of lithium. Gravity caused clouds of these early elements to coalesce into stars, and it was inside these stars that heavier elements were formed

Our solar system began to form around 5 billion years ago, roughly 8.7 billion years after the Big Bang. A **solar system** consists of a collection of objects orbiting one or more central stars. All solar systems start out the same way. They begin in a cloud of gas and dust called a **nebula**. Nebulae are some of the most beautiful objects that have been photographed in space, with vibrant colors from the gases and dust they contain, and brilliant twinkling from the many stars that have formed within them (Figure 2.1.2). The gas consists largely of hydrogen and helium, and the dust consists of tiny mineral grains, ice crystals, and organic particles.



Figure 2.1.2 Photograph of a nebula. The Pillars of Creation within the Eagle Nebula viewed in visible light (left) and near infrared light (right). Near infrared light captures heat from stars, and allows us to view stars that would otherwise be hidden by dust. This is why the picture on the right appears to have more stars than the picture on the left [NASA, ESA, and the Hubble Heritage Team (STScI/AURA) http://bit.ly/1Dm2X5a].

A solar system begins to form when a small patch within a nebula (small by the standards of the universe, that is) begins to collapse upon itself. Exactly how this starts isn't clear, although it might be triggered by the violent behavior of nearby stars as they progress through their life cycles. Energy and matter released by these stars might compress the gas and dust in nearby neighborhoods within the nebula. Once it is triggered, the collapse of gas and dust within that patch continues for two reasons. One of those reasons is that gravitational force pulls gas molecules and dust particles together. But early in the process, those particles are very small, so the gravitational force between them isn't strong. So how do they come together? The answer is that dust first accumulates in loose clumps for the same reason dust bunnies form under your bed: static electricity. As the small patch within a nebula condenses, a star begins to form from material drawn into the center of the patch, and the remaining dust and gas settle into a disk that rotates around the star. The disk is where planets eventually form, so it's called a **protoplanetary disk**. In Figure 2.1.3 the image in the upper left shows an artist's impression of a protoplanetary disk, and the image in the upper right shows an actual protoplanetary disk surrounding the star HL Tauri. Notice the dark rings in the protoplanetary disk. These are gaps where planets are beginning to form. The rings are there because incipient planets are beginning to collect the dust and gas in their orbits. There is an analogy for this in our own solar system, because the dark rings are akin to the gaps in the rings of Saturn (Figure 2.1.3, lower left), where moons can be found (Figure 2.1.3, lower right).

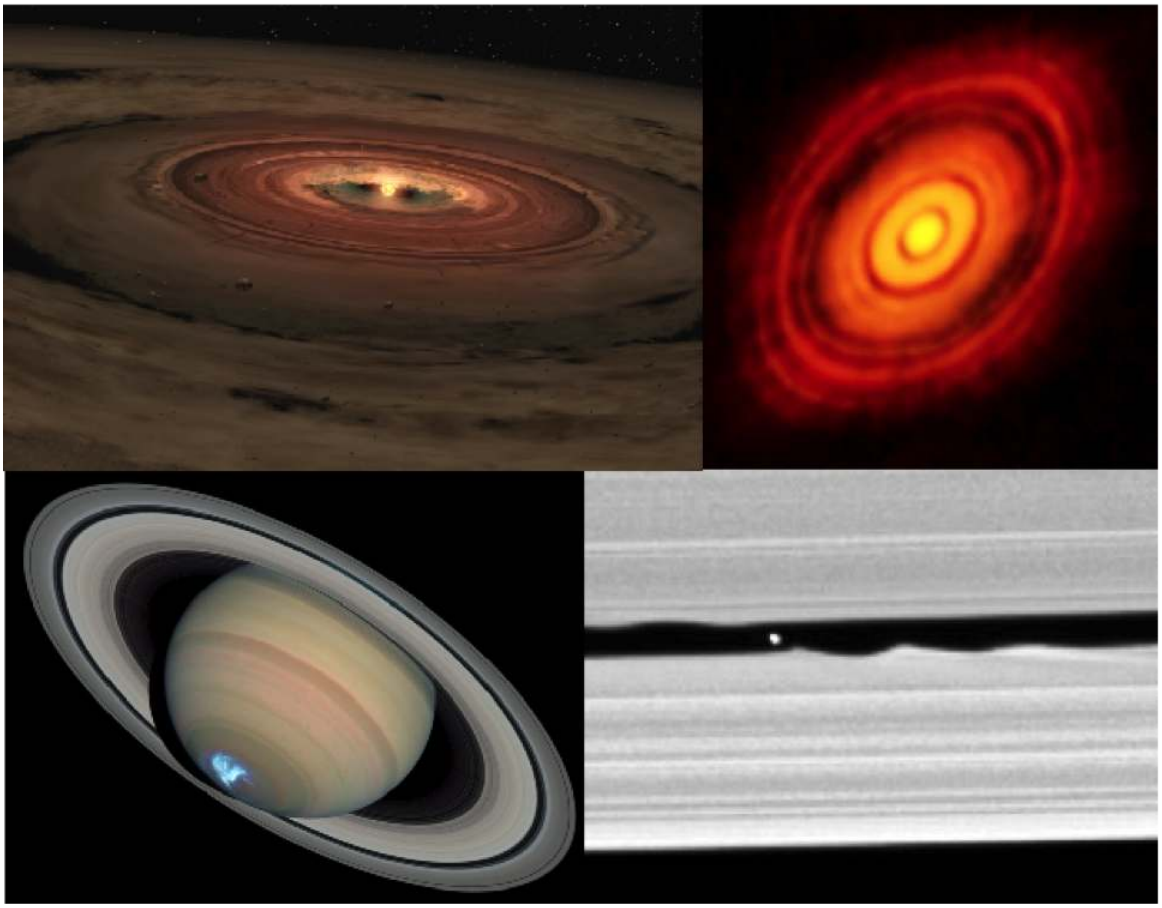


Figure 2.1.3 Protoplanetary disks and Saturn's rings. Upper left: An artists impression of a protoplanetary disk containing gas and dust, surrounding a new star. [NASA/ JPL-Caltech, http://1.usa.gov/1E5tFJR] Upper right: A photograph of the protoplanetary disk surrounding HL Tauri. The dark rings within the disk are thought to be gaps where newly forming planets are sweeping up dust and gas. [ALMA (ESO/NAOJ/NRAO) http://bit.ly/1KNCq0e]. Lower left: A photograph of Saturn showing similar gaps within its rings. The bright spot at the bottom is an aurora, similar to the northern lights on Earth. [NASA, ESA, J. Clarke (Boston University), and Z. Levay (STScI) http://bit.ly/1IfSCX5] Lower right: a close-up view of a gap in Saturn's rings showing a small moon as a white dot. [NASA/JPL/Space Science Institute, http://1.usa.gov/1g2EeYw].

In general, planets can be classified into three categories based on what they are made of (Figure 2.1.4). **Terrestrial planets** are those planets like Earth, Mercury, Venus, and Mars that have a core of metal surrounded by rock. **Jovian planets** (also called **gas giants**) are those planets like Jupiter and Saturn that consist predominantly of hydrogen and helium. **Ice giants** are planets such as Uranus and Neptune that consist largely of water ice, methane (CH_4) ice, and ammonia (NH_3) ice, and have rocky cores. Often, the ice giant planets Uranus and Neptune are grouped with Jupiter and Saturn as gas giants; however, Uranus and Neptune are very different from Jupiter and Saturn.

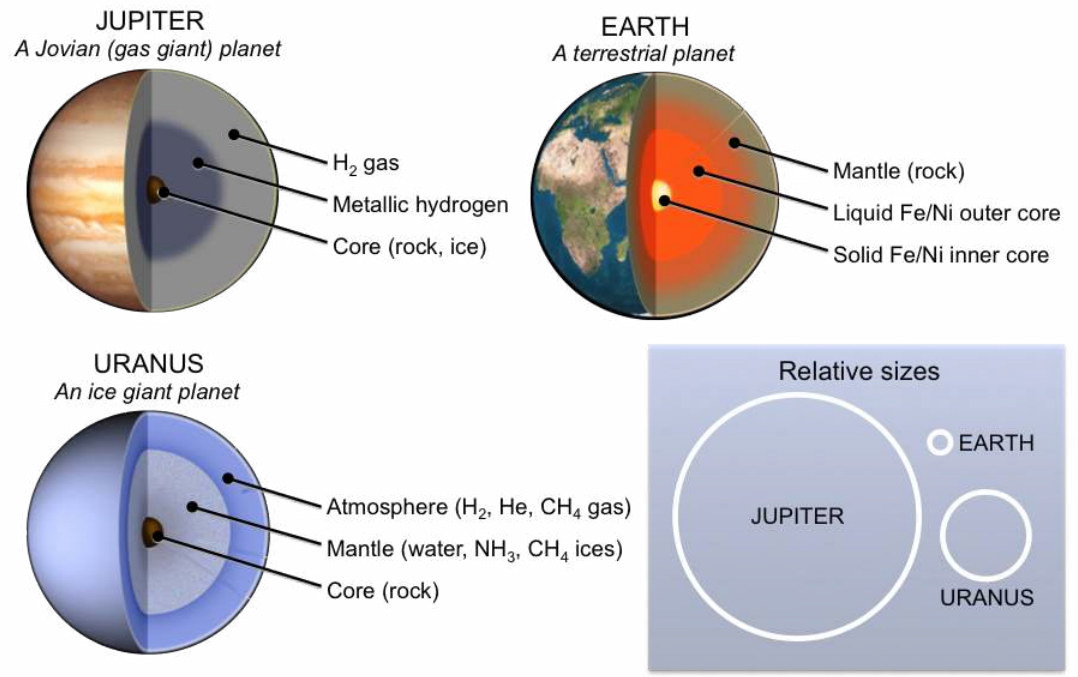


Figure 2.1.4 Three types of planets. Jovian (or gas giant) planets such as Jupiter consist mostly of hydrogen and helium. They are the largest of the three types. Ice giant planets such as Uranus are the next largest. They contain water, ammonia, and methane ice. Terrestrial planets such as Earth are the smallest, and they have metal cores covered by rocky mantles. [KP, after public domain images by Francesco A, Wolfman SF (http://bit.ly/1eP75P4), and NASA (http://1.usa.gov/1gFVsf6, http://1.usa.gov/1M89jI3)].

These three types of planets are not mixed together randomly within our solar system. Instead they occur in a systematic way, with terrestrial planets closest to the sun, followed by the Jovian planets and then the ice giants. Part of the reason for this arrangement is the **frost line** (also referred to as the **snow line**). The frost line separated the inner part of the protoplanetary disk closer to the sun, where it was too hot to permit anything but silicate minerals and metal to crystallize, from the outer part of the disk farther from the Sun, where it was cool enough to allow ice to form. As a result, the objects that formed in the inner part of the protoplanetary disk consist largely of rock and metal, while the objects that formed in the outer part consist largely of gas and ice. The young sun also blasted the solar system with raging **solar winds** (winds made up of energetic particles), which helped to drive lighter molecules toward the outer part of the protoplanetary disk.

The objects in our solar system formed by **accretion**. Early in this process, mineral and rock particles collected in fluffy clumps because of static electricity. As the mass of the clumps increased, gravity became more important, pulling material from farther away and growing these solid masses into larger and larger bodies. Eventually the mass of the objects became large enough that their gravity was strong enough to hang onto gas molecules, because gas molecules are very light.

Our Earth formed though this process of accretion about 4.6 billion years ago. The early Earth was very hot and had a molten, fluid composition, with lost of geological and volcanic activity on the surface. The Earth's heat came from a variety of processes:

- Heat came from the decay of radioactive elements within the Earth, specifically the decay of 235U, 238U, 40K, and 232Th, which are primarily present in the mantle. The total heat produced that way has been decreasing over time (because these isotopes are getting used up), and is now roughly 25% of what it was when Earth formed. This means that Earth's interior is slowly becoming cooler.
- Heat came from the thermal energy already contained within the objects that accreted to form the Earth.
- Heat came from collisions. When objects hit Earth, some of the energy from their motion went into deforming Earth, and some of it was transformed into heat. (The very worst collision that Earth experienced was with a planet named Theia, which was approximately the size of Mars. Not long after Earth formed, Theia struck Earth. When Theia slammed into Earth, Theia's metal core merged with Earth's core, and debris from the outer silicate layers was cast into space, forming a ring of rubble around Earth. The material within the ring coalesced into a new body in orbit around Earth, giving us our moon. Remarkably, the debris may have coalesced in 10 years or fewer! This scenario for the formation of the moon is called the **giant impact hypothesis**.)

- As Earth became larger, its gravitational force became stronger. This increased Earth's ability to draw objects to it, but it also caused the material making Earth to be compressed, rather like Earth giving itself a giant gravitational hug. Compression causes materials to heat up.

Heating had a very important consequence for Earth's structure. As Earth grew, it collected a mixture of silicate mineral grains as well as iron and nickel. These materials were scattered throughout Earth. That changed when Earth began to heat up: it got so hot that both the silicate minerals and the metals melted. The metal melt was much denser than the silicate mineral melt, so the metal melt sank to Earth's center to become its core, and the silicate melt rose upward to become Earth's crust and mantle. In other words, Earth unmixed itself. The separation of silicate minerals and metals into a rocky outer layer and a metallic core, respectively, is called **differentiation**. Gravity has since pulled Earth into an almost spherical shape with a radius of 6371 km, and a circumference of about 40,000 km. However, it is not a perfect sphere, as the Earth's rotation causes an equatorial bulge, so that the Earth's circumference is 21 km (0.3%) wider at the equator than it is pole to pole. Thus it is technically an "oblate spheroid."

If we were to take an inventory of the elements that make up Earth, we would find that 95% of Earth's mass comes from only four elements: oxygen, magnesium, silicon, and iron. Most of the remaining 5% comes from aluminum, calcium, nickel, hydrogen, and sulphur. We know that the Big Bang made hydrogen, helium, and lithium, but where did the rest of the elements come from? The answer is that the other elements were made by stars. The heat and pressure within stars cause smaller atoms to smash together and fuse into new, larger atoms. For example, when hydrogen atoms smash together and fuse, helium is formed. Large amounts of energy are released when some atoms fuse and that energy is what causes stars to shine.

It takes larger stars to make elements as heavy as iron and nickel. Our Sun is an average star; after it uses up its hydrogen fuel to make helium, and then some of that helium is fused to make small amounts of beryllium, carbon, nitrogen, oxygen, and fluorine, it will be at the end of its life. It will stop making atoms and will cool down and bloat until its middle reaches the orbit of Mars. In contrast, large stars end their lives in spectacular fashion, exploding as supernovae and casting off newly formed atoms —including the elements heavier than iron — into space. It took many generations of stars creating heavier elements and casting them into space before heavier elements were abundant enough to form planets like Earth.

---

*"Physical Geology" by Steven Earle used under a CC-BY 4.0 international license. Download this book for free at http://open.bccampus.ca

2.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **3.1: Origin of Earth and the Solar System** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

2.2: Origin of the Elements

The Earth is composed of 90 chemical elements, of which 81 have at least one stable isotope. Most of these elements have also been detected in stars. Where did these elements come from? The accepted scenario is that the first major element to condense out of the primordial soup was helium, which still comprises about one-quarter of the mass of the known universe.

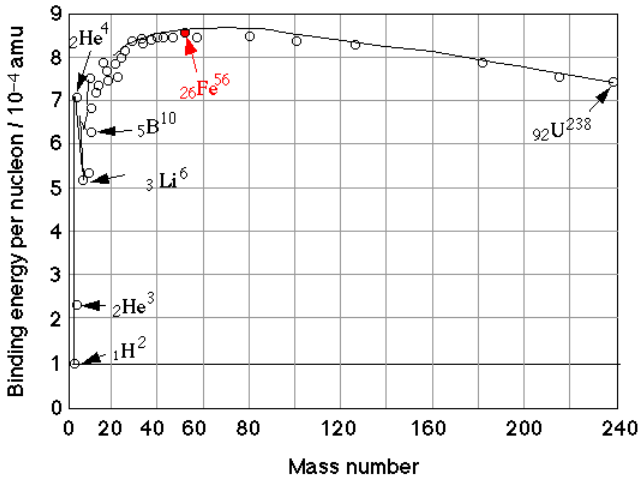


Figure 2.2.1: Hydrogen is the least thermodynamically stable of the elements, and at very high temperatures will combine with itself in a reaction known as *nuclear fusion* to form the next element, $_2He^4$. "Heavier" nuclei (that is , those having high atomic numbers, indicated here by the subscript preceding the element symbol), are more stable than "lighter" ones, so this fusion process can continue up to $_{56}Fe$, which is the most energetically stable of all the nuclides. Beyond this point, heavier nuclei slowly become less stable, so fission becomes more likely. FIssion, however, is not considered an important mechanism of primordial *nucleosynthesis*, so other processes are invoked, as discussed farther below.

Primordial Chemistry

According to the *"big bang" theory* for which there is now overwhelming evidence, the universe as we know it (that is, all space, time, and matter) had its origin in a point source or *singularity* that began an explosive expansion about 12-15 billion years ago, and which is still continuing.

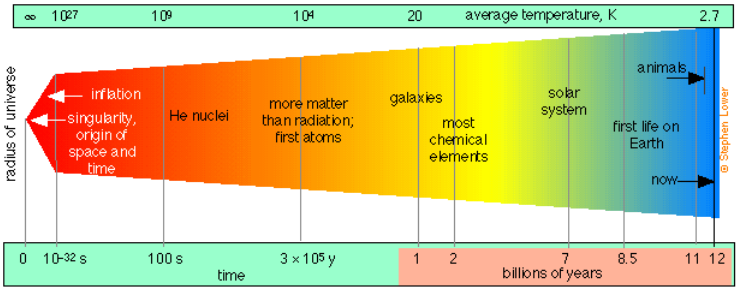


Figure 2.2.2

Following a brief period of extremely rapid expansion called *inflation*, protons and neutrons condensed out of the initial quantum soup after about 10^{-32} s. Helium and hydrogen became stable during the first few minutes, along with some of the very lightest nuclides up to $_7Li$, which were formed through various fusion and neutron-absorption processes. Formation of most heavier elements was delayed for about 10^6 years until nucleosynthesis commenced in the first stars. Hydrogen still accounts for about 93% of the atoms in the universe.

Stellar Nucleosynthesis

All elements beyond hydrogen were formed in regions where the concentration of matter was large, and the temperature was high; in other words, in stars. The formation of a star begins when the gravitational forces due to a large local concentration of hydrogen bring about a contraction and compression to densities of around 10^5 g cm^{-3}. This is a highly exothermic process in which the

gravitational potential energy is released as heat, about 1200 kJ per gram, raising the temperature to about 10^7 K. Under these conditions, hydrogen nuclei possess sufficient kinetic energy to overcome their electrostatic repulsion and undergo nuclear fusion:

$$4\,{}^{1}_{1}\mathrm{H} \longrightarrow {}^{4}_{2}\mathrm{He} + 2\,\beta^{+} + 2\,\gamma^{+}\ 2\,\mathrm{n} \tag{2.2.1}$$

Hydrogen burning

There will be a net mass loss in above process, which will therefore be highly exothermic and is known as “hydrogen burning”. As hydrogen burning proceeds, the helium collects in the core of the star, raising the density to 10^8 g cm^{-3} and the temperature to 10^8 K. This temperature is high enough to initiate *helium burning*, which proceeds in several steps:

$$2\,{}^{4}_{2}\mathrm{He} \longrightarrow 4\,\mathrm{Be}_8 + \mathrm{g} \tag{2.2.2}$$

The first product, ${}^{8}_{4}\mathrm{Be}$ has a half life of only 10^{-16} sec, but a sufficient amount accumulates to drive the following two reactions:

$${}^{8}_{4}\mathrm{Be} + {}^{4}_{2}\mathrm{He} \longrightarrow {}^{12}_{6}\mathrm{C} + \gamma \tag{2.2.3}$$

$${}^{12}_{6}\mathrm{C} + {}^{1}_{1}\mathrm{H} \longrightarrow {}^{13}_{1}\mathrm{N} + {}^{13}_{1}\mathrm{C} + \beta^{+} + \gamma \tag{2.2.4}$$

The size of a star depends on the balance between the kinetic energy of its matter and the gravitational attraction of its mass. As the helium burning runs its course, the temperature drops and the star begins to contract. The course of further nucleosynthesis and the subsequent fate of the star itself depends on the star’s mass.

Small stars

If the mass of the star is no greater than 1.4 times the mass of our sun, the star collapses to a white dwarf, and eventually cools to a dark, dense dead star.

Big stars

In larger stars, the gravitational forces are sufficiently strong to overcome the normal repulsion between atoms, and so gravitational collapse continues. The gravitational energy released in this process produces temperatures of 6 $\rightarrow$ 10^8 K, which are sufficient to initiate a complex series of nuclear reactions known as the *carbon-nitrogen cycle*. The net reaction of this cycle is the further fusion of hydrogen to helium, in which C^{12} acts as a catalyst, and various nuclides of nitrogen and oxygen are intermediates. The temperature is sufficiently high, however, to initiate fusion reactions of some of these intermediates:

$_6C^{12} + {}_6C^{12} \rightarrow {}_{10}Ne^{20} + {}_2He^4$

$2\ {}_8O^{16} \rightarrow {}_{14}Si^{28} + {}_2He^4$

$2\ {}_8O^{16} \rightarrow {}_{16}S^{31} + {}_0n^1$

Supernovas

The intense gamma radiation that is produced in some of these reactions breaks some of the product nuclei into smaller fragments, which can then fuse into a variety of heavier species, up to the limit of $_{26}Fe^{56}$, beyond which fusion is no longer exothermic. The greater relative abundance of elements such as $_6C^{12}$, $_8O^{16}$, and $_{10}Ne^{20}$ which differ by a $_2He^4$ nucleus, reflects the participation of the latter species in these processes. These exothermic reactions eventually produce temperatures of 8 $\rightarrow$ 10^9 K, while contraction continues until the central core is essentially a ball of neutrons having a radius of about 10 km and a density of 10^{14} g cm^{-3}. At the same time the outer shell of the star is blasted away in an explosion known as a *supernova*.

Note

Only six supernovas have been observed in our galaxy. The supernova of 1987 was the most recent; the one before this occurred in 1604, prior to the invention of the telescope. Tycho Brahe's observation of a supernova in 1572 was crucial in overturning the Aristotelian tradition of the immutability of the "fixed stars", or "firmament". The remains of these supernovas have been detected and studied by X-ray observations.

Thus all of the elements in our solar system that are heavier than iron are the recycled remnants of former stars.

The Chandra X-ray observatory page on **supernovas**

Elements heavier than iron

Since $_{26}Fe^{56}$ has the highest binding energy per nucleon of any nuclide, there are no exothermic processes which can lead to the formation of heavier elements. Fusion into heavier species is also precluded by the electrostatic repulsion of the highly charged nuclei. However, the process of *neutron capture* can still take place (this is the same process that is used to make synthetic elements). The neutrons are by-products of a large variety of stellar processes, and are present in a wide range of energies. Two general types of neutron capture processes are recognized. In an “s” (slow) process, only a single neutron is absorbed and the product usually decomposes by b-decay into a more proton-rich species:

$$_{26}Fe^{56} + {}_{0}n^{1} \rightarrow {}_{26}Fe^{56} \rightarrow {}_{26}Fe^{59} \rightarrow {}_{27}Co^{59} + {}_{-1}e^{0}$$

This process occurs at rates of about 10^{5} yr^{-1} , and accounts for the lighter isotopes of many elements. The other process (the “r”, or rapid process) occurs in regions of high neutron density and involves multiple captures at rates of 0.1-10 sec^{-1}:

$$_{26}Fe^{56} + 13\ {}_{0}n^{1} \rightarrow {}_{26}Fe^{56} \rightarrow {}_{27}Co^{59} + {}_{-1}e^{0}$$

This mechanism favors the heavier, neutron-rich isotopes and the heaviest elements.

Other elements

A few nuclei are not accounted for by any of the processes mentioned. These are all low-abundance species, and they probably result from processes having low rates. Examples are Sn^{112} and Sn^{114} which may be produced through proton-capture, and H^{2}, Li^{6}, Li^{7}, Be, B^{10} and B^{11}, which may come from spallation processes resulting from collisions of cosmic ray particles with heavier elements.

Contributors

- Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook

This page titled 2.2: Origin of the Elements is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

2.3: Formation and evolution of the Earth

The solar system is believed to have formed about 5 billion years ago as a result of aggregation of cosmic dust and interstellar atoms in a region of space in which the density of such material happened to be greater than average. Over 99.8% of this mass, which consisted mostly of hydrogen, collapsed into a proto-sun; the gravitational energy released in this process raised the temperature sufficiently to initiate the hydrogen fusion reactions discussed above.

The planets

The remaining material probably formed a disk that rotated around the sun. As the temperature dropped to around 2000K, some of the most stable combinations of the elements began to condense out. These substances might have been calcium aluminum silicates, followed by the more volatile iron-nickel system, and then magnesium silicates. The further aggregation of these materials, together with the other constituents of the cooling disk, is now believed to be the origin of the planets. Density estimates indicate that the planets closest to the sun are predominantly rocky in nature, and probably condensed first. The outer planets (Uranus, Neptune and Pluto) appear to consist largely of water ice, methane, and ammonia, with a smaller rocky core.

Formation of the Earth

The Earth formed by accretion of solid and particulate material that remained after the much more massive amounts of hydrogen and helium present in the original protoplanets had been dispersed out of the solar system. Gradually, the heat produced by decay of radioactive elements brought about partial melting of the silicate rocks; these lower density molten materials migrated upward, leaving the more dense, iron-containing minerals below. This process, which took about 2 million years, was the first of the three stages into which the chemical evolution of the earth is usually divided:

1. **Primary differentiation** of the elements between the core and mantle.
2. **Secondary differentiation** of the elements, reflecting relative ionic sizes, bonding properties, and solubilities (influencing phase behavior such as fractional crystallization, etc.)
3. **Tertiary differentiation**, still operative, involving the interaction of the crust with the hydrosphere and atmosphere.

The above listing should not be taken too literally; all three kinds of processes have probably proceeded simultaneously, and over a number of cycles. Since the earth is losing approximately four times as much heat as is generated by radioactive decay, the principal driving force of primary and secondary differentiation has gradually slowed down. Partial melting of the upper mantle brought about further fractionation as silicon-containing materials of low density migrated outward to form a crust. In its early stages the stronger granitic rocks had not yet appeared, and the crust was mechanically weak. Upwelling flows of lava would break the surface, and the weight of the solidified lava would cause the crust to subside. In some places, magma would solidify underground, forming low-density rock (batholiths) that would eventually rise by buoyancy and push up overlying crust. These mountain-building periods probably occurred in 6-8 major episodes, each lasting about 800 million years.

Rain

At the same time, outgassing of solids released large amounts of HCl, CO, CO_2, H_2S, CH_4 , SO_2 , and SO_3 into the primitive atmosphere. Large amounts of water were present in the primeval rocks in the form of hydrates, which were broken down as the result of the heating. Eventually when the outer crust cooled enough to permit condensation of the water vapor as rain, a new stage of chemical evolution began. The rain was initially highly acidic, equivalent to about 1M HCl; this reacted readily with the basic rocks having high contents of K, Na, Mg, and Ca, leaching them away and forming what would eventually evolve into the oceans. The partial dissolution of the rocks also resulted in large amounts of sediments, which played their own role in the transformation of the earth's surface.

The continents

Within the crust, the lighter materials, being in isostatic equilibrium with the upper mantle, floated higher, and gradually became the nuclei of continents, which grew by accumulating similar material around their boundaries. This picture of continental development is supported by isotopic ratio studies which indicate that the nucleus of the North American continent, the Canadian Shield, is over 2.5 billion years old, while the peripheral parts are less than 0.6 billion years of age.

Primary differentiation of the elements

The more traditional geochemical view of primary differentiation begins with the assumption that the core of the earth is in a chemically reduced state, while the metallic elements constituting the mantle are almost entirely oxidized to their lower free energy cationic forms. Oxygen and sulfur acted as the major electron acceptors in this process, but the abundance of these elements was insufficient to oxidize much of the nickel or iron.

Iron as a reductant

Iron itself is believed to have played a crucial role in the primary differentiation of other metals and of oxidized metallic elements that iron is able to reduce. As the dense molten iron migrated in toward the core, it dissolved (formed a liquid alloy with) any other metals with which it came in contact, and it reduced (donated electrons to) those metallic cations that are less “active” metals than iron under these conditions. The resulting metal would then mix with more of the migrating liquid iron, and be carried along with it into the core.

Redox power of the elements

Accordingly, elements whose reduction potentials are more positive than iron (i.e., are lower-free energy electron sinks) are called *siderophiles*; these elements have a low abundance in the crust and upper mantle. The other two important classes of solid-forming elements are *lithophile* and *chalcophile* (see below.) These generally have more negative reduction potentials than iron, and are distinguished mainly by their relative affinity for oxygen or sulfur. The chalcophiles, of which Cu, Cd, and Sb are examples, tend to form larger, more polarizable ions which can associate with the sulfide ion. The lithophiles comprise those elements such as K, Al, Mn, and Si, which have smaller ions and which combine preferentially with oxygen. This broad classification is reflected in the dominant forms in which many of these elements occur in nature.

Secondary differentiation of the elements

The differential distribution of the elements within one of the main regions of the earth has been studied in detail only in that portion that is accessible, namely the upper crust. It is clear that fractional crystallization from the cooling magma has played an important role. The relative temperatures at which minerals crystallize is determined in large part by their lattice energies, which are in turn related to ionic sizes and charges. Minerals with small, highly charged ions will have higher melting points and should crystallize first. Thus the sodium-containing feldspar albite ($NaAlSi_2O_8$) is found nearer the surface than is its calcium analog anorthite ($CaAlSi_2O_8$). The less abundant elements often do not form minerals of their own, but may replace the ion of a more abundant mineral in its crystal lattice. This is known as *isomorphous replacement*, and it naturally depends on the relative ionic radii. Some ion pairs that undergo isomorphous replacement in minerals are K^+ and Ba^{2+}, Si^{4+} and Ge^{4+}.

Phase behavior

The Phase Rule can be invoked to explain in a very rough way the differentiation of the elements into distinct solid phases.

$P = C + 2 - F$

Taking the degrees of freedom as 2 (fixed temperature and pressure), the six major elemental components (O, Si, Al, Fe, Mg and Na) can form up to six phases. Actually, more than 99% of igneous rocks comprise seven principal mineral phases. These are: the silica minerals, feldspars, feldspathoids, olivine, pyroxenes, amphiboles and micas.

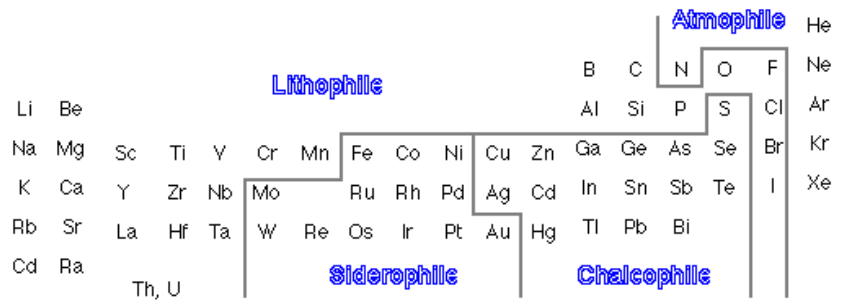
The differential deposition of minerals is also influenced by the temperature-composition phase relations as exemplified by the ordinary two-component phase diagram. If the mineral that is rich in one component and which first crystallizes out is also more dense, then the richer ore will occur near the bottom of the deposit, while a more mixed ore (approaching the eutectic) will remain near the top.

Geochemical classification of the elements

Whether an element is concentrated in the crust or elsewhere depends on its chemical behavior and on the physical properties of its stable compounds. Geochemists have found it convenient to establish the following general classifications:

- *lithophiles* ("rock-loving") elements are those such as Fe, Al, and Si which tend to occur as oxides (and to a lesser extent as chlorides and carbonates.) Elements in this, the largest of all the groups, are concentrated in the crust.

- *chalcophiles* also occur in the crust, but mainly in combination with sulfur and the other chalcogen elements of Groups 15-16.
- *siderophiles* refer to the elements such as Ni which have concentrated in the core along with Fe.
- *atmosphiles* consist of N, H and their volatile compounds
  and the noble gases which concentrate in the atmosphere.



References

"TheTalk" site contains numerous links to discussions of geochronology and other evidence for the age of the earth, and on the pseudoscientific concept of a "young Earth".

Unravelling geological time (Columbia University)

Formation of the Earth's core - description of theories based on high pressure studies

Page last modified: 21.01.2008

© 1998 by Stephen Lower

For information about this Web site or to contact the author,
please see the*Chem1 Virtual Textbook home page*.

SOME RIGHTS RESERVED

This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License.

This page titled 2.3: Formation and evolution of the Earth is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

2.4: The Earth's crust

The structure and composition of the outer part of the lithosphere has been profoundly affected by interactions with the atmosphere over one-quarter of the surface area of the earth, and with the hydrosphere over the remaining area. Further modification of the outermost parts of the crust has occurred as the result of the activities of living organisms. These changes have transformed much of the outermost parts of the crust into an unconsolidated surface region called the regolith. Further weathering and translocation of soluble substances often results in a sequence of horizons consisting of sediments, soils, or evaporites.

Chemically, the earth's crust consists of about 80 elements distributed in approximately 2000 compounds or minerals, many of which are of variable composition. Over 99% of the mass of the crustal material is made up of only eight of these elements, however:

Table 2.4.1: Average amounts of elements in crustal rocks, mg/g.

O	466,000	Na	28,300
Si	277,200	K	25,900
Al	81,300	Mg	20,900
Fe	50,000	Ti	4,400
Ca	36300	H	1,400

The crust has its origin in the upwelling convection currents that bring mantle material near to the surface at the mid-ocean ridges. The reduced pressure causes it to melt into *magma*. The magma may solidify before it reaches the surface, forming basalt, or it may energe from the surface in a volcanic eruption. The oceanic crust consists mostly of the simpler silicate minerals, which are said to be *basic* or *mafic*. The more evolved, silicon-rich rocks found in the continental crust are known as *acidic* or *sialic*.

Oceanic crust is continually being extruded from regions of the plastic mantle that intrude upward to just beneath the ocean's floor at the mid-ocean ridges. A corresponding amount of this crust is being returned to the lithosphere at subduction zones off the West coasts of the Americas, and in the process pushing up the mountain ranges that lie along these coasts. The subducted oceanic crust is reheated and combined with sedimentary material to undergo partial remelting and reworking; this is believed by some to be the origin of *granite*. Subduction proceeds at a rate of a few cm per year, and the complete cycle time is on the order of a few hundred million years.

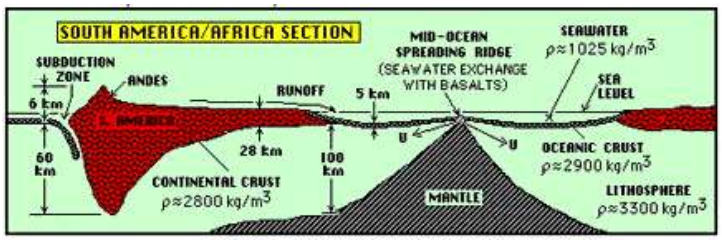


Figure 2.4.1: The continental and oceanic crusts. This schematic cross section runs in a west-to-east direction, from just of the West coast of South America on the left to the African continent on the right. Oceanic crust is shown in black, with with continental crust in red. (From Cosmos and Earth Study Guide, NYU)

Both oceanic and contintental crusts float on the more dense upper lithosphere, and gradually shift their positions as they push against each other, and in response to the slow convective motions in the medium that supports them. The **continental crust** is thicker than the oceanic crust, but it is also less dense, which allows it to float higher (and thus to differentiate continents from oceans.) The lower density also prevents it from being subducted. Recycling can occur indirectly as continental material erodes and

is deposited as sediments on the ocean floor, but this is a much slower process and one that takes billions instead of millions of years. Some of the very oldest rocks, found in Greenland and Labrador, have been dated at 3.9 billion years, and thus approach the age of the Earth itself.

Chemistry of the crust

When magma crystallizes it forms igneous rock, the major component of the Earth's crust. The crystallization is a complex process which is not entirely understood, due largely to the lack of sufficient thermodynamic data on the various components at high temperatures and pressures. It is known that the different components of magma have differing melting points and densities, and that the phase behavior of multicomponent systems based on some of these substances is quite complex, involving binary and ternary eutectics, solid solutions, the presence of dissolved water (under pressure), and incongruent melting. One consequence of this complexity is that the composition of the magma will change as crystallization takes place; different substances will crystallize at various stages, and the resulting solids may migrate toward the top or bottom of the region if their densities differ greatly from that of the magma

It is well known that larger crystals form when a melt cools more slowly. This principle affords a simple distinction between the coarser-grained plutonic rocks, which are believed to have been formed by gradual cooling of magma pockets within the crust, and the fine-grained volcanic rocks such as basalt. Under the influence of heat and pressure, particularly at plate boundaries, solid crustal material may undergo partial or complete remelting, followed by cooling and transformation into metamorphic rocks such as gneiss, micas, quartzite, and possibly granites.

> ***Granite*** *was once thought to be an igneous rock, originating from the crystallization of a particular kind of magma. The association of granitic rocks with mountainous regions, and the similarity of their compositions in widely scattered regions, lends credence to the more recent hypothesis that granitic rocks are of metamorphic origin.*

Another class of rock is *sedimentary rock*, formed from the consolidation of material produced by weathering and other chemical, and biological processes. Sedimentary rocks cover about three-quarters of the land area of the earth; 80% are shales, 15% sandstones and 5% limestones.

Composition of rock

The chemical composition of rocks tends to be complex and variable, and can only be specified in a precise way at the structural level. The traditional way of expressing rock compositions is in terms of the mass percent of the oxides of the elements present in the rock.

Table 2.4.1: **: Chemical composition of a typical rock** (quartz-feldspar-biotite gneiss). The figures are in percent by weight.

oxide	common name	fresh rock	weathered rock
SiO_2	silica	71.54	70.30
Al_2O_3	alumina	14.62	18.34
Fe_2O_3	ferric oxide	0.69	1.55
FeO	ferrous oxide	1.64	0.22
MgO	magnesia	0.77	0.21
CaO	lime	2.08	0.10
Na_2O	soda	3.84	0.09
K_2O	potash	0.32	5.88
H_2O	water	0.32	5.88
others		0.65	0.54
total		100.07	99.70

https://geo.libretexts.org/@go/page/20855

oxide	common name	fresh rock	weathered rock
total			

This does not mean that these oxides, or the structural units they represent, are actually present as such in a rock. In the chemical analysis of rocks, oxygen is generally not determined separately. When it is, however, it is found in an amount that would be expected to combine stoichiometrically with the other elements present. Thus the composition of albite can be written as either $NaAlSi_3O_8$ or $Na_2O{\cdot}Al_2O_3{\cdot}6SiO_2$. Some rocks contain varying ratios of certain elements. For example olivine, which can be considered a solid solution of Mg_2SiO_4 and Fe_2SiO_4, can be represented by $(Mg,Fe)_2SiO_4$; this implies that the ratio of metal to silica is constant, and that magnesium is ordinarily present in greater amount than iron.

The major structural elements of rock (both in the crust and in the mantle) are the silicate minerals, built from silicon atoms surrounded tetrahedrally by four oxygens. The simplest of these consist just of SiO_4^{4-} tetrahedra interspersed with positive ions to achieve electroneutrality; olivine, $(Mg,Fe)_2SiO_4$ is a well known example. More commonly, the silicate groups polymerize by sharing one or more oxygen atoms at adjacent tetrahedral corners. Depending on the number of joined corners per silicate unit, this can lead to the formation of a wide variety of chains (pyroxenes, amphiboles) and sheets (micas), culminating in the complete tetrahedral polymerization that produces quartz, SiO_2.

Higher degrees of polymerization are associated with higher ratios of Si to O, smaller quantities of positive ions, and higher melting points. Thus when magma cools, the first silicates to crystallize are the olivines, followed by chain and sheet minerals having progressively higher degrees of polymerization and smaller fractions of cations of metals such as Fe and Mg.

Distribution of elements; ores

Although some elements are distributed fairly uniformly throughout the crust, others occur at greatly enhanced concentrations in localized areas. There are two general processes that result in these localized excesses, which are called *ores* when their extraction and refining is economically feasible

The first of these relates to how well a metallic ion can fit into the silicate lattice structure. Ions having the right charge and size can readily enter this structure, displacing the more common ions of Fe, Al and Mg. Such ions (of which Ga^{3+} is an example) are readily soluble in other minerals and thus are widely disributed and do not concentrate into ores. Other ions may be too large (Cs and La), too small (Li, Be, B) or too highly charged (Nb, Ta, W) to be accommodated in silicate mineral structures; these elements tend to remain in the magma as it solidifies, finally forming solid minerals only in the last stages of cooling.

The other major source of ores is *hydrothermal* formation. Magma contains some water itself, and additional water from the surface is able to reach the heated rock near magma chambers. At the very high temperatures and pressures that prevail in these regions, the water can dissolve many compounds such as sulfides which are normally considered highly insoluble. When these superheated solutions rise to the surface the solids are re-deposited, often in highly concentrated form. Ores of Cu, Sn, W, and possibly some iron ores, as well as some native metals such as gold, are believed to be formed in this way.



Figure 2.4.2: The hydrothermal vent photographed here emits both black and white "smoke".

Hydrothermal vents known as “black smokers” have been observed at sites of sea-floor spreading; the “smoke” consists of metallic sulfides which precipitate in the cold seawater. The veins of pyrites (FeS_2) and similar sulfide minerals that are often observed in rock formations are the result of hydrothermal solutions that once penetrated cracks and fissures in the rock.

Deep-sea hydrothermal vents (many more pictures) Theories of ore formation (a more technical tutorial)

Chemical weathering

The weathering of rocks at the earth’s surface is a complex process involving both physical and chemical changes. The latter tend in principle to be rather simple kinds of reactions involving dissolution, reaction with carbon dioxide, hydrolysis, hydration, and oxidation. The difficulty in studying them and in arriving at a quantitative description is that these reactions occur very slowly and may never reach an equilibrium state. A comparison of the two rightmost columns in Table 2 on page 14 provides some illustration of the overall effect of these changes, although it must be emphasized that these are relative composition data, and thus cannot show how much of a given component has been lost. In general, sodium, calcium and magnesium seem to be lost more rapidly than potassium and silicon, while iron and aluminum decrease very slowly. Individual rates are of course dependent on the particular structural units containing the element, and also vary somewhat with grain size and condition of the surface.

Action of water

Water is undoubtedly the most important weathering agent. Not only does it act as a solvent for ionic dissolution products, but it also brings other active agents such as carbon dioxide and oxygen into intimate contact with the rock material. As water percolates into the outermost layers of the crust, it extends the zone of weathering beneath the surface; the effects of this are quite noticeable in a number of buried sedimentary materials such as Paleozoic sandstones, which tend to be depleted of all but the most resistant minerals. Dissolution, the simplest of all the weathering processes, usually results in ionic species, some of which may react with water to yield acidic or alkaline solutions. Dissolution of silica, however, results in the neutral species H_4SiO_4. Reactions involving hydration and dehydration are very common, and since the free energy changes tend to be small, these reactions can usually take place in either direction under slightly different conditions. Thus gypsum and anhydrite are interconvertable at observable rates under common environmental conditions:

$CaSO_4 \cdot 2H_2O \rightarrow CaSO_4 + 2\ H_2O$

In many cases, however, the reaction products are not very well characterized, thermodynamic data is lacking, and the reactions proceed so slowly that they are not entirely understood. For example, both hydrous and anhydrous iron oxides can be found in similar geologic environments, but the little is known about the interconversion process, represented approximately as

$Fe_2O_3 + H_2O \rightarrow 2\ FeOOH$

Solid carbonates tend to dissolve in acidic solutions, including those produced when atmospheric carbon dioxide dissolves in water. Thus the major surface limestone deposits (largely $CaCO_3$, with some admixture of $MgCO_3$) tend to be highly eroded in non-arid

regions, and the local groundwater may have Ca^{2+} as high as 0.1-0.2M. Thermodynamics can unambiguously predict the most stable oxidation state of a metal ion under given conditions of pH and oxidant concentration. The mechanisms tend to be very uncertain, however. For one thing, both the reactant and product can often exist in various states of hydration, and the dissolved species (which probably undergo the actual oxidation) often consist of polycations and complexed species.

Oxidation of iron

Compounds of Fe(II), for example, will always tend to oxidize to Fe(III) in the presence of air; the various oxides of iron are responsible for the bright colors seen in many geological formations, and in certain soils. Some of the net reactions that probably occur are

$Fe_2SiO_4 + 1/2\ O_2 + H_2O \rightarrow Fe_2O_3 + H_2SiO_4$

$2\ FeCO_3 + 1/2\ O_2 \rightarrow Fe_2O_3 + 2\ H_2CO_3$

An environmental side effect of the first process is the release of hydrated silica. Also, where both starting materials are present, the carbonic acid produced in the second reaction is believed to promote the dissolution of ferrous silicate, creating a source of Fe(II) ions that can be rapidly oxidized:

$Fe_2SiO_4 + 4\ H_2CO_3 \rightarrow 2\ Fe^{2+} + 4\ HCO_3^- + H_4SiO_4$

$Fe^{2+} + 4\ HCO_3^- + 1/2\ O_2 \rightarrow Fe_2O_3 + 4\ H_2CO_3$

The oxidation of sulfides can produce strongly acidic solutions:

$2\ FeS_2 + 15/2\ O_2 + 4\ H_2O \rightarrow Fe_2O_3 + 2\ SO_4^{2-} + 8\ H^+$

The effects of this can be seen in formations containing outcrops of pyrite veins, where the surrounding rocks are heavily stained with yellow and brown Fe(III) oxides, and the groundwater tends to be highly acidic. This process is mediated by microorganisms, and is an important source of acid pollution associated with mines and mine tailings.

Sequence of weathering

The various components of rocks weather at different rates. The more basic components such as CaO and MgO tend to disappear first, especially if in contact with groundwaters containing high CO2 concentrations. For rocks in general, the first reaction is usually hydration, followed by hydrolysis which can be summarized by

$4\ KAlSi_3O_8(s) + 22\ H_2O \rightarrow Al_4Si_4O_{10}(OH)_8(s) + 8\ H_4SiO_4(aq) + 4\ K^+ + 4\ OH^-$

in which other Group 1 or 2 cations might replace potassium. The product $Al_4Si_4O_{10}(OH)_8$ is *kaolinite*, a form of clay (see below).

In general, the rocks which crystallized first from the magma (the Ca-feldspars and olivines) weather more rapidly than do the lower-melting rocks.

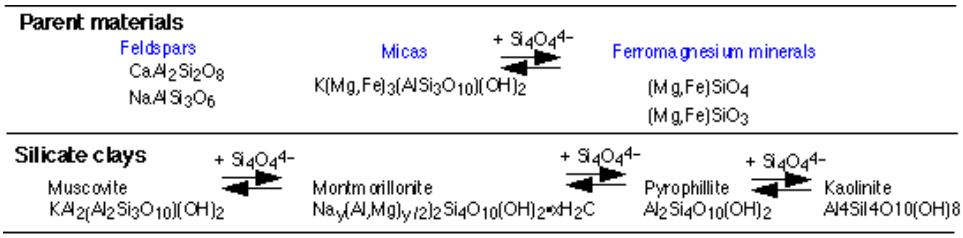


Figure 2.4.3:

Clays

Clays are the solid end products of the weathering of rocks. They are basically composed of alternating sheets of "SiO_2" and "AlO_6" units in ratios of 1:1 (kaolinite), 2:1 (montmorillonite and vermiculite) and 2:2 (chlorite). In between the sheets, and

holding them together by hydrogen bonding are water molecules. Also present are cations such as K^+, Ca^{2+} and Mg^{2+} which act to neutralize the negative charges of the oxide ions.

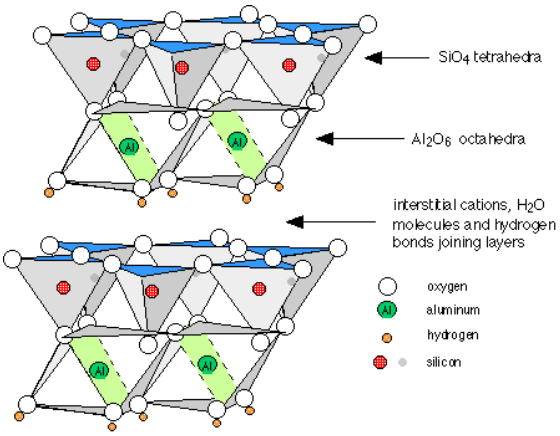


Figure 2.4.4: Structure of a clay, showing two layers of the stacked sheets of kaolinite

Physical weathering

The major agents of physical weathering of exposed rocks are rapid changes in temperature (promoting fracture by differential expansion), the abrasive action of windborne material and glacier movement, and especially by the penetration of water into cracks and its subsequent freezing. The expansion of water on freezing can exert a pressure of 150 kg cm^{-2}, whereas the tensile strength of a typical rock is around 120 kg cm^{-2} . The roots of some plants are able to penetrate rock quite effectively, producing comparable expansive pressures in subsurface rocks.

Composition and structure of soils

Soils are a product of the interaction of water, air, and living organisms with exposed rocks or sediments at the earth's surface. A typical soil contains about 45% inorganic solids and 5% organic solids by volume. Water and air each make up about 20-30%.

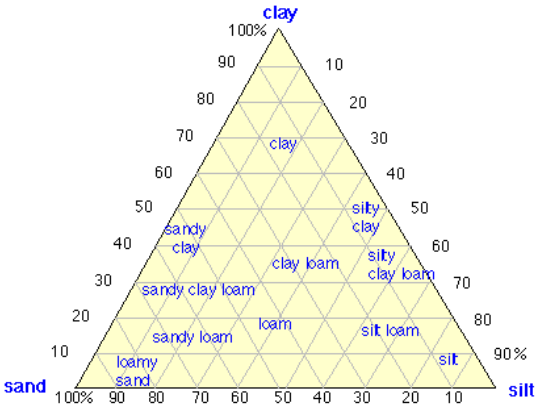


Figure 2.4.5: A simple way of classifying soils is based on the relative quantities of clay, silt and sand in the solid component. For ordinary agricultural purposes, loams are considered the most ideal soil type.

Mineral Components of soil

The primary inorganic components of soils consist of sand and silt particles that come directly from the parent rocks. This fraction is dominated by quartz and feldspars (aluminosilicates). Secondary components are formed by chemical changes within the soil itself, or in sediments from which the soil derived. These are most commonly clays, but may also include calcite, gypsum, and sulfide minerals such as pyrites; the latter are formed by bacterial action under reducing conditions in the presence of organic matter.

The clays have an especially important effect on both the physical properties of the soil, and on its ability to store plant nutrients, including trace nutrients such as Mo and Mn. These properties are due to the high ion-exchange capacity of clays. The more highly charged cations such as Al3+ and Fe3+ tend to be more strongly absorbed within the inter-sheet regions than do Mg^{2+} or K^{+}. As plants withdraw these latter cations from the soil water, more are released by the clay components, which thus act as nutrient reservoirs.

The ion-exchange properties of clays also help to maintain the pH balance of soils, through the exchange of H^{+} and cations such as Ca^{2+}. The soil pH, in turn, strongly affects the solubility of nutrient cations, and thus their availability to plants. For example, the uptake of phosphorus (in the form of $H_2PO_4^{2-}$, is only efficient within the rather narrow pH range between 6 and 7. Below 6, dihydrogen phosphates of Fe and Al are precipitated, while insoluble $Ca_3(PO_4)_2$ forms at higher pH's.

Organic Components

Part of the organic matter of soil consists of organisms (mainly bacteria and fungi) and roots and root hairs. The remainder is largely in the form of fulvic and *humic* acids. These substances of indefinite composition are classified on the basis of their solubility behavior; fulvic acids remain in solution at pH 2, but humic acids, having molecular weights of 20,000 to 100,000, are precipitated. Both are flexible polyelectrolytes that interact strongly with their own kind and with inorganic ions.

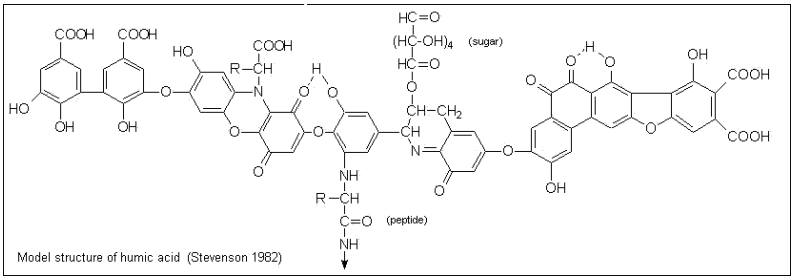


Figure 2.4.6: Model structure of humic acid (Stevenson 1982)

Associated with the fulvic and humic fractions are a wide variety smaller molecules such as alkanes, amino acids, amino sugars, sulfur and phosphorus derivatives of sugars, etc. Part of the organic carbon in a fertile soil is recycled in 1-2 years; plant residues, which are the major source of soil organic matter, have a half-life of days to months. Once carbon gets incorporated into humic substances, it is locked into a much slower recycling process; the turnover times of fulvic acids are a hundred years or more, while those for fulvic acids are around a thousand years. For this reason, humic substances are the major reservoir of organic carbon in soil. Organic matter, particularly polysaccharides, binds strongly to the cation components of clay colloids; the two together act as cementing agents and strongly influence the consistency and structure of the soil.

Water

Soil water is held by capillary action and adsorption with varying degrees of tenacity. This water binding strength is traditionally expressed in terms of the pressure, or "tension" that would be required to force the water out of the soil. The tension of capillary water varies over a wide range of 0.1-32 atm; only in the lower half of this range will it be available to plants, which can exert an osmotic pressure of up to about 15 atm. Water in excess of the capillary capacity fills larger voids and is called gravitational water. Its presence in surface soils corresponds to a flooded condition that inhibits plant growth by reducing soil aeration.

Air

The gas phase within soil pores generally has a CO_2 content of 5-50 times that of the atmosphere due to the action of organisms. O_2 tends to be depleted to roughly the extent that CO_2 is present in excess. Under conditions of poor aeration (i.e., exchange with the atmosphere), considerable quantities of N_2O, NO, H_2, CH_4, C_2H_4, and H_2S may also be present.

What is soil? An introductory survey

Soils interactive textbook

Page last modified: 21.01.2008

© 1998 by Stephen Lower

For information about this Web site or to contact the author,
please see the*Chem1 Virtual Textbook home page*.

This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License.

O	466,000
Si	277,200
Al	81,300
Fe	50,000
Ca	36300

This page titled 2.4: The Earth's crust is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

SECTION OVERVIEW

2.5: Origin of the atmosphere and the oceans

2.5.1: Origin of the Oceans

2.5.2: Origin and Evolution of the Atmosphere

2.5: Origin of the atmosphere and the oceans is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.5.1: Origin of the Oceans

So how did the oceans form in the first place? Remember from section 3.1 that the early Earth was formed through the accretion of various materials, and that a period of melting and intense volcanic activity followed. The materials that accreted on the early Earth contained the components that would eventually become our oceans and atmosphere. Under the high pressures found in the Earth's interior, gases remain dissolved in magma. As these magmas rise to the surface through volcanic activity, the pressure is reduced and the gases are released through a process called **outgassing**. Volcanic activity releases many different gases, including water vapor, carbon dioxide (CO_2), sulfur dioxide (SO_2), carbon monoxide (CO), hydrogen sulfide (H_2S), hydrogen gas, nitrogen, and methane (CH_4). Lighter gases such as hydrogen and helium dissipated into space, but the heavier gases remained and formed Earth's early atmosphere.

The rise of atmospheric oxygen

> It's worth noting that the early atmosphere lacked free oxygen (O_2), the form of oxygen that we breathe. We know this in part because prior to 2 billion years ago, there were no sedimentary beds stained red from oxidized iron minerals. Iron minerals were present, but not in oxidized form. At that time, O_2 was produced in the atmosphere when the Sun's ultraviolet rays split water molecules apart; however, chemical reactions removed the oxygen as quickly as it was produced. It wasn't until the appearance of life that Earth's atmosphere began to become oxygenated. Photosynthetic organisms used the abundant CO_2 in the atmosphere to manufacture their food, and released O_2 as a by-product. At first all of the oxygen was consumed by chemical reactions, but eventually the organisms released so much O_2 that it overwhelmed the chemical reactions and oxygen began to accumulate in the atmosphere, although present levels of 21% oxygen didn't occur until about 350 Ma. Today the part of our atmosphere that isn't oxygen consists largely of nitrogen (78%). The oxygen-rich atmosphere on our planet is life's signature. If geologic process were the only processes controlling our atmosphere, it would consist mostly of carbon dioxide, like the atmosphere of Venus.

As the early Earth cooled, the water vapor in the atmosphere condensed and fell as rain. By about 4 billion years ago, the first permanent accumulations of water were present on Earth, forming the oceans and other bodies of water. Water moves between these different reservoirs through the **hydrological cycle**. Water is evaporated from the oceans, lakes, streams, the surface of the land, and plants (transpiration) by solar energy (Figure 2.5.1.1). It is moved through the atmosphere by winds and condenses to form clouds of water droplets or ice crystals. It comes back down as rain or snow and then flows through streams and rivers, into lakes, and eventually back to the oceans. Water on the surface and in streams and lakes infiltrates the ground to become groundwater. Groundwater slowly moves through the rock and surface materials; some returns to other streams and lakes, and some goes directly back to the oceans.

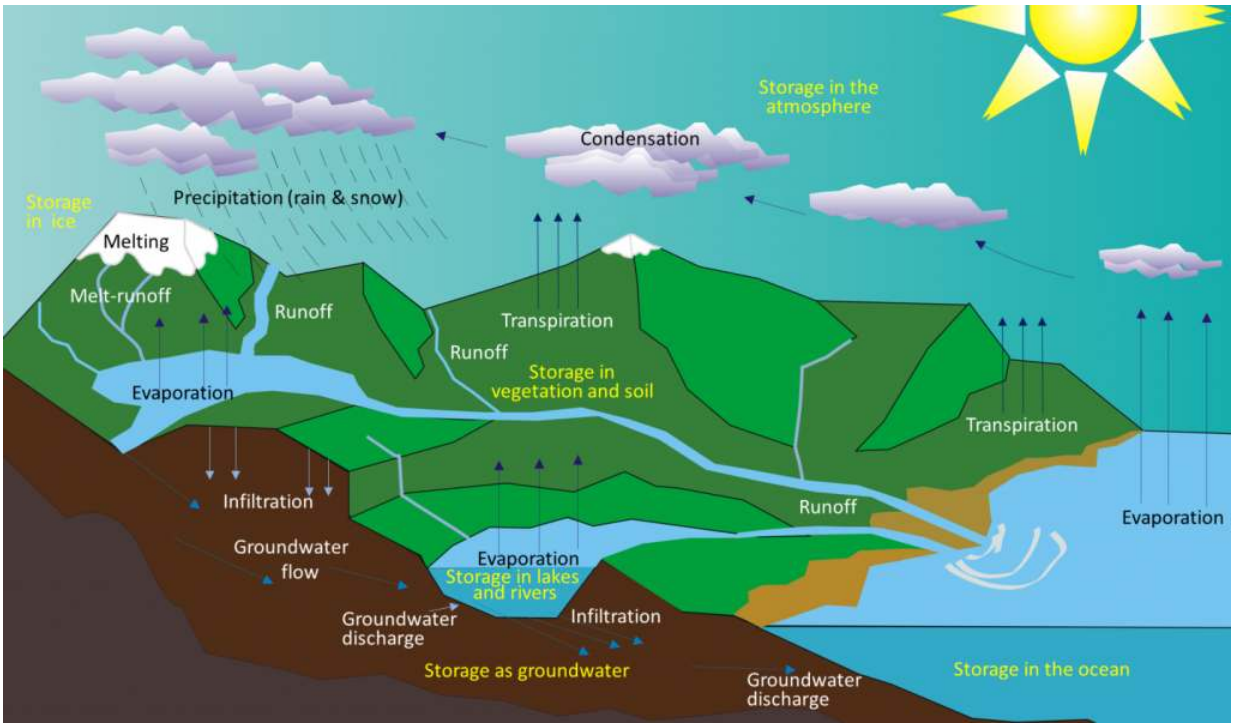


Figure 2.5.1.1 Earth's hydrological cycle (Steven Earle, "Physical Geology").

Water is stored in various reservoirs as it moves through this cycle. The largest, by far, is the oceans, accounting for 97% of the volume (Figure 2.5.1.2). Of course, that water is salty. The remaining 3% is fresh water. Two-thirds of our fresh water is stored in the ground and one-third is stored in ice. The remaining fresh water — about 0.03% of the total — is stored in lakes, streams, vegetation, and the atmosphere.

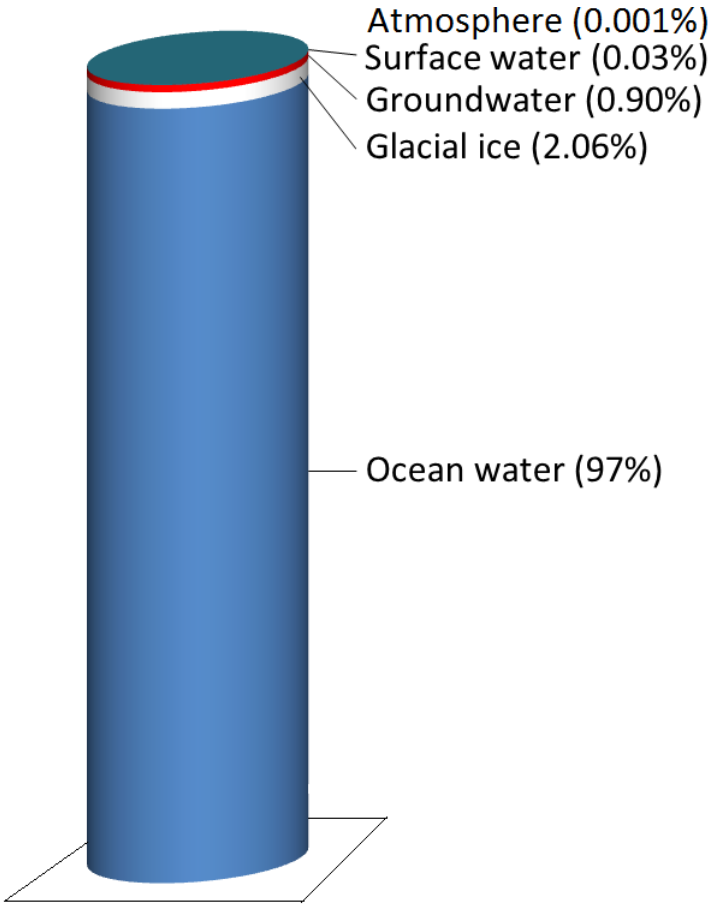


Figure 2.5.1.2 Proportions of Earth's water found in the various reservoirs (Steven Earle, "Physical Geology").

To put that in perspective, let's think about putting all of Earth's water into a 1 L jug. We start by almost filling the jug with 970 ml of water and 34 g of salt. Then we add one regular-sized (~20 mL) ice cube (representing glacial ice) and two teaspoons (~10 mL) of groundwater. All of the water that we see around us in lakes and streams and up in the sky can be represented by adding three more drops from an eyedropper.

Although the proportion of Earth's water that is in the atmosphere is tiny, the actual volume is huge. At any given time, there is the equivalent of approximately 13,000 km^3 of water in the air in the form of water vapor and water droplets in clouds. Water is evaporated from the oceans, vegetation, and lakes at a rate of 1,580 km^3 per day, and just about exactly the same volume falls as rain and snow every day, over both the oceans and land. The precipitation that falls on land goes back to the ocean in the form of stream flow (117 km^3/day) and groundwater flow (6 km^3/day).

How did the oceans get salty?

Outgassing was responsible for ocean formation, but how did the ocean water get salty? Most of the salts and dissolved elements in the ocean were probably outgassed along with the water vapor, so the ocean has probably always been about as salty as it is now. But we know that rainfall and other processes weather rocks on the Earth's surface, and runoff carries dissolved substances into the ocean, contributing to its salinity. Yet despite this constant input, the ocean's salt composition remains essentially the same. Therefore, the rate of input of new material must be balanced by the rate of removal; in other words, the oceans are in a **steady state** in regards to salinity.

There are multiple pathways through which dissolved ions enter the ocean; runoff from streams and rivers, volcanic activity, hydrothermal vents (see section 4.11), dissolution or decay of substances in the ocean, and groundwater input. Ions are removed from seawater as they are incorporated by living organisms (for example in shell production) or sediments, sea spray, percolation of water into the crust, or when sea water gets isolated from the ocean and evaporates.

The relationship between the input and removal of an ion can be examined through the concept of **residence time**, which is the average length of time a single atom of an element remains in the ocean before being removed. Residence time is calculated as:

$$\text{Residence time} = \frac{\text{amount of the substance in the ocean}}{\text{the rate at which the substance is added or removed}}$$

There is great variation in residence times for different substances (Table 5.2.1). Generally speaking, substances that are readily used in biological processes have short residence times, as they are used up as they become available. Substances with longer residence times are less reactive, and may be a part of long-scale geological cycles.

[table id=4 /]

So what about lakes? They are subjected to runoff and river input, so why aren't they salty like the oceans? One reason is that compared to the oceans, lakes and ponds are relatively temporary phenomena, so they do not last long enough to accumulate the same levels of ions as the oceans. Furthermore, lakes often have rivers flowing both into and out of them, so many ions are removed through the outflow, eventually finding their way to the oceans. The oceans only receive river input; there are no rivers flowing out of the ocean to remove these materials, so they are found in greater abundance in sea water. It should be noted that there are some lakes that contain water whose salt content may rival or exceed that of the ocean; these lakes usually lack river outflow. The Great Salt Lake in the western United States is an example.

---

*"Physical Geology" by Steven Earle used under a CC-BY 4.0 international license. Download this book for free at http://open.bccampus.ca

This page titled 2.5.1: Origin of the Oceans is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Paul Webb via source content that was edited to the style and standards of the LibreTexts platform.

- **5.2: Origin of the Oceans** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

2.5.2: Origin and Evolution of the Atmosphere

Prebiotic atmosphere

The atmosphere of the Earth (and also of Venus and Mars) is generally believed to have its origin in relatively volatile compounds that were incorporated into the solids from which these planets accreted. Such compounds could include nitrides (a source of N_2), water (which can be taken up by silica, for example), carbides, and hydrogen compounds of nitrogen and carbon. Many of these compounds (and also some noble gases) can form clathrate complexes with water and some minerals which are fairly stable at low temperatures. The high temperatures developed during the later stages of accretion as well as subsequent heating produced by decay of radioactive elements presumably released these gases the surface. Even at the present time, large amounts of CO_2 , water vapor, N_2, HCl, SO_2 and H_2S are emitted from volcanos. The more reactive of these gases would be selectively removed from the atmosphere by reaction with surface rocks or dissolution in the ocean, leaving an atmosphere enriched in its present major components with the exception of oxygen which is discussed in the next section.

Any hydrogen present would tend to escape into space, causing the atmosphere to gradually became less reducing. However there is now some doubt that hydrogen and other volatiles (mainly the inert gases) were present in the newly accreted planets in anything like their cosmic abundance. The main evidence against this is the observation that gases such as helium, neon and argon, which are among the ten most abundant elements in the universe, are depleted on the earth by factors of 10^{-7} to 10^{-11} . This implies that there was a selective removal of these volatiles prior to or during the planetary accretion process.

The overall oxidation state of the earth's mantle is not consistent with what one would expect from equilibration with highly reduced volatiles, and there is no evidence to suggest that the composition of the mantle has not remained the same. If this is correct, then the primitive atmosphere may well have had about the same composition as the gases emitted by volcanos at the present time. These consist mainly of water and CO_2, together with small amounts of N_2, H_2, H_2S, SO_2, CO, CH_4, NH_3, HCl and HF. If water vapor was a major component of outgassing of the accreted earth, it must have condensed quite rapidly into rain. Any significant concentration of water vapor in the atmosphere would have led to a runaway greenhouse effect, resulting in temperatures as high as 400°C.

Origin of Atmospheric Oxygen

Free oxygen is never more than a trace component of most planetary atmospheres. Thermodynamically, oxygen is much happier when combined with other elements as oxides; the pressure of O_2 in equilibrium with basaltic magmas is only about 10^{-7} atm. Photochemical decomposition of gaseous oxides in the upper atmosphere is the major source of O_2 on most planets. On Venus, for example, CO_2 is broken down into CO and O_2 . On the earth, the major inorganic source of O_2 is the photolysis of water vapor; most of the resulting hydrogen escapes into space, allowing the O2 concentration to build up. An estimated 2 $\times$ 10^{11} g of O_2 per year is generated in this way. Integrated over the earth's history, this amounts to less than 3% of the present oxygen abundance. The partial pressure of O_2 in the prebiotic atmosphere is estimated to be no more than 10^{-3} atm, and may have been several orders of magnitude less. The major source of atmospheric oxygen on the earth is photosynthesis carried out by green plants and certain bacteria:

$$H_2O + CO_2 \longrightarrow (CH_2O)_x + O_2 \tag{2.5.2.1}$$

A historical view of the buildup of atmospheric oxygen concentration since the beginning of the sedimentary record (3.7 $\times$ 10^9 ybp) can be worked out by making use of the fact that the carbon in the product of the above reaction has a slightly lower C^{13} content than does carbon of inorganic origin. Isotopic analysis of carbon-containing sediments thus provides a measure of the amounts of photosynthetic O_2 produced at various times in the past.

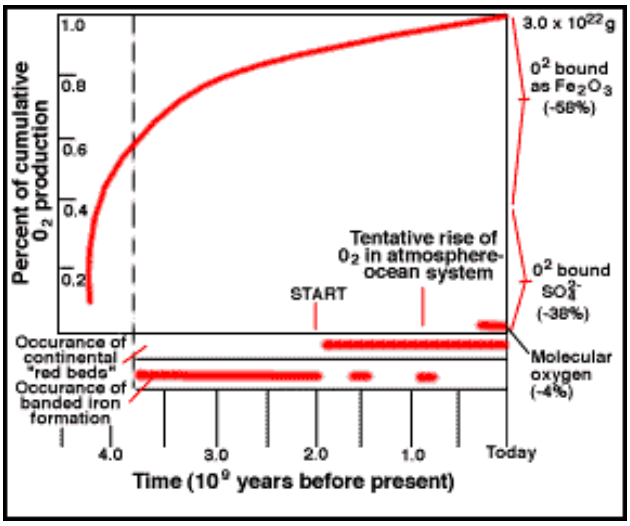


Figure 2.5.2.1: ***Evolution of atmospheric oxygen content.*** *Note carefully that the curve plots cumulative O_2 production, but that until a few hundred million years ago, most of this was taken up by Fe(II) compounds in the crust and by reduced sulfur; only after this massive "oxygen sink" became filled did free O_2 begin to accumulate in the atmosphere.*

Carbon Dioxide

Carbon dioxide has probably always been present in the atmosphere in the relatively small absolute amounts now observed (around 54 x 10^{15} mol = 54 Pmol). The reaction of CO_2 with silicate-containing rocks to form precambrian limestones suggest a possible moderating influence on its atmospheric concentration throughout geological time.

$$\mathrm{CaSiO_3 + CO_2 \rightarrow CaCO_3 + SiO_2} \tag{2.5.2.2}$$

About ten percent of the atmospheric CO_2 is taken up each year by photosynthesis. Of this, all except 0.05 percent is returned by *respiration*, almost entirely due to microorganisms. The remainder leaks into the slow part of the geochemical cycle, mostly as buried carbonate sediments.

Since the advent of the industrial revolution in 1860, the amount of CO_2 in the atmosphere has been increasing. Isotopic analysis shows that most of this has been due to fossil-fuel combustion; in recent years, the mass of carbon released to the atmosphere by this means has been more than ten times the estimated rate of natural removal into sediments. The large-scale destruction of tropical forests, which has accelerated greatly in recent years, is believed to exacerbate this effect by removing a temporary sink for CO_2.

The oceans have a large absorptive capacity for CO_2 owing to its reaction with carbonate:

$$\mathrm{CO_2 + {CO_3}^{2-} \rightleftharpoons 2\,HCO_3^-} \tag{2.5.2.3}$$

There is about 60 times as much inorganic carbon in the oceans as in the atmosphere. However, efficient transfer takes place only into the topmost (100 m) wind-mixed layer of the ocean, which contains only about one atmosphere equivalent of CO_2. Further uptake is limited by the very slow transport of water into the deep ocean, which takes around 1000 years. For this reason, the buffering effect of the oceans on atmospheric CO_2 is not very effective; only about ten percent of the added CO_2 is taken up by the oceans.

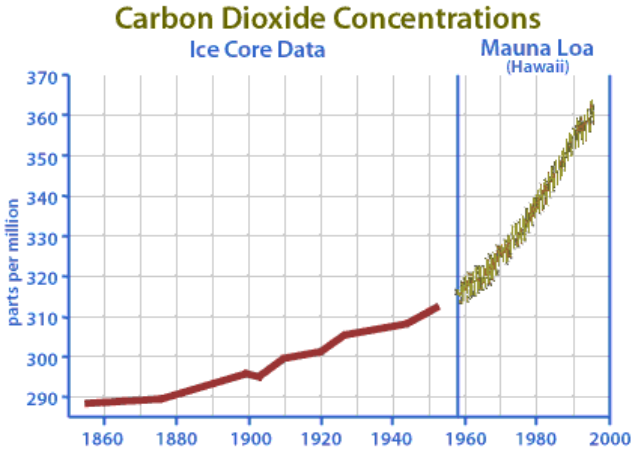


Figure. This set of plots, taken from an ETE page with data from NASA, shows the dramatic increase in atmospheric CO_2since the beginning of the industrial age. The "squiggles" on the Mauna Loa data reflect seasonal variations; more $CO_2$is taken up during the longer-daylight periods of the summer.

Contributors and Attributions

- Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook

---

This page titled 2.5.2: Origin and Evolution of the Atmosphere is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

SECTION OVERVIEW

2.6: Earthlike Planets - Venus and Mars

The Moon and Mercury are geologically dead. In contrast, the larger terrestrial planets—Earth, Venus, and Mars—are more active and interesting worlds. We have already discussed Earth, and we now turn to Venus and Mars. These are the nearest planets and the most accessible to spacecraft. Not surprisingly, the greatest effort in planetary exploration has been devoted to these fascinating worlds. In the chapter, we discuss some of the results of more than four decades of scientific exploration of Mars and Venus. Mars is exceptionally interesting, with evidence that points to habitable conditions in the past. Even today, we are discovering things about Mars that make it the most likely place where humans might set up a habitat in the future. However, our robot explorers have clearly shown that neither Venus nor Mars has conditions similar to Earth. How did it happen that these three neighboring terrestrial planets have diverged so dramatically in their evolution?

2.6.1: The Nearest Planets - An Overview

2.6.2: The Geology of Venus

2.6.3: The Massive Atmosphere of Venus

2.6.4: The Geology of Mars

2.6.5: Water and Life on Mars

2.6.6: Divergent Planetary Evolution

2.6.E: Earthlike Planets - Venus and Mars (Exercises)

This page titled 2.6: Earthlike Planets - Venus and Mars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

2.6.1: The Nearest Planets - An Overview

Learning Objectives

By the end of this section, you will be able to:

- Explain why it's difficult to learn about Venus from Earth-based observation alone
- Describe the history of our interest in Mars before the Space Age
- Compare the basic physical properties of Earth, Mars, and Venus, including their orbits

As you might expect from close neighbors, Mars and Venus are among the brightest objects in the night sky. The average distance of Mars from the Sun is 227 million kilometers (1.52 AU), or about half again as far from the Sun as Earth. Venus' orbit is very nearly circular, at a distance of 108 million kilometers (0.72 AU) from the Sun. Like Mercury, Venus sometimes appears as an "evening star" and sometimes as a "morning star." Venus approaches Earth more closely than does any other planet: at its nearest, it is only 40 million kilometers from us. The closest Mars ever gets to Earth is about 56 million kilometers.

Appearance

Venus appears very bright, and even a small telescope reveals that it goes through phases like the Moon. Galileo discovered that Venus displays a full range of phases, and he used this as an argument to show that Venus must circle the Sun and not Earth. The planet's actual surface is not visible because it is shrouded by dense clouds that reflect about 70% of the sunlight that falls on them, frustrating efforts to study the underlying surface, even with cameras in orbit around the planet (Figure 2.6.1.1).

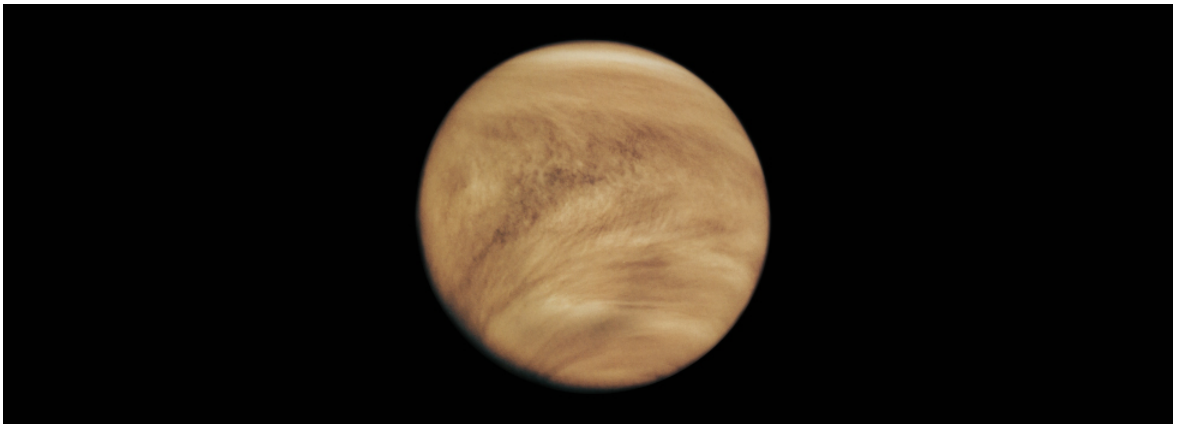


Figure 2.6.1.1 Venus as Photographed by the Pioneer Venus Orbiter. This ultraviolet image shows an upper-atmosphere cloud structure that would be invisible at visible wavelengths. Note that there is not even a glimpse of the planet's surface. (credit: modification of work by NASA)

In contrast, Mars is more tantalizing as seen through a telescope (Figure 2.6.1.2). The planet is distinctly red, due (as we now know) to the presence of iron oxides in its soil. This color may account for its association with war (and blood) in the legends of early cultures. The best resolution obtainable from telescopes on the ground is about 100 kilometers, or about the same as what we can see on the Moon with the unaided eye. At this resolution, no hint of topographic structure can be detected: no mountains, no valleys, not even impact craters. On the other hand, bright polar ice caps can be seen easily, together with dusky surface markings that sometimes change in outline and intensity from season to season.

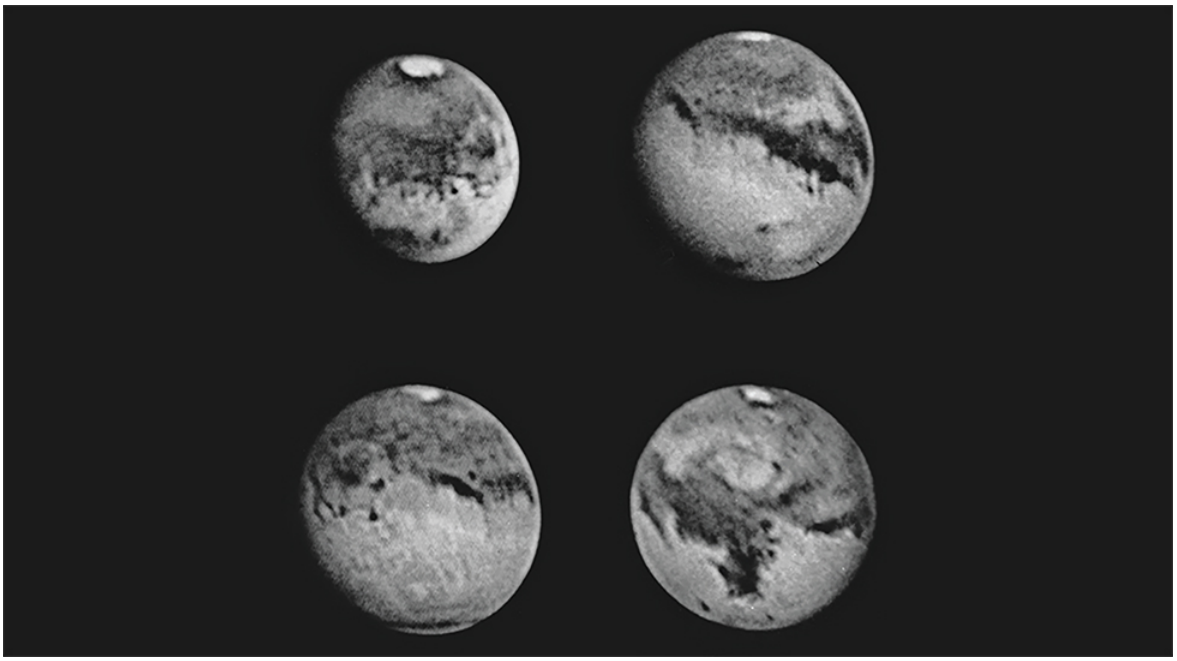


Figure 2.6.1.2 Mars as Seen from Earth's Surface. These are among the best Earth-based photos of Mars, taken in 1988 when the planet was exceptionally close to Earth. The polar caps and dark surface markings are evident, but not topographic features.

For a few decades around the turn of the twentieth century, some astronomers believed that they saw evidence of an intelligent civilization on Mars. The controversy began in 1877, when Italian astronomer Giovanni Schiaparelli (1835–1910) announced that he could see long, faint, straight lines on Mars that he called *canale*, or channels. In English-speaking countries, the term was mistakenly translated as "canals," implying an artificial origin.

Even before Schiaparelli's observations, astronomers had watched the bright polar caps change size with the seasons and had seen variations in the dark surface features. With a little imagination, it was not difficult to picture the canals as long fields of crops bordering irrigation ditches that brought water from the melting polar ice to the parched deserts of the red planet. (They assumed the polar caps were composed of water ice, which isn't exactly true, as we will see shortly.)

Until has death in 1916, the most effective proponent of intelligent life on Mars was Percival Lowell, a self-made American astronomer and member of the wealthy Lowell family of Boston (see the feature box on Percival Lowell: Dreaming of an Inhabited Mars). A skilled author and speaker, Lowell made what seemed to the public to be a convincing case for intelligent Martians, who had constructed the huge canals to preserve their civilization in the face of a deteriorating climate (Figure 2.6.1.3).

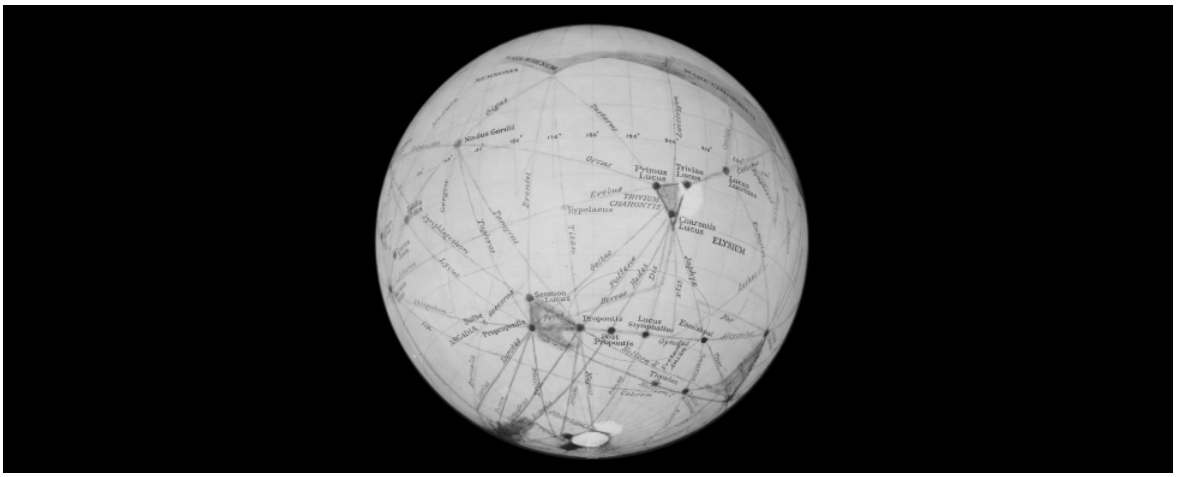


Figure 2.6.1.3 Lowell's Mars Globe. One of the remarkable globes of Mars prepared by Percival Lowell, showing a network of dozens of canals, oases, and triangular water reservoirs that he claimed were visible on the red planet.

The argument for a race of intelligent Martians, however, hinged on the reality of the canals, a matter that remained in serious dispute among astronomers. The canal markings were always difficult to study, glimpsed only occasionally because atmospheric conditions caused the tiny image of Mars to shimmer in the telescope. Lowell saw canals everywhere (even a few on Venus), but many other observers could not see them at all and remained unconvinced of their existence. When telescopes larger than Lowell's failed to confirm the presence of canals, the skeptics felt vindicated. Now it is generally accepted that the straight lines were an optical illusion, the result of the human mind's tendency to see order in random features that are glimpsed dimly at the limits of the eye's resolution. When we see small, dim dots of surface markings, our minds tend to connect those dots into straight lines.

PERCIVAL LOWELL: DREAMING OF AN INHABITED MARS

Percival Lowell was born into the well-to-do Massachusetts family about whom John Bossidy made the famous toast:

> *And this is good old Boston,*
>
> *The home of the bean and the cod, Where the Lowells talk to the Cabots And the Cabots talk only to God.*

Percival's brother Lawrence became president of Harvard University, and his sister, Amy, became a distinguished poet. Percival was already interested in astronomy as a boy: he made observations of Mars at age 13. His undergraduate thesis at Harvard dealt with the origin of the solar system, but he did not pursue this interest immediately. Instead, he entered the family business and traveled extensively in Asia. In 1892, however, he decided to dedicate himself to carrying on Schiaparelli's work and solving the mysteries of the martian canals.

In 1894, with the help of astronomers at Harvard but using his own funds, Lowell built an observatory on a high plateau in Flagstaff, Arizona, where he hoped the seeing would be clear enough to show him Mars in unprecedented detail. He and his assistants quickly accumulated a tremendous number of drawings and maps, purporting to show a vast network of martian canals (see Figure 2.6.1.3). He elaborated his ideas about the inhabitants of the red planet in several books, including *Mars* (1895) and *Mars and Its Canals* (1906), and in hundreds of articles and speeches.

As Lowell put it,

> *A mind of no mean order would seem to have presided over the system we see—a mind certainly of considerably more comprehensiveness than that which presides over the various departments of our own public works. Party politics, at all events, have had no part in them; for the system is planet-wide. . . . Certainly what we see hints at the existence of beings who are in advance of, not behind us, in the journey of life.*

Lowell's views captured the public imagination and inspired many novels and stories, the most famous of which was H. G. Wells' *War of the Worlds* (1897). In this famous "invasion" novel, the thirsty inhabitants of a dying planet Mars (based entirely on Lowell's ideas) come to conquer Earth with advanced technology.

Although the Lowell Observatory first became famous for its work on the martian canals, both Lowell and the observatory eventually turned to other projects as well. He became interested in the search for a ninth (and then undiscovered) planet in the solar system. In 1930, Pluto was found at the Lowell Observatory, and it is not a coincidence that the name selected for the new planet starts with Lowell's initials. It was also at the Lowell Observatory that the first measurements were made of the great speed at which galaxies are moving away from us, observations that would ultimately lead to our modern view of an expanding universe.

Lowell (Figure 2.6.1.4) continued to live at his observatory, marrying at age 53 and publishing extensively. He relished the debate his claims about Mars caused far more than the astronomers on the other side, who often complained that Lowell's work was making planetary astronomy a less respectable field. At the same time, the public fascination with the planets fueled by Lowell's work (and its interpreters) may, several generations later, have helped fan support for the space program and the many missions whose results grace the pages of our text.

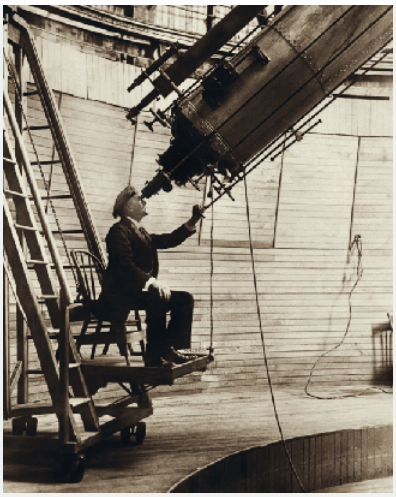


Figure 2.6.1.4 Percival Lowell (1855–1916). This 1914 photograph shows Percival Lowell observing Venus with his 24-inch telescope at Flagstaff, Arizona.

In October 1938, the Mercury Theater of the Air on radio dramatized *The War of the Worlds* as a series of radio news reports. This broadcast scared many people into thinking that Lowell's Martians were really invading New Jersey, and caused something of a panic. You can listen to the original radio broadcast if you scroll down to "War of the Worlds."

Rotation of the Planets

Astronomers have determined the rotation period of Mars with great accuracy by watching the motion of permanent surface markings; its sidereal day is 24 hours 37 minutes 23 seconds, just a little longer than the rotation period of Earth. This high precision is not obtained by watching Mars for a single rotation, but by noting how many turns it makes over a long period of time. Good observations of Mars date back more than 200 years, a period during which tens of thousands of martian days have passed. As a result, the rotation period can be calculated to within a few hundredths of a second.

The rotational axis of Mars has a tilt of about 25°, similar to the tilt of Earth's axis. Thus, Mars experiences seasons very much like those on Earth. Because of the longer martian year (almost two Earth years), however, each season there lasts about six of our months.

The situation with Venus is different. Since no surface detail can be seen through Venus' clouds, its rotation period can be found only by bouncing radar signals off the planet (as explained for Mercury in the Cratered Worlds chapter). The first radar observations of Venus' rotation were made in the early 1960s. Later, topographical surface features were identified on the planet that showed up in the reflected radar signals. The rotation period of Venus, precisely determined from the motion of such "radar features" across its disk, is 243 days. Even more surprising than how *long* Venus takes to rotate is the fact that it spins in a backward or retrograde direction (east to west).

Stop for a moment and think about how odd this slow rotation makes the calendar on Venus. The planet takes 225 Earth days to orbit the Sun and 243 Earth days to spin on its axis. So the day on Venus (as defined by its spinning once) is longer than the year! As a result, the time the Sun takes to return to the same place in Venus' sky—another way we might define the meaning of a day—turns out to be 117 Earth days. (If you say "See you tomorrow" on Venus, you'll have a long time to wait.) Although we do not know the reason for Venus' slow backward rotation, we can guess that it may have suffered one or more extremely powerful collisions during the formation process of the solar system.

Basic Properties of Venus and Mars

Before discussing each planet individually, let us compare some of their basic properties with each other and with Earth (Table 2.6.1.1). Venus is in many ways Earth's twin, with a mass 0.82 times the mass of Earth and an almost identical density. The average amount of geological activity has been also relatively high, almost as high as on Earth. On the other hand, with a surface pressure nearly 100 times greater than ours, Venus' atmosphere is not at all like that of Earth. The surface of Venus is also remarkably hot, with a temperature of 730 K (over 850 °F), hotter than the self-cleaning cycle of your oven. One of the major challenges presented

by Venus is to understand why the atmosphere and surface environment of this twin have diverged so sharply from those of our own planet.

Table 2.6.1.1: Properties of Earth, Venus, and Mars

Property	Earth	Venus	Mars
Semimajor axis (AU)	1.00	0.72	1.52
Period (year)	1.00	0.61	1.88
Mass (Earth = 1)	1.00	0.82	0.11
Diameter (km)	12,756	12,102	6,790
Density (g/cm3)	5.5	5.3	3.9
Surface gravity (Earth = 1)	1.00	0.91	0.38
Escape velocity (km/s)	11.2	10.4	5.0
Rotation period (hours or days)	23.9 h	243 d	24.6 h
Surface area (Earth = 1)	1.00	0.90	0.28
Atmospheric pressure (bar)	1.00	90	0.007

Mars, by contrast, is rather small, with a mass only 0.11 times the mass of Earth. It is larger than either the Moon or Mercury, however, and, unlike them, it retains a thin atmosphere. Mars is also large enough to have supported considerable geological activity in the distant past. But the most fascinating thing about Mars is that long ago it probably had a thick atmosphere and seas of liquid water—the conditions we associate with development of life. There is even a chance that some form of life persists today in protected environments below the martian surface.

Key Concepts and Summary

Venus, the nearest planet, is a great disappointment through the telescope because of its impenetrable cloud cover. Mars is more tantalizing, with dark markings and polar caps. Early in the twentieth century, it was widely believed that the “canals” of Mars indicated intelligent life there. Mars has only 11% the mass of Earth, but Venus is nearly our twin in size and mass. Mars rotates in 24 hours and has seasons like Earth; Venus has a retrograde rotation period of 243 days. Both planets have been extensively explored by spacecraft.

This page titled 2.6.1: The Nearest Planets - An Overview is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **10.1: The Nearest Planets - An Overview** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.6.2: The Geology of Venus

Learning Objectives

By the end of this section, you will be able to:

- Describe the general features of the surface of Venus
- Explain what the study of craters on Venus tells us about the age of its surface
- Compare tectonic activity and volcanoes on Venus with those of Earth
- Explain why the surface of Venus is inhospitable to human life

Since Venus has about the same size and composition as Earth, we might expect its geology to be similar. This is partly true, but Venus does not exhibit the same kind of *plate tectonics* as Earth, and we will see that its lack of erosion results in a very different surface appearance.

Spacecraft Exploration of Venus

Nearly 50 spacecraft have been launched to Venus, but only about half were successful. Although the 1962 US Mariner 2 flyby was the first, the Soviet Union launched most of the subsequent missions to Venus. In 1970, Venera 7 became the first probe to land and broadcast data from the surface of Venus. It operated for 23 minutes before succumbing to the high surface temperature. Additional Venera probes and landers followed, photographing the surface and analyzing the atmosphere and soil.

To understand the geology of Venus, however, we needed to make a global study of its surface, a task made very difficult by the perpetual cloud layers surrounding the planet. The problem resembles the challenge facing air traffic controllers at an airport, when the weather is so cloudy or smoggy that they can't locate the incoming planes visually. The solution is similar in both cases: use a radar instrument to probe through the obscuring layer.

The first global radar map was made by the US Pioneer Venus orbiter in the late 1970s, followed by better maps from the twin Soviet Venera 15 and 16 radar orbiters in the early 1980s. However, most of our information on the geology of Venus is derived from the US *Magellan* spacecraft, which mapped Venus with a powerful *imaging radar*. *Magellan* produced images with a resolution of 100 meters, much better than that of previous missions, yielding our first detailed look at the surface of our sister planet (Figure). (The *Magellan* spacecraft returned more data to Earth than all previous planetary missions combined; each 100 minutes of data transmission from the spacecraft provided enough information, if translated into characters, to fill two 30-volume encyclopedias.)

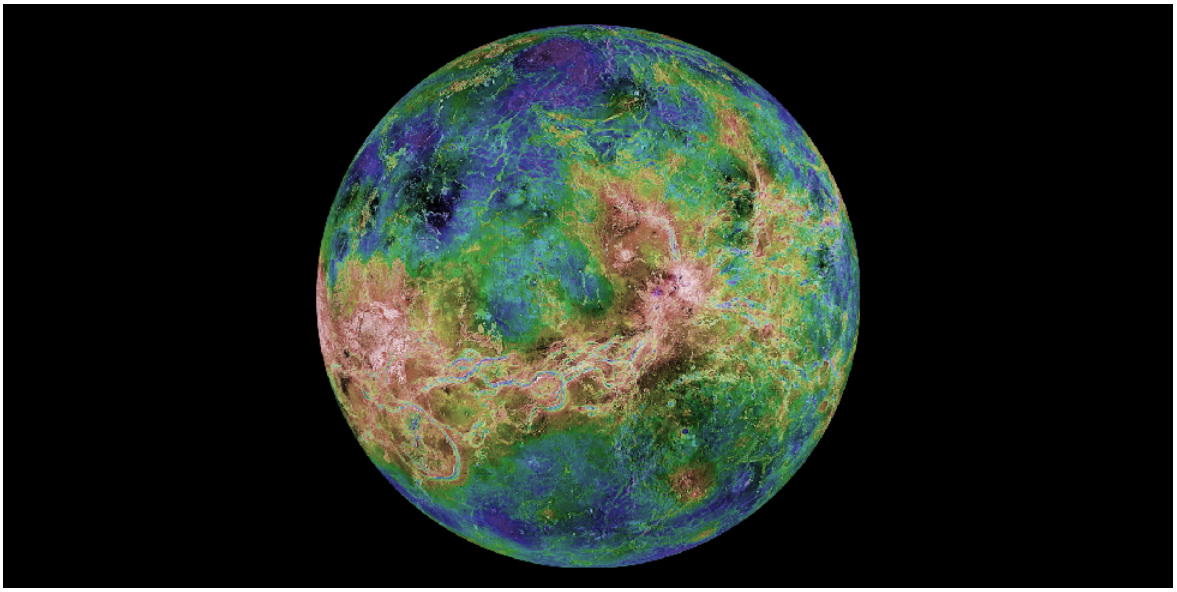


Figure 2.6.2.1 Radar Map of Venus. This composite image has a resolution of about 3 kilometers. Colors have been added to indicate elevation,

with blue meaning low and brown and white high. The large continent Aphrodite stretches around the equator, where the bright (therefore rough) surface has been deformed by tectonic forces in the crust of Venus.

Consider for a moment how good *Magellan*'s resolution of 100 meters really is. It means the radar images from Venus can show anything on the surface larger than a football field. Suddenly, a whole host of topographic features on Venus became accessible to our view. As you look at the radar images throughout this chapter, bear in mind that these are constructed from radar reflections, not from visible-light photographs. For example, bright features on these radar images are an indication of rough terrain, whereas darker regions are smoother.

Probing Through the Clouds of Venus

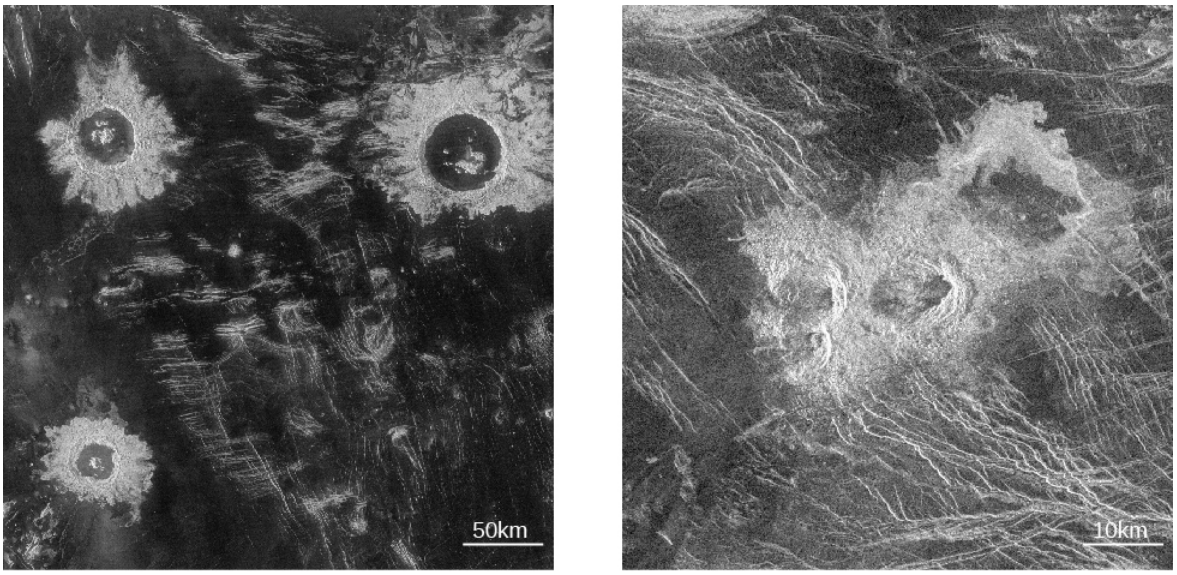
The radar maps of Venus reveal a planet that looks much the way Earth might look if our planet's surface were not constantly being changed by erosion and deposition of sediment. Because there is no water or ice on Venus and the surface wind speeds are low, almost nothing obscures or erases the complex geological features produced by the movements of Venus' crust, by volcanic eruptions, and by impact craters. Having finally penetrated below the clouds of Venus, we find its surface to be naked, revealing the history of hundreds of millions of years of geological activity.

About 75% of the surface of Venus consists of lowland lava plains. Superficially, these plains resemble the basaltic ocean basins of Earth, but they were not produced in quite the same way. There is no evidence of subduction zones on Venus, indicating that, unlike Earth, this planet never experienced plate tectonics. Although *convection* (the rising of hot materials) in its mantle generated great stresses in the crust of Venus, they did not start large continental plates moving. The formation of the lava plains of Venus more nearly resembles that of the lunar maria. Both were the result of widespread lava eruptions without the crustal spreading associated with plate tectonics.

Rising above the lowland lava plains are two full-scale continents of mountainous terrain. The largest continent on Venus, called Aphrodite, is about the size of Africa (you can see it stand out in Figure 2.6.2.1). Aphrodite stretches along the equator for about one-third of the way around the planet. Next in size is the northern highland region Ishtar, which is about the size of Australia. Ishtar contains the highest region on the planet, the Maxwell Mountains, which rise 11 kilometers above the surrounding lowlands. (The Maxwell Mountains are the only feature on Venus named after a man. They commemorate James Clerk Maxwell, whose theory of electromagnetism led to the invention of radar. All other features are named for women, either from history or mythology.)

Craters and the Age of the Venus Surface

One of the first questions astronomers addressed with the high-resolution *Magellan* images was the age of the surface of Venus. Remember that the age of a planetary surface is rarely the age of the world it is on. A young age merely implies an active geology in that location. Such ages can be derived from counting impact craters. Figure 2.6.2.2 is an example of what these craters look like on the Venus radar images. The more densely cratered the surface, the greater its age. The largest crater on Venus (called Mead) is 275 kilometers in diameter, slightly larger than the largest known terrestrial crater (Chicxulub), but much smaller than the lunar impact basins.



(a) (b)

Figure 2.6.2.2 Impact Craters on Venus. (a) These large impact craters are in the Lavinia region of Venus. Because they are rough, the crater rims and ejecta appear brighter in these radar images than do the smoother surrounding lava plains. The largest of these craters has a diameter of 50 kilometers. (b) This small, complex crater is named after writer Gertrude Stein. The triple impact was caused by the breaking apart of the incoming asteroid during its passage through the thick atmosphere of Venus. The projectile had an initial diameter of between 1 and 2 kilometers.

You might think that the thick atmosphere of Venus would protect the surface from impacts, burning up the projectiles long before they could reach the surface. But this is the case for only smaller projectiles. Crater statistics show very few craters less than 10 kilometers in diameter, indicating that projectiles smaller than about 1 kilometer (the size that typically produces a 10-kilometer crater) were stopped by the atmosphere. Those craters with diameters from 10 to 30 kilometers are frequently distorted or multiple, apparently because the incoming projectile broke apart in the atmosphere before it could strike the ground as shown in the Stein crater in Figure 2.6.2.2. If we limit ourselves to impacts that produce craters with diameters of 30 kilometers or larger, however, then crater counts are as useful on Venus for measuring surface age as they are on airless bodies such as the Moon.

The large craters in the venusian plains indicate an average surface age that is only between 300 and 600 million years. These results indicate that Venus is indeed a planet with persistent geological activity, intermediate between that of Earth's ocean basins (which are younger and more active) and that of its continents (which are older and less active).

Almost all of the large craters on Venus look fresh, with little degradation or filling in by either lava or windblown dust. This is one way we know that the rates of erosion or sediment deposition are very low. We have the impression that relatively little has happened since the venusian plains were last resurfaced by large-scale volcanic activity. Apparently Venus experienced some sort of planet-wide volcanic convulsion between 300 and 600 million years ago, a mysterious event that is unlike anything in terrestrial history.

Volcanoes on Venus

Like Earth, Venus is a planet that has experienced widespread volcanism. In the lowland plains, volcanic eruptions are the principal way the surface is renewed, with large flows of highly fluid lava destroying old craters and generating a fresh surface. In addition, numerous younger volcanic mountains and other structures are associated with surface hot spots—places where convection in the planet's mantle transports the interior heat to the surface.

The largest individual volcano on Venus, called Sif Mons, is about 500 kilometers across and 3 kilometers high—broader but lower than the Hawaiian volcano Mauna Loa. At its top is a volcanic crater, or *caldera*, about 40 kilometers across, and its slopes show individual lava flows up to 500 kilometers long. Thousands of smaller volcanoes dot the surface, down to the limit of visibility of the *Magellan* images, which correspond to cones or domes about the size of a shopping mall parking lot. Most of these seem similar to terrestrial volcanoes. Other volcanoes have unusual shapes, such as the "pancake domes" illustrated in Figure 2.6.2.3.

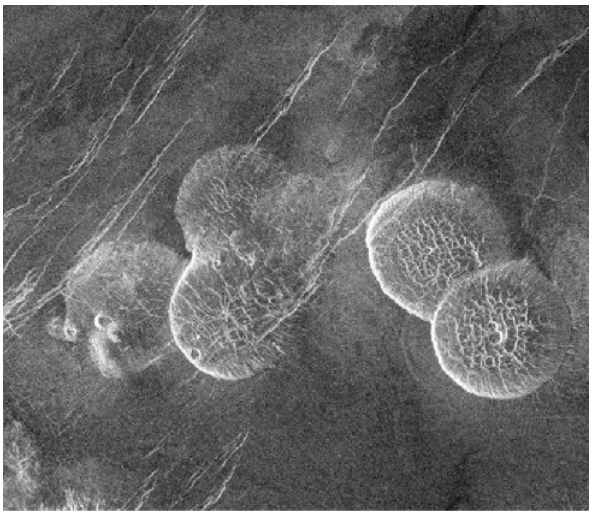


Figure 2.6.2.3 Pancake-Shaped Volcanoes on Venus. These remarkable circular domes, each about 25 kilometers across and about 2 kilometers tall, are the result of eruptions of highly viscous (sludgy) lava that spreads out evenly in all directions.

All of the volcanism is the result of eruption of lava onto the surface of the planet. But the hot lava rising from the interior of a planet does not always make it to the surface. On both Earth and Venus, this upwelling lava can collect to produce bulges in the crust. Many of the granite mountain ranges on Earth, such as the Sierra Nevada in California, involve such subsurface volcanism. These bulges are common on Venus, where they produce large circular or oval features called *coronae* (singular: corona) (Figure 2.6.2.4).



Figure 2.6.2.4 The “Miss Piggy” Corona. Fotla Corona is located in the plains to the south of Aphrodite Terra. Curved fracture patterns show where the material beneath has put stress on the surface. A number of pancake and dome volcanoes are also visible. Fotla was a Celtic fertility goddess. Some students see a resemblance between this corona and Miss Piggy of the Muppets (her left ear, at the top of the picture, is the pancake volcano in the upper center of the image).

Tectonic Activity

Convection currents of molten material in the mantle of Venus push and stretch the crust. Such forces are called tectonic, and the geological features that result from these forces are called *tectonic features*. On Venus’ lowland plains, tectonic forces have broken the lava surface to create remarkable patterns of ridges and cracks (Figure 2.6.2.5). In a few places, the crust has even torn apart to generate rift valleys. The circular features associated with coronae are tectonic ridges and cracks, and most of the mountains of Venus also owe their existence to tectonic forces.

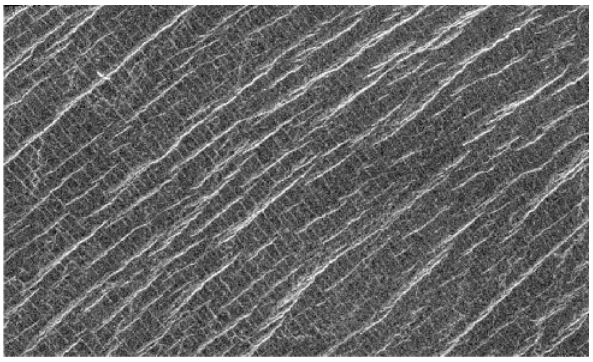


Figure 2.6.2.5 Ridges and Cracks. This region of the Lakshmi Plains on Venus has been fractured by tectonic forces to produce a cross-hatched grid of cracks and ridges. Be sure to notice the fainter linear features that run perpendicular to the brighter ones. As this is a radar image, the brightness of the ridges indicates their relative height. This image shows a region about 80 kilometers wide and 37 kilometers high. Lakshmi is a Hindu goddess of prosperity.

The Ishtar continent, which has the highest elevations on Venus, is the most dramatic product of these tectonic forces. Ishtar and its tall Maxwell Mountains resemble the Tibetan Plateau and Himalayan Mountains on Earth. Both are the product of compression of the crust, and both are maintained by the continuing forces of mantle convection.

On Venus' Surface

The successful Venera landers of the 1970s found themselves on an extraordinarily inhospitable planet, with a surface pressure of 90 bars and a temperature hot enough to melt lead and zinc. Despite these unpleasant conditions, the spacecraft were able to photograph their surroundings and collect surface samples for chemical analysis before their instruments gave out. The diffuse sunlight striking the surface was tinted red by the clouds, and the illumination level was equivalent to a heavy overcast on Earth.

The probes found that the rock in the landing areas is igneous, primarily basalts. Examples of the Venera photographs are shown in Figure. Each picture shows a flat, desolate landscape with a variety of rocks, some of which may be ejecta from impacts. Other areas show flat, layered lava flows. There have been no further landings on Venus since the 1970s.

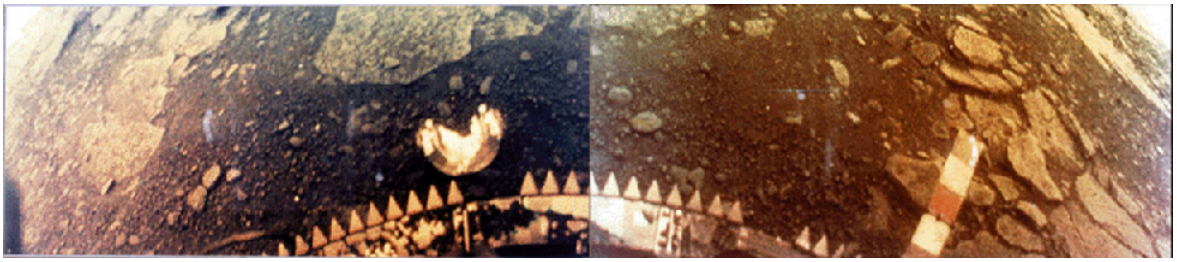


Figure 2.6.2.6 Surface of Venus. These views of the surface of Venus are from the Venera 13 spacecraft. Everything is orange because the thick atmosphere of Venus absorbs the bluer colors of light. The horizon is visible in the upper corner of each image. (credit: NASA)

Key Concepts and Summary

Venus has been mapped by radar, especially with the *Magellan* spacecraft. Its crust consists of 75% lowland lava plains, numerous volcanic features, and many large coronae, which are the expression of subsurface volcanism. The planet has been modified by widespread tectonics driven by mantle convection, forming complex patterns of ridges and cracks and building high continental regions such as Ishtar. The surface is extraordinarily inhospitable, with pressure of 90 bars and temperature of 730 K, but several Russian Venera landers investigated it successfully.

Glossary

tectonic

geological features that result from stresses and pressures in the crust of a planet; tectonic forces can lead to earthquakes and motion of the crust

This page titled 2.6.2: The Geology of Venus is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **10.2: The Geology of Venus** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.6.3: The Massive Atmosphere of Venus

Learning Objectives

By the end of this section, you will be able to:

- Describe the general composition and structure of the atmosphere on Venus
- Explain how the greenhouse effect has led to high temperatures on Venus

The thick atmosphere of Venus produces the high surface temperature and shrouds the surface in a perpetual red twilight. Sunlight does not penetrate directly through the heavy clouds, but the surface is fairly well lit by diffuse light (about the same as the light on Earth under a heavy overcast). The weather at the bottom of this deep atmosphere remains perpetually hot and dry, with calm winds. Because of the heavy blanket of clouds and atmosphere, one spot on the surface of Venus is similar to any other as far as weather is concerned.

Composition and Structure of the Atmosphere

The most abundant gas on Venus is carbon dioxide (CO_2), which accounts for 96% of the atmosphere. The second most common gas is nitrogen. The predominance of carbon dioxide over nitrogen is not surprising when you recall that Earth's atmosphere would also be mostly carbon dioxide if this gas were not locked up in marine sediments (see the discussion of Earth's atmosphere in Earth as a Planet).

Table 2.6.3.1 compares the compositions of the atmospheres of Venus, Mars, and Earth. Expressed in this way, as percentages, the proportions of the major gases are very similar for Venus and Mars, but in total quantity, their atmospheres are dramatically different. With its surface pressure of 90 bars, the venusian atmosphere is more than 10,000 times more massive than its martian counterpart. Overall, the atmosphere of Venus is very dry; the absence of water is one of the important ways that Venus differs from Earth.

Table 2.6.3.1: Atmospheric Composition of Earth, Venus, and Mars

Gas	Earth	Venus	Mars
Carbon dioxide (CO_2)	0.03%	96%	95.3%
Nitrogen (N_2)	78.1%	3.5%	2.7%
Argon (Ar)	0.93%	0.006%	1.6%
Oxygen (O_2)	21.0%	0.003%	0.15%
Neon (Ne)	0.002%	0.001%	0.0003%

The atmosphere of Venus has a huge troposphere (region of convection) that extends up to at least 50 kilometers above the surface (Figure 2.6.3.1). Within the troposphere, the gas is heated from below and circulates slowly, rising near the equator and descending over the poles. Being at the base of the atmosphere of Venus is something like being a kilometer or more below the ocean surface on Earth. There, the mass of water evens out temperature variations and results in a uniform environment—the same effect the thick atmosphere has on Venus.

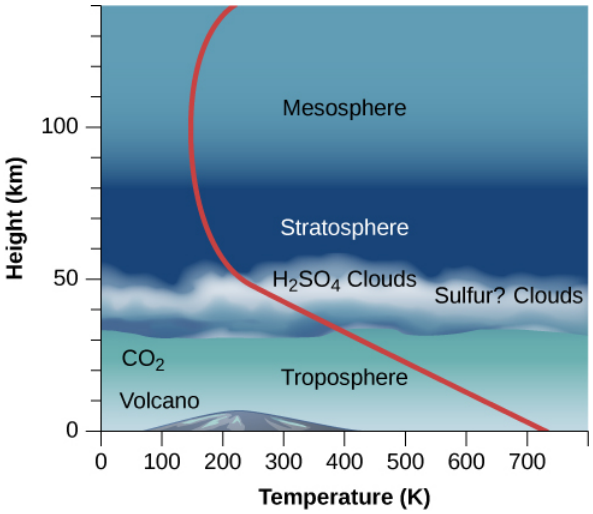


Figure 2.6.3.1 Venus' Atmosphere. The layers of the massive atmosphere of Venus shown here are based on data from the Pioneer and Venera entry probes. Height is measured along the left axis, the bottom scale shows temperature, and the red line allows you to read off the temperature at each height. Notice how steeply the temperature rises below the clouds, thanks to the planet's huge greenhouse effect.

In the upper troposphere, between 30 and 60 kilometers above the surface, a thick cloud layer is composed primarily of sulfuric acid droplets. Sulfuric acid (H_2SO_4) is formed from the chemical combination of sulfur dioxide (SO_2) and water (H_2O). In the atmosphere of Earth, sulfur dioxide is one of the primary gases emitted by volcanoes, but it is quickly diluted and washed out by rainfall. In the dry atmosphere of Venus, this unpleasant substance is apparently stable. Below 30 kilometers, the Venus atmosphere is clear of clouds.

Surface Temperature on Venus

The high surface temperature of Venus was discovered by radio astronomers in the late 1950s and confirmed by the Mariner and Venera probes. How can our neighbor planet be so hot? Although Venus is somewhat closer to the Sun than is Earth, its surface is hundreds of degrees hotter than you would expect from the extra sunlight it receives. Scientists wondered what could be heating the surface of Venus to a temperature above 700 K. The answer turned out to be the *greenhouse effect*.

The greenhouse effect works on Venus just as it does on Earth, but since Venus has so much more CO_2—almost a million times more—the effect is much stronger. The thick CO_2 acts as a blanket, making it very difficult for the infrared (heat) radiation from the ground to get back into space. As a result, the surface heats up. The energy balance is only restored when the planet is radiating as much energy as it receives from the Sun, but this can happen only when the temperature of the lower atmosphere is very high. One way of thinking of greenhouse heating is that it must raise the surface temperature of Venus until this energy balance is achieved.

Has Venus always had such a massive atmosphere and high surface temperature, or might it have evolved to such conditions from a climate that was once more nearly earthlike? The answer to this question is of particular interest to us as we look at the increasing levels of CO_2 in Earth's atmosphere. As the greenhouse effect becomes stronger on Earth, are we in any danger of transforming our own planet into a hellish place like Venus?

Let us try to reconstruct the possible evolution of Venus from an earthlike beginning to its present state. Venus may once have had a climate similar to that of Earth, with moderate temperatures, water oceans, and much of its CO_2 dissolved in the ocean or chemically combined with the surface rocks. Then we allow for modest additional heating—by gradual increase in the energy output of the Sun, for example. When we calculate how Venus' atmosphere would respond to such effects, it turns out that even a small amount of extra heat can lead to increased evaporation of water from the oceans and the release of gas from surface rocks.

This in turn means a further increase in the atmospheric CO_2 and H_2O, gases that would amplify the greenhouse effect in Venus' atmosphere. That would lead to still more heat near Venus' surface and the release of further CO_2 and H_2O. Unless some other processes intervene, the temperature thus continues to rise. Such a situation is called the **runaway greenhouse effect**.

We want to emphasize that the runaway greenhouse effect is not just a large greenhouse effect; it is an evolutionary *process*. The atmosphere evolves from having a small greenhouse effect, such as on Earth, to a situation where greenhouse warming is a major

factor, as we see today on Venus. Once the large greenhouse conditions develop, the planet establishes a new, much hotter equilibrium near its surface.

Reversing the situation is difficult because of the role water plays. On Earth, most of the CO_2 is either chemically bound in the rocks of our crust or dissolved by the water in our oceans. As Venus got hotter and hotter, its oceans evaporated, eliminating that safety valve. But the water vapor in the planet's atmosphere will not last forever in the presence of ultraviolet light from the Sun. The light element hydrogen can escape from the atmosphere, leaving the oxygen behind to combine chemically with surface rock. The loss of water is therefore an irreversible process: once the water is gone, it cannot be restored. There is evidence that this is just what happened to the water once present on Venus.

We don't know if the same runaway greenhouse effect could one day happen on Earth. Although we are uncertain about the point at which a stable greenhouse effect breaks down and turns into a runaway greenhouse effect, Venus stands as clear testament to the fact that a planet cannot continue heating indefinitely without a major change in its oceans and atmosphere. It is a conclusion that we and our descendants will surely want to pay close attention to.

Summary

The atmosphere of Venus is 96% CO_2. Thick clouds at altitudes of 30 to 60 kilometers are made of sulfuric acid, and a CO_2, which is greenhouse effect maintains the high surface temperature. Venus presumably reached its current state from more earthlike initial conditions as a result of a runaway greenhouse effect, which included the loss of large quantities of water.

Glossary

runaway greenhouse effect

the process by which the greenhouse effect, rather than remaining stable or being lessened through intervention, continues to grow at an increasing rate

---

This page titled 2.6.3: The Massive Atmosphere of Venus is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **10.3: The Massive Atmosphere of Venus** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.6.4: The Geology of Mars

Learning Objectives

- Discuss the main missions that have explored Mars
- Explain what we have learned from examination of meteorites from Mars
- Describe the various features found on the surface of Mars
- Compare the volcanoes and canyons on Mars with those of Earth
- Describe the general conditions on the surface of Mars

Mars is more interesting to most people than Venus because it is more hospitable. Even from the distance of Earth, we can see surface features on Mars and follow the seasonal changes in its polar caps (Figure 2.6.4.1). Although the surface today is dry and cold, evidence collected by spacecraft suggests that Mars once had blue skies and lakes of liquid water. Even today, it is the sort of place we can imagine astronauts visiting and perhaps even setting up permanent bases.



Figure 2.6.4.1 Mars Photographed by the Hubble Space Telescope. This is one of the best photos of Mars taken from our planet, obtained in June 2001 when Mars was only 68 million kilometers away. The resolution is about 20 kilometers—much better than can be obtained with ground-based telescopes but still insufficient to reveal the underlying geology of Mars. (credit: modification of work by NASA and the Hubble Heritage Team (STScI/AURA))

Spacecraft Exploration of Mars

Mars has been intensively investigated by spacecraft. More than 50 spacecraft have been launched toward Mars, but only about half were fully successful. The first visitor was the US Mariner 4, which flew past Mars in 1965 and transmitted 22 photos to Earth. These pictures showed an apparently bleak planet with abundant impact craters. In those days, craters were unexpected; some people who were romantically inclined still hoped to see canals or something like them. In any case, newspaper headlines sadly announced that Mars was a “dead planet.”

In 1971, NASA’s Mariner 9 became the first spacecraft to orbit another planet, mapping the entire surface of Mars at a resolution of about 1 kilometer and discovering a great variety of geological features, including volcanoes, huge canyons, intricate layers on the polar caps, and channels that appeared to have been cut by running water. Geologically, Mars didn’t look so dead after all.

The twin Viking spacecraft of the 1970s were among the most ambitious and successful of all planetary missions. Two *orbiters*surveyed the planet and served to relay communications for two *landers* on the surface. After an exciting and sometimes frustrating search for a safe landing spot, the Viking 1 lander touched down on the surface of Chryse Planitia (the Plains of Gold) on July 20, 1976, exactly 7 years after Neil Armstrong’s historic first step on the Moon. Two months later, Viking 2 landed with

equal success in another plain farther north, called Utopia. The landers photographed the surface with high resolution and carried out complex experiments searching for evidence of life, while the orbiters provided a global perspective on Mars geology.

Mars languished unvisited for two decades after Viking. Two more spacecraft were launched toward Mars, by NASA and the Russian Space Agency, but both failed before reaching the planet.

The situation changed in the 1990s as NASA began a new exploration program using spacecraft that were smaller and less expensive than Viking. The first of the new missions, appropriately called Pathfinder, landed the first wheeled, solar-powered rover on the martian surface on July 4, 1997 (Figure 2.6.4.2). An orbiter called *Mars Global Surveyor* (MGS) arrived a few months later and began high-resolution photography of the entire surface over more than one martian year. The most dramatic discovery by this spacecraft, which is still operating, was evidence of gullies apparently cut by surface water, as we will discuss later. These missions were followed in 2003 by the NASA *Mars Odyssey* orbiter, and the ESA *Mars Express* orbiter, both carrying high-resolution cameras. A gamma-ray spectrometer on *Odyssey* discovered a large amount of subsurface hydrogen (probably in the form of frozen water). Subsequent orbiters included the NASA *Mars Reconnaissance Orbiter* to evaluate future landing sites, MAVEN to study the upper atmosphere, and India's *Mangalayaan*, also focused on study of Mars' thin layers of air. Several of these orbiters are also equipped to communicate with landers and rovers on the surface and serve as data relays to Earth.

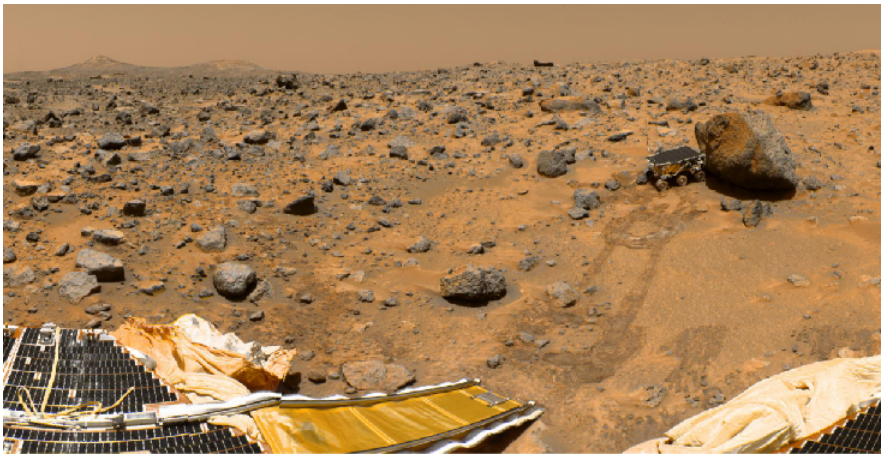
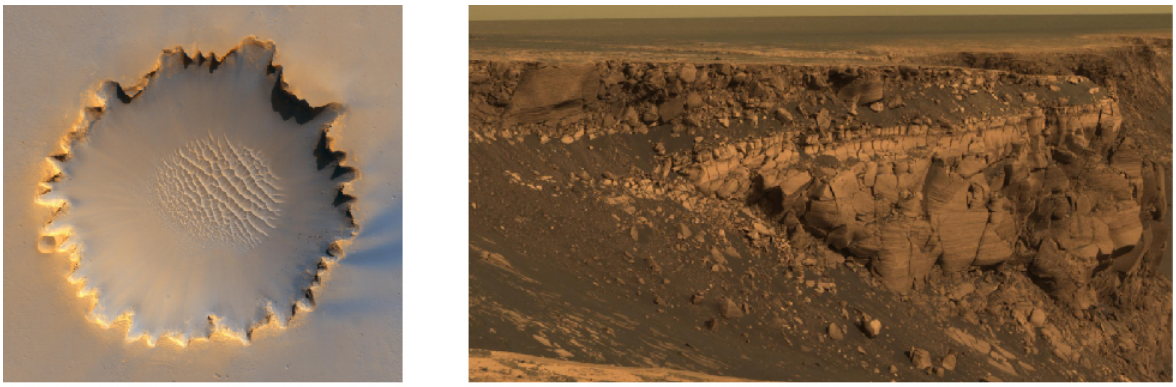


Figure 2.6.4.2 Surface View from Mars Pathfinder. The scene from the Pathfinder lander shows a windswept plain, sculpted long ago when water flowed out of the martian highlands and into the depression where the spacecraft landed. The Sojourner rover, the first wheeled vehicle on Mars, is about the size of a microwave oven. Its flat top contains solar cells that provided electricity to run the vehicle. You can see the ramp from the lander and the path the rover took to the larger rock that the mission team nicknamed "Yogi." (credit: NASA/JPL)

In 2003, NASA began a series of highly successful Mars landers. Twin Mars Exploration Rovers (MER), named *Spirit* and *Opportunity*, have been successful far beyond their planned lifetimes. The design goal for the rovers was 600 meters of travel; in fact, they have traveled jointly more than 50 kilometers. After scouting around its rim, *Opportunity* drove down the steep walls into an impact crater called Victoria, then succeeded with some difficulty in climbing back out to resume its route (Figure 2.6.4.3). Dust covering the rovers' solar cells caused a drop in power, but when a seasonal dust storm blew away the dust, the rovers resumed full operation. In order to survive winter, the rovers were positioned on slopes to maximize solar heating and power generation. In 2006, *Spirit* lost power on one of its wheels, and subsequently became stuck in the sand, where it continued operation as a fixed ground station. Meanwhile, in 2008, *Phoenix* (a spacecraft "reborn" of spare parts from a previous Mars mission that had failed) landed near the edge of the north polar cap, at latitude 68°, and directly measured water ice in the soil.



(a) (b)

Figure 2.6.4.3 Victoria Crater. (a) This crater in Meridiani Planum is 800 meters wide, making it slightly smaller than Meteor crater on Earth. Note the dune field in the interior. (b) This image shows the view from the Opportunity rover as it scouted the rim of Victoria crater looking for a safe route down into the interior.

In 2011, NASA launched its largest (and most expensive) Mars mission since Viking. The 1-ton rover *Curiosity*, the size of a subcompact car, has plutonium-powered electrical generators, so that it is not dependent on sunlight for power. *Curiosity* made a pinpoint landing on the floor of Gale crater, a site selected for its complex geology and evidence that it had been submerged by water in the past. Previously, Mars landers had been sent to flat terrains with few hazards, as required by their lower targeting accuracy. The scientific goals of *Curiosity* include investigations of climate and geology, and assessment of the habitability of past and present Mars environments. In 2018, NASA's InSight Lander touched down on Mars, carrying a suite of scientific instruments. These include a package (nicknamed "the mole") that will dig into the surface of Mars 1 mm at a time, hoping to reach a depth of 5 meters with heat sensors. Neither of these missions carries a specific life detection instrument, however. So far, scientists have not been able to devise a simple instrument that could distinguish living from nonliving materials on Mars.

> The *Curiosity* rover required a remarkably complex landing sequence and NASA made a video about it called "7 Minutes of Terror" that went viral on the Internet.
>
> A dramatic video summary of the first two years of *Curiosity*'s exploration of the martian surface can be viewed as well.

Martian Samples

Much of what we know of the Moon, including the circumstances of its origin, comes from studies of lunar samples, but spacecraft have not yet returned martian samples to Earth for laboratory analysis. It is with great interest, therefore, that scientists have discovered that samples of martian material are nevertheless already here on Earth, available for study. These are all members of a rare class of *meteorites* (Figure 2.6.4.4)—rocks that have fallen from space.



Figure 2.6.4.4 Martian Meteorite. This fragment of basalt, ejected from Mars in a crater-forming impact, eventually arrived on Earth's surface.

How would rocks have escaped from Mars? Many impacts have occurred on the red planet, as shown by its heavily cratered surface. Fragments blasted from large impacts can escape from Mars, whose surface gravity is only 38% of Earth's. A long time later (typically a few million years), a very small fraction of these fragments collide with Earth and survive their passage through our atmosphere, just like other meteorites. (We'll discuss meteorites in more detail in the chapter on Cosmic Samples and the

Origin of the Solar System.) By the way, rocks from the Moon have also reached our planet as meteorites, although we were able to demonstrate their lunar origin only by comparison with samples returned by the Apollo missions

Most of the martian meteorites are volcanic basalts; most of them are also relatively young—about 1.3 billion years old. We know from details of their composition that they are not from Earth or the Moon. Besides, there was no volcanic activity on the Moon to form them as recently as 1.3 billon years ago. It would be very difficult for ejecta from impacts on Venus to escape through its thick atmosphere. By the process of elimination, the only reasonable origin seems to be Mars, where the Tharsis volcanoes were active at that time.

The martian origin of these meteorites was confirmed by the analysis of tiny gas bubbles trapped inside several of them. These bubbles match the atmospheric properties of Mars as first measured directly by Viking. It appears that some atmospheric gas was trapped in the rock by the shock of the impact that ejected it from Mars and started it on its way toward Earth.

One of the most exciting results from analysis of these martian samples has been the discovery of both water and organic (carbon-based) compounds in them, which suggests that Mars may once have had oceans and perhaps even life on its surface. As we have already hinted, there is other evidence for the presence of flowing water on Mars in the remote past, and even extending to the present.

In this and the following sections, we will summarize the picture of Mars as revealed by all these exploratory missions and by about 40 samples from Mars.

Global Properties of Mars

Mars has a diameter of 6790 kilometers, just over half the diameter of Earth, giving it a total surface area very nearly equal to the continental (land) area of our planet. Its overall density of 3.9 g/cm^3 suggests a composition consisting primarily of silicates but with a small metal core. The planet has no global magnetic field, although there are areas of strong surface magnetization that indicate that there was a global field billions of years ago. Apparently, the red planet has no liquid material in its core today that would conduct electricity.

Thanks to the *Mars Global Surveyor*, we have mapped the entire planet, as shown in Figure 2.6.4.5. A laser altimeter on board made millions of separate measurements of the surface topography to a precision of a few meters—good enough to show even the annual deposition and evaporation of the polar caps. Like Earth, the Moon, and Venus, the surface of Mars has continental or highland areas as well as widespread volcanic plains. The total range in elevation from the top of the highest mountain (Olympus Mons) to the bottom of the deepest basin (Hellas) is 31 kilometers.

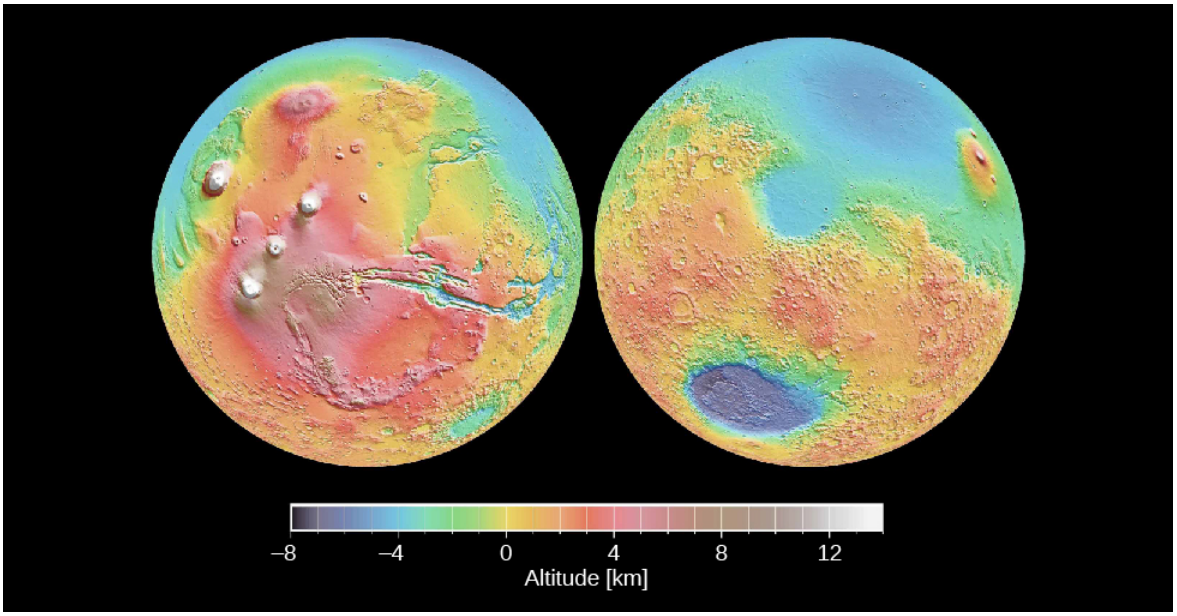


Figure 2.6.4.5 Mars Map from Laser Ranging. These globes are highly precise topographic maps, reconstructed from millions of individual elevation measurements made with the Mars Global Surveyor. Color is used to indicate elevation. The hemisphere on the left

includes the Tharsis bulge and Olympus Mons, the highest mountain on Mars; the hemisphere on the right includes the Hellas basin, which has the lowest elevation on Mars.

Approximately half the planet consists of heavily cratered highland terrain, found primarily in the southern hemisphere. The other half, which is mostly in the north, contains younger, lightly cratered volcanic plains at an average elevation about 5 kilometers lower than the highlands. Remember that we saw a similar pattern on Earth, the Moon, and Venus. A geological division into older highlands and younger lowland plains seems to be characteristic of all the terrestrial planets except Mercury.

Lying across the north-south division of Mars is an uplifted continent the size of North America. This is the 10-kilometer-high Tharsis bulge, a volcanic region crowned by four great volcanoes that rise still higher into the martian sky.

Volcanoes on Mars

The lowland plains of Mars look very much like the lunar maria, and they have about the same density of impact craters. Like the lunar maria, they probably formed between 3 and 4 billion years ago. Apparently, Mars experienced extensive volcanic activity at about the same time the Moon did, producing similar basaltic lavas.

The largest volcanic mountains of Mars are found in the Tharsis area (you can see them in Figure 2.6.4.5), although smaller volcanoes dot much of the surface. The most dramatic volcano on Mars is Olympus Mons (Mount Olympus), with a diameter larger than 500 kilometers and a summit that towers more than 20 kilometers above the surrounding plains—three times higher than the tallest mountain on Earth (Figure 2.6.4.6). The volume of this immense volcano is nearly 100 times greater than that of Mauna Loa in Hawaii. Placed on Earth's surface, Olympus would more than cover the entire state of Missouri.



Figure 2.6.4.6 Olympus Mons. The largest volcano on Mars, and probably the largest in the solar system, is Olympus Mons, illustrated in this computer-generated rendering based on data from the Mars Global Surveyor's laser altimeter. Placed on Earth, the base of Olympus Mons would completely cover the state of Missouri; the caldera, the circular opening at the top, is 65 kilometers across, about the size of Los Angeles.

Images taken from orbit allow scientists to search for impact craters on the slopes of these volcanoes in order to estimate their age. Many of the volcanoes show a fair number of such craters, suggesting that they ceased activity a billion years or more ago. However, Olympus Mons has very, very few impact craters. Its present surface cannot be more than about 100 million years old; it may even be much younger. Some of the fresh-looking lava flows might have been formed a hundred years ago, or a thousand, or a million, but geologically speaking, they are quite young. This leads geologists to the conclusion that Olympus Mons possibly remains intermittently active today—something future Mars land developers may want to keep in mind.

Martian Cracks and Canyons

The Tharsis bulge has many interesting geological features in addition to its huge volcanoes. In this part of the planet, the surface itself has bulged upward, forced by great pressures from below, resulting in extensive tectonic cracking of the crust. Among the most spectacular tectonic features on Mars are the canyons called the Valles Marineris (or Mariner Valleys, named after Mariner 9, which first revealed them to us), which are shown in Figure 2.6.4.7. They extend for about 5000 kilometers (nearly a quarter of the way around Mars) along the slopes of the Tharsis bulge. If it were on Earth, this canyon system would stretch all the way from Los Angeles to Washington, DC. The main canyon is about 7 kilometers deep and up to 100 kilometers wide, large enough for the Grand Canyon of the Colorado River to fit comfortably into one of its side canyons.

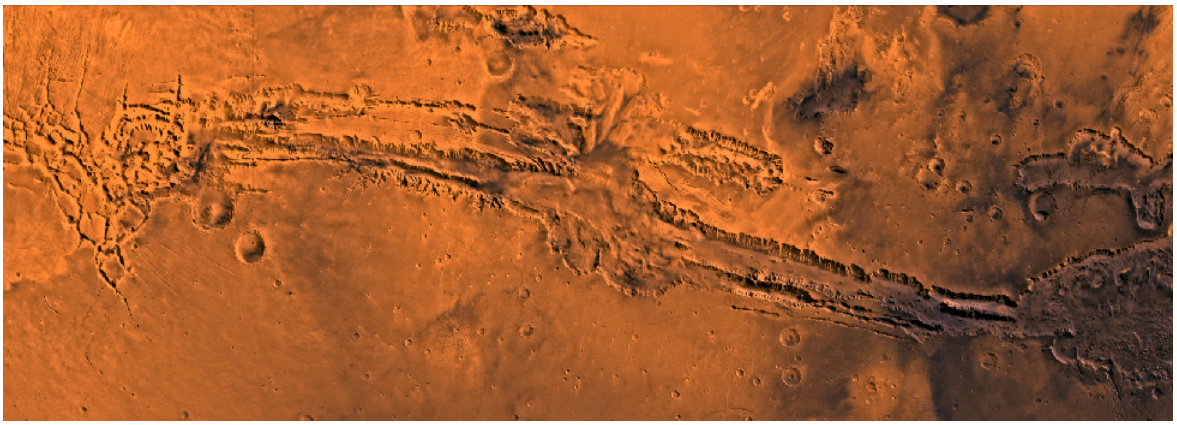


Figure 2.6.4.7 Heavily Eroded Canyonlands on Mars. This image shows the Valles Marineris canyon complex, which is 3000 kilometers wide and 8 kilometers deep.

> An excellent 4-minute video tour of Valles Marineris, narrated by planetary scientist Phil Christensen, is available for viewing.

The term “canyon” is somewhat misleading here because the Valles Marineris canyons have no outlets and were not cut by running water. They are basically tectonic cracks, produced by the same crustal tensions that caused the Tharsis uplift. However, water has played a later role in shaping the canyons, primarily by seeping from deep springs and undercutting the cliffs. This undercutting led to landslides that gradually widened the original cracks into the great valleys we see today (Figure 2.6.4.8). Today, the primary form of erosion in the canyons is probably wind.

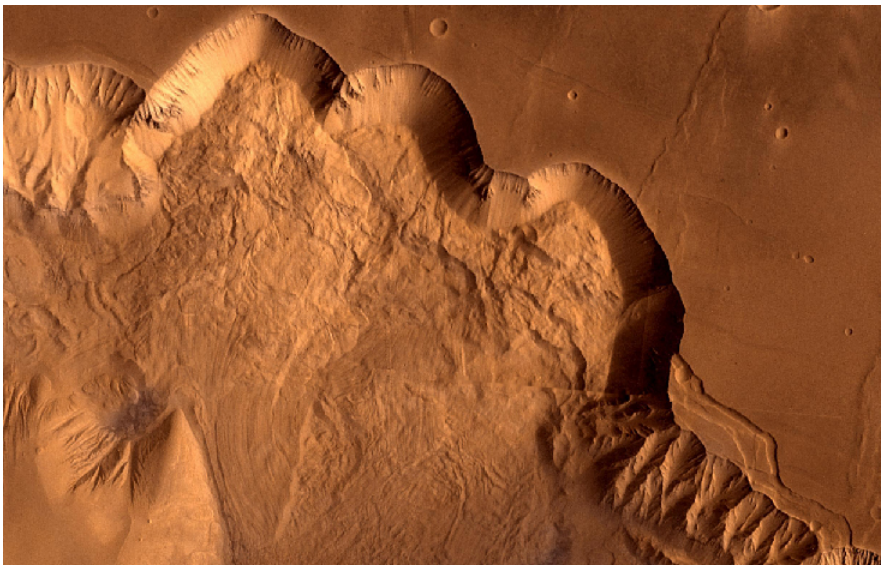


Figure 2.6.4.8 Martian Landslides. This Viking orbiter image shows Ophir Chasma, one of the connected valleys of the Valles Marineris canyon system. Look carefully and you can see enormous landslides whose debris is piled up underneath the cliff wall, which tower up to 10 kilometers above the canyon floor.

While the Tharsis bulge and Valles Marineris are impressive, in general, we see fewer tectonic structures on Mars than on Venus. In part, this may reflect a lower general level of geological activity, as would be expected for a smaller planet. But it is also possible that evidence of widespread faulting has been buried by wind-deposited sediment over much of Mars. Like Earth, Mars may have hidden part of its geological history under a cloak of soil.

The View on the Martian Surface

The first spacecraft to land successfully on Mars were Vikings 1 and 2 and Mars Pathfinder. All sent back photos that showed a desolate but strangely beautiful landscape, including numerous angular rocks interspersed with dune like deposits of fine-grained,

reddish soil (Figure 2.6.4.9).



Figure 2.6.4.9 Three Martian Landing Sites. The Mars landers Viking 1 in Chryse, Pathfinder in Ares Valley, and Viking 2 in Utopia, all photographed their immediate surroundings. It is apparent from the similarity of these three photos that each spacecraft touched down on a flat, windswept plain littered with rocks ranging from tiny pebbles up to meter-size boulders. It is probable that most of Mars looks like this on the surface.

All three of these landers were targeted to relatively flat, lowland terrain. Instruments on the landers found that the soil consisted of clays and iron oxides, as had long been expected from the red color of the planet. All the rocks measured appeared to be of volcanic origin and roughly the same composition. Later landers were targeted to touch down in areas that apparently were flooded sometime in the past, where sedimentary rock layers, formed in the presence of water, are common. (Although we should note that nearly all the planet is blanketed in at least a thin layer of wind-blown dust).

The Viking landers included weather stations that operated for several years, providing a perspective on martian weather. The temperatures they measured varied greatly with the seasons, due to the absence of moderating oceans and clouds. Typically, the summer maximum at Viking 1 was 240 K (–33 °C), dropping to 190 K (–83 °C) at the same location just before dawn. The lowest air temperatures, measured farther north by Viking 2, were about 173 K (–100 °C). During the winter, Viking 2 also photographed water frost deposits on the ground (Figure 2.6.4.10). We make a point of saying “water frost” here because at some locations on Mars, it gets cold enough for carbon dioxide (dry ice) to freeze out of the atmosphere as well.

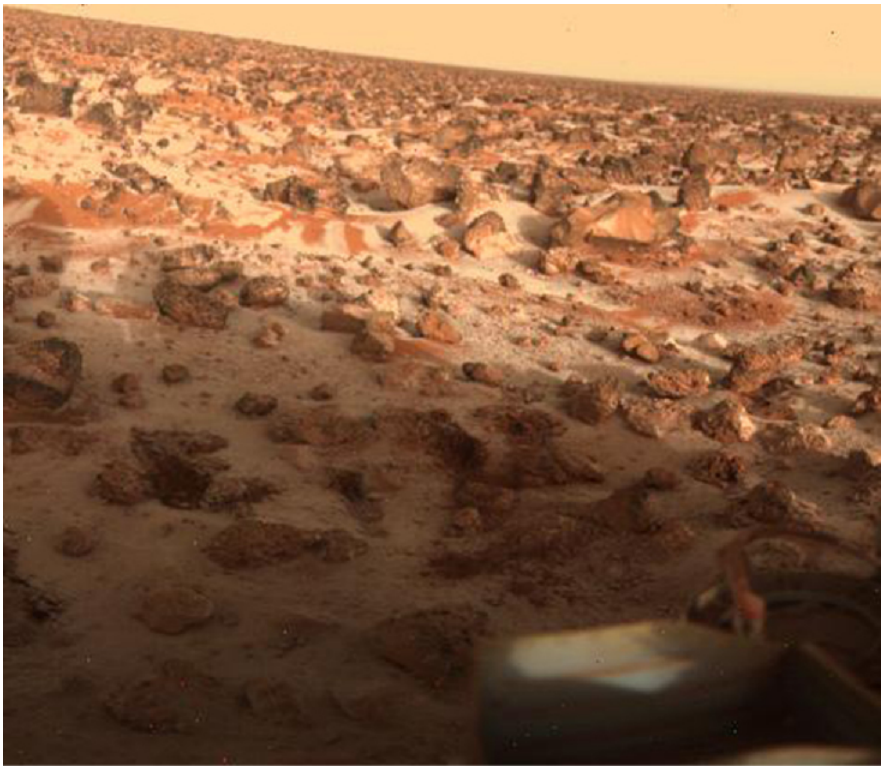
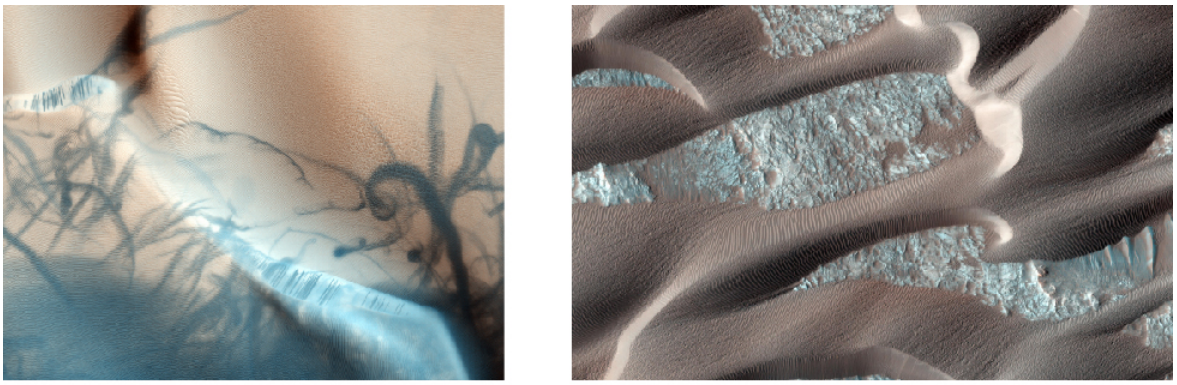


Figure 2.6.4.10 Water Frost in Utopia. This image of surface frost was photographed at the Viking 2 landing site during late winter.

Most of the winds measured on Mars are only a few kilometers per hour. However, Mars is capable of great windstorms that can shroud the entire planet with windblown dust. Such high winds can strip the surface of some of its loose, fine dust, leaving the rock exposed. The later rovers found that each sunny afternoon the atmosphere became turbulent as heat rose off the surface. This turbulence generated dust devils, which play an important role in lifting the fine dust into the atmosphere. As the dust devils strip off the top layer of light dust and expose darker material underneath, they can produce fantastic patterns on the ground (Figure 2.6.4.11).

Wind on Mars plays an important role in redistributing surface material. Figure 2.6.4.11 shows a beautiful area of dark sand dunes on top of lighter material. Much of the material stripped out of the martian canyons has been dumped in extensive dune fields like this, mostly at high latitudes.



(a) (b)

Figure 2.6.4.11
Dust Devil Tracks and Sand Dunes. (a) This high-resolution photo from the Mars Global Surveyor shows the dark tracks of several dust devils that have stripped away a thin coating of light-colored dust. This view is of an area about 3 kilometers across. Dust devils are one of the most important ways that dust gets redistributed by the martian winds. They may also help keep the solar panels of our rovers free of dust. (b) These windblown sand dunes on Mars overlay a lighter sandy surface. Each dune in this high-resolution view is about 1 kilometer across.

Key Concepts and Summary

Most of what we know about Mars is derived from spacecraft: highly successful orbiters, landers, and rovers. We have also been able to study a few martian rocks that reached Earth as meteorites. Mars has heavily cratered highlands in its southern hemisphere, but younger, lower volcanic plains over much of its northern half. The Tharsis bulge, as big as North America, includes several huge volcanoes; Olympus Mons is more than 20 kilometers high and 500 kilometers in diameter. The Valles Marineris canyons are tectonic features widened by erosion. Early landers revealed only barren, windswept plains, but later missions have visited places with more geological (and scenic) variety. Landing sites have been selected in part to search for evidence of past water.

This page titled 2.6.4: The Geology of Mars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **10.4: The Geology of Mars** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.6.5: Water and Life on Mars

Learning Objectives

By the end of this section, you will be able to:

- Describe the general composition of the atmosphere on Mars
- Explain what we know about the polar ice caps on Mars and how we know it
- Describe the evidence for the presence of water in the past history of Mars
- Summarize the evidence for and against the possibility of life on Mars

Of all the planets and moons in the solar system, Mars seems to be the most promising place to look for life, both fossil microbes and (we hope) some forms of life deeper underground that still survive today. But where (and how) should we look for life? We know that the one requirement shared by all life on Earth is liquid water. Therefore, the guiding principle in assessing habitability on Mars and elsewhere has been to "follow the water." That is the perspective we take in this section, to follow the water on the red planet and hope it will lead us to life.

Atmosphere and Clouds on Mars

The atmosphere of Mars today has an average surface pressure of only 0.007 bar, less than 1% that of Earth. (This is how thin the air is about 30 kilometers above Earth's surface.) Martian air is composed primarily of carbon dioxide (95%), with about 3% nitrogen and 2% argon. The proportions of different gases are similar to those in the atmosphere of Venus (see Table 10.3.1), but a lot less of each gas is found in the thin air on Mars.

While winds on Mars can reach high speeds, they exert much less force than wind of the same velocity would on Earth because the atmosphere is so thin. The wind is able, however, to loft very fine dust particles, which can sometimes develop planet-wide dust storms. It is this fine dust that coats almost all the surface, giving Mars its distinctive red color. In the absence of surface water, wind erosion plays a major role in sculpting the martian surface (Figure 2.6.5.1).

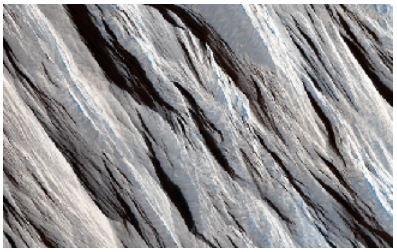


Figure 2.6.5.1 Wind Erosion on Mars. These long straight ridges, called yardangs, are aligned with the dominant wind direction. This is a high-resolution image from the Mars Reconnaissance Orbiter and is about 1 kilometer wide.

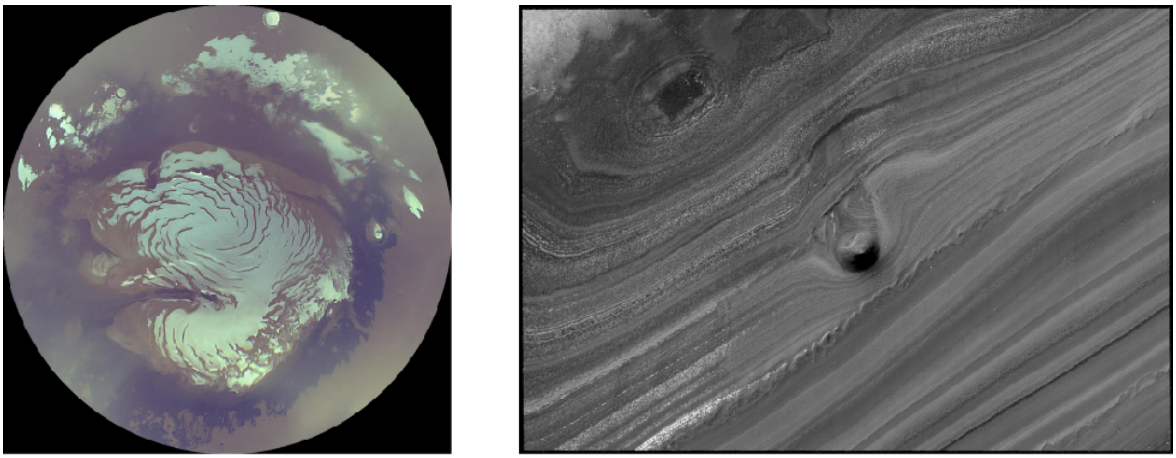
> The issue of how strong the winds on Mars can be plays a big role in the 2015 hit movie *The Martian* in which the main character is stranded on Mars after being buried in the sand in a windstorm so great that his fellow astronauts have to leave the planet so their ship is not damaged. Astronomers have noted that the martian winds could not possibly be as forceful as depicted in the film. In most ways, however, the depiction of Mars in this movie is remarkably accurate.

Although the atmosphere contains small amounts of water vapor and occasional clouds of water ice, liquid water is not stable under present conditions on Mars. Part of the problem is the low temperatures on the planet. But even if the temperature on a sunny summer day rises above the freezing point, the low pressure means that liquid water still cannot exist on the surface, except at the lowest elevations. At a pressure of less than 0.006 bar, the boiling point is as low or lower than the freezing point, and water changes directly from solid to vapor without an intermediate liquid state (as does "dry ice," carbon dioxide, on Earth). However, salts dissolved in water lower its freezing point, as we know from the way salt is used to thaw roads after snow and ice forms during winter on Earth. Salty water is therefore sometimes able to exist in liquid form on the martian surface, under the right conditions.

Several types of clouds can form in the martian atmosphere. First there are dust clouds, discussed above. Second are water-ice clouds similar to those on Earth. These often form around mountains, just as happens on our planet. Finally, the CO2 of the atmosphere can itself condense at high altitudes to form hazes of dry ice crystals. The CO2 clouds have no counterpart on Earth, since on our planet temperatures never drop low enough (down to about 150 K or about 125 °C) for this gas to condense.

The Polar Caps

Through a telescope, the most prominent surface features on Mars are the bright polar caps, which change with the seasons, similar to the seasonal snow cover on Earth. We do not usually think of the winter snow in northern latitudes as a part of our polar caps, but seen from space, the thin winter snow merges with Earth's thick, permanent ice caps to create an impression much like that seen on Mars (Figure 2.6.5.2).



(a) (b)

Figure 2.6.5.2 Martian North Polar Cap. (a) This is a composite image of the north pole in summer, obtained in October 2006 by the Mars Reconnaissance Orbiter. It shows the mostly water-ice residual cap sitting atop light, tan-colored, layered sediments. Note that although the border of this photo is circular, it shows only a small part of the planet. (b) Here we see a small section of the layered terrain near the martian north pole. There is a mound about 40 meters high that is sticking out of a trough in the center of the picture.

The *seasonal caps* on Mars are composed not of ordinary snow but of frozen CO_2 (dry ice). These deposits condense directly from the atmosphere when the surface temperature drops below about 150 K. The caps develop during the cold martian winters and extend down to about 50° latitude by the start of spring.

Quite distinct from these thin seasonal caps of CO_2 are the *permanent* or *residual caps* that are always present near the poles. The southern permanent cap has a diameter of 350 kilometers and is composed of frozen CO_2 deposits together with a great deal of water ice. Throughout the southern summer, it remains at the freezing point of CO_2, 150 K, and this cold reservoir is thick enough to survive the summer heat intact.

The northern permanent cap is different. It is much larger, never shrinking to a diameter less than 1000 kilometers, and is composed of water ice. Summer temperatures in the north are too high for the frozen CO_2 to be retained. Measurements from the *Mars Global Surveyor* have established the exact elevations in the north polar region of Mars, showing that it is a large basin about the size of our own Arctic Ocean basin. The ice cap itself is about 3 kilometers thick, with a total volume of about 10 million km3(similar to that of Earth's Mediterranean Sea). If Mars ever had extensive liquid water, this north polar basin would have contained a shallow sea. There is some indication of ancient shorelines visible, but better images will be required to verify this suggestion.

Images taken from orbit also show a distinctive type of terrain surrounding the permanent polar caps, as shown in Figure 2.6.5.2. At latitudes above 80° in both hemispheres, the surface consists of recent layered deposits that cover the older cratered ground below. Individual layers are typically ten to a few tens of meters thick, marked by alternating light and dark bands of sediment. Probably the material in the polar deposits includes dust carried by wind from the equatorial regions of Mars.

What do these terraced layers tell us about Mars? Some cyclic process is depositing dust and ice over periods of time. The time scales represented by the polar layers are tens of thousands of years. Apparently the martian climate experiences periodic changes

at intervals similar to those between ice ages on Earth. Calculations indicate that the causes are probably also similar: the gravitational pull of the other planets produces variations in Mars' orbit and tilt as the great clockwork of the solar system goes through its paces.

The *Phoenix* spacecraft landed near the north polar cap in summer (Figure 2.6.5.3). Controllers knew that is would not be able to survive a polar winter, but directly measuring the characteristics of the polar region was deemed important enough to send a dedicated mission. The most exciting discovery came when the spacecraft tried to dig a shallow trench under the spacecraft. When the overlying dust was stripped off, they saw bright white material, apparently some kind of ice. From the way this ice sublimated over the next few days, it was clear that it was frozen water.

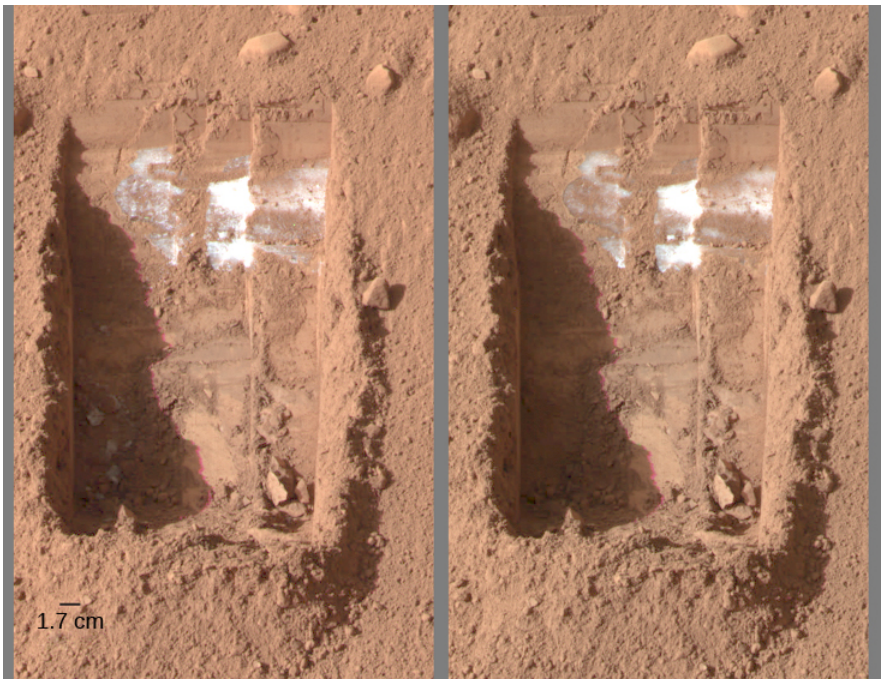


Figure 2.6.5.3 Evaporating Ice on Mars. We see a trench dug by the Phoenix lander in the north polar region four martian days apart in June 2008. If you look at the shadowed region in the bottom left of the trench, you can see three spots of ice in the left image which have sublimated away in the right image. (credit: modification of work by NASA/JPL-Caltech/University of Arizona/Texas A&M University)

2.6.5.1 Example : Comparing the Amount of Water on Mars and Earth

It is interesting to estimate the amount of water (in the form of ice) on Mars and to compare this with the amount of water on Earth. In each case, we can find the total volume of a layer on a sphere by multiplying the area of the sphere ($4\pi R^2$) by the thickness of the layer. For Earth, the ocean water is equivalent to a layer 3 km thick spread over the entire planet, and the radius of Earth is 6.378×10^6 m (see Appendix F). For Mars, most of the water we are sure of is in the form of ice near the poles. We can calculate the amount of ice in one of the residual polar caps if it is (for example) 2 km thick and has a radius of 400 km (the area of a circle is πR^2).

Solution

The volume of Earth's water is therefore the area $4\pi R^2$

$$4\pi \left(6.378 \times 10^6 \text{ m}\right)^2 = 5.1 \times 10^{14} \text{ m2} \backslash \text{nonumer} \tag{2.6.5.1}$$

multiplied by the thickness of 3000 m:

$$5.1 \times 10^{14} \text{ m}^2 \times 3000 \text{ m} = 1.5 \times 10^{18} \text{ m}^3$$

This gives 1.5×10^{18} m^3 of water. Since water has a density of 1 ton per cubic meter (1000 kg/m^3), we can calculate the mass:

$$1.5 \times 10^{18} \text{ m}^3 \times 1 \text{ ton/m}^3 = 1.5 \times 10^{18} \text{ tons}$$

For Mars, the ice doesn't cover the whole planet, only the caps; the polar cap area is

$$\pi R^2 = \pi \left(4 \times 10^5 \text{ m}\right)^2 = 5 \times 10^{11} \text{ m}^2$$

(Note that we converted kilometers to meters.)

The volume = area × height, so we have:

$$\left(2 \times 10^3 \text{ m}\right)\left(5 \times 10^{11} \text{ m}^2\right) = 1 \times 10^{15} \text{ m}^3 = 10^{15} \text{ m}^3$$

Therefore, the mass is:

$$10^{15} \text{ m}^3 \times 1 \text{ ton/m}^3 = 10^{15} \text{ tons}$$

This is about 0.1% that of Earth's oceans.

2.6.5.1 Exercise

A better comparison might be to compare the amount of ice in the Mars polar ice caps to the amount of ice in the Greenland ice sheet on Earth, which has been estimated as 2.85×10^{15} m^3. How does this compare with the ice on Mars?

Answer

The Greenland ice sheet has about 2.85 times as much ice as in the polar ice caps on Mars. They are about the same to the nearest power of 10.

Channels and Gullies on Mars

Although no bodies of liquid water exist on Mars today, evidence has accumulated that rivers flowed on the red planet long ago. Two kinds of geological features appear to be remnants of ancient watercourses, while a third class—smaller gullies—suggests intermittent outbreaks of liquid water even today. We will examine each of these features in turn.

In the highland equatorial plains, there are multitudes of small, sinuous (twisting) channels—typically a few meters deep, some tens of meters wide, and perhaps 10 or 20 kilometers long (Figure 2.6.5.4). They are called runoff channels because they look like what geologists would expect from the surface runoff of ancient rain storms. These runoff channels seem to be telling us that the planet had a very different climate long ago. To estimate the age of these channels, we look at the cratering record. Crater counts show that this part of the planet is more cratered than the lunar maria but less cratered than the lunar highlands. Thus, the runoff channels are probably older than the lunar maria, presumably about 4 billion years old.

The second set of water-related features we see are *outflow channels* (Figure 2.6.5.4) are much larger than the runoff channels. The largest of these, which drain into the Chryse basin where Pathfinder landed, are 10 kilometers or more wide and hundreds of kilometers long. Many features of these outflow channels have convinced geologists that they were carved by huge volumes of running water, far too great to be produced by ordinary rainfall. Where could such floodwater have come from on Mars?



(a)



(b)

Figure 2.6.5.4 Runoff and Outflow Channels. (a) These runoff channels in the old martian highlands are interpreted as the valleys of ancient rivers fed by either rain or underground springs. The width of this image is about 200 kilometers. (b) This intriguing channel, called Nanedi Valles, resembles Earth riverbeds in some (but not all) ways. The tight curves and terraces seen in the channel certainly suggest the sustained flow of a fluid like water. The channel is about 2.5 kilometers across. (credit a: modification of work by Jim Secosky/NASA; credit b: modification of work by Jim Secosky/NASA)

As far we can tell, the regions where the outflow channels originate contained abundant water frozen in the soil as permafrost. Some local source of heating must have released this water, leading to a period of rapid and catastrophic flooding. Perhaps this heating was associated with the formation of the volcanic plains on Mars, which date back to roughly the same time as the outflow channels.

Note that neither the runoff channels nor the outflow channels are wide enough to be visible from Earth, nor do they follow straight lines. They could not have been the “canals” Percival Lowell imagined seeing on the red planet.

The third type of water feature, the smaller *gullies*, was discovered by the *Mars Global Surveyor* (Figure 2.6.5.5). The *Mars Global Surveyor’s* camera images achieved a resolution of a few meters, good enough to see something as small as a truck or bus on the surface. On the steep walls of valleys and craters at high latitudes, there are many erosional features that look like gullies carved by flowing water. These gullies are very young: not only are there no superimposed impact craters, but in some instances, the gullies seem to cut across recent wind-deposited dunes. Perhaps there is liquid water underground that can occasionally break out to produce short-lived surface flows before the water can freeze or evaporate.

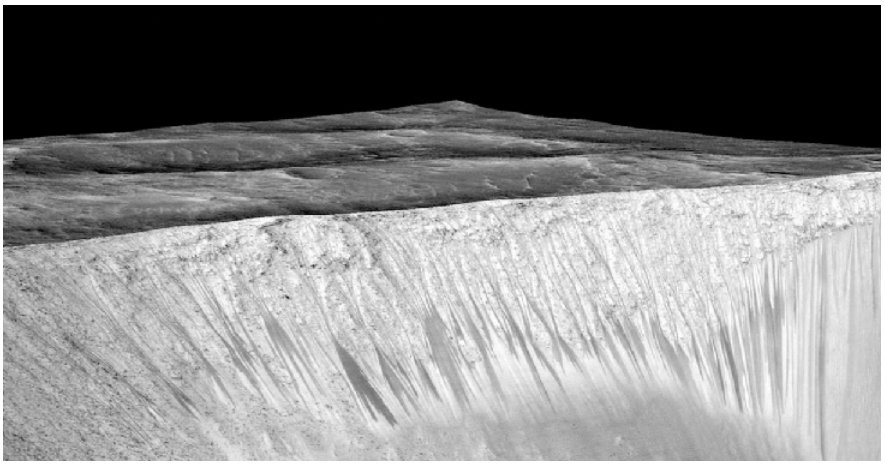


Figure 2.6.5.5 Gullies on the Wall of Garni Crater. This high-resolution image is from the Mars Reconnaissance Orbiter. The dark streaks, which are each several hundred meters long, change in a seasonal pattern that suggests they are caused by the temporary flow of surface water. (credit: NASA/JPL-Caltech/University of Arizona)

The gullies also have the remarkable property of changing regularly with the martian seasons. Many of the dark streaks (visible in Figure 2.6.5.5) elongate within a period of a few days, indicating that something is flowing downhill—either water or dark sediment. If it is water, it requires a continuing source, either from the atmosphere or from springs that tap underground water layers (aquifers.) Underground water would be the most exciting possibility, but this explanation seems inconsistent with the fact that many of the dark streaks start at high elevations on the walls of craters.

Additional evidence that the dark streaks (called by the scientists *recurring slope lineae*) are caused by water was found in 2015 when spectra were obtained of the dark streaks (Figure 2.6.5.6). These showed the presence of hydrated salts produced by the evaporation of salty water. If the water is salty, it could remain liquid long enough to flow downstream for distances of a hundred meters or more, before it either evaporates or soaks into the ground. However, this discovery still does not identify the ultimate source of the water.

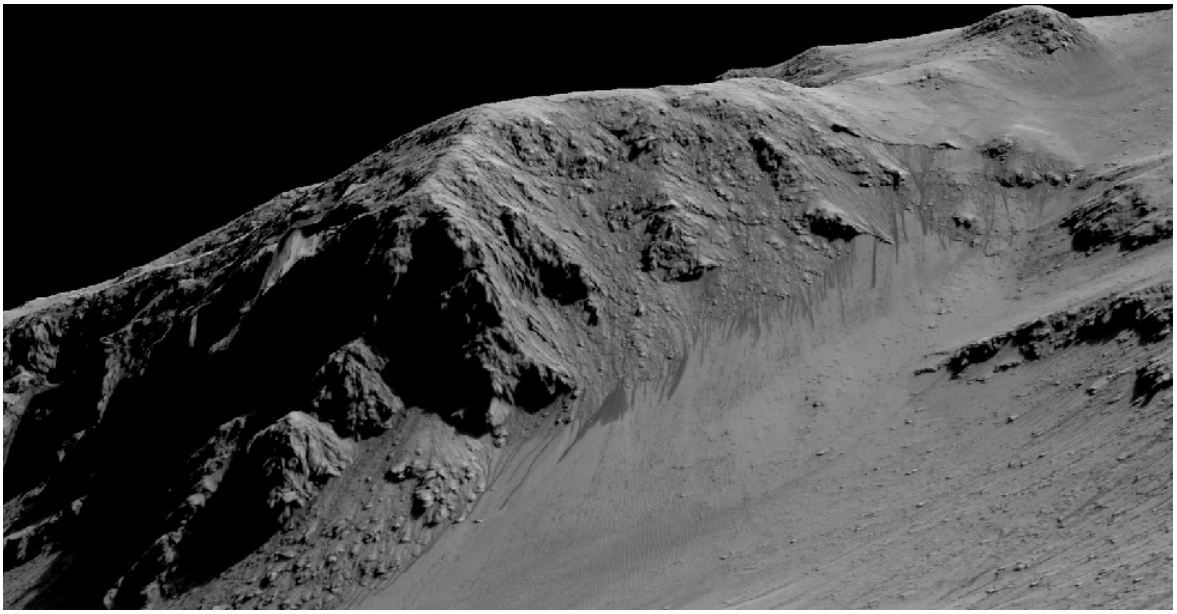


Figure 2.6.5.6 Evidence for Liquid Water on Mars. The dark streaks in Horowitz crater, which move downslope, have been called recurring slope lineae. The streaks in the center of the image go down the wall of the crater for about a distance of 100 meters. Spectra taken of this region indicate that these are locations where salty liquid water flows on or just below the surface of Mars. (The vertical dimension is exaggerated by a factor of 1.5 compared to horizontal dimensions.) (credit: NASA/JPL-Caltech/University of Arizona)

Ancient Lakes and Glaciers

The rovers (*Spirit*, *Opportunity*, and *Curiosity*) that have operated on the surface of Mars have been used to hunt for additional evidence of water. They could not reach the most interesting sites, such as the gullies, which are located on steep slopes. Instead, they explored sites that might be dried-out lake beds, dating back to a time when the climate on Mars was warmer and the atmosphere thicker—allowing water to be liquid on the surface.

Spirit was specifically targeted to explore what looked like an ancient lake-bed in Gusev crater, with an outflow channel emptying into it. However, when the spacecraft landed, it found that the former lakebed had been covered by thin lava flows, blocking the rover from access to the sedimentary rocks it had hoped to find. However, *Opportunity* had better luck. Peering at the walls of a small crater, it detected layered sedimentary rock. These rocks contained chemical evidence of evaporation, suggesting there had been a shallow salty lake in that location. In these sedimentary rocks were also small spheres that were rich in the mineral hematite, which forms only in watery environments. Apparently this very large basin had once been underwater.

> The small spherical rocks were nicknamed “blueberries” by the science team and the discovery of a whole “berry-bowl” of them was announced in this interesting news release from NASA.

The *Curiosity* rover landed inside Gale crater, where photos taken from orbit also suggested past water erosion. It discovered numerous sedimentary rocks, some in the form of mudstones from an ancient lakebed; it also found indications of rocks formed by the action of shallow water at the time the sediment formed (Figure 2.6.5.7).

Even today there is evidence of large quantities of ice just below the surface of Mars. In the mid-latitudes, high-resolution photos from orbit have revealed glaciers covered with dirt and dust. In some cliffs, the ice is observed directly (see Figure 2.6.5.7). These glaciers are thought to have formed during warm periods, when the atmospheric pressure was greater and snow and ice could precipitate. They also suggest readily available frozen water that could support future human exploration of the planet.



(a) (b)

Figure 2.6.5.7
Gale Crater and Underground Ice Deposits. (a) This scene, photographed by the *Curiosity* rover, shows an ancient lakebed of cracked mudstones. (b) Geologists working with the *Curiosity* rover interpret this image of cross-bedded sandstone in Gale crater as evidence of liquid water passing over a loose bed of sediment at the time this rock formed. (c) Ice bands a hundred meters tall are visible in blue in a cliff-face on Mars, suggesting large deposits of frozen water buried just a few meters below the surface. Note that the blue color has been exaggerated in this photo, taken by the Mars Reconnaissance Orbiter spacecraft. (credit a: modification of work by NASA/JPL-Caltech/MSSS; credit b: modification of work by NASA/JPL-Caltech/MSSS; credit c: modification of work by NASA/JPL-Caltech/UA/USGS)

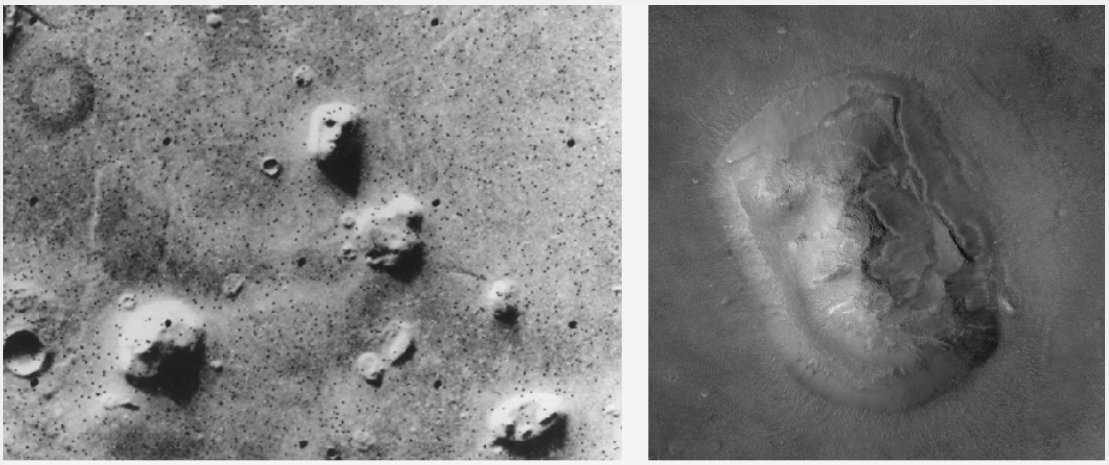
astronomy and pseudoscience: the "face on Mars"

People like human faces. We humans have developed great skill in recognizing people and interpreting facial expressions. We also have a tendency to see faces in many natural formations, from clouds to the man in the Moon. One of the curiosities that emerged from the Viking orbiters’ global mapping of Mars was the discovery of a strangely shaped mesa in the Cydonia region that resembled a human face. Despite later rumors of a cover-up, the “Face on Mars” was, in fact, recognized by Viking scientists and included in one of the early mission press releases. At the low resolution and oblique lighting under which the Viking image was obtained, the mile-wide mesa had something of a Sphinx-like appearance.

Unfortunately, a small band of individuals decided that this formation was an artificial, carved sculpture of a human face placed on Mars by an ancient civilization that thrived there hundreds of thousands of years ago. A band of “true believers” grew around the face and tried to deduce the nature of the “sculptors” who made it. This group also linked the face to a variety of other pseudoscientific phenomena such as crop circles (patterns in fields of grain, mostly in Britain, now known to be the work of pranksters).

Members of this group accused NASA of covering up evidence of intelligent life on Mars, and they received a great deal of help in publicizing their perspective from tabloid media. Some of the believers picketed the Jet Propulsion Laboratory at the time of the failure of the *Mars Observer* spacecraft, circulating stories that the “failure” of the *Mars Observer* was itself a fake, and that its true (secret) mission was to photograph the face.

The high-resolution *Mars Observer* camera (MOC) was reflown on the *Mars Global Surveyor* mission, which arrived at Mars in 1997. On April 5, 1998, in Orbit 220, the MOC obtained an oblique image of the face at a resolution of 4 meters per pixel, a factor-of-10 improvement in resolution over the Viking image. Another image in 2001 had even higher resolution. Immediately released by NASA, the new images showed a low mesa-like hill cut crossways by several roughly linear ridges and depressions, which were misidentified in the 1976 photo as the eyes and mouth of a face. Only with an enormous dose of imagination can any resemblance to a face be seen in the new images, demonstrating how dramatically our interpretation of geology can change with large improvements in resolution. The original and the higher resolution images can be seen in Figure 2.6.5.8.



(a) (b)

Figure 2.6.5.8 Face on Mars. The so-called “Face on Mars” is seen (a) in low resolution from Viking (the “face” is in the upper part of the picture) and (b) with 20 times better resolution from the Mars Global Surveyor.

After 20 years of promoting pseudoscientific interpretations and various conspiracy theories, can the “Face on Mars” believers now accept reality? Unfortunately, it does not seem so. They have accused NASA of faking the new picture. They also suggest that the secret mission of the *Mars Observer*included a nuclear bomb used to destroy the face before it could be photographed in greater detail by the *Mars Global Surveyor*.

Space scientists find these suggestions incredible. NASA is spending increasing sums for research on life in the universe, and a major objective of current and upcoming Mars missions is to search for evidence of past microbial life on Mars. Conclusive evidence of extraterrestrial life would be one of the great discoveries of science and incidentally might well lead to increased funding for NASA. The idea that NASA or other government agencies would (or could) mount a conspiracy to suppress such welcome evidence is truly bizarre.

Alas, the “Face on Mars” story is only one example of a whole series of conspiracy theories that are kept before the public by dedicated believers, by people out to make a fast buck, and by irresponsible media attention. Others include the “urban legend” that the Air Force has the bodies of extraterrestrials at a secret base, the widely circulated report that UFOs crashed near Roswell, New Mexico (actually it was a balloon carrying scientific instruments to find evidence of Soviet nuclear tests), or the

notion that alien astronauts helped build the Egyptian pyramids and many other ancient monuments because our ancestors were too stupid to do it alone.

In response to the increase in publicity given to these “fiction science” ideas, a group of scientists, educators, scholars, and magicians (who know a good hoax when they see one) have formed the Committee for Skeptical Inquiry. Two of the original authors of your book are active on the committee. For more information about its work delving into the rational explanations for paranormal claims, see their excellent magazine, *The Skeptical Inquirer*, or check out their website at www.csicop.org/.

Climate Change on Mars

The evidence about ancient rivers and lakes of water on Mars discussed so far suggests that, billions of years ago, martian temperatures must have been warmer and the atmosphere must have been more substantial than it is today. But what could have changed the climate on Mars so dramatically?

We presume that, like Earth and Venus, Mars probably formed with a higher surface temperature thanks to the greenhouse effect. But Mars is a smaller planet, and its lower gravity means that atmospheric gases could escape more easily than from Earth and Venus. As more and more of the atmosphere escaped into space, the temperature on the surface gradually fell.

Eventually Mars became so cold that most of the water froze out of the atmosphere, further reducing its ability to retain heat. The planet experienced a sort of *runaway refrigerator effect*, just the opposite of the runaway greenhouse effect that occurred on Venus. Probably, this loss of atmosphere took place within less than a billion years after Mars formed. The result is the cold, dry Mars we see today.

Conditions a few meters below the martian surface, however, may be much different. There, liquid water (especially salty water) might persist, kept warm by the internal heat of Mars or the insulating layers solid and rock. Even on the surface, there may be ways to change the martian atmosphere temporarily.

Mars is likely to experience long-term climate cycles, which may be caused by the changing orbit and tilt of the planet. At times, one or both of the polar caps might melt, releasing a great deal of water vapor into the atmosphere. Perhaps an occasional impact by a comet might produce a temporary atmosphere that is thick enough to permit liquid water on the surface for a few weeks or months. Some have even suggested that future technology might allow us to *terraform* Mars—that is, to engineer its atmosphere and climate in ways that might make the planet more hospitable for long-term human habitation.

The Search for Life on Mars

If there was running water on Mars in the past, perhaps there was life as well. Could life, in some form, remain in the martian soil today? Testing this possibility, however unlikely, was one of the primary objectives of the Viking landers in 1976. These landers carried miniature biological laboratories to test for microorganisms in the martian soil. Martian soil was scooped up by the spacecraft’s long arm and placed into the experimental chambers, where it was isolated and incubated in contact with a variety of gases, radioactive isotopes, and nutrients to see what would happen. The experiments looked for evidence of *respiration* by living animals, *absorption* of *nutrients* offered to organisms that might be present, and an *exchange of gases* between the soil and its surroundings for any reason whatsoever. A fourth instrument pulverized the soil and analyzed it carefully to determine what organic (carbon-bearing) material it contained.

The Viking experiments were so sensitive that, had one of the spacecraft landed anywhere on Earth (with the possible exception of Antarctica), it would easily have detected life. But, to the disappointment of many scientists and members of the public, no life was detected on Mars. The soil tests for absorption of nutrients and gas exchange did show some activity, but this was most likely caused by chemical reactions that began as water was added to the soil and had nothing to do with life. In fact, these experiments showed that martian soil seems much more chemically active than terrestrial soils because of its exposure to solar ultraviolet radiation (since Mars has no ozone layer).

The organic chemistry experiment showed no trace of organic material, which is apparently destroyed on the martian surface by the sterilizing effect of this ultraviolet light. While the possibility of life on the surface has not been eliminated, most experts consider it negligible. Although Mars has the most earthlike environment of any planet in the solar system, the sad fact is that nobody seems to be home today, at least on the surface.

However, there is no reason to think that life could not have begun on Mars about 4 billion years ago, at the same time it started on Earth. The two planets had very similar surface conditions then. Thus, the attention of scientists has shifted to the search for *fossil*life on Mars. One of the primary questions to be addressed by future spacecraft is whether Mars once supported its own life forms and, if so, how this martian life compared with that on our own planet. Future missions will include the return of martian samples selected from sedimentary rocks at sites that once held water and thus perhaps ancient life. The most powerful searches for martian life (past or present) will thus be carried out in our laboratories here on Earth.

planetary protection

When scientists begin to search for life on another planet, they must make sure that we do not contaminate the other world with life carried from Earth. At the very beginning of spacecraft exploration on Mars, an international agreement specified that all landers were to be carefully sterilized to avoid accidentally transplanting terrestrial microbes to Mars. In the case of Viking, we know the sterilization was successful. Viking's failure to detect martian organisms also implies that these experiments did not detect hitchhiking terrestrial microbes.

As we have learned more about the harsh conditions on the martian surface, the sterilization requirements have been somewhat relaxed. It is evident that no terrestrial microbes could grow on the martian surface, with its low temperature, absence of water, and intense ultraviolet radiation. Microbes from Earth might survive in a dormant, dried state, but they cannot grow and proliferate on Mars.

The problem of contaminating Mars will become more serious, however, as we begin to search for life below the surface, where temperatures are higher and no ultraviolet light penetrates. The situation will be even more daunting if we consider human flights to Mars. Any humans will carry with them a multitude of terrestrial microbes of all kinds, and it is hard to imagine how we can effectively keep the two biospheres isolated from each other if Mars has indigenous life. Perhaps the best situation could be one in which the two life-forms are so different that each is effectively invisible to the other—not recognized on a chemical level as living or as potential food.

The most immediate issue of public concern is not with the contamination of Mars but with any dangers associated with returning Mars samples to Earth. NASA is committed to the complete biological isolation of returned samples until they are demonstrated to be safe. Even though the chances of contamination are extremely low, it is better to be safe than sorry.

Most likely there is no danger, even if there is life on Mars and alien microbes hitch a ride to Earth inside some of the returned samples. In fact, Mars is sending samples to Earth all the time in the form of the Mars meteorites. Since some of these microbes (if they exist) could probably survive the trip to Earth inside their rocky home, we may have been exposed many times over to martian microbes. Either they do not interact with our terrestrial life, or in effect our planet has already been inoculated against such alien bugs.

More than any other planet, Mars has inspired science fiction writers over the years. You can find scientifically reasonable stories about Mars in a subject index of such stories online. If you click on Mars as a topic, you will find stories by a number of space scientists, including William Hartmann, Geoffrey Landis, and Ludek Pesek.

Key Concepts and Summary

The martian atmosphere has a surface pressure of less than 0.01 bar and is 95% CO_2. It has dust clouds, water clouds, and carbon dioxide (dry ice) clouds. Liquid water on the surface is not possible today, but there is subsurface permafrost at high latitudes. Seasonal polar caps are made of dry ice; the northern residual cap is water ice, whereas the southern permanent ice cap is made predominantly of water ice with a covering of carbon dioxide ice. Evidence of a very different climate in the past is found in water erosion features: both runoff channels and outflow channels, the latter carved by catastrophic floods. Our rovers, exploring ancient lakebeds and places where sedimentary rock has formed, have found evidence for extensive surface water in the past. Even more exciting are the gullies that seem to show the presence of flowing salty water on the surface today, hinting at near-surface aquifers. The Viking landers searched for martian life in 1976, with negative results, but life might have flourished long ago. We have found evidence of water on Mars, but following the water has not yet led us to life on that planet.

This page titled 2.6.5: Water and Life on Mars is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **10.5: Water and Life on Mars** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.6.6: Divergent Planetary Evolution

Learning Objectives

By the end of this section, you will be able to:

- Compare the planetary evolution of Venus, Earth, and Mars

Venus, Mars, and our own planet Earth form a remarkably diverse triad of worlds. Although all three orbit in roughly the same inner zone around the Sun and all apparently started with about the same chemical mix of silicates and metals, their evolutionary paths have diverged. As a result, Venus became hot and dry, Mars became cold and dry, and only Earth ended up with what we consider a hospitable climate.

We have discussed the runaway greenhouse effect on Venus and the runaway refrigerator effect on Mars, but we do not understand exactly what started these two planets down these separate evolutionary paths. Was Earth ever in danger of a similar fate? Or might it still be diverted onto one of these paths, perhaps due to stress on the atmosphere generated by human pollutants? One of the reasons for studying Venus and Mars is to seek insight into these questions.

Some people have even suggested that if we understood the evolution of Mars and Venus better, we could possibly reverse their evolution and restore more earthlike environments. While it seems unlikely that humans could ever make either Mars or Venus into a replica of Earth, considering such possibilities is a useful part of our more general quest to understand the delicate environmental balance that distinguishes our planet from its two neighbors. In Cosmic Samples and the Origin of the Solar System, we return to the comparative study of the terrestrial planets and their divergent evolutionary histories.

Summary

Earth, Venus, and Mars have diverged in their evolution from what may have been similar beginnings. We need to understand why if we are to protect the environment of Earth.

This page titled 2.6.6: Divergent Planetary Evolution is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **10.6: Divergent Planetary Evolution** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.6.E: Earthlike Planets - Venus and Mars (Exercises)

For Further Exploration

Articles

Venus

- Dorminey, B. “Cool Science on a Hot World.” *Astronomy* (February 2006): 46. Five-page overview of Venus and the Venus Express mission plans.
- Kargel, J. “Rivers of Venus.” *Sky & Telescope* (August 1997): 32. On lava channels.
- Robertson, D. “Parched Planet.” *Sky & Telescope* (April 2008): 26. Overview of our understanding of the planet.
- Robinson, C. “Magellan Reveals Venus.” *Astronomy* (February 1995): 32.
- Stofan, E. “The New Face of Venus.” *Sky & Telescope* (August 1993): 22.
- Zimmerman, R. “Taking Venus by Storm.” *Astronomy* (October 2008): 66. On results from the Venus Express mission.

Mars

- Albee, A. “The Unearthly Landscapes of Mars.” *Scientific American* (June 2003): 44. Results from the Mars Global Surveyor and Mars Odyssey missions and an overview.
- Bell, J. “A Fresh Look at Mars.” *Astronomy* (August 2015): 28. Nice summary of recent spacecraft results and how they are revising our understanding of Mars.
- Bell, J. “Uncovering Mars’ Secret Past.” *Sky & Telescope* (July 2009): 22. How rovers and orbiters are helping us to understand Mars history and the role of water.
- Bell, J. “The Red Planet’s Watery Past.” *Scientific American* (December 2006): 62. Rovers are furnishing proof that ancient Mars was wet.
- Burnham, R. “Red Planet Rendezvous.” *Astronomy* (May 2006): 68. About Mariner Valley and a flyover film constructed from many still images.
- Christensen, P. “The Many Faces of Mars.” *Scientific American* (July 2005): 32. Results from the Rover mission; evidence that Mars was once wet in places.
- Lakdawalla, E. “The History of Water on Mars.” *Sky & Telescope* (September 2013): 16. Clear review of our current understanding of the role of water on Mars in different epochs.
- Malin, M. “Visions of Mars.” *Sky & Telescope* (April 1999): 42. A geological tour of the red planet, with new Mars Global Surveyor images.
- McEwen, A. “Mars in Motion.” *Scientific American* (May 2013): 58. On gullies and other surface changes.
- McKay, C. & Garcia, V. “How to Search for Life on Mars.” *Scientific American* (June 2014): 44. Experiments future probes could perform.
- Naeye, R. “Europe’s Eye on Mars.” *Sky & Telescope* (December 2005): 30. On the Mars Express mission and the remarkable close-up images it is sending.
- Talcott, R. “Seeking Ground Truth on Mars.” *Astronomy* (October 2009): 34. How rovers and orbiters are helping scientists understand the red planet’s surface.

Websites

- European Space Agency Mars Express Page: www.esa.int/Our_Activities/Sp...e/Mars_Express.
- European Space Agency Venus Express Page: www.esa.int/Our_Activities/Sp.../Venus_Express.
- High Resolution Imaging Science Experiment: http://hirise.lpl.arizona.edu/.
- Jet Propulsion Lab Mars Exploration Page: http://mars.jpl.nasa.gov/.
- Mars Globe HD app: https://itunes.apple.com/us/app/mars...376020224?mt=8.
- Mars Rover 360° Panorama: http://www.360cities.net/image/curio....10,26.50,70.0. Interactive.
- NASA Center for Mars Exploration: www.nasa.gov/mission_pages/ma...ain/index.html.
- NASA Solar System Exploration Mars Page: http://solarsystem.nasa.gov/planets/mars.
- NASA Solar System Exploration Venus Page: http://solarsystem.nasa.gov/planets/venus.
- NASA’s apps about Mars for phones and tablets can be found at: http://mars.nasa.gov/mobile/info/.

- NASA's Magellan Mission to Venus: http://www2.jpl.nasa.gov/magellan/.
- Russian (Soviet) Venus Missions and Images: http://mentallandscape.com/C_CatalogVenus.htm.
- Venus Atlas app: https://itunes.apple.com/us/app/venu...317310503?mt=8.
- Venus Express Results Article: http://www.mpg.de/798302/F002_Focus_026-033.pdf.

Videos

- 50 Years of Mars Exploration: http://www.jpl.nasa.gov/video/details.php?id=1395. NASA's summary of all missions through *MAVEN*; good quick overview (4:08).
- Being a Mars Rover: What It's Like to be an Interplanetary Explorer: https://www.youtube.com/watch?v=nRpCOEsPD54. 2013 talk by Dr. Lori Fenton about what it's like on the surface of Mars (1:07:24).
- Magellan Maps Venus: www.bbc.co.uk/science/space/s...probe#p005y07s. BBC clip with Dr. Ellen Stofan on the radar images of Venus and what they tell us (3:06).
- Our *Curiosity*: https://www.youtube.com/watch?v=XczKXWvokm4. Mars *Curiosity* rover 2-year anniversary video narrated by Neil deGrasse Tyson and Felicia Day (6:01).
- Planet Venus: The Deadliest Planet, Venus Surface and Atmosphere: https://www.youtube.com/watch?v=HqFVxWfVtoo. Quick tour of Venus' atmosphere and surface (2:04).
- Planetary Protection and Hitchhikers in the Solar System: The Danger of Mingling Microbes: https://www.youtube.com/watch?v=6iGC3uO7jBI. 2009 talk by Dr. Margaret Race on preventing contamination between worlds (1:28:50).

Collaborative Group Activities

1. Your group has been asked by high NASA officials to start planning the first human colony on Mars. Begin by making a list of what sorts of things humans would need to bring along to be able to survive for years on the surface of the red planet.
2. As a publicity stunt, the mayor of Venus, Texas (there really is such a town), proposes that NASA fund a mission to Venus with humans on board. Clearly, the good mayor neglected to take an astronomy course in college. Have your group assemble a list of as many reasons as possible why it is unlikely that humans will soon land on the surface of Venus.
3. Even if humans would have trouble surviving on the surface of Venus, this does not mean we could not learn a lot more about our veiled sister planet. Have your group brainstorm a series of missions (pretend cost is no object) that would provide us with more detailed information about Venus' atmosphere, surface, and interior.
4. Sometime late in the twenty-first century, when travel to Mars has become somewhat routine, a very wealthy couple asks you to plan a honeymoon tour of Mars that includes the most spectacular sights on the red planet. Constitute your group as the Percival Lowell Memorial Tourist Agency, and come up with a list of not-to-be missed tourist stops on Mars.
5. In the popular book and film, called *The Martian*, the drama really begins when our hero is knocked over and loses consciousness as he is half buried by an intense wind storm on Mars. Given what you have learned about Mars' atmosphere in this chapter, have your group discuss how realistic that scenario is. (By the way, the author of the book has himself genially acknowledged in interviews and talks that this is a reasonable question to ask.)
6. Astronomers have been puzzled and annoyed about the extensive media publicity that was given the small group of "true believers" who claimed the "Face on Mars" was not a natural formation (see the Astronomy and Pseudoscience: The "Face on Mars" feature box in Section 10.5). Have your group make a list of the reasons many of the media were so enchanted by this story. What do you think astronomers could or should do to get the skeptical, scientific perspective about such issues before the public?
7. Your group is a special committee of scientists set up by the United Nations to specify how any Mars samples should be returned to Earth so that possible martian microbes do not harm Earth life. What precautions would you recommend, starting at Mars and going all the way to the labs that analyze the martian samples back on Earth?
8. Have your group brainstorm about Mars in popular culture. How many movies, songs or other music, and products can you think of connected with Mars? What are some reasons that Mars would be a popular theme for filmmakers, songwriters, and product designers?

Review Questions

1. List several ways that Venus, Earth, and Mars are similar, and several ways they are different.

2. Compare the current atmospheres of Earth, Venus, and Mars in terms of composition, thickness (and pressure at the surface), and the greenhouse effect
3. How might Venus' atmosphere have evolved to its present state through a runaway greenhouse effect?
4. Describe the current atmosphere on Mars. What evidence suggests that it must have been different in the past?
5. Explain the runaway refrigerator effect and the role it may have played in the evolution of Mars.
6. What evidence do we have that there was running (liquid) water on Mars in the past? What evidence is there for water coming out of the ground even today?
7. What evidence is there that Venus was volcanically active about 300–600 million years ago?
8. Why is Mars red?
9. What is the composition of clouds on Mars?
10. What is the composition of the polar caps on Mars?
11. Describe two anomalous features of the rotation of Venus and what might account for them.
12. How was the *Mars Odyssey* spacecraft able to detect water on Mars without landing on it?

Thought Questions

1. What are the advantages of using radar imaging rather than ordinary cameras to study the topography of Venus? What are the relative advantages of these two approaches to mapping Earth or Mars?
2. Venus and Earth are nearly the same size and distance from the Sun. What are the main differences in the geology of the two planets? What might be some of the reasons for these differences?
3. Why is there so much more carbon dioxide in the atmosphere of Venus than in that of Earth? Why so much more carbon dioxide than on Mars?
4. If the Viking missions were such a rich source of information about Mars, why have we sent the Pathfinder, *Global Surveyor*, and other more recent spacecraft to Mars? Make a list of questions about Mars that still puzzle astronomers.
5. Compare Mars with Mercury and the Moon in terms of overall properties. What are the main similarities and differences?
6. Contrast the mountains on Mars and Venus with those on Earth and the Moon.
7. We believe that all of the terrestrial planets had similar histories when it comes to impacts from space. Explain how this idea can be used to date the formation of the martian highlands, the martian basins, and the Tharsis volcanoes. How certain are the ages derived for these features (in other words, how do we check the ages we derive from this method)?
8. Is it likely that life ever existed on either Venus or Mars? Justify your answer in each case.
9. Suppose that, decades from now, NASA is considering sending astronauts to Mars and Venus. In each case, describe what kind of protective gear they would have to carry, and what their chances for survival would be if their spacesuits ruptured.
10. We believe that Venus, Earth, and Mars all started with a significant supply of water. Explain where that water is now for each planet.
11. One source of information about Mars has been the analysis of meteorites from Mars. Since no samples from Mars have ever been returned to Earth from any of the missions we sent there, how do we know these meteorites are from Mars? What information have they revealed about Mars?
12. The runaway greenhouse effect and its inverse, the runaway refrigerator effect, have led to harsh, uninhabitable conditions on Venus and Mars. Does the greenhouse effect always cause climate changes leading to loss of water and life? Give a reason for your answer.
13. In what way is the high surface temperature of Venus relevant to concerns about global warming on Earth today?
14. What is a dust devil? Would you expect to feel more of a breeze from a dust devil on Mars or on Earth? Explain.
15. Near the martian equator, temperatures at the same spot can vary from an average of –135 °C at night to an average of 30 °C during the day. How can you explain such a wide difference in temperature compared to that on Earth?

Figuring for Yourself

1. Estimate the amount of water there could be in a global (planet-wide) region of subsurface permafrost on Mars (do the calculations for two permafrost thicknesses, 1 and 10 km, and a concentration of ice in the permafrost of 10% by volume). Compare the two results you get with the amount of water in Earth's oceans calculated in Example 10.5.1.
2. At its nearest, Venus comes within about 41 million km of Earth. How distant is it at its farthest?
3. If you weigh 150 lbs. on the surface of Earth, how much would you weigh on Venus? On Mars?

4. Calculate the relative land area—that is, the amount of the surface not covered by liquids—of Earth, the Moon, Venus, and Mars. (Assume that 70% of Earth is covered with water.)
5. The closest approach distance between Mars and Earth is about 56 million km. Assume you can travel in a spaceship at 58,000 km/h, which is the speed achieved by the New Horizons space probe that went to Pluto and is the fastest speed so far of any space vehicle launched from Earth. How long would it take to get to Mars at the time of closest approach?

This page titled 2.6.E: Earthlike Planets - Venus and Mars (Exercises) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- **10.E: Earthlike Planets - Venus and Mars (Exercises)** by OpenStax is licensed CC BY 4.0. Original source: https://openstax.org/details/books/astronomy.

2.7: Summary



Under Construction



2.7: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

3: The Atmosphere

3.1: Introduction
3.2: Structure and Composition of the Atmosphere
3.3: Origin and Evolution of the Atmosphere
3.4: General Circulation
3.4.1: A Simplified Description of the Global Circulation
3.4.2: Radiative Differential Heating
3.4.3: Pressure Profiles
3.4.4: Geostrophic Wind and Geostrophic Adjustment
3.4.5: Thermal Wind Effect
3.4.6: Horizontal Circulation
3.4.7: Extratropical Ridges and Troughs (Rossby Waves)
3.4.8: Three-band Global Circulation
3.4.9: Ekman Spiral of Ocean Currents
3.4.10: Explaining the General Circulation
3.4.11: Jet Streams
3.4.12: Types of Vorticity
3.4.13: Key Terms
3.4.14: Review
3.4.15: Homework Exercises
3.5: Atmospheric Composition
3.5.1: Atmospheric Composition
3.5.2: Changes in Atmospheric Composition
3.5.3: Other Trace Gases
3.5.4: Stratospheric Ozone Formation
3.5.5: The Story of the Atmosphere's PAC-MAN
3.5.6: Where do Cloud Condensation Nuclei (CCN) come from?
3.5.7: Summary and Final Tasks
3.6: Biogeochemical reactions in the troposphere
3.7: Atmospheric deposition
3.7.1: Acid Deposition
3.8: Biogeochemical reactions in the stratosphere
3.9: Models
3.10: Impacts
3.11: Summary

3: The Atmosphere is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.1: Introduction

The Atmosphere

The **atmosphere**, the gaseous layer that surrounds the earth, formed over four billion years ago and is held in place by the attractive forces of gravity. During the evolution of the solid earth, volcanic eruptions released gases into the developing atmosphere. Assuming the outgasing was similar to that of modern volcanoes, the gases released included: water vapor (H2O), carbon monoxide (CO), carbon dioxide (CO2), hydrochloric acid (HCl), methane (CH4), ammonia (NH3), nitrogen (N2) and sulfur gases. The atmosphere was reducing because there was no free oxygen. Most of the hydrogen and helium that outgassed would have eventually escaped into outer space due to the inability of the earth's gravity to hold on to their small masses. There may have also been significant contributions of volatiles from the massive meteoritic bombardments known to have occurred early in the earth's history.

Water vapor in the atmosphere condensed and rained down, eventually forming lakes and oceans. The oceans provided homes for the earliest organisms which were probably similar to cyanobacteria. Oxygen was released into the atmosphere by these early organisms, and carbon became sequestered in sedimentary rocks. This led to our current oxidizing atmosphere, which is mostly comprised of nitrogen (roughly 78 percent) and oxygen (roughly 21 percent) with 0.9% argon (Ar) and 0.04% carbon dioxide (CO_2). The atmosphere also contains highly variable concentrations of water vapor, which can range from only 0.01% in frigid winter air in the Arctic to 5% in warm, humid, tropical air. The atmosphere also contains several gases in trace amounts, such as helium, neon, methane and nitrous oxide. One very important trace gas is ozone (O_3), which absorbs harmful UV radiation from the sun.

The earth's atmosphere extends outward to about 1,000 kilometers where it transitions to interplanetary space. However, most of the mass of the atmosphere (greater than 99 percent) is located within the first 40 kilometers. On average, the total weight of the atmospheric mass exerts a pressure at sea level of around 1.0×10^5 pascals (Pa; or one atmosphere), which is equivalent to 1.0 kg per cm^2. The density of the atmospheric mass is much greater close to the surface and decreases rapidly with increasing altitude. The vertical temperature profile of the atmosphere is variable and depends upon the types of radiation that affect each atmospheric layer. This, in turn, depends upon the chemical composition of that layer (mostly involving trace gases). Based on these factors, the atmosphere can be divided into four distinct layers: the troposphere, stratosphere, mesosphere, and thermosphere (Figure 3.1.1). The boundaries of the layers are inexact because they may vary over time and space. Beyond the atmosphere is outer space, an immeasurably vast region where the Earth exerts no detectable chemical or thermal influences.

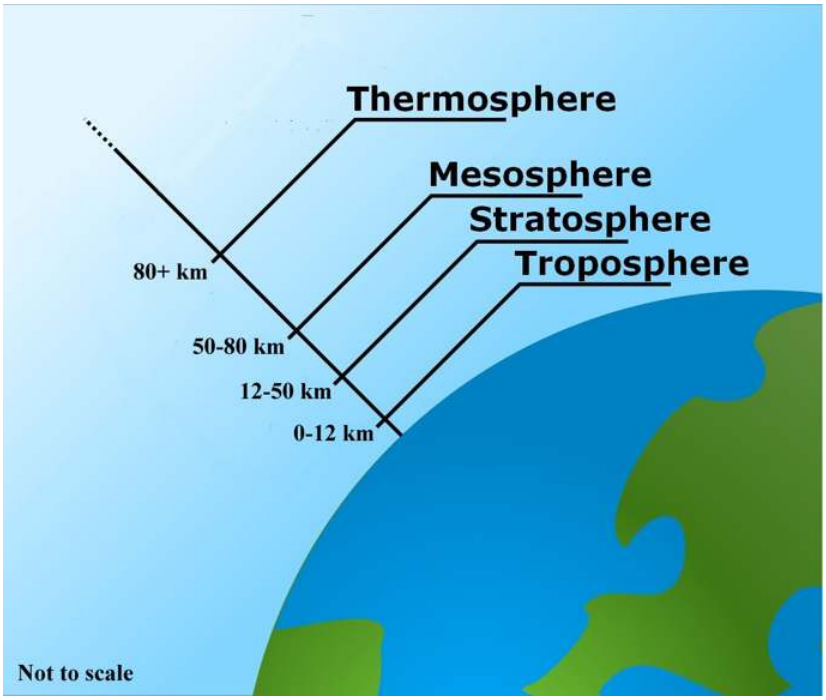


Figure 3.1.1**:** The layers of the atmosphere. the toposphere is the closest to the Earth's surface (0-12 km). Next, is the stratosphere (12-50 km), mesophere (50-80 km), and thermosphere (80+ km). The outermost layer (the exosphere) is not shown. **Ground-level ozone** in the troposphere is a form of air pollution, but the **ozone layer** in the stratosphere helps filter UV rays. Image by GFDL (CC-BY-SA).

The **troposphere** (or lower atmosphere) contains 85-90% of the atmospheric mass and extends from the surface to an altitude of 8-20 km. It is thinner at high latitudes, and thicker at equatorial latitudes, but also varies seasonally, at any place being thicker during the summer than in the winter. Because convective air currents (winds) are common in the troposphere (the name troposphere means “region of mixing”) it is sometimes referred to as the “weather layer.” The troposphere contains the largest percentage of the mass of the atmosphere relative to the other layers. It also contains some 99 percent of the total water vapor of the atmosphere. The temperature of the troposphere is warm (roughly 17º C) near the surface of the earth. This is due to the absorption of infrared radiation from the surface by water vapor and other greenhouse gases (e.g. carbon dioxide, nitrous oxide and methane) in the troposphere. The concentration of these gases decreases with altitude, and therefore, the heating effect is greatest near the surface. The temperature in the troposphere decreases at a rate of roughly 6.5º C per kilometer of altitude. The temperature at its upper boundary is very cold (roughly -60º C). Water vapor evaporated from the earth's surface condenses in the cooler upper regions of the troposphere and falls back to the surface as rain. Dust and pollutants injected into the troposphere become well mixed in the layer, but are eventually washed out by rainfall. The troposphere is therefore self cleaning. A narrow zone at the top of the troposphere is called the **tropopause**. It effectively separates the underlying troposphere and the overlying stratosphere. The temperature in the tropopause is relatively constant. Strong eastward winds, known as the **jet stream**, also occur here.

The **stratosphere** extends from the troposphere to as high as about 50 km, depending on the season and latitude. Within the stratosphere there are few convective air currents. The temperature profile of the stratosphere is quite different from that of the troposphere. The temperature remains relatively constant up to roughly 25 kilometers and then gradually increases up to the upper boundary of the layer. The amount of water vapor in the stratosphere is very low, so it is not an important factor in the temperature regulation of the layer. Instead, it is ozone (O3) that causes the observed temperature inversion. Most of the ozone in the atmosphere is contained in a layer of the stratosphere from roughly 20 to 30 kilometers. This ozone layer absorbs solar energy in the form of ultraviolet radiation (UV), and the energy is ultimately dissipated as heat in the stratosphere. This heat leads to the rise in temperature. Stratospheric ozone is also very important for living organisms on the surface of the earth as it protects them by absorbing most of the harmful UV radiation from the sun. The upper boundary of the stratosphere is known as the **stratopause**, which is marked by a sudden decrease in temperature.

The third layer in the earth's atmosphere is called the **mesosphere**. It extends from the stratopause (about 50 kilometers) to roughly 85 kilometers above the earth's surface. Because the mesosphere has negligible amounts of water vapor and ozone for generating heat, the temperature drops across this layer. It is warmed from the bottom by the stratosphere. The air is very thin in this region with a density about 1/1000 that of the surface. With increasing altitude this layer becomes increasingly dominated by lighter gases, and in the outer reaches, the remaining gases become stratified by molecular weight.

The fourth layer, the **thermosphere**, extends outward from about 85 kilometers to about 600 kilometers. Its upper boundary is ill defined. The temperature in the thermosphere increases with altitude, up to 1500º C or more. The high temperatures are the result of absorption of intense solar radiation by the last remaining oxygen molecules. The temperature can vary substantially depending upon the level of solar activity. The lower region of the thermosphere (up to about 550 kilometers) is also known as the **ionosphere**. Because of the high temperatures in this region, gas particles become ionized. The ionosphere is important because it reflects radio waves from the earth's surface, allowing long-distance radio communication. The visual atmospheric phenomenon known as the northern lights also occurs in this region. The outer region of the atmosphere is known as the **exosphere**. The exosphere represents the final transition between the atmosphere and interplanetary space. It extends about 1000 kilometers and contains mainly helium and hydrogen. Most satellites operate in this region.

Climate and Weather

A common misconception about global climate change is that a specific weather event occurring in a particular region (for example, a very cool week in June in central Indiana) is evidence of global climate change. However, a cold week in June is a weather-related event and not a climate-related one. These misconceptions often arise because of confusion over the terms climate and weather.

Climate refers to the long-term, predictable atmospheric conditions of a specific area. The climate of a biome is characterized by having consistent temperature and annual rainfall ranges. Climate does not address the amount of rain that fell on one particular day in a biome or the colder-than-average temperatures that occurred on one day. Climate scientists often average over decades to determine climate in a given location.

In contrast, weather refers to the conditions of the atmosphere during a short period of time. Weather forecasts are usually made for 48-hour cycles. Long-range weather forecasts are available but can be unreliable.

To better understand the difference between climate and weather, imagine that you are planning an outdoor event in northern Wisconsin. You would be thinking about *climate* when you plan the event in the summer rather than the winter because you have long-term knowledge that any given Saturday in the months of May to August would be a better choice for an outdoor event in Wisconsin than any given Saturday in January. However, you cannot determine the specific day that the event should be held on because it is difficult to accurately predict the weather on a specific day. Climate can be considered “average” weather.

What is Climate Change?

Climate change refers to any significant change in the measures of climate lasting for an extended period of time. In other words, climate change includes major changes in temperature, precipitation, or wind patterns, among other effects, that occur over several decades or longer.

Contributors and Attributions

Modified by Kyle Whittinghill, Melissa Ha, and Rachel Schleiger from the following sources:

- Connie Rye (East Mississippi Community College), Robert Wise (University of Wisconsin, Oshkosh), Vladimir Jurukovski (Suffolk County Community College), Jean DeSaix (University of North Carolina at Chapel Hill), Jung Choi (Georgia Institute of Technology), Yael Avissar (Rhode Island College) among other contributing authors. Original content by OpenStax (CC BY 4.0; Download for free at http://cnx.org/contents/185cbf87-c72...f21b5eabd@9.87).

- Kyle Whittinghill (University of Pittsburgh)
- Contributed by Melissa Ha and Rachel Schleiger Faculty (Biological Sciences) at Yuba College & Butte College
- Climate and the Effects of Global Climate Change from *General Biology* by OpenStax (licensed under CC-BY)
- Climate Change from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Carbon Cycle from *Biology* by John W. Kimball (licensed under CC-BY)
- The Physical World from *Environmental Science: A Canadian Perspective* by Bill Freedman
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0
- The Atmosphere from *AP Environmental Science* by University of California College Prep accessed through CNX.

3.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **44.5: Climate and the Effects of Global Climate Change** by OpenStax is licensed CC BY 4.0.
- **10.2: Ozone Depletion** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

3.2: Structure and Composition of the Atmosphere

Life as we know it on the Earth is entirely dependent on the tenuous layer of gas that clings to the surface of the globe, adding about 1% to its diameter and an insignificant amount to its total mass. And yet the atmosphere serves as the earth's window and protective shield, as a medium for the transport of heat and water, and as source and sink for exchange of carbon, oxygen, and nitrogen with the biosphere. The atmosphere acts as a compressible fluid tied to the earth by gravitation; as a receptor of solar energy and a thermal reservoir, it constitutes the working fluid of a heat engine that transports and redistributes matter and energy over the entire globe. The atmosphere is also a major temporary repository of a number of chemical elements that move in a cyclic manner between the hydrosphere, atmosphere, and the upper lithosphere. Finally, the atmosphere is a site for a large variety of complex photochemically initiated reactions involving both natural and anthropogenic substances.

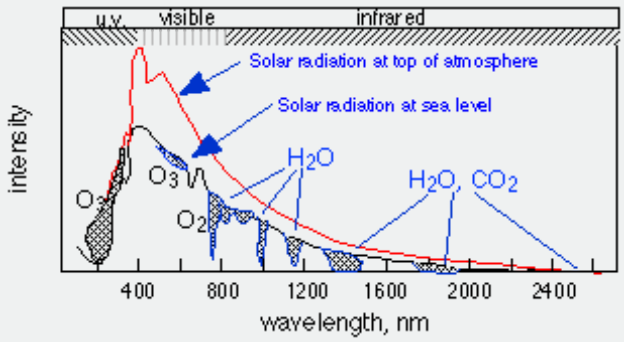
> *On the scale of cubic meters the air is a homogeneous mixture of its constituent gases, but on a larger scale the atmosphere is anything but uniform. Variations of temperature, pressure, and moisture content in the layers of air near the earth's surface give rise to the dynamic effects we know as the weather.*

Although the density of the atmosphere decreases without limit with increasing height, for most practical purposes one can roughly place its upper boundary at about 500 km. However, half the mass of the atmosphere lies within 5 km, and 99.99% within 80 km of the surface. The average atmospheric pressure at sea level is 1.01 x 10^5 pascals, or 1010 millibars. A 1-cm^2 cross section of the earth's surface supports a column weighing 1030 g; the total mass of the atmosphere is about 5.27 x 10^{21} g.

About 80% of the mass of the atmosphere resides in the first 10 km; this well-mixed region of fairly uniform composition is known as the *troposphere*.

Solar irradiation of the Earth

The gases ozone, water vapor, and carbon dioxide are only minor components of the atmosphere, but they exert a huge effect on the Earth by absorbing radiation in the ranges indicated by the shading. Ozone in the upper atmosphere filters out the ultraviolet light below about 360 nm that is destructive of life. O_2, H_2O, CO_2 and CH_4 are "greenhouse" gases that trap some of the heat absorbed from the Sun and prevent it from re-radiating back into space.

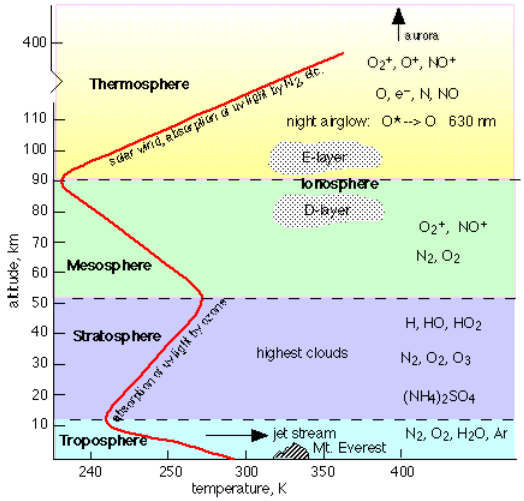


Structure of the atmosphere

We commonly think of gas molecules as moving about in a completely random manner, but the Earth's gravitational field causes downward motions to be very slightly favored so that the molecules in any thin layer of the air collide more frequently with those in the layer below. This gives rise to a pressure gradient that is the most predictable and well known structural characteristic of the atmosphere. This gradient is described by an exponential law which predicts that the atmospheric pressure should decrease by 50% for every 6 km increase in altitude. This law also predicts that the composition of a gas mixture will change with altitude, the lower-molecular weight components being increasingly favored at higher altitudes. However, this gravitational fractionation effect is completely obliterated below about 160 km owing to turbulence and convective flows (winds).

The atmosphere is divided vertically into several major regions which are distinguished by the sign of the temperature gradient. In the lowermost region, the troposphere, the temperature falls with increasing altitude. The major source of heat input into this part of the atmosphere is long-wave radiation from the earth's surface, while the major loss is radiation into space.

At higher elevations the temperature begins to rise with altitude as we move into a region in which heat is produced by exothermic chemical reactions, mainly the decomposition of ozone that is formed photochemically from dioxygen in the stratosphere. At still higher elevations the ozone gives out and the temperature begins to drop; this is the mesosphere, which is finally replaced by the thermosphere which consists largely of a plasma (gaseous ions). This outer section of the atmosphere which extends indefinitely to perhaps 2000 km is heated by absorption of intense u.v. radiation from the Sun and also from the solar wind, a continual rain of electrons, protons, and other particles emitted from the Sun's surface.



Structure of the atmosphere. *The main divisions of the atmosphere are defined by the elevations at which the sign of the temperature gradient changes. The chemical formulas at the right show the major species of interest in the various regions. The shaded D- and E-layers are regions of high ion concentrations that reflect radio waves and are important in long-distance communication.*

Composition of the atmosphere

Except for water vapor, whose atmospheric abundance varies from practically zero up to 4%, the fractions of the major atmospheric components N_2, O_2, and Ar are remarkably uniform below about 100 km. At greater heights, diffusion becomes the principal transport process, and the lighter gases become relatively more abundant. In addition, photochemical processes result in the formation of new species whose high reactivities would preclude their existence in significant concentrations at the higher pressures found at lower elevations.

The atmospheric gases fall into three abundance categories: major, minor, and trace. Nitrogen, the most abundant component, has accumulated over time as a result of its geochemical inertness; a very small fraction of it passes into the other phases as a result of biological activity and natural fixation by lightning. It is believed that denitrifying bacteria in marine sediments may provide the major route for the return of N_2 to the atmosphere. Oxygen is almost entirely of biological origin, and cycles through the hydrosphere, the biosphere, and sedimentary rocks. Argon consists mainly of Ar^{40} which is a decay product of K^{40} in the mantle and crust.

Table 1: Major components

nitrogen	N_2	78.08 %
oxygen	O_2	20.95 %
argon	Ar	0.93 %

The most abundant of the minor gases aside from water vapor is carbon dioxide, about which more will be said below. Next in abundance are neon and helium. Helium is a decay product of radioactive elements in the earth, but neon and the other inert gases are primordial, and have probably been present in their present relative abundances since the earth's formation. Two of the minor gases, ozone and carbon monoxide, have abundances that vary with time and location. A variable abundance implies an imbalance between the rates of formation and removal. In the case of carbon monoxide, whose major source is anthropogenic (a small amount is produced by biological action), the variance is probably due largely to localized differences in fuel consumption, particularly in internal combustion engines. The nature of the carbon monoxide sink (removal mechanism) is not entirely clear; it may be partly microbial.

Table 2: Minor components

water	H_2O	0-4 %
carbon dioxide	CO_2	325 ppm
neon	Ne	18 ppm
helium	He	5 ppm
methane	CH_4	2 ppm
krypton	Kr	1 ppm
hydrogen	H_2	0.5 ppm
nitrous oxide	N_2O	0.3 ppm
carbon monoxide	CO	0.05-0.2 ppm
ozone	O_3	0.02 - 10 ppm
xenon	Xe	0.08 ppm

Table 3:Trace components

ammonia	NH_3	4 ppb
nitrogen oxide	NO	1 ppb
sulfur dioxide	SO_2	1 ppb
hydrogen sulfide	H_2S	0.05 ppb

Ozone

Ozone is formed by the reaction of O_2 with oxygen atoms produced photochemically. As a consequence the abundance of ozone varies with the time of day, the concentration of O atoms from other sources (photochemical smog, for example), and particularly with altitude; at 30 km, the ozone concentration reaches a maximum of 12 ppm.

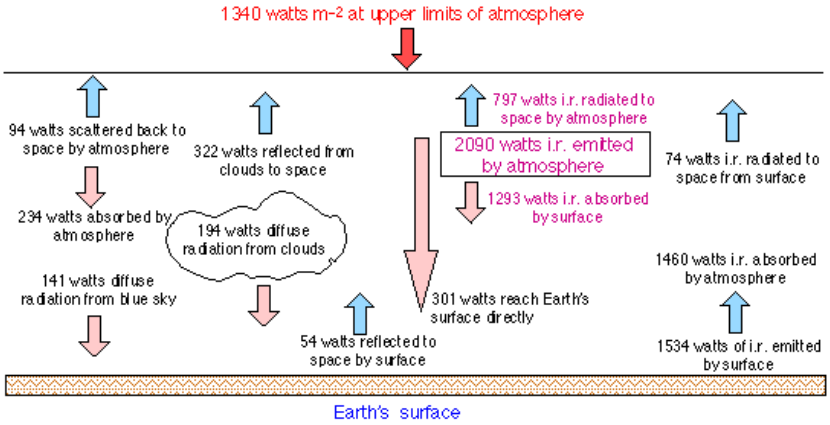
Carbon dioxide

The concentration of atmospheric carbon dioxide, while fairly uniform globally, is increasing at a rate of 0.2-0.7% per year as a result of fossil fuel burning. The present CO_2 content of the atmosphere is about 129 $\times 10^{18}$ g. Most of the CO^2, however, is of natural origin, and represents the smallest part of the total carbonate reservoir that includes oceanic CO_2, HCO_3^-, and carbonate sediments. The latter contain about 600 times as much CO_2 as the atmosphere, and the oceans contain about 50 times as much. These relative amounts are controlled by the rates of the reactions that interconvert the various forms of carbonate.

The surface conditions on the earth are sensitively dependent on the atmospheric CO_2 concentration. This is due mainly to the strong infrared absorption of CO_2, which promotes the absorption and trapping of solar heat (see below). Since CO_2 acts as an acid in aqueous solution, the pH of the oceans is also dependent on the concentration of CO_2 in the atmosphere; it has been estimated that if only 1% of the carbonate presently in sediments were still in the atmosphere, the pH of the oceans would be 5.9, instead of the present 8.2.

Energy balance of the atmosphere and earth

The amount of energy (the solar flux) impinging on the outer part of the atmosphere is 1367 watts m^{-2} . About 30% of this is reflected or scattered back into space by clouds, dust, and the atmospheric gas molecules themselves, and by the earth's surface. About 19% of the radiation is absorbed by clouds or the atmosphere (mainly by and O_3 , but not CO_2), leaving 51% of the incident energy available for absorption by the earth's surface. If one takes into account the uneven illumination of the earth's surface and the small flux of internal heat to the surface, the assumption of thermal equilibrium requires that the earth emit about 240 watts m^{-2}. This corresponds to the power that would be emitted by a black body at 255 K, or –18°C, which is the average temperature of the atmosphere at an altitude of 5 km. The observed mean global surface temperature of the earth is 13°C, and is presumably the temperature required to maintain thermal equilibrium between the earth and the atmosphere.



The greenhouse effect

The energy radiated by the earth has a longer wavelength (maximum 12m) than the incident radiation. Most gases absorb radiation in this range quite efficiently, including those gases such as CO_2 and N_2O that do not absorb the incident radiation. The energy absorbed by atmospheric gases is re-radiated in all directions; some of it therefore escapes into space, but a portion returns to the earth and is reabsorbed, thus raising its temperature.This is commonly called the greenhouse effect. If the amount of an infrared-absorbing gas such as carbon dioxide increases, a larger fraction of the incident solar radiation is trapped, and the mean temperature of the earth will increase.

Any significant increase in the temperature of the oceans would increase the atmospheric concentrations of both water and CO_2, producing the possibility of a runaway process that would be catastrophic from a human perspective. Fossil fuel combustion and deforestation during the last two hundred years have increased the atmospheric CO_2 concentration by 25%, and this increase is continuing. The same combustion processes responsible for the increasing atmospheric CO_2 concentration also introduce considerable quantities of particulate materials into the upper atmosphere. The effect of these would be to scatter more of the incoming solar radiation, reducing the amount that reaches and heats the earth's surface. The extent to which this process counteracts the greenhouse effect is still a matter of controversy; all that is known for sure is that the average temperature of the Earth is increasing.

Nitrous oxide

Carbon dioxide is not the only atmospheric gas of anthropogenic origin that can affect the heat balance of the earth; other examples are SO_2 and N_2O. Nitrous oxide is of particular interest, since its abundance is fairly high, and is increasing at a rate of about 0.5% per year. It is produced mainly by bacteria, and much of the increase is probably connected with introduction of increased nitrate into the environment through agricultural fertilization and sewage disposal. Besides being a strong infrared absorber, N_2O is

photochemically active, and can react with ozone. Any significant depletion of the ozone content of the upper atmosphere would permit more ultraviolet radiation to reach the earth. This would have numerous deleterious effects on present life forms, as well as contributing to a temperature increase. The warming effect attributed to anthropogenic additions of greenhouse gases to the atmosphere is estimated to be about 2 watts per m^2 , or about 1.5% of the 150 watts per m^2trapped by clouds and atmospheric gases. This is a relatively large perturbation compared to the maximum variation in solar output of 0.5 watts per m^2 that has been observed during the past century. Continuation of greenhouse gas emission at present levels for another century could increase the atmospheric warming effect by 6-8 watts per m^2.

> *A less-appreciated side effect of the increase in atmospheric carbon dioxide (and of other plant nutrients such as nitrates) may be reduction in plant species diversity by selectively encouraging the growth of species which are ordinarly held in check by other species that are able to grow well with fewer nutrients. This effect, for which there is already some evidence, could be especially pronounced when the competing species utilize the C_3 and C_4 photosynthetic pathways that differ in their sensitivity to CO_2.*

Contributors and Attributions

Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook

This page titled 3.2: Structure and Composition of the Atmosphere is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

3.3: Origin and Evolution of the Atmosphere

Prebiotic atmosphere

The atmosphere of the Earth (and also of Venus and Mars) is generally believed to have its origin in relatively volatile compounds that were incorporated into the solids from which these planets accreted. Such compounds could include nitrides (a source of N_2), water (which can be taken up by silica, for example), carbides, and hydrogen compounds of nitrogen and carbon. Many of these compounds (and also some noble gases) can form clathrate complexes with water and some minerals which are fairly stable at low temperatures. The high temperatures developed during the later stages of accretion as well as subsequent heating produced by decay of radioactive elements presumably released these gases the surface. Even at the present time, large amounts of CO_2 , water vapor, N_2, HCl, SO_2 and H_2S are emitted from volcanos. The more reactive of these gases would be selectively removed from the atmosphere by reaction with surface rocks or dissolution in the ocean, leaving an atmosphere enriched in its present major components with the exception of oxygen which is discussed in the next section.

Any hydrogen present would tend to escape into space, causing the atmosphere to gradually became less reducing. However there is now some doubt that hydrogen and other volatiles (mainly the inert gases) were present in the newly accreted planets in anything like their cosmic abundance. The main evidence against this is the observation that gases such as helium, neon and argon, which are among the ten most abundant elements in the universe, are depleted on the earth by factors of 10^{-7} to 10^{-11} . This implies that there was a selective removal of these volatiles prior to or during the planetary accretion process.

The overall oxidation state of the earth's mantle is not consistent with what one would expect from equilibration with highly reduced volatiles, and there is no evidence to suggest that the composition of the mantle has not remained the same. If this is correct, then the primitive atmosphere may well have had about the same composition as the gases emitted by volcanos at the present time. These consist mainly of water and CO_2, together with small amounts of N_2, H_2, H_2S, SO_2, CO, CH_4, NH_3, HCl and HF. If water vapor was a major component of outgassing of the accreted earth, it must have condensed quite rapidly into rain. Any significant concentration of water vapor in the atmosphere would have led to a runaway greenhouse effect, resulting in temperatures as high as 400°C.

Origin of Atmospheric Oxygen

Free oxygen is never more than a trace component of most planetary atmospheres. Thermodynamically, oxygen is much happier when combined with other elements as oxides; the pressure of O_2 in equilibrium with basaltic magmas is only about 10^{-7} atm. Photochemical decomposition of gaseous oxides in the upper atmosphere is the major source of O_2 on most planets. On Venus, for example, CO_2 is broken down into CO and O_2 . On the earth, the major inorganic source of O_2 is the photolysis of water vapor; most of the resulting hydrogen escapes into space, allowing the O2 concentration to build up. An estimated 2 $\times$ 10^{11} g of O_2 per year is generated in this way. Integrated over the earth's history, this amounts to less than 3% of the present oxygen abundance. The partial pressure of O_2 in the prebiotic atmosphere is estimated to be no more than 10^{-3} atm, and may have been several orders of magnitude less. The major source of atmospheric oxygen on the earth is photosynthesis carried out by green plants and certain bacteria:

$$H_2O + CO_2 \longrightarrow (CH_2O)_x + O_2 \tag{3.3.1}$$

A historical view of the buildup of atmospheric oxygen concentration since the beginning of the sedimentary record (3.7 $\times$ 10^9 ybp) can be worked out by making use of the fact that the carbon in the product of the above reaction has a slightly lower C^{13} content than does carbon of inorganic origin. Isotopic analysis of carbon-containing sediments thus provides a measure of the amounts of photosynthetic O_2 produced at various times in the past.

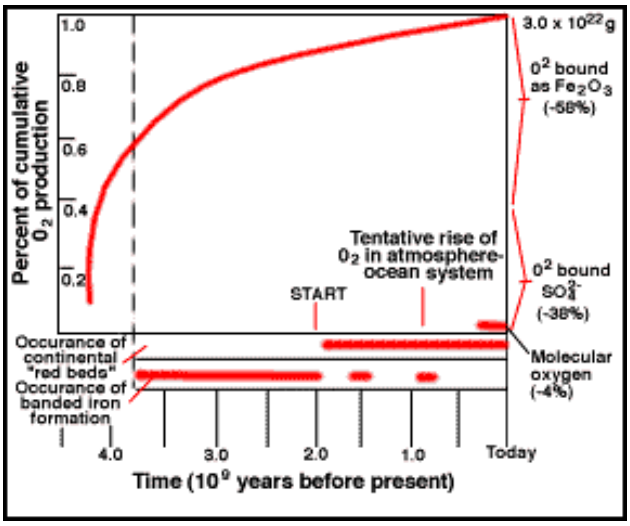


Figure 3.3.1: ***Evolution of atmospheric oxygen content.*** *Note carefully that the curve plots cumulative O_2 production, but that until a few hundred million years ago, most of this was taken up by Fe(II) compounds in the crust and by reduced sulfur; only after this massive "oxygen sink" became filled did free O_2 begin to accumulate in the atmosphere.*

Carbon Dioxide

Carbon dioxide has probably always been present in the atmosphere in the relatively small absolute amounts now observed (around 54 x 10^{15} mol = 54 Pmol). The reaction of CO_2 with silicate-containing rocks to form precambrian limestones suggest a possible moderating influence on its atmospheric concentration throughout geological time.

$$\mathrm{CaSiO_3 + CO_2 \rightarrow CaCO_3 + SiO_2} \tag{3.3.2}$$

About ten percent of the atmospheric CO_2 is taken up each year by photosynthesis. Of this, all except 0.05 percent is returned by *respiration*, almost entirely due to microorganisms. The remainder leaks into the slow part of the geochemical cycle, mostly as buried carbonate sediments.

Since the advent of the industrial revolution in 1860, the amount of CO_2 in the atmosphere has been increasing. Isotopic analysis shows that most of this has been due to fossil-fuel combustion; in recent years, the mass of carbon released to the atmosphere by this means has been more than ten times the estimated rate of natural removal into sediments. The large-scale destruction of tropical forests, which has accelerated greatly in recent years, is believed to exacerbate this effect by removing a temporary sink for CO_2.

The oceans have a large absorptive capacity for CO_2 owing to its reaction with carbonate:

$$\mathrm{CO_2 + {CO_3}^{2-} \rightleftharpoons 2\,HCO_3^-} \tag{3.3.3}$$

There is about 60 times as much inorganic carbon in the oceans as in the atmosphere. However, efficient transfer takes place only into the topmost (100 m) wind-mixed layer of the ocean, which contains only about one atmosphere equivalent of CO_2. Further uptake is limited by the very slow transport of water into the deep ocean, which takes around 1000 years. For this reason, the buffering effect of the oceans on atmospheric CO_2 is not very effective; only about ten percent of the added CO_2 is taken up by the oceans.

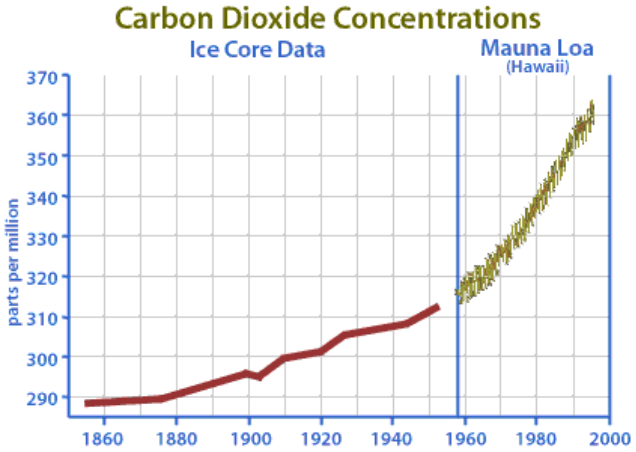


Figure. This set of plots, taken from an ETE page with data from NASA, shows the dramatic increase in atmospheric CO_2since the beginning of the industrial age. The "squiggles" on the Mauna Loa data reflect seasonal variations; more $CO_2$is taken up during the longer-daylight periods of the summer.

Contributors and Attributions

- Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook

---

This page titled 3.3: Origin and Evolution of the Atmosphere is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

SECTION OVERVIEW

3.4: General Circulation

A spatial imbalance between radiative inputs and outputs exists for the earth-ocean-atmosphere system. The earth loses energy at all latitudes due to outgoing infrared (IR) radiation. Near the tropics, more solar radiation enters than IR leaves, hence there is a net input of radiative energy. Near Earth's poles, incoming solar radiation is too weak to totally offset the IR cooling, allowing a net loss of energy. The result is **differential heating**, creating warm equatorial air and cold polar air (Fig. 11.1a).

This imbalance drives the global-scale **general circulation** of winds. Such a circulation is a fluiddynamical analogy to **Le Chatelier's Principle** of chemistry. Namely, an imbalanced system reacts in a way to partially counteract the imbalance. The continued destabilization by radiation causes a general circulation of winds that is unceasing.

Because buoyancy causes warmer air to rise and colder air to sink, you might guess that equator-topole overturning would exist (Fig. 11.1b). Instead, the real general circulation has three bands of circulations in the Northern Hemisphere (Fig. 11.1c), and three in the Southern. In this chapter, we will identify characteristics of the general circulation, explain why they exist, and learn how they work.

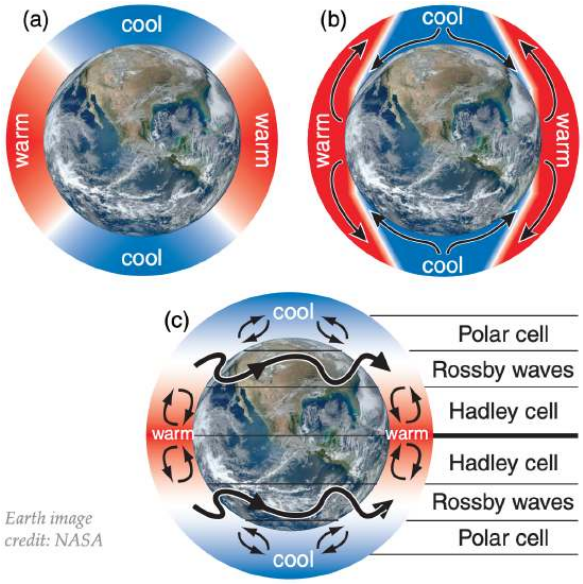


Figure 11.1 Radiative imbalances create (a) warm tropics and cold poles, inducing (b) buoyant circulations. Add Earth's rotation, and (c) three circulation bands form in each hemisphere.

3.4.1: A Simplified Description of the Global Circulation

3.4.2: Radiative Differential Heating

3.4.3: Pressure Profiles

3.4.4: Geostrophic Wind and Geostrophic Adjustment

3.4.5: Thermal Wind Effect

3.4.6: Horizontal Circulation

3.4.7: Extratropical Ridges and Troughs (Rossby Waves)

3.4.8: Three-band Global Circulation

3.4.9: Ekman Spiral of Ocean Currents

3.4.10: Explaining the General Circulation

3.4.11: Jet Streams

3.4.12: Types of Vorticity

3.4.13: Key Terms

3.4.14: Review

3.4.15: Homework Exercises

This page titled 3.4: General Circulation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

3.4.1: A Simplified Description of the Global Circulation

This section summarizes “what” happens. The subsequent sections explain “why” and “how”.

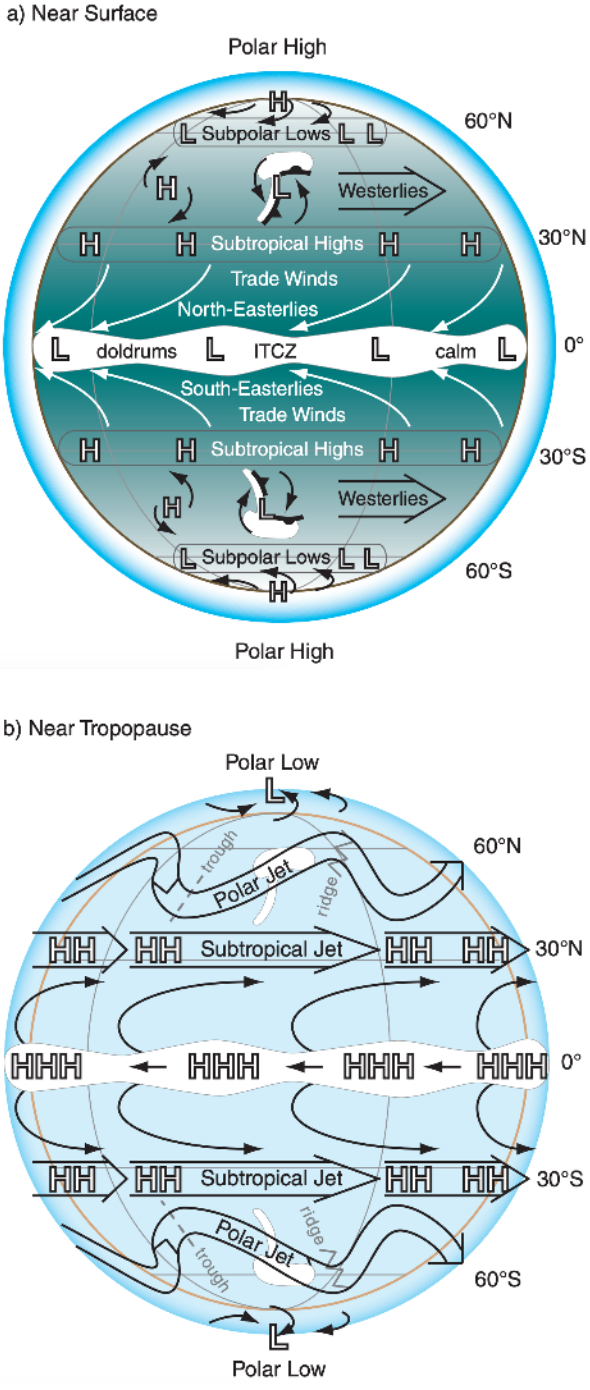
Consider a hypothetical rotating planet with no contrast between continents and oceans. The **climatological average** (average over 30 years; see the Climate chapter) winds in such a simplified planet would have characteristics as sketched in Figs. 11.3. Actual winds on any day could differ from this climatological average due to transient weather systems that perturb the average flow. Also, monthly-average conditions tend to shift toward the summer hemisphere (e.g., the circulation bands shift northward during April through September).

11.2.1. Near the Surface

Near-surface average winds are sketched in Fig. 11.3a. At low latitudes are broad bands of persistent easterly winds ($U \approx -7$ m s^{-1}) called **trade winds**, named because the **easterlies** allowed sailing ships to conduct transoceanic trade in the old days.

These trade winds also blow toward the equator from both hemispheres, and the equatorial belt of convergence is called the **intertropical convergence zone** (**ITCZ**). On average, the air at the ITCZ is hot and humid, with low pressure, strong upward air motion, heavy convective (thunderstorm) precipitation, and light to calm winds except in thunderstorms. This **equatorial trough** (low-pressure belt) was called the **doldrums** by sailors whose sailing ships were becalmed there for many days.

Figure 11.3 Simplified global circulation in the troposphere: (a) near the surface, and (b) near the tropopause. H and L indicate high and low pressures, and HHH means very strong high pressure. White indicates precipitating clouds. CAUTION: high and low pressures are RELATIVE to the average pressure AT THE SAME ALTITUDE. All absolute pressures near the tropopause are actually lower than the near-surface pressures.



At 30° latitude are belts of high surface pressure called subtropical highs (Fig. 11.3a). In these belts are hot, dry, cloud-free air descending from higher in the troposphere. Surface winds in these belts are also calm on average. In the old days, becalmed sailing ships would often run short of drinking water, causing horses on board to die and be thrown overboard. Hence, sailors called these miserable places the horse latitudes. On land, many of the world's deserts are near these latitudes.

In mid-latitudes are transient centers of low pressure (**mid-latitude cyclones, L**) and high pressure (**anticyclones, H**). Winds around lows **converge** (come together) and circulate **cyclonically** — counterclockwise in the N. Hemisphere, and clockwise in the

S. Hemisphere. Winds around highs **diverge** (spread out) and rotate **anticyclonically** — clockwise in the N. Hemisphere, and counterclockwise in the S. Hemisphere. The cyclones are regions of bad weather (clouds, rain, high humidity, strong winds) and fronts. The anticyclones are regions of good weather (clear skies or fair-weather clouds, no precipitation, dry air, and light winds).

The high- and low-pressure centers move on average from west to east, driven by large-scale winds from the west. Although these **westerlies** dominate the general circulation at mid-latitudes, the surface winds are quite variable in time and space due to the sum of the westerlies plus the transient circulations around the highs and lows.

Near 60° latitude are belts of low surface pressure called **subpolar lows**. Along these belts are light to calm winds, upward air motion, clouds, cool temperatures, and precipitation (as snow in winter).

Near each pole is a climatological region of high pressure called a **polar high**. In these regions are often clear skies, cold dry descending air, light winds, and little snowfall. Between each polar high (at 90°) and the subpolar low (at 60°) is a belt of weak easterly winds, called the **polar easterlies**.

11.2.2. Upper-troposphere

The stratosphere is strongly statically stable, and acts like a lid to the troposphere. Thus, vertical circulations associated with our weather are mostly trapped within the troposphere. These vertical circulations couple the average near-surface winds with the average upper-tropospheric (near the tropopause) winds described here (Fig. 11.3b).

In the tropics is a belt of very strong equatorial high pressure along the tops of the ITCZ thunderstorms. Air in this belt blows from the east, due to easterly inertia from the trade winds being carried upward in the thunderstorm convection. Diverging from this belt are winds that blow toward the north in the N. Hemisphere, and toward the south in the S. Hemisphere. As these winds move away from the equator, they turn to have an increasingly westerly component as they approach 30° latitude.

Near 30° latitude in each hemisphere is a persistent belt of strong westerly winds at the tropopause called the **subtropical jet**. This jet meanders north and south a bit. Pressure here is very high, but not as high as over the equator.

In mid-latitudes at the tropopause is another belt of strong westerly winds called the **polar jet**. The centerline of the polar jet meanders north and south, resulting in a wave-like shape called a **Rossby wave** (or **planetary wave**), as sketched in Fig. 11.1c. The equatorward portions of the wave are known as low-pressure **troughs**, and poleward portions are known as high-pressure **ridges**. These ridges and troughs are very transient, and generally shift from west to east relative to the ground.

Near 60° at the tropopause is a belt of low to medium pressure. At each pole is a low-pressure center near the tropopause, with winds at high latitudes generally blowing from the west causing a cyclonic circulation around the **polar low**. Thus, contrary to near-surface conditions, the near-tropopause average winds blow from the west at all latitudes (except near the equator).

11.2.3. Vertical Circulations

Vertical circulations of warm rising air in the tropics and descending air in the subtropics are called **Hadley cells** or **Hadley circulations** (Fig. 11.4). At the bottom of the Hadley cell are the trade winds. At the top, near the tropopause, are divergent winds. The updraft portion of the Hadley circulation often contains thunderstorms and heavy precipitation at the ITCZ. This vigorous convection in the troposphere causes a high tropopause (15 - 18 km altitude) and a belt of heavy rain in the tropics.

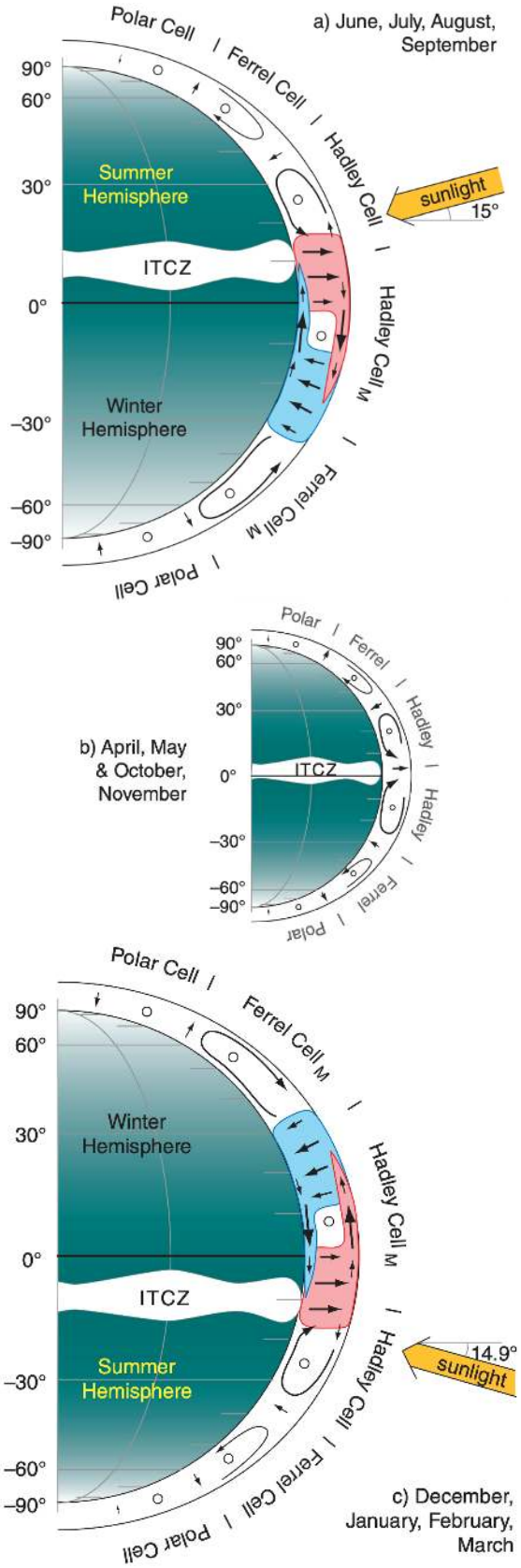


Figure 11.4 (at left) Vertical cross section of Earth's global circulation in the troposphere. (a) N. Hemisphere summer. (b) Transition months. (c) S. Hemisphere summer. The major (subscript $_M$) Hadley cell is colored in red and blue. Minor circulations have no subscript.

https://geo.libretexts.org/@go/page/20931

The summer- and winter-hemisphere Hadley cells are strongly asymmetric (Fig. 11.4). The major Hadley circulation (denoted with subscript "M") crosses the equator, with rising air in the summer hemisphere and descending air in the winter hemisphere. The updraft is often between 0° and 15° latitudes in the summer hemisphere, and has average core vertical velocities of 6 mm s^{-1}. The broader downdraft is often found between 10° and 30° latitudes in the winter hemisphere, with average velocity of about –4 mm s^{-1} in downdraft centers. Connecting the up- and downdrafts are meridional wind components of 3 m s^{-1} at the cell top & bottom.

The major Hadley cell changes direction and shifts position between summer and winter. During June-July-August-September, the average solar declination angle is 15°N, and the updraft is in the Northern Hemisphere (Fig. 11.4a). Out of these four months, the most well-defined circulation occurs in August and September. At this time, the ITCZ is centered at about 9°N, but varies with longitude.

During December-January-February-March, the average solar declination angle is 14.9°S, and the major updraft is in the Southern Hemisphere (Fig. 11.4c). Out of these four months, the strongest circulation is during February and March, and the ITCZ is centered at roughly 6°S, but varies with longitude. The major Hadley cell transports significant heat away from the tropics, and also from the summer to the winter hemisphere.

During the transition months (April-May and October-November) between summer and winter, the Hadley circulation has nearly symmetric Hadley cells in both hemispheres (Fig. 11.4b). During this transition, the intensities of the Hadley circulations are weak.

When averaged over the whole year, the strong but reversing major Hadley circulation partially cancels itself, resulting in an annual average circulation that is somewhat weak and looks like Fig. 11.4b. This weak annual average is deceiving, and does not reflect the true movement of heat, moisture, and momentum by the winds. Hence, climate experts prefer to look at months JJA and DJF separately to give **seasonal averages**.

In the winter hemisphere, one or more jet streams circle the earth at mid-latitudes while meandering north and south as Rossby waves (Fig. 11.1c). When averaged around latitude bands, the net effect is a weak vertical circulation called a **Ferrel cell**. At high latitudes is a modest **polar cell**.

In the summer hemisphere, all the circulations are weaker. There are minor Hadley and Ferrel cells (Fig. 11.4). Summer-hemisphere circulations are weaker because the temperature contrast between the tropics and poles are weaker.

11.2.4. Monsoonal Circulations

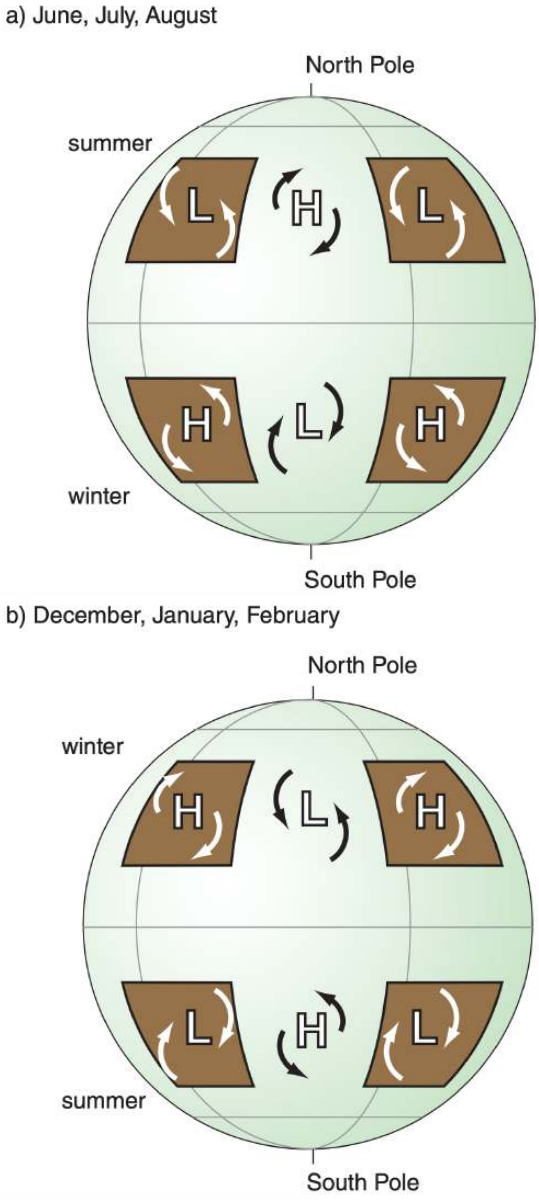


Figure 11.5 Idealized seasonal-average monsoon circulations near the surface. Continents are shaded dark brown; oceans are light green. H and L are surface high- and low-pressure centers.

Monsoon circulations are continental-scale circulations driven by continent-ocean temperature contrasts, as sketched in Figs. 11.5. In summer, high-pressure centers (anticyclones) are over the relatively warm oceans, and low-pressure centers (cyclones) are over the hotter continents. In winter, low-pressure centers are over the cool oceans, and high-pressure centers are over the colder continents.

These monsoon circulations represent average conditions over a season. The actual weather on any given day can be variable, and can deviate from these seasonal averages.

Our Earth has a complex arrangement of continents and oceans. As a result, seasonally-varying monsoonal circulations are superimposed on the seasonally-varying planetary-scale circulation to yield a complex and varying global-circulation pattern.

At this point, you have a descriptive understanding of the global circulation. But what drives it?

This page titled 3.4.1: A Simplified Description of the Global Circulation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.2: A Simplified Description of the Global Circulation** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.2: Radiative Differential Heating

The general circulation is driven by differential heating. Incoming solar radiation (**insolation**) nearly balances the outgoing infrared (IR) radiation when averaged over the whole globe. However, at different latitudes are significant imbalances (Fig. 11.6), which cause the differential heating.

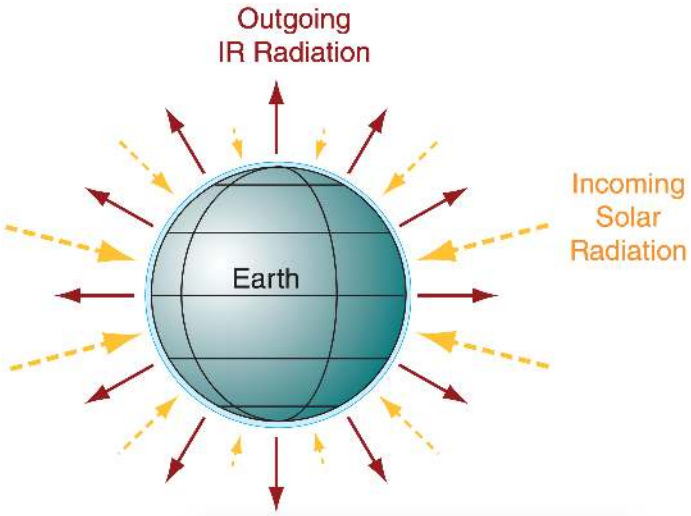


Figure 11.6 Annual average incoming solar radiation (yellow dashed arrows) and outgoing infrared (IR) radiation (solid dark red arrows), where arrow length indicates relative magnitude. [Because the Earth rotates and exposes all locations to the sun at one time or another during the year, the incoming solar radiation is sketched as approaching all locations on the Earth's surface.]

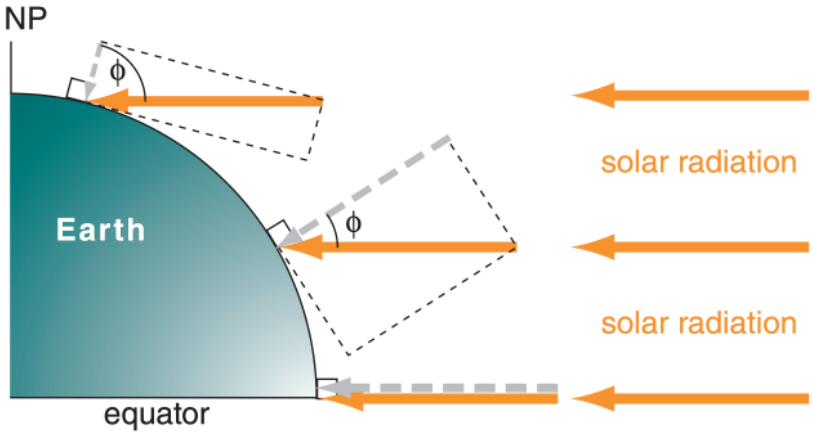


Figure 11.7 Of the solar radiation approaching the Earth (thick solid yellow arrows), the component (dashed grey arrow) that is perpendicular to the top of the atmosphere is proportional to the cosine of the latitude ϕ (during the equinox).

Recall from the Solar & Infrared Radiation chapter that the flux of solar radiation incident on the top of the atmosphere depends more or less on the cosine of the latitude, as shown in Fig. 11.7. The component of the incident ray of sunlight that is perpendicular to the Earth's surface is small in polar regions, but larger toward the equator (grey dashed arrows in Figs. 11.6 and 11.7). The incoming energy adds heat to the Earth-atmosphere-ocean system.

Heat is lost due to infrared (IR) radiation emitted from the Earth-ocean-atmosphere system to space. Since all locations near the surface in the Earthocean-atmosphere system are relatively warm compared to absolute zero, the Stefan-Boltzmann law from the Solar & Infrared Radiation chapter tells us that the emission rates are also more or less uniform around the Earth. This is sketched by the solid red arrows in Fig. 11.6.

Thus, at low latitudes, more solar radiation is absorbed than leaves as IR, causing net warming. At high latitudes, the opposite is true: IR radiative losses exceed solar heating, causing net cooling. This differential heating drives the global circulation.

The general circulation can't instantly eliminate all the global north-south temperature differences. What remains is a meridional temperature gradient — the focus of the next subsection.

11.3.1. North-South Temperature Gradient

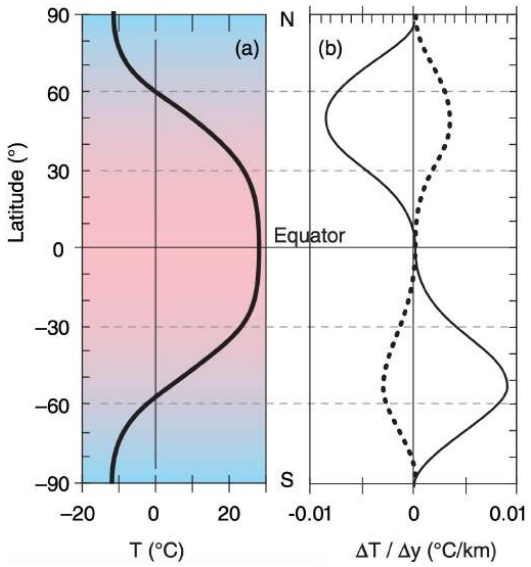


Figure 11.8 Idealized variation of annual-average temperature (a) and meridional temperature gradient (b) with latitude, averaged around latitude belts (i.e., zonal averages). Solid curve represents Earth's surface (= sea-level because the toy model neglects terrain) and dotted curve is for 15 km altitude above sea level.

To create a first-order toy model, neglect monthly variations, monsoonal variations and mountains. Instead, focus on surface temperatures averaged around separate latitude belts and over one year. Those latitude belts near the equator are warmer, and those near the poles are colder (Fig. 11.8a). One equation that roughly approximates the variation of zonally-averaged surface temperature T with latitude ϕ is:

Erroneous nesting of equation structures

where $a \approx -12°C$ is an offset and $b \approx 40°C$ is a temperature difference between equator and pole.

The equation above applies only to the surface. At higher altitudes the north-south temperature difference is smaller, and even becomes negative in the stratosphere (i.e., warm over the poles and cold over the tropics). To account for this altitude variation, b cannot be a constant. Instead, use:

Erroneous nesting of equation structures

where average tropospheric depth is $z_T \approx 11$ km, parameter $b_1 = 40°C$, and z is height above the surface.

A similarly crude but useful generalization of parameter a can be made so that it too changes with altitude z above sea level:

Erroneous nesting of equation structures

where $a_1 = -12°C$ and $\gamma = 3.14\ °C\ km^{-1}$.

In Fig. 11.8a, the temperature curve looks like a flattened cosine wave. The flattened curve indicates somewhat uniformly warm temperatures between ±30° latitude, caused the strong mixing and heat transport by the Hadley circulations.

With uniform tropical temperatures, the remaining change to colder temperature is pushed to the mid-latitude belts. The slope of the Fig. 11.8a curve is plotted in Fig. 11.8b. This slope is the north-south **(meridional) temperature gradient**:

Erroneous nesting of equation structures

where y is distance in the north-south direction, $c = 1.18 \times 10^{-3}\ km^{-1}$ is a constant valid at all heights, and b is given by eq. (11.2) which causes the gradient to change sign at higher altitudes (e.g., at z = 15 km).

Because the meridional temperature gradient results from the interplay of differential radiative heating and advection by the global circulation, let us now look at radiative forcings in more detail.

Sample Application

What is the value of annual zonal average temperature and meridional temperature gradient at 50°S latitude for the surface and for z = 15 km.

Find the Answer

Given: ϕ = –50°, (a) z = 0 (surface), and (b) z = 15 km

Find: T_{sfc} = ? °C, ΔT/Δy = ? °C km^{-1}

a) z = 0. Apply eq. (11.2): b = (40°C)·(1–0) = 40°C.

Apply eq. (11.3): a = (–12°C)–0°C = –12°C.

Apply eq. (11.1) :

$$T_o \approx -12^\circ\mathrm{C} + (40^\circ\mathrm{C}) \cdot \left[\cos^3(-50^\circ) \cdot \left(1 + \frac{3}{2} \cdot \sin^2(-50^\circ)\right)\right] = 7.97^\circ\mathrm{C}$$

Apply eq. (11.4):

$$\frac{\Delta T}{\Delta y} \approx -(40^\circ\mathrm{C}) \cdot \left(1.18 \times 10^{-3}\right) \cdot \cos^2(-50^\circ) \cdot \sin^3(-50^\circ) = 0.0087^\circ\mathrm{C km}^{-1}$$

b) z = 15 km.

Apply eq. (11.2): b = (40°C)·[1–(15/11)] = –14.55°C.

Apply (11.3): a = (–12°C)–(3.14°C/km)·(15km) = –59.1°C.

Apply eq. (11.1) :

$$T_{15km} \approx -59.1^\circ\mathrm{C} - 14.6^\circ\mathrm{C} \cdot \left[\cos^3(-50^\circ) \cdot \left(1 + \frac{3}{2} \cdot \sin^2(-50^\circ)\right)\right] = -66.4^\circ\mathrm{C}$$

Apply eq. (11.4) at z = 15 km:

$$\frac{\Delta T}{\Delta y} \approx +(14.55^\circ\mathrm{C}) \cdot \left(1.18 \times 10^{-3}\right) \cdot \cos^2(-50^\circ) \cdot \sin^3(-50^\circ) = -0.0032^\circ\mathrm{C km}^{-1}$$

Check: Phys. & units reasonable. Agrees with Fig. 11.8 for Southern Hemisphere.

Exposition: Cold at z = 15 km even though warm at z = 0. Gradient signs would be opposite in N. Hem.

11.3.2. Global Radiation Budgets

HIGHER MATH • Derivation of the NorthSouth Temperature Gradient

The goal is to find ∂T/∂y for the toy model.

a) First, expand the derivative: Erroneous nesting of equation structures

We will look at factors ∂T/∂ϕ and ∂ϕ/∂y separately:

b) Factor ∂ϕ/∂y describes the meridional gradient of latitude. Consider a circumference of the Earth that passes through both poles. The total latitude change around this circle is Δϕ = 2π radians. The total circumference of this circle is Δy = 2πR for average Earth radius of R = 6371 km. Hence:

Erroneous nesting of equation structures

c) For factor ∂T/∂ϕ, start with eq. (11.1) for the toy model:

Erroneous nesting of equation structures

and take its derivative vs. latitude: $\frac{\partial T}{\partial \phi} = b \cdot \left(\frac{3}{2}\right)$.

$$\left[(2\sin\phi \cdot \cos\phi)\cos^3\phi - 3\left(\frac{2}{3} + \sin^2\phi\right)\cos^2\phi \cdot \sin\phi\right]$$

Next, take the common term ($\sin\phi \cdot \cos^2\phi$) out of [ ]:

$$\frac{\partial T}{\partial \phi} = b\left(\frac{3}{2}\right)\sin\phi \cdot \cos^2\phi \cdot [2\cos^2\phi - 2 - 3\sin^2\phi]$$

Use the trig. identity: $\cos_2\phi = 1 - \sin 2\phi$.

Hence: $2\cos^2\phi = 2 - 2\sin^2\phi$. Substituting this in into the previous full-line equation gives:

Erroneous nesting of equation structures

d) Plug equations (b) and (c) back into eq. (a):

Erroneous nesting of equation structures

Define:

$$c = \left(\frac{15}{2}\cdot\frac{1}{R}\right) = \left(\frac{15}{2}\cdot\frac{1}{6371\text{km}}\right) = 1.177\times10^{-3}\text{km}^{-1}$$

Thus, the final answer is:

Erroneous nesting of equation structures

Check: Fig. 11.8 shows the curves that were calculated from eqs. (11.1) and (11.4). The fact that the sign and shape of the curve for $\Delta T/\Delta y$ is consistent with the curve for T(y) suggests the answer is reasonable.

Alert: Eqs. (11.4) and (11.1) are based on a highly idealized "toy model" of the real atmosphere. They were designed only to illustrate first-order effects.

11.3.2.1. Incoming Solar Radiation

Because of the tilt of the Earth's axis and the change of seasons, the actual flux of incoming solar radiation is not as simple as was sketched in Fig. 11.6. But this complication was already discussed in the Solar & Infrared Radiation chapter, where we saw an equation to calculate the incoming solar radiation (**insolation**) as a function of latitude and day. The resulting insolation figure is reproduced below (Fig. 11.9a).

If you take the spreadsheet data from the Solar & Infrared Radiation chapter that was used to make this figure, and average rows of data (i.e., average over all months for any one latitude), you can find the annual average insolation E_{insol} for each latitude (Fig. 11.9b). Insolation in polar regions is not small.

The curve in Fig. 11.9b is simple, and in the spirit of a toy model can be nicely approximated by:

Erroneous nesting of equation structures

where the empirical parameters are E_o = 298 W m^{-2}, E_1 = 123 W m^{-2}, and ϕ is latitude. This curve and the data points it approximates are plotted in Fig. 11.10.

But not all the radiation incident on the top of the atmosphere is absorbed by the Earth-ocean-atmosphere system. Some is reflected back into space from snow and ice on the surface, from the oceans, and from light-colored land. Some is reflected from cloud top. Some is scattered off of air molecules.

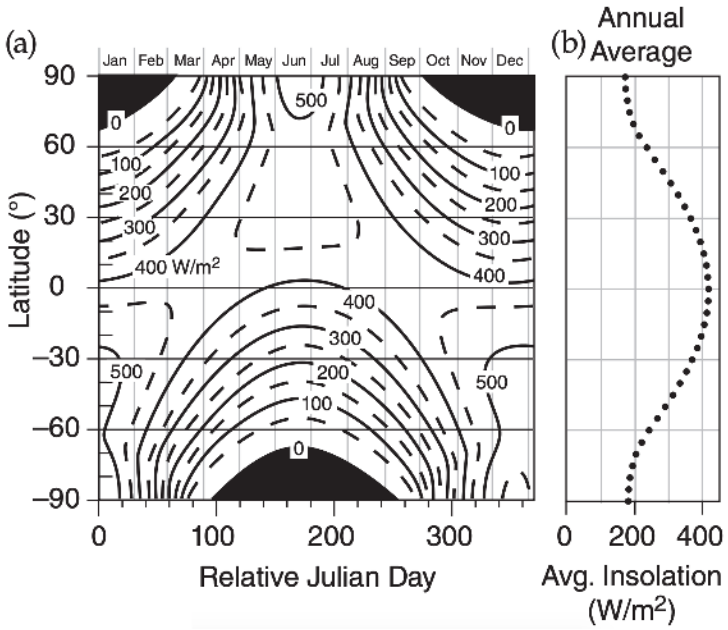


Figure 11.9 (a) Solar radiation (W m^{-2}) incident on the top of the atmosphere for different latitudes and months (copied from the Solar & Infrared Radiation chapter). The slight asymmetry about the equator is because Earth's orbit is closer to the sun during S. Hemisphere summer. (b) Meridional variation of insolation, found by averaging the data from the left figure over all months for each separate latitude (i.e., averages for each row of data).

The amount of insolation that is NOT absorbed is surprisingly constant with latitude at about $E_2 \approx 110$ W m^{-2}. Thus, the amount that IS absorbed is:

Erroneous nesting of equation structures

where E_{in} is the incoming flux (W m^{-2}) of solar radiation absorbed into the Earth-ocean-atmosphere system (Fig. 11.10). This absorbed radiation causes heating.

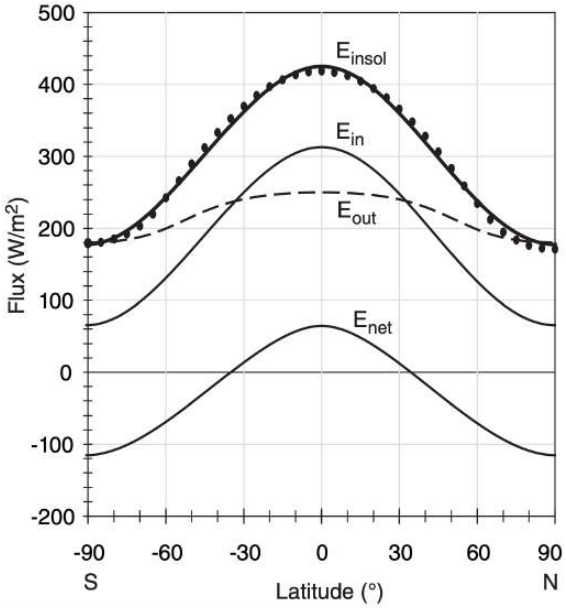


Figure 11.10 Data points are insolation vs. latitude from Fig. 11.9b. Eq. 11.5 approximates this insolation E_{insol} (thick black line). E_{in} is the solar radiation that is absorbed (thin solid line, from eq. 11.6). E_{out} is outgoing terrestrial (IR) radiation (dashed; from eq. 11.7). Net flux $E_{net} = E_{in} - E_{out}$. Positive Enet causes heating; negative causes cooling.

Sample Application

Estimate the annual average solar energy absorbed at the latitude of the Eiffel Tower in Paris, France.

Find the Answer

Given: $\phi = 48.8590°$ (at the Eiffel tower)

Find: $E_{in} = ?\ W\ m^{-2}$

Use eq. (11.5):

$E_{insol} = (298\ W\ m^{-2}) + (123\ W\ m^{-2})\cdot\cos(2 \cdot 48.8590°)$

$= (298\ W\ m^{-2}) – (16.5\ W\ m^{-2}) = 281.5\ W\ m^{-2}$

Use eq. (11.6):

$E_{in} = 281.5\ W\ m^{-2} – 110.0\ W\ m^{-2} = 171.5\ W\ m^{-2}$

Check: Units OK. Agrees with Fig. 11.10.

Exposition: The actual annual average E_{in} at the Eiffel Tower would probably differ from this zonal avg.

11.3.2.2. Outgoing Terrestrial Radiation

As you learned in the Satellites & Radar chapter, infrared radiation emission and absorption in the atmosphere are very complex. At some wavelengths the atmosphere is mostly transparent, while at others it is mostly opaque. Thus, some of the IR emissions to space are from the Earth's surface, some from cloud top, and some from air at middle altitudes in the atmosphere.

In the spirit of a toy model, suppose that the net IR emissions are characteristic of the absolute temperature T_m near the middle of the troposphere (at about $z_m = 5.5$ km). Approximate the outbound flux of radiation E_{out} (averaged over a year and averaged around latitude belts) by the Stefan-Boltzmann law (see the Solar & Infrared Radiation chapter):

Erroneous nesting of equation structures

where the effective emissivity is $\varepsilon \approx 0.9$ (see the Climate chapter), and the Stefan-Boltzmann constant is $\sigma_{SB} = 5.67 \times 10^{-8}\ W \cdot m^{-2} \cdot K^{-4}$. When you use $z = z_m = 5.5$ km in eqs. (11.1 - 11.3) to get T_m vs. latitude for use in eq. (11.7), the result is E_{out} vs. ϕ, as plotted in Fig. 11.10.

11.3.2.3. Net Radiation

For an air column over any square meter of the Earth's surface, the radiative input minus output gives the net radiative flux:

Erroneous nesting of equation structures

which is plotted in Fig. 11.10 for our toy model.

Sample Application

What is E_{net} at the Eiffel Tower latitude?

Find the Answer

Given: $\phi = 48.8590°$, $z = 5.5$ km, $E_{in} = 171.5\ W\ m^{-2}$ from previous Sample Application

Find: $E_{net} = ?\ W\ m^{-2}$

Apply eq. (11.2): $b = (40°C)\cdot[\ 1 – (5.5\ km/11\ km)] = 20°C$

Apply eq.(11.3): $a=(-12°C)–(3.14°C\ km^{-1})\cdot(5.5\ km) = -29.27°C$

Apply eq. (11.1): $T_m = –18.73\ °C = 254.5\ K$

Apply eq.(11.7): $E_{out}=(0.9)\cdot(5.67 \times 10^{-8} W \cdot m^{-2} \cdot K^{-4})\cdot(254.5\ K)^4 = 213.8\ W\ m^{-2}$

Apply eq. (11.8): $E_{net} = (171.5\ W\ m^{-2}) – (213.8\ W\ m^{-2}) =$ **$\underline{–42.3\ W\ m^{-2}}$**

Check: Units OK. Agrees with Fig. 11.10

Exposition: The net radiative heat loss at Paris latitude must be compensated by winds blowing heat in.

11.3.3. Radiative Forcing by Latitude Belt

Do you notice anything unreasonable about E_{net} in Fig. 11.10? It appears that the negative area under the curve is much greater than the positive area, which would cause the Earth to get colder and colder — an effect that is not observed.

Don't despair. Eq. (11.8) is correct, but we must remember that the circumference $[2\pi \cdot R_{Earth} \cdot \cos(\phi)]$ of a parallel (a constant latitude circle) is smaller near the poles than near the equator. ϕ is latitude and the average Earth radius R_{Earth} is 6371 km.

The solution is to multiply eqs. (11.6 - 11.8) by the circumference of a parallel. The result

Erroneous nesting of equation structures

can be applied for $E = E_{in}$, or $E = E_{out}$, or $E = E_{net}$. This converts from E in units of W m^{-2} to $E\phi$ in units of W m^{-1}, where the distance is north-south distance. You can interpret $E\phi$ as the power being transferred to/from a one-meter-wide sidewalk that encircles the Earth along a parallel.

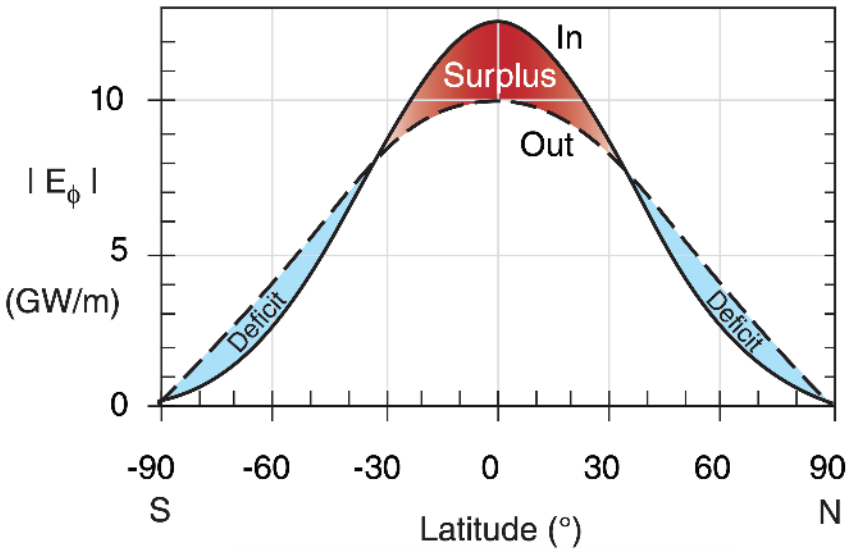


Figure 11.11 Zonally-integrated radiative forcings for absorbed incoming solar radiation (solid line) and emitted net outgoing terrestrial (IR) radiation (dashed line). The surplus balances the deficit.

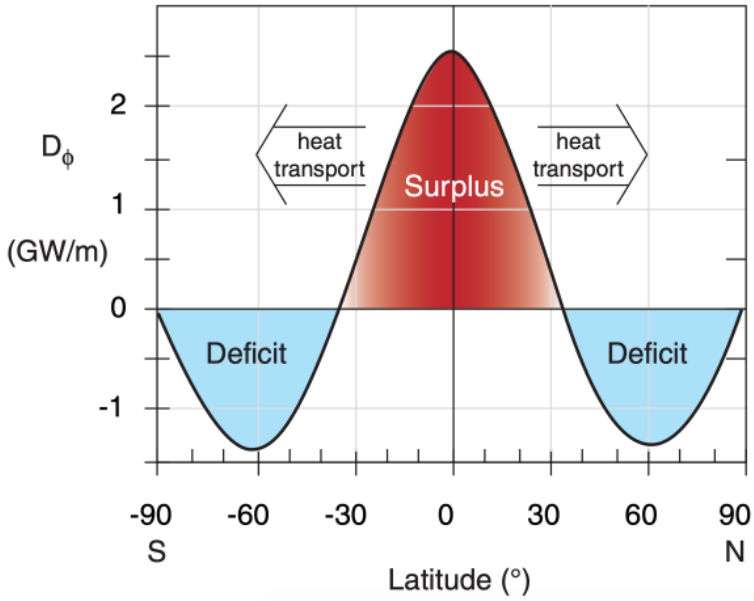


Figure 11.12 Net (incoming minus outgoing) zonally-integrated radiative forcings on the Earth. Surplus balances deficits. This differential heating imposed on the Earth must be compensated by heat transport by the global circulation; otherwise, the tropics would keep getting hotter and the polar regions colder.

Figs. (11.11) and (11.12) show the resulting incoming, outgoing, and net radiative forcings vs. latitude. At most latitudes there is nonzero net radiation. We can define E_{net} as **differential heating** D_ϕ :

Erroneous nesting of equation structures

Our despair is now quelled, because the surplus and deficit areas in Figs. 11.11 and 11.12 are almost exactly equal in magnitude to each other. Hence, we anticipate that Earth's climate should be relatively steady (neglecting global warming for now).

11.3.4. General Circulation Heat Transport

Nonetheless, the imbalance of net radiation between equator and poles in Fig. 11.12 drives atmospheric and oceanic circulations. These circulations act to undo the imbalance by removing the excess heat from the equator and depositing it near the poles (as per Le Chatelier's Principle). First, we can use the radiative differential heating to find how much global-circulation heat transport is needed. Then, we can examine the actual heat transport by atmospheric and oceanic circulations.

11.3.4.1. The Amount of Transport Required

Meridional transport at each pole is zero, because the pole is a singularity where all meridians converge. If we sum D_ϕ from Fig. 11.12 over all latitude belts from the North Pole to any other latitude ϕ , we can find the total transport Tr required for the global circulation to compensate all the radiative imbalances north of that latitude:

Erroneous nesting of equation structures

where the width of any latitude belt is Δy.

[Meridional distance Δy is related to latitude change $\Delta\phi$ by: $\Delta y(\text{km}) = (111\ \text{km}/^\circ) \cdot \Delta\phi\ (^\circ)$.]

The resulting "needed transport" is shown in Fig. 11.13, based on the simple "toy model" temperature and radiation curves of the past few sections. The magnitude of this curve peaks at about 5.6 PW (1 petawatt equals 10^{15} W) at latitudes of about 35° North and South (positive Tr means northward transport).

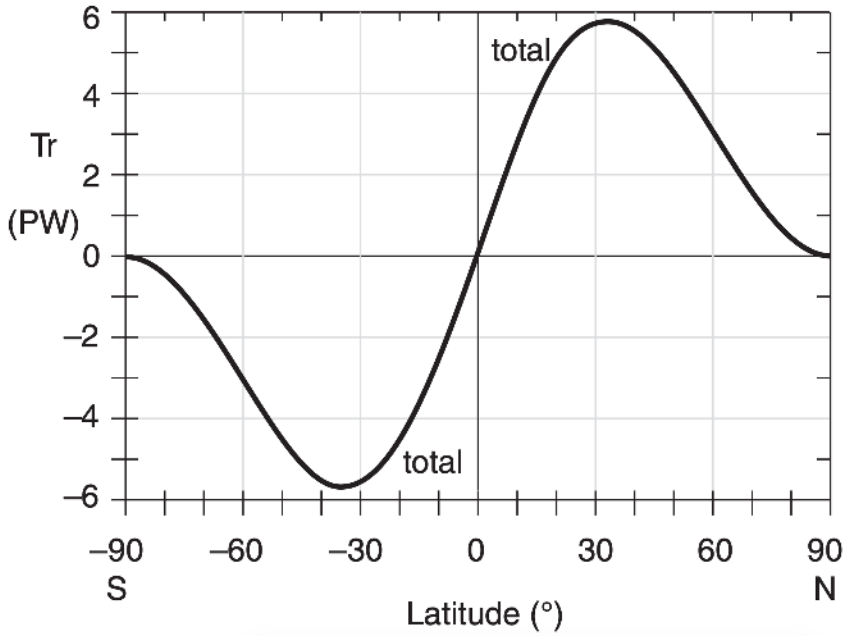


Figure 11.13 Required heat transport Tr by the global circulation to compensate radiative differential heating, based on a simple "toy model". Agrees very well with actual achieved transport in Fig. 11.14.

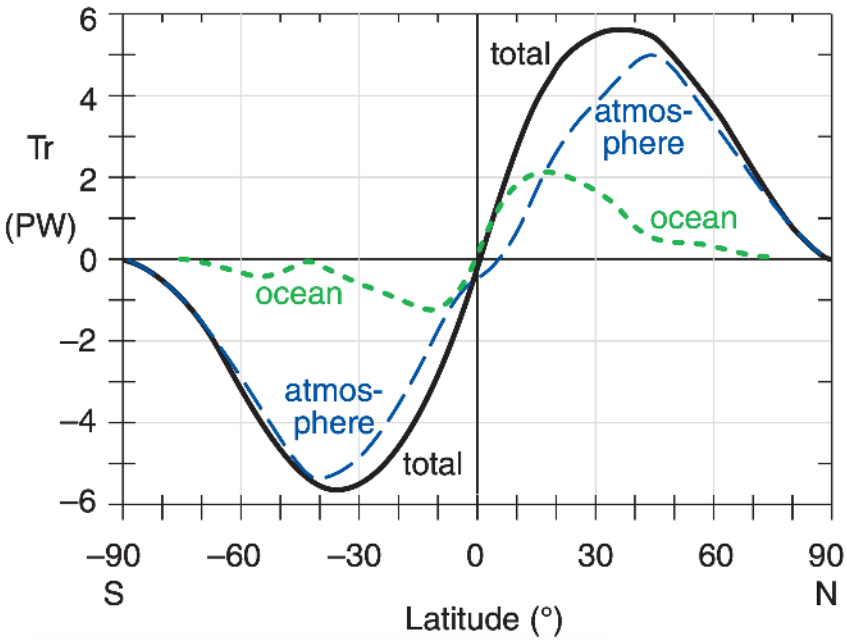


Figure 11.14 Meridional heat transports: Satellite-observed total (solid line) & ocean estimates (dotted). Atmospheric (dashed) is found as a residual. 1 PW = 1 petaWatt = 10^{15} W. [Data from K. E. Trenberth and J. M. Caron, 2001: "J. Climate", **14**, 3433-3443.]

11.3.4.2. Transport Achieved

Satellite observations of radiation to and from the Earth, estimates of heat fluxes to/from the ocean based on satellite observations of sea-surface temperature, and in-situ measurements of the atmosphere provide some of the transport data needed. Numerical forecast models are then used to tie the observations together and fill in the missing pieces. The resulting estimate of heat transport achieved by the atmosphere and ocean is plotted in Fig. 11.14.

Ocean currents dominate the total heat transport only at latitudes 0 to 17°, and remain important up to latitudes of ± 40°. Asymmetry of the ocean curve across the equator is due to the different ocean basin shapes and currents. In the atmosphere, the Hadley circulation is a dominant contributor in the tropics and subtropics, while the Rossby waves dominate atmospheric transport at mid-latitudes.

Knowing that global circulations undo the heating imbalance raises another question. How does the differential heating in the atmosphere drive the winds in those circulations? That is the subject of the next three sections.

Sample Application

What total heat transports by the atmosphere and ocean circulations are needed at 50°N latitude to compensate for all the net radiative cooling between that latitude and the North Pole? The differential heating as a function of latitude is given in the following table (based on the toy model):

Lat (°)	D_ϕ (GW m^{-1})	Lat (°)	D_ϕ (GW m^{-1})
90	0	65	–1.380
85	–0.396	60	–1.403
80	–0.755	55	–1.331
75	–1.049	50	–1.164
70	–1.261	45	–0.905

Find the Answer

Given: ϕ = 50°N.

D_ϕ data in table above.

Find: Tr = ? PW

Use eq. (11.11). Use sidewalks (latitude belts) each of width $\Delta\phi$ = 5°. Thus, Δy (m) = (111,000 m/°) · (5°) = 555,000 m is the sidewalk width. If one sidewalk spans 85 - 90°, and the next spans 80 - 85° etc, then the values in the table above give D_ϕ along the edges of the sidewalk, not along the middle. A better approximation is to average the D_ϕ values from each edge to get a value representative of the whole sidewalk. Using a bit of algebra, this works out to:

$Tr = -(555000\ m)\cdot[(0.5)\cdot 0.0 - 0.396 - 0.755 - 1.049 - 1.261 - 1.38 - 1.403 - 1.331 - (0.5)\cdot 1.164]\ (GW\ m^{-1})$

$Tr = (555000\ m)\cdot[8.157\ GW\ m^{-1}] =$ **4.527 PW**

Check: Units OK (106 GW = 1 PW). Agrees with Fig. 11.14.

Exposition: This northward heat transport warms all latitudes north of 50°N, not just one sidewalk. The warming per sidewalk is $\Delta Tr/\Delta\phi$.

This page titled 3.4.2: Radiative Differential Heating is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.3: Radiative Differential Heating** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.3: Pressure Profiles

A SCIENTIFIC PERSPECTIVE • Residuals

If something you cannot measure contributes to things you can measure, then you can estimate the unknown as the **residual** (i.e., difference) from all the knowns. This is a valid scientific approach. It was used in Fig. 11.14 to estimate the atmospheric portion of global heat transport.

CAUTION: When using this approach, your residual not only includes the desired signal, but it also includes the sums of all the errors from the items you measured. These errors can easily accumulate to cause a "noise" that is larger than the signal you are trying to estimate. For this reason, **error estimation** and **error propagation** (see Appendix A) should always be done when using the method of residuals.

The following fundamental concepts can help you understand how the global circulation works:

- non-hydrostatic pressure couplets caused by horizontal winds and vertical buoyancy,
- hydrostatic thermal circulations,
- geostrophic adjustment, and
- the thermal wind.

The first two concepts are discussed in this section. The last two are discussed in subsequent sections.

11.4.1. Non-hydrostatic Pressure Couplets

Consider a background reference environment with no vertical acceleration (i.e., hydrostatic). Namely, the pressure-decrease with height causes an upward pressure-gradient force that exactly balances the downward pull of gravity, causing zero net vertical force on the air (see Fig. 1.12 and eq. 1.25).

Next, suppose that immersed in this environment is a column of air that might experience a different pressure decrease (Fig. 11.15); i.e., **non-hydrostatic pressures**. At any height, let $p' = P_{column} - P_{hydrostatic}$ be the deviation of the actual pressure in the column from the theoretical hydrostatic pressure in the environment. Often a positive p' in one part of the atmospheric column is associated with negative p' elsewhere. Taken together, the positive and negative p's form a **pressure couplet**.

Non-hydrostatic p' profiles are often associated with non-hydrostatic vertical motions through Newton's second law. These non-hydrostatic motions can be driven by horizontal convergence and divergence, or by buoyancy (a vertical force). These two effects create opposite pressure couplets, even though both can be associated with upward motion, as explained next.

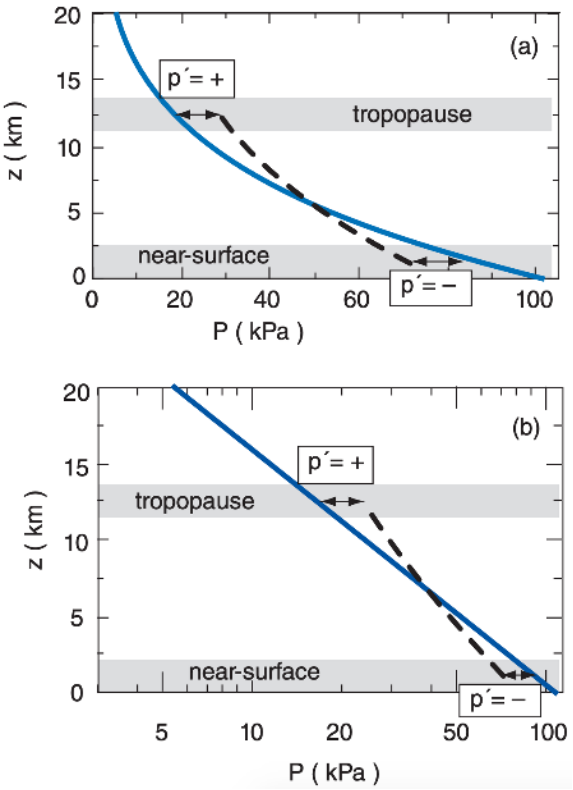


Figure 11.15 Background hydrostatic pressure (solid line), and non-hydrostatic column of air (dashed line) with pressure perturbation p' that deviates from the hydrostatic pressure $P_{hydrostatic}$ at most heights z. In this example, even though p' is positive (+) near the tropopause, the total pressure in the column ($P_{column} = P_{hydrostatic} + p'$) at the tropopause is still less than the surface pressure. The same curve is plotted as (a) linear and as (b) semilog.

https://geo.libretexts.org/@go/page/20933

11.4.1.1. Horizontal Convergence/Divergence

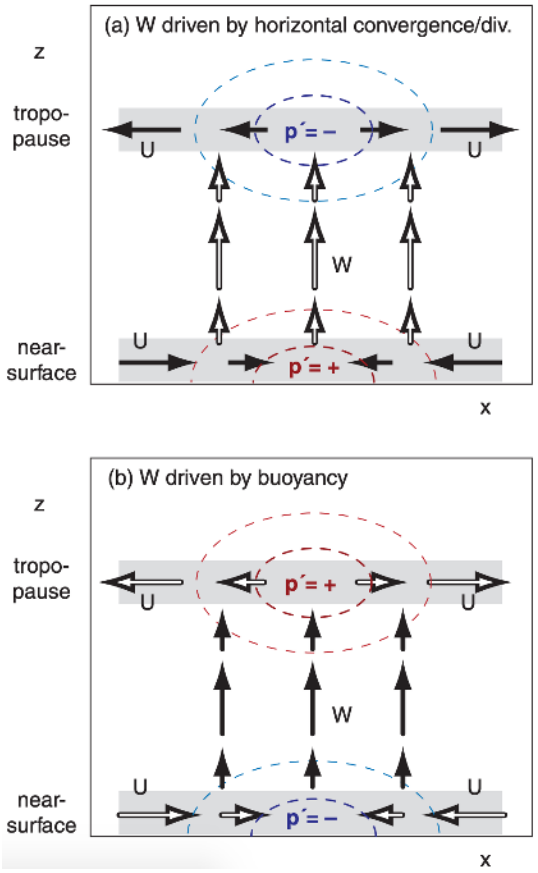


Figure 11.16 (a) Vertical motions driven by pressure gradients caused by horizontal convergence or divergence. (b) Vertical motions driven by buoyancy. In both figures, black arrows indicate the cause (the driving force), and white arrows are the effect (the response). p' is the pressure perturbation (deviation from hydrostatic), and thin dashed lines are isobars of p'. U and W are horizontal and vertical velocities. In both figures, the responding flow (white arrows) is driven down the pressure-perturbation gradient, away from positive p' and toward negative p'.

If external forcings cause air near the ground to converge horizontally, then air molecules accumulate. As density ρ increases according to eq. (10.60), the ideal gas law tells us that p' will also become positive (Fig. 11.16a). Non-zero p' implies that a pressure gradient exists, which will drive winds.

Positive p' does two things: it (1) decelerates the air that was converging horizontally, and (2) accelerates air vertically in the column. Thus, the pressure perturbation causes **mass continuity** (horizontal inflow near the ground balances vertical outflow).

Similarly, an externally imposed horizontal divergence at the top of the troposphere would lower the air density and cause negative p', which would also accelerate air in the column upward. Hence, we expect upward motion (W = positive) to be driven by a p' couplet, as shown in Fig. 11.16a.

11.4.1.2. Buoyant Forcings

For a different scenario, suppose air in a column is positively buoyant, such as in a thunderstorm where water-vapor condensation releases lots of latent heat. This vertical buoyant force creates upward motion (i.e., warm air rises, as in Fig. 11.16b).

As air in the thunderstorm column moves away from the ground, it removes air molecules and lowers the density and the pressure; hence, p' is negative near the ground. This suction under the updraft causes air near the ground to horizontally converge, thereby conserving mass.

Conversely, at the top of the troposphere where the thunderstorm updraft encounters the even warmer environmental air in the stratosphere, the upward motion rapidly decelerates, causing air molecules to accumulate, making p' positive. This pressure perturbation drives air to diverge horizontally near the tropopause, causing the outflow in the anvil-shaped tops of thunderstorms.

The resulting pressure-perturbation p' couplet in Fig. 11.16b is opposite that in Fig. 11.16a, yet both are associated with upward vertical motion. The reason for this pressure-couplet difference is the difference in driving mechanism: imposed horizontal convergence/divergence vs. imposed vertical buoyancy.

Similar arguments can be made for downward motions. For either upward or downward motions, the pressure couplets that form depend on the type of forcing. We will use this process to help explain the pressure patterns at the top and bottom of the troposphere.

11.4.2. Hydrostatic Thermal Circulations

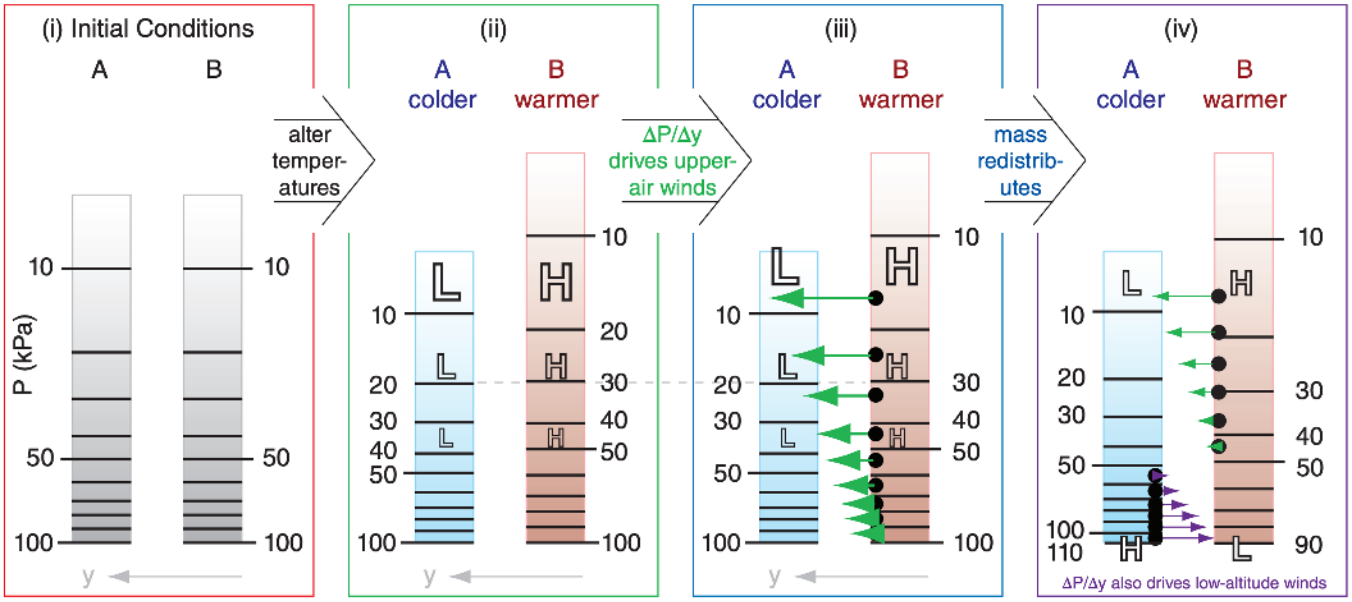


Figure 11.17 Formation of a thermal circulation. The response of two columns of air that are heated differently is that the warmer air column develops a low pressure perturbation at the surface and a high pressure perturbation aloft. Response of the cold column is the opposite. Notation: H = high pressure perturbation (i.e., relative to the average pressure at that altitude). L = low pressure perturbation (relative to the average pressure at that same altitude). Black dots represents air parcels. Thin arrows are winds (i.e., movement of air parcels).

Cold columns of air tend to have high surface pressures, while warm columns have low surface pressures. Figs. 11.17 illustrate how this happens.

Consider initial conditions (Fig. 11.17i) of two equal columns of air at the same temperature and with the same number of air molecules in each column. Since pressure is related to the mass of air above, this means that both columns A and B have the same initial pressure (100 kPa) at the surface.

Next, suppose that some process heats one column relative to the other (Fig. 11.17ii). Perhaps condensation in a thunderstorm cloud causes latent heating of column B, or infrared radiation cools column A. After both columns have finished expanding or contracting due to the temperature change, they will reach new hydrostatic equilibria for their respective temperatures.

The hypsometric equation (see Chapter 1) says that pressure decreases more rapidly with height in cold air than in warm air. Thus, although both columns have the same surface pressure because they contain the same number of molecules, the higher you go above the surface, the greater is the pressure difference between warm and cold air. In Fig. 11.17ii, the printed size of the "H" and "L" indicate the relative magnitudes of the high- and low-pressure perturbations p'.

The horizontal pressure gradient ΔP/Δy aloft between the warm and cold air columns drives horizontal winds from high toward low pressure (Fig. 11.17iii). Since winds are the movement of air molecules, this means that molecules leave the regions of high pressure-perturbation and accumulate in the regions of low. Namely, they leave the warm column, and move into the cold column.

Since there are now more molecules (i.e., more mass) in the cold column, it means that the surface pressure must be greater (H) in the cold column (Fig. 11.17iv). Similarly, mass lost from the warm column results in lower (L) surface pressure. This is called a **thermal low**.

A result (Fig. 11.17iv) is that, near the surface, high pressure in the cold air drives winds toward the low pressure in warm air. Aloft, high pressureperturbation in the warm air drives winds towards low pressure-perturbation in the cold air. The resulting **thermal circulation** causes each column to gain as many air molecules as they lose; hence, they are in mass equilibrium.

This equilibrium circulation also transports heat. Air from the warm air column mixes into the cold column, and vice versa. This intermixing reduces the temperature contrast between the two columns, causing the corresponding equilibrium circulations to weaken. Continued destabilization (more latent heating or radiative cooling) would be needed to maintain the circulation.

The circulations and mass exchange described above can be realized at the equator. At other latitudes, the exchange of mass is often slower (near the surface) or incomplete (aloft) because Coriolis force turns the winds to some angle away from the pressure-gradient direction. This added complication, due to geostrophic wind and geostrophic adjustment, is described next.

This page titled 3.4.3: Pressure Profiles is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.4: Pressure Profiles** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.4: Geostrophic Wind and Geostrophic Adjustment

Ageostrophic Winds at the Equator

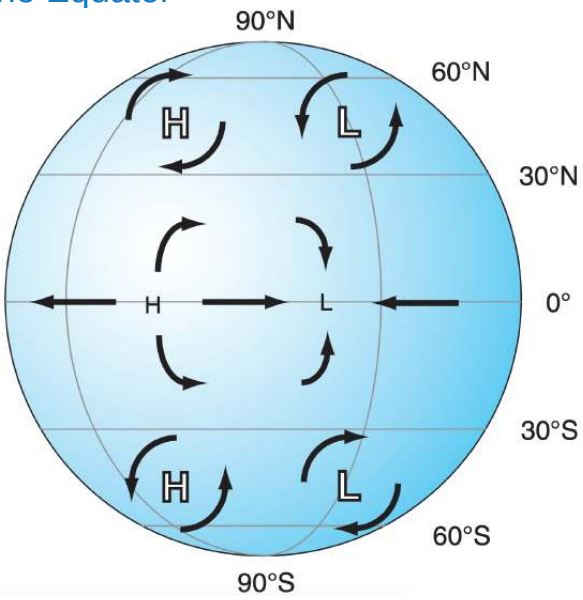


Figure 11.18 At the equator, winds flow directly from high (H) to low (L) pressure centers. At other latitudes, Coriolis force causes the winds to circulate around highs and lows. Smaller size font for H and L at the equator indicate weaker pressure gradients at that latitude.

Air at the equator can move directly from high (H) to low (L) pressure (Fig. 11.18 - center part) under the influence of pressure-gradient force. Zero Coriolis force at the equator implies infinite geostrophic winds. But actual winds have finite speed, and are thus **ageostrophic** (not geostrophic).

Because such flows can happen very easily and quickly, equatorial air tends to quickly flow out of highs into lows, causing the pressure centers to neutralize each other. Indeed, weather maps at the equator show very little pressure variations zonally. One exception is at continent-ocean boundaries, where continental-scale differential heating can continually regenerate pressure gradients to compensate the pressure-equalizing action of the wind. Thus, very small pressure gradients can cause continental-scale (5000 km) monsoon circulations near the equator. Tropical forecasters focus on winds, not pressure.

If the large-scale pressure is uniform in the horizontal near the equator (away from monsoon circulations), then the horizontal pressure gradients disappear. With no horizontal pressure-gradient force, no large-scale winds can be driven there. However, winds can exist at the equator due to inertia — if the winds were first created geostrophically at nonzero latitude and then coast across the equator.

But at most other places on Earth, Coriolis force deflects the air and causes the wind to approach geostrophic or gradient values (see the Forces & Winds chapter). Non-accelerating geostrophic and gradient winds do not cross isobars, so they cannot transfer mass from highs to lows. Thus, significant pressure patterns (e.g., strong high and low centers, Fig. 11.18) can be maintained for long periods at mid-latitudes in the global circulation.

Definitions

A SCIENTIFIC PERSPECTIVE • The Scientific Method Revisited

"Like other exploratory processes, [the scientific method] can be resolved into a dialogue between fact and fancy, the actual and the possible; between what could be true and what is in fact the case. The purpose of scientific enquiry is not to compile an inventory of factual information, nor to build up a totalitarian world picture of Natural Laws in which every event that is not compulsory is forbidden. We should think of it rather as a logically articulated structure of justifiable beliefs about a Possible

> World — a story which we invent and criticize and modify as we go along, so that it ends by being, as nearly as we can make it, a story about real life."
>
> - by Nobel Laureate Sir Peter Medawar (1982) *Pluto's Republic*. Oxford Univ. Press.

A **temperature field** is a map showing how temperatures are spatially distributed. A **wind field** is a map showing how winds are distributed. A **mass field** represents how air mass is spatially distributed. But we don't routinely measure air mass.

In the first Chapter, we saw how pressure at any one altitude depends on (is a measure of) all the air mass above that altitude. So we can use the **pressure field** at a fixed altitude as a surrogate for the mass field. Similarly, the **height field** on a map of constant pressure is another surrogate.

But the temperature and mass fields are coupled via the hypsometric equation (see Chapter 1).

Namely, if the temperature changes in the horizontal (defined as a **baroclinic** atmosphere), then pressures at any fixed altitude must also change in the horizontal. Later in this chapter, we will also see how the wind and temperature fields are coupled (via the thermal wind relationship).

Geostrophic Adjustment - Part 1

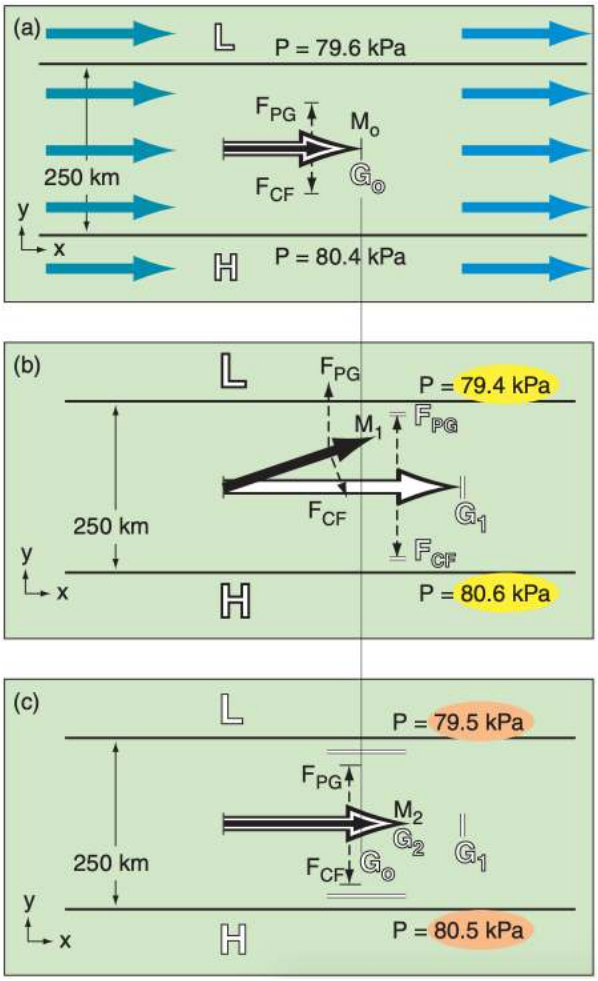


Figure 11.19 Example of geostrophic adjustment in the N. Hemisphere (not at equator). (a) Initial conditions, with the actual wind M (thick black arrow) in equilibrium with (equal to) the theoretical geostrophic value G (white arrow with black outline). (b) Transition. (c) End result at a new equilibrium. Dashed lines indicate forces F. Each frame focuses on the region of disturbance. Note that both the wind AND the pressure gradient have adjusted.

The tendency of non-equatorial winds to approach geostrophic values (or gradient values for curved isobars) is a very strong process in the Earth's atmosphere. If the actual winds are not in geostrophic balance with the pressure patterns, then both the winds

and the pressure patterns tend to change to bring the winds back to geostrophic (another example of Le Chatelier's Principle). **Geostrophic adjustment** is the name for this process.

Picture a wind field (grey arrows in Fig. 11.19a) initially in geostrophic equilibrium ($M_o = G_o$) at altitude 2 km above sea level (thus, no drag at ground). We will focus on just one of those arrows (the black arrow in the center), but all the wind vectors will march together, performing the same maneuvers.

Next, suppose an external process increases the horizontal pressure gradient to the value shown in Fig. 11.19b, with the associated faster theoretical geostrophic wind speed G_1. Inertia prevents instantaneous response of the actual wind M. During this transition, pressure-gradient force F_{PG} is greater than Coriolis force F_{CF}. This force imbalance turns the wind M_1 slightly toward low pressure and accelerates the air (Fig. 11.19b).

The component of wind M_1 from high to low pressure horizontally moves air molecules, weakening the pressure field and thereby reducing the theoretical geostrophic wind (G_2 in Fig. 11.19c). Namely, the **wind field changed the mass (i.e., pressure)** field. Simultaneously the actual wind accelerates to M_2. Thus, the **mass field also changed the wind field**. After both fields have adjusted, the result is $M_2 > M_o$ and $G_2 < G_1$, with $M_2 = G_2$. These changes are called **geostrophic adjustments**.

Define a "disturbance" as the region that was initially forced out of equilibrium. As you move further away from the disturbance, the amount of geostrophic adjustment diminished. The e-folding distance for this reduction is

Erroneous nesting of equation structures

where λ_R is known as the **internal Rossby deformation radius**. The Coriolis parameter is f_c, tropospheric depth is Z_T, and the Brunt-Väisälä frequency is N_{BV} (see eq. on page 372). λ_R is roughly 1300 km.

For a given wavelength λ of the initial disturbance, eq. (11.12) can be used as follows. For large disturbances ($\lambda > \lambda_R$), the wind field experiences the greatest adjustment. For smaller λ, the pressure and temperature fields have the greatest adjustment.

Sample Application

Find the internal Rossby radius of deformation in a standard atmosphere at 45°N.

Find the Answer

Given: $\phi = 45°$. Standard atmosphere from Chapter 1: T(z = Z_T =11 km) = –56.5°C, T(z=0) = 15°C.

Find: λ_R = ? km

First, find $f_c = (1.458 \times 10^{-4}\ s^{-1}) \cdot \sin(45°) = 1.031 \times 10^{-4}\ s^{-1}$

Next, find the average temperature and temperature difference across the depth of the troposphere:

$T_{avg} = 0.5 \cdot (-56.5 + 15.0)°C = -20.8°C = 252\ K$

$\Delta T = (-56.5 - 15.0)\ °C = -71.5°C$ across $\Delta z = 11$ km

Find the Brunt-Väisälä frequency (see the Atmos. Stability chapter). $N_{BV} = \sqrt{\frac{(9.8 m/s)}{252 K}\left[\frac{-71.5 K}{11000 m} + 0.0098 \frac{K}{m}\right]} = 0.0113 s^{-1}$

where the temperature differences in square brackets can be expressed in either °C or Kelvin.

Finally, use eq. (11.12): $\lambda_R = (0.0113\ s^{-1}) \cdot (11\ km) / (1.031 \times 10^{-4}\ s^{-1}) =$ **1206 km**

Check: Physics, units & magnitude are reasonable.

Exposition: When a cold-front over the Pacific approaches the steep mountains of western Canada, the front feels the influence of the mountains 1200 to 1300 km before reaching the coast, and begins to slow down.

This page titled 3.4.4: Geostrophic Wind and Geostrophic Adjustment is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.5: Geostrophic Wind and Geostrophic Adjustment** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.5: Thermal Wind Effect

Recall that horizontal temperature gradients cause vertically varying horizontal pressure gradients (Fig. 11.17), and that horizontal pressure gradients drive geostrophic winds. We can combine those concepts to see how horizontal temperature gradients drive vertically varying geostrophic winds. This is called the **thermal wind effect**.

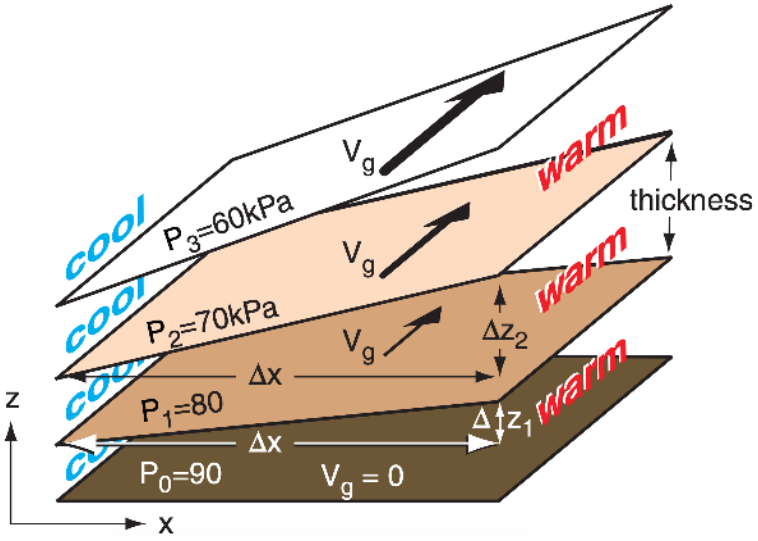


Figure 11.20 Isobaric surfaces are shaded brown. A zonal (west-east) temperature gradient causes the isobaric surfaces to tilt more and more with increasing altitude. Greater tilt causes stronger geostrophic winds (V_g) meridionally (south-north), as plotted with the black vectors for the N. Hemisphere. Geostrophic winds are reversed in S. Hemisphere.

This effect can be pictured via the slopes of isobaric surfaces (Fig. 11.20). The hypsometric equation from Chapter 1 describes how there is greater thickness between any two isobaric (constant pressure) surfaces in warm air than in cold air. This causes the tilt of the isobaric surfaces to change with altitude. But as described by eq. (10.29), tilting isobaric surface imply a pressure-gradient force that can drive the geostrophic wind (U_g, V_g).

The relationship between the horizontal temperature gradient and the changing geostrophic wind with altitude is known as the **thermal wind effect**. After a bit of manipulation (see the Higher Math box on the next page), one finds that:

Erroneous nesting of equation structures

Erroneous nesting of equation structures

where the Coriolis parameter is f_c, virtual absolute temperature is T_v, and gravitational acceleration magnitude is $|g| = 9.8\ m{\cdot}s^{-2}$. Thus, the meridional temperature gradient causes the zonal geostrophic winds to change with altitude, and the zonal temperature gradient causes the meridional geostrophic winds to change with altitude.

Above the atmospheric boundary layer, actual winds are nearly equal to geostrophic or gradient winds. Thus, eqs. (11.13) provide a first-order estimate of the variation of actual winds with altitude.

Sample Application

While driving north a distance of 500 km, your car thermometer shows the outside air temperature decreasing from 20°C to 10°C. How does the theoretical geostrophic wind change with altitude?

Find the Answer

Given: $\Delta y = 500$ km, $\Delta T = 10°C - 20°C = -10°C$, average $T = 0.5{\cdot}(10+20°C) = 15°C = 288$ K.

Find: $\Delta U_g/\Delta z = ?\ (m\ s^{-1})/km$

Given lack of other data, for simplicity assume: $f_c = 10^{-4}\ s^{-1}$, and $T_v = T$ (i.e., air is dry).

Apply eq. (11.13a):

$$\frac{\Delta U_g}{\Delta z} \approx \frac{-\left(9.8\mathrm{m/s^2}\right)}{(288\mathrm{K})\cdot\left(10^{-4}\mathrm{s^{-1}}\right)} \cdot \frac{(-10^{\circ}\mathrm{C})}{(500\mathrm{km})} = 6.81\ \left(\mathrm{ms^{-1}}\right)/\mathrm{km}$$

Check: Physics & units reasonable. Agrees with Fig.

Exposition: Even if the wind were zero at the ground, the answer indicates that a west wind will increase in speed by 6.81 m s–1 for each 1 km of altitude gained.

Given no info about temperature change in the x direction, assume uniform temperature, which implies $\Delta V_g/\Delta z = 0$ from eq. (11.13b). Namely, the south wind is constant with altitude.

11.6.1. Definition of Thickness

HIGHER MATH • Thermal Wind Effect

Problem: Derive Thermal Wind eq. (11.13a).

Find the Answer Start with the definitions of geostrophic wind (10.26a) and hydrostatic balance (1.25b):

$$U_g = -\frac{1}{\rho \cdot f_c}\frac{\partial P}{\partial y} \quad \text{and} \quad \rho \cdot |g| = -\frac{\partial P}{\partial z}$$

Replace the density in both eqs using the ideal gas law (1.20). Thus:

$$\frac{U_g \cdot f_c}{T_v} = -\frac{\Re_d}{P}\frac{\partial P}{\partial y} \text{ and } \frac{|g|}{T_v} = -\frac{\Re_d}{P}\frac{\partial P}{\partial z}$$

Use (1/P)·∂P = ∂ln(P) from calculus to rewrite both:

$$\frac{U_g \cdot f_c}{T_v} = -\Re_d\frac{\partial \ln(P)}{\partial y} \quad \text{and} \quad \frac{|g|}{T_v} = -\Re_d\frac{\partial \ln(P)}{\partial z}$$

Differentiate the left eq. with respect to z:

$$\frac{\partial}{\partial z}\left(\frac{U_g \cdot f_c}{T_v}\right) = -\Re_d\frac{\partial \ln(P)}{\partial y \partial z}$$

and the right eq. with respect to y:

$$\frac{\partial}{\partial y}\left(\frac{|g|}{T_v}\right) = -\Re_d\frac{\partial \ln(P)}{\partial y \partial z}$$

But the right side of both eqs are identical, thus we can equate the left sides to each other:

$$\frac{\partial}{\partial z}\left(\frac{U_g \cdot f_c}{T_v}\right) = \frac{\partial}{\partial y}\left(\frac{|g|}{T_v}\right)$$

Next, do the indicated differentiations, and rearrange to get the exact relationship for thermal wind:

$$\frac{\partial U_g}{\partial z} = -\frac{|g|}{T_v \cdot f_c}\frac{\partial T_v}{\partial y} + \frac{U_g}{T_v}\left(\frac{\partial T_v}{\partial z}\right)$$

The last term depends on the geostrophic wind speed and the lapse rate, and has magnitude of 0 to 30% of the first term on the right. If we neglect the last term, we get the approximate thermal wind relationship:

Erroneous nesting of equation structures

Exposition: A **barotropic** atmosphere is when the geostrophic wind does not vary with height. Using the exact equation above, we see that this is possible only when the two terms on the right balance.

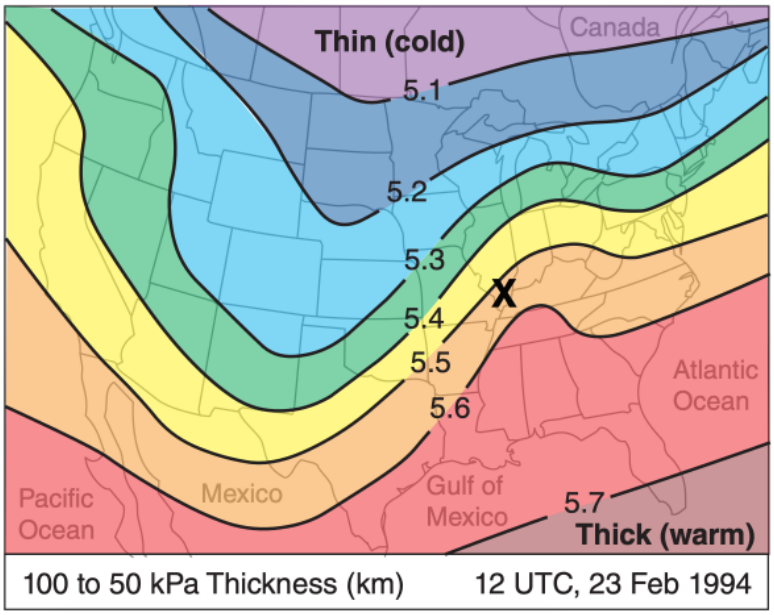


Figure 11.21 Thickness chart based on a US National Weather Service 24- hour forecast. The surface cyclone center is at the "X".

In Fig. 11.20, focus on two isobaric surfaces, such as P = 90 kPa (dark brown in that figure), and P = 80 kPa (medium brown). These two surfaces are at different altitudes z, and the altitude difference is called the "thickness". For our example, we focused on the "90 to 80 kPa thickness". For any two isobaric surfaces P_1 and P_2 having altitudes z_{P1} and z_{P2} , the thickness is defined as

Erroneous nesting of equation structures

The hypsometric equation from Chapter 1 tells us that the thickness is proportional to the average absolute virtual-temperature within that layer. Colder air has thinner thickness. Thus, horizontal changes in temperature must cause horizontal changes in thickness.

Weather maps of "100 to 50 kPa thickness" such as Fig. 11.21 are often created by forecast centers. Larger thickness on this map indicates warmer air within the lowest 5 km of the atmosphere.

11.6.2. Thermal-wind Components

Suppose that we use thickness TH as a surrogate for absolute virtual temperature. Then we can combine eqs. (11.14) and (10.29) to yield:

Erroneous nesting of equation structures

Erroneous nesting of equation structures

where U_{TH} and V_{TH} are **components of the thermal wind**, |g| = magnitude of gravitational-acceleration, f_c = Coriolis parameter, (U_{G1}, V_{G1}) are geostrophic-wind components on the P1 isobaric surface, and (U_{G2}, V_{G2}) are geostrophic-wind components on the P2 isobaric surface.

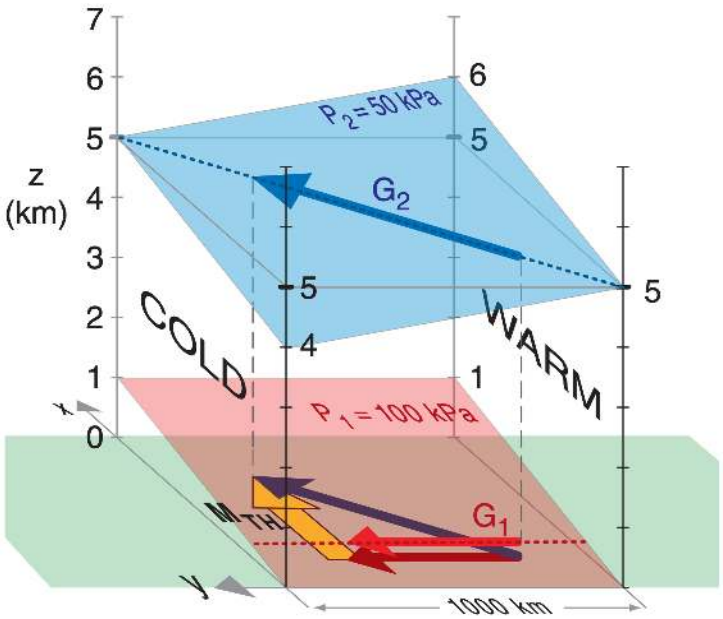


Figure 11.22 Relationship between the thermal wind M_{TH} and the geostrophic winds G on isobaric surfaces P. View is from the west northwest, in the Northern Hemisphere.

The horizontal vector defined by (U_{TH}, V_{TH}) is the difference between the geostrophic wind vector on the P_2 surface and the geostrophic wind vector on the P_1 surface, as Fig. 11.22 demonstrates. The corresponding **magnitude of the thermal wind** M_{TH} is:

Erroneous nesting of equation structures

To illustrate this, consider two isobaric surfaces P_2 = 50 kPa (shaded blue in Figure 11.22) and P_1 = 100 kPa (shaded red). The P_1 surface has higher height to the east (toward the back of this sketch). If you conceptually roll a ball bearing down this red surface to find the direction of the pressure gradient, and then recall that the geostrophic wind in the N. Hemisphere is 90° to the right of that direction, then you would anticipate a geostrophic wind G_1 direction as shown by the red arrow. Namely, it is parallel to a constant height contour (dotted red line) pointing in a direction such that low heights are to the vector's left.

Suppose cold air in the north (left side of this sketch) causes a small thickness of only 4 km between the red and blue surfaces. Warm air to the south causes a larger thickness of 5 km between the two isobaric surfaces. Adding those thicknesses to the heights of the P_1 surface (red) give the heights of the P_2 surface (blue). The blue P_2 surface tilts more steeply than the red P_1 surface, hence the geostrophic wind G_2 is faster (blue arrow) and is parallel to a constant height contour (blue dotted line).

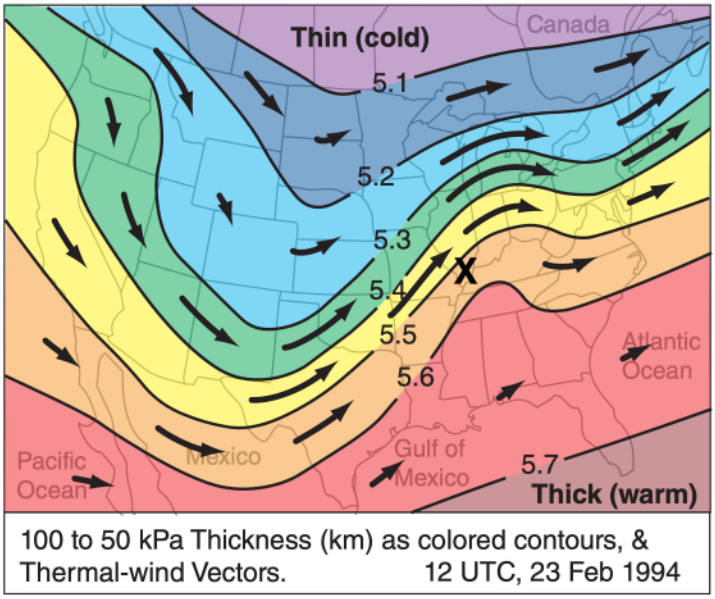


Figure 11.23 Arrows indicate thermal-wind vectors, where longer arrows indicate thermal winds that are stronger. Colored contours are 100 to 50 kPa thicknesses.

Projecting the G_1 and G_2 vectors to the ground (green in Fig. 11.22), the vector difference is shown in yellow, and is labeled as the thermal wind M_{TH}. It is parallel to the contours of thickness (i.e., perpendicular to the temperature gradient between cold and warm air) pointing in a direction with cold air (thin thicknesses) to its left (Fig. 11.23).

Sample Application. For Fig. 11.22, what are the thermal-wind components. Assume $f_c = 1.1 \times 10^{-4}\ s^{-1}$.

Find the Answer. Given: $TH_2 = 4$ km, $TH_1 = 5$ km, $\Delta y = 1000$ km from the figure, $f_c = 1.1 \times 10^{-4}\ s^{-1}$.

Find: $U_{TH} = ?\ m\ s^{-1}$, $V_{TH} = ?\ m\ s^{-1}$

Apply eq. (11.15a): $U_{TH} = \dfrac{-|g|}{f_c}\dfrac{\Delta TH}{\Delta y} = \dfrac{-(9.8 \mathrm{ms}^{-2}) \cdot (4-5)\mathrm{km}}{(1.1 \times 10^{-4}\mathrm{s}^{-1}) \cdot (5000\mathrm{km})} = 17.8 \mathrm{ms}^{-1}$

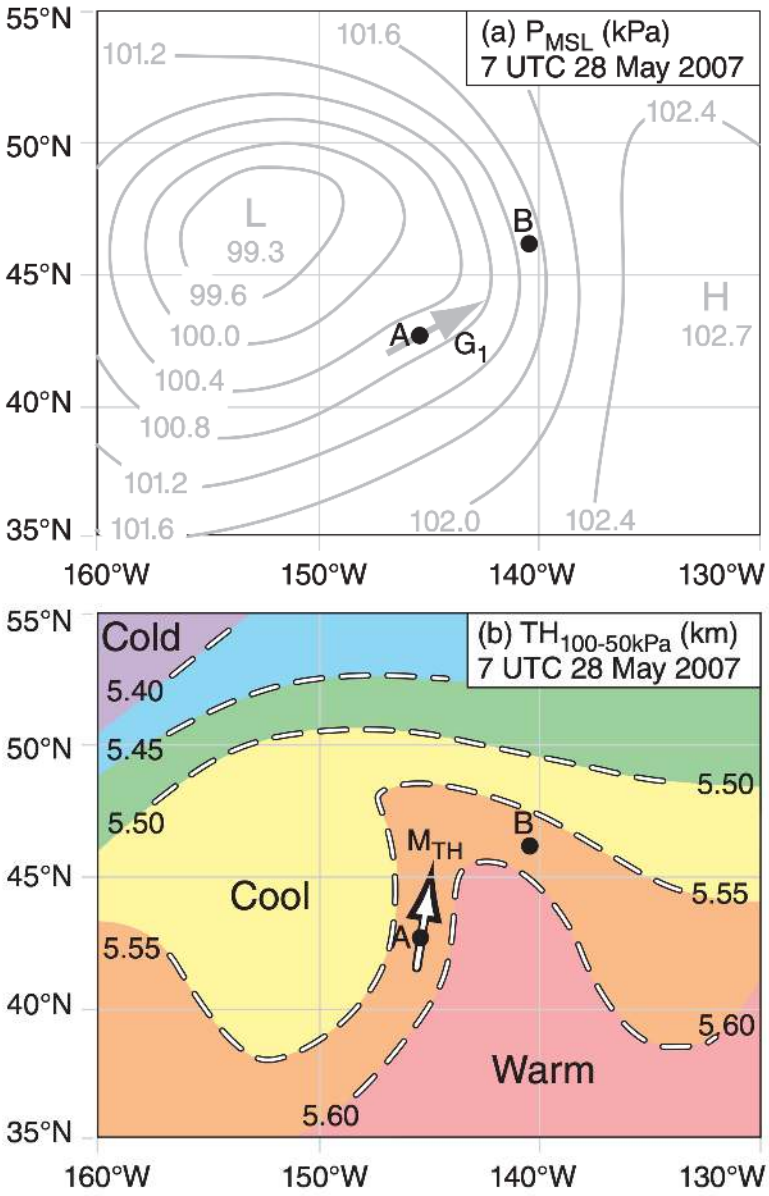
Check: Physics & units are reasonable. Positive sign for U_{TH} indicates wind toward positive x direction, as in Fig.

Exposition: With no east-to-west temperature gradient, and hence no east-to-west thickness gradient, we would expect zero north-south thermal wind;, hence, $V_{TH} = 0$.

Thermal-wind magnitude is stronger where the thickness gradient is greater. Thus, regions on a weather map (Fig. 11.23) where thickness contours are closer together (i.e., have **tighter packing**) indicates faster thermal winds. The relationship between thermal winds and thickness contours is analogous to the relationship between geostrophic winds and height contours. But never forget that no physically realistic wind can equal the thermal wind, because the thermal wind represents the difference or shear between two geostrophic winds.

Nonetheless, you will find the thermal-wind concept useful because it helps you anticipate how geostrophic wind will change with altitude. Actual winds tend towards being geostrophic above the atmospheric boundary layer, hence the thermal-wind concept allows you to anticipate real wind shears.

11.6.3. Case Study



https://geo.libretexts.org/@go/page/20935

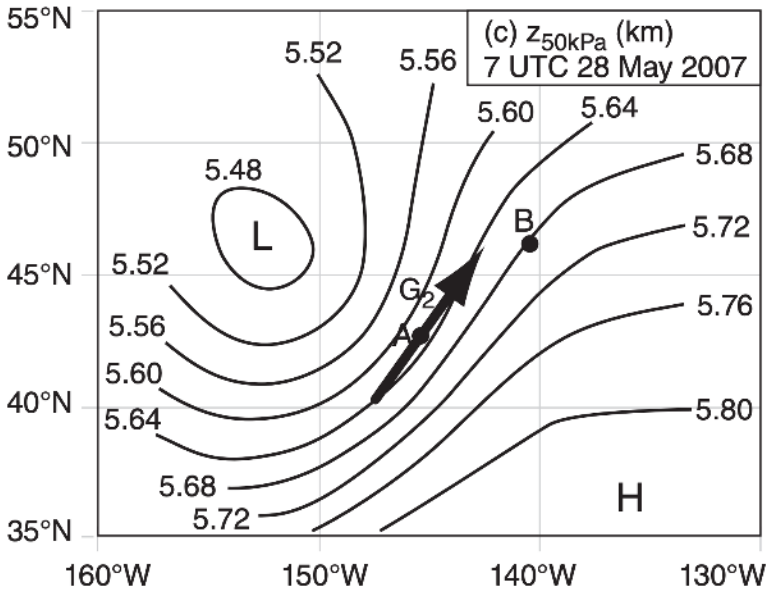


Figure 11.24 Weather maps for a thermal-wind case-study. (a) Mean sealevel pressure (kPa), as a surrogate for height of the 100 kPa surface. (b) Thickness (km) of the layer between 100 kPa to 50 kPa isobaric surfaces. (c) Geopotential heights (km) of the 50 kPa isobaric surface.

Figs. 11.24 show how geostrophic winds and thermal winds can be found on weather maps, and how to interpret the results. These maps may be copied onto transparencies and overlain.

Fig. 11.24a is a weather map of pressure at sea level in the N. Hemisphere, at a location over the northeast Pacific Ocean. As usual, L and H indicate low- and high-pressure centers. At point A, we can qualitatively draw an arrow (grey) showing the theoretical geostrophic (G_1) wind direction; namely, it is parallel to the isobars with low pressure to its left.

Recall that pressures on a constant height surface (such as at height z = 0 at sea level) are closely related to geopotential heights on a constant pressure surface. So we can be confident that a map of 100 kPa heights would look very similar to Fig. 11.24a.

Fig. 11.24b shows the 100 - 50 kPa thickness map, valid at the same time and place. The thickness between the 100 and the 50 kPa isobaric surfaces is about 5.6 km in the warm air, and only 5.4 km in the cold air. The white arrow qualitatively shows the thermal wind M_{TH}, as being parallel to the thickness lines with cold temperatures to its left.

Fig. 11.24c is a weather map of geopotential heights of the 50 kPa isobaric surface. L and H indicate low and high heights. The black arrow at A shows the geostrophic wind (G_2), drawn parallel to the height contours with low heights to its left.

If we wished, we could have calculated quantitative values for G_1, G_2, and M_{TH}, utilizing the scale that 5° of latitude equals 555 km. [ALERT: *This scale does not apply to longitude, because the meridians get closer together as they approach the poles. However, once you have determined the scale (map mm : real km) based on latitude, you can use it to good approximation in any direction on the map*.]

Back to the thermal wind: if you add the geostrophic vector from Fig. 11.24a with the thermal wind vector from Fig. 11.24b, the result should equal the geostrophic wind vector in Fig. 11.24c. This is shown in the Sample Application.

Although we will study much more about weather maps and fronts in the next few chapters, I will interpret these maps for you now.

Point A on the maps is near a cold front. From the thickness chart, we see cold air west and northwest of point A. Also, knowing that winds rotate counterclockwise around lows in the N. Hemisphere (see Fig. 11.24a), I can anticipate that the cool air will advance toward point A. Hence, this is a region of **cold-air advection**. Associated with this cold-air advection is **backing of the wind** (i.e., turning counterclockwise with increasing height), which we saw was fully explained by the thermal wind.

Point B is near a warm front. I inferred this from the weather maps because warmer air is south of point B (see the thickness chart) and that the counterclockwise winds around lows are causing this warm air to advance toward point B. **Warm air advection** is

associated with **veering of the wind** (i.e., turning clockwise with increasing height), again as given by the thermal wind relationship. I will leave it to you to draw the vectors at point B to prove this to yourself.

Sample Application

For Fig. 11.24, qualitatively verify that when vector M_{TH} is added to vector G_1, the result is vector G_2.

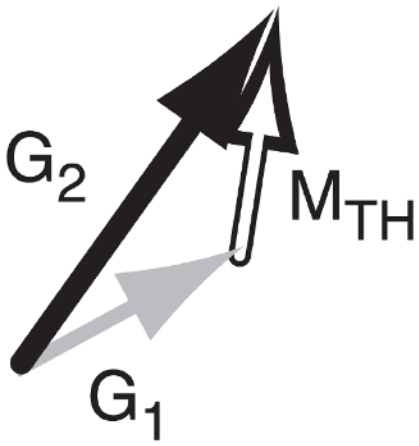
Find the Answer

Given: the arrows from Fig. 11.24 for point A.

These are copied and pasted here.

Find: The vector sum of $G_1 + M_{TH} = ?$

Recall that to do a vector sum, move the tail of the second vector (M_{TH}) to be at the arrow head of the first vector (G_1). The vector sum is then the vector drawn from the tail of the first vector to the tip (head) of the second vector.



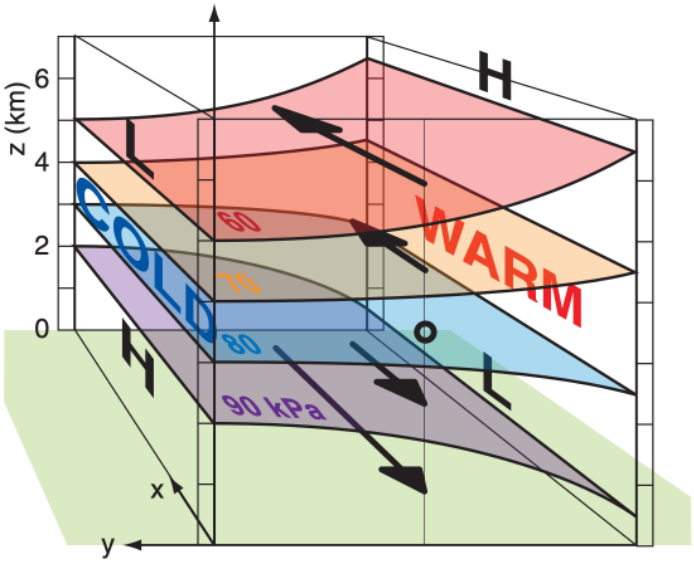
Check: Sketch is reasonable.

Exposition: The vector sum indeed equals vector G_2, as predicted by the thermal wind.

11.6.4. Thermal Wind & Geostrophic Adjustment - Part 2

Figure 11.25 Typical equilibrium state of the pressure, temperature and wind fields after it has adjusted geostrophically. Isobaric surfaces are shaded with color, and recall that high heights of isobaric surfaces correspond to regions of high pressure on constant altitude surfaces. Black arrows give the geostrophic wind vectors.

Consider the red-shaded isobaric surface representing P = 60 kPa. That surface has high (H) heights to the south (to the right in this figure), and low (L) heights to the north. In the Northern Hemisphere, the geostrophic wind would be parallel to a constant height contour in a direction with lower pressure to its left. A similar interpretation can be made for the purpleshaded isobaric surface at P = 90 kPa. At a middle altitude in this sketch there is no net pressure gradient (i.e., zero slope of an isobaric surface), hence no geostrophic wind.



As geostrophic winds adjust to changes in pressure gradients, they move mass to alter the pressure gradients. Eventually, an equilibrium is approached (Fig. 11.25) based on the combined effects of geostrophic adjustment and the thermal wind. This figure is much more realistic than Fig. 11.17(iv) because Coriolis force prevents the winds from flowing directly from high to low pressure.

With these concepts of:

- differential heating,
- nonhydrostatic pressure couplets due to horizontal winds and vertical buoyancy,
- hydrostatic thermal circulations,
- geostrophic adjustment, and
- the thermal wind,

we can now go back and explain why the global circulation works the way it does.

This page titled 3.4.5: Thermal Wind Effect is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.6: Thermal Wind Effect** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.6: Horizontal Circulation

Consider a closed shape of finite area (Fig. 11.47a). Pick any starting point on the perimeter, and hypothetically travel counterclockwise around the perimeter until you return to the starting point. As you travel each increment of distance Δl, observe the local winds along that increment, and get the average tangential component of wind velocity M_t.

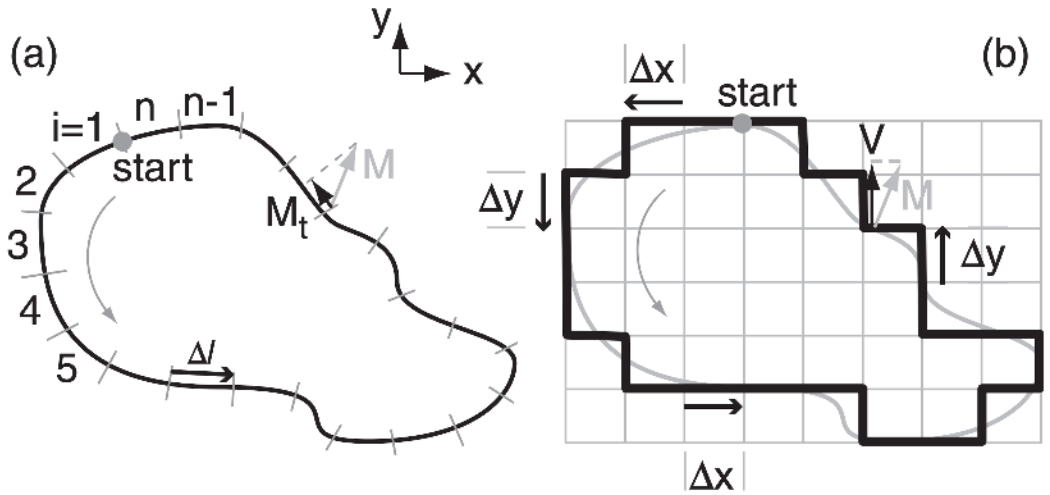


Figure 11.47 Method for finding the circulation. M is the wind vector. Mt and V are projections onto the perimeter. (a) Stepping in increments of Δl around an arbitrary shape. (b) Stepping around a Cartesian (gridded) approximation to the shape in (a).

The horizontal **circulation** C is defined as the product of this tangential velocity times distance increment, summed over all the increments around the whole perimeter:

Erroneous nesting of equation structures

where i is the index of each increment, and n is the number of increments needed to complete one circuit around the perimeter. Take care that the sign of M_t is such that it is positive if the tangential wind is in the same direction as you are traveling, and negative if opposite. The units of circulation are $m^2 \cdot s^{-1}$.

If we approximate the perimeter by Cartesian line segments (Fig. 11.47b), then eq. (11.28) becomes:

Erroneous nesting of equation structures

The sign of Δx is (+) if you travel in the positive x-direction (toward the East), and (–) if opposite. Similar rules apply for Δy (+ toward North).

To better understand circulation, consider idealized cases (Figs. 11.48a & b). For winds of tangential velocity M_t rotating counterclockwise around a circle of radius R, the circulation is $C = 2\pi R \cdot M_t$. For clockwise circular rotation, the circulation is $C = -2\pi R \cdot M_t$, namely, the circulation value is negative. From these two equations, we see that a fast speed around a small circle (such as a tornado) can give the same circulation magnitude as a slower speed around a larger circle (e.g., a mid-latitude cyclone).

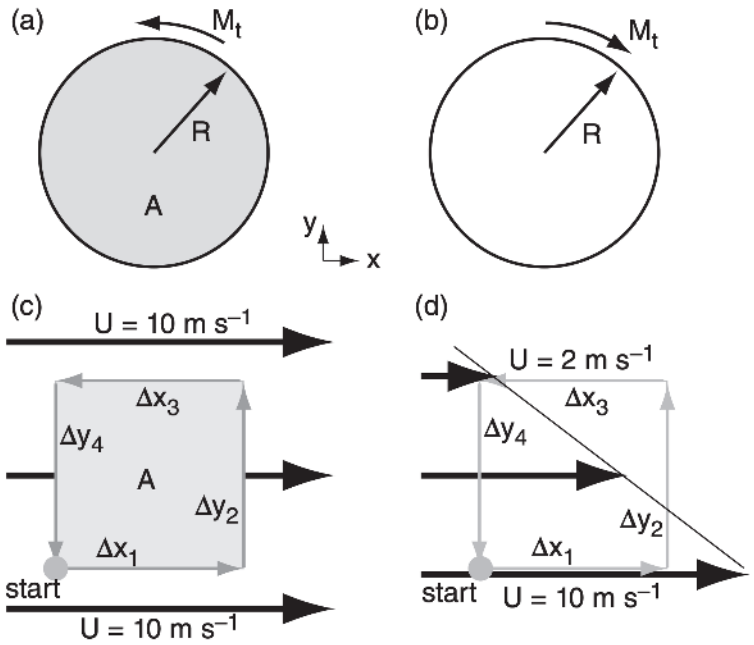


Figure 11.48 Circulation examples. (a) Counterclockwise rotation around a circle. (b) Clockwise rotation around a circle. (c) Uniform wind. (d) Uniform shear. Area A enclosed by circulation is shaded.

Consider two more cases (Figs. 11.48c & d). For a circuit in a constant wind field of any speed, the circulation is C = 0. For a circuit within a region of uniform shear such as $\Delta U/\Delta y$, the circulation is $C = -(\Delta U/\Delta y)\cdot(\Delta y\cdot\Delta x)$. Comparing these last two cases, we see that the wind speed is irrelevant for the circulation, but the wind shear is very important.

In the last equation above, $(\Delta y\cdot\Delta x) = A$ is the area enclosed by the circulation of Fig. 11.48d. In general, for uniform U and V shear across a region, the horizontal circulation is:

Erroneous nesting of equation structures

But the term in parentheses is the relative vorticity ζ_r . This gives an important relationship between horizontal circulation and vorticity:

Erroneous nesting of equation structures

Vorticity is defined at any one point in a fluid, while circulation is defined around a finite-size area. Thus, eq. (11.31) is valid only in the limit as A becomes small, or for the special case of a fluid having uniform vorticity within the whole circulation area.

The horizontal circulation C defined by eq. (11.30 & 11.31) is also known as the **relative circulation** C_r . An **absolute circulation** C_a can be defined as

Erroneous nesting of equation structures

where f_c is the Coriolis parameter. The absolute circulation is the circulation that would be seen from a fixed point in space looking down on the atmosphere rotating with the Earth.

For the special case of a frictionless **barotropic atmosphere** (where isopycnics are parallel to isobars), **Kelvin's circulation theorem** states that Ca is constant with time.

For a more **realistic baroclinic** atmosphere containing horizontal temperature gradients, the **Bjerknes circulation theorem**:

Erroneous nesting of equation structures

says relative circulation varies with the torque applied to the fluid (via the component of pressure forces in the direction of travel, summed around the perimeter of the circulation area) minus the Earth's rotation effects in a changing circulation area. The units of $\Delta P/\rho$ are J kg^{-1}, which are equivalent to the m$^2\cdot$s^{-2} units of $\Delta C_r/\Delta t$. The pressure term in eq. (11.33) is called the **solenoid term**.

Sample Application

Find the horizontal circulation for Fig. 11.48d. Assume $\Delta x = \Delta y = 1$ km. Relate to shear and rel. vorticity.

Find the Answer

Given: U_{bottom}=10 m s^{-1}, U_{top}=2 m s^{-1}, V=0, $\Delta x = \Delta y$=1 km.

Find: C = ? m^2 s^{-1}, ζ_r = ? s^{-1}

Use eq. (11.29) from start point:

$C = (U \cdot \Delta x)_1 + (V \cdot \Delta y)_2 + (U \cdot \Delta x)_3 + (V \cdot \Delta x)_4$ (10m s^{-1})·(1km) + 0 + (2m s^{-1})·(–1km) + 0

C = 8 (m s^{-1})·km = **8000 m^2 s^{-1}**

Use eq. (11.30): with area $A = \Delta x \cdot \Delta y$ = 1 km^2

$U_{shear} = \Delta U/\Delta y = (U_{top} – U_{bottom})/\Delta y$ = [(2 – 10 m s^{-1})/(1 km)] = –8 (m s^{-1})/km

$C = [0 – U_{shear}] \cdot A$ = –[–8 (m s^{-1})/km]· (1 km^2) = 8 (m s^{-1})·km = **8000 m^2 s^{-1}**.

Use eq. (11.31): $\zeta_r = C/A$ = 8 (m s^{-1})/km = **0.008 s^{-1}** .

Check: Physics, magnitude & units are reasonable.

Exposition: Strong shear. Large circ. Large vorticity.

This page titled 3.4.6: Horizontal Circulation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.10: Horizontal Circulation** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.7: Extratropical Ridges and Troughs (Rossby Waves)

The atmosphere is generally warm near the equator and cool near the poles. This meridional temperature gradient drives a west-to-east wind having increasing speed with increasing altitude within the troposphere, as described by the thermal-wind effect. The resulting fast wind near the tropopause is called the jet stream. To first order, we would expect this jet stream to encircle the globe (Fig. 11.49a) along the zone between the warm and cool airmasses, at roughly 50 to 60° latitude in winter.

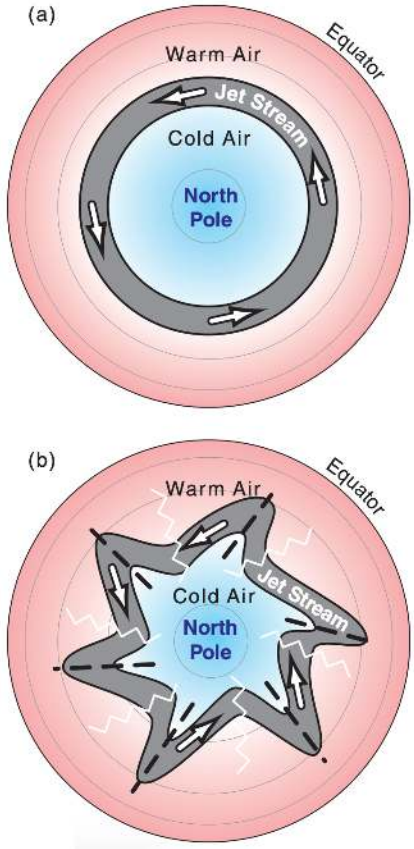


Figure 11.49 (a) Jet stream in the N. Hemisphere on a world with no instabilities. (b) Jet stream with barotropic or baroclinic instabilities, creating a meandering jet. Troughs are marked with a black dashed line, and ridges with a white zig-zag line.

However, this flow is unstable, allowing small disturbances (e.g., flow over mountain ranges) to grow into large north-south meanders (Fig. 11.49b) of the jet stream. These meanders are called **Rossby waves** or **planetary waves**. Typical wavelengths are 3 - 4 Mm. Given the circumference of a parallel at those latitudes, one typically finds 3 to 13 waves around the globe, with a normal **zonal wavenumber** of 7 to 8 waves.

The equatorward region of any meander is called a **trough** (pronounced like “troff”) and is associated with low pressure or low geopotential height. The poleward portion of a meander is called a **ridge**, and has high pressure or height. The turning of winds around troughs and ridges are analogous to the turning around closed lows and highs, respectively. The trough center or **trough axis** is labeled with a dashed line, while the **ridge axis** is labeled with a zig-zag symbol (Fig. 11.49b).

Like many waves or oscillations in nature, Rossby waves result from the interplay between inertia (trying to make the jet stream continue in the direction it was deflected) and a restoring force (acting opposite to the deflection). For Rossby waves, the restoring force can be explained by the conservation of potential vorticity, which depends on both the Coriolis parameter and the layer thickness (related to layer static stability). **Baroclinic instability** considers both restoring factors, while **barotropic instability** is a simpler approximation that considers only the Coriolis effect.

11.11.1. Barotropic Instability

Consider tropospheric air of constant depth Δz ($\approx$ 11 km). For this situation, the conservation of potential vorticity can be written as

$$\left[\frac{M}{R}+f_c\right]_{\text{initial}}=\left[\frac{M}{R}+f_c\right]_{\text{later}} \tag{11.34}$$

where jet-stream wind speed M divided by radius of curvature R gives the relative vorticity, and f_c is the Coriolis parameter (which is a function of latitude).

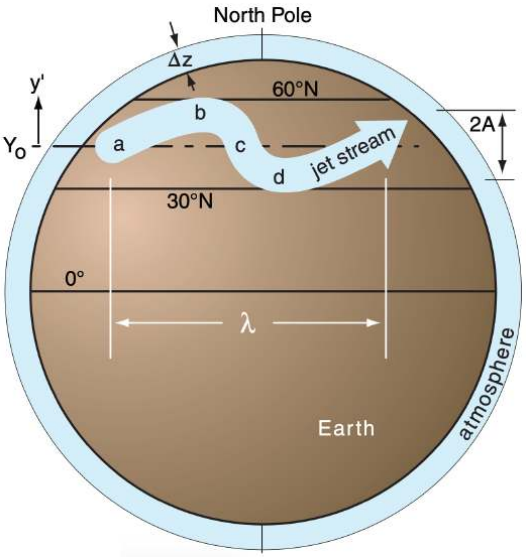


Figure 11.50 Illustration of barotropic instability.

(Location **a** in Fig. 11.50) Consider a jet stream at initial latitude Y_o moving in a straight line from the southwest. At that latitude it has a certain value of the Coriolis parameter, but no relative vorticity (M/R = 0, because R = ∞ for a straight line). But f_c increases as the air moves poleward, thus the M/R term on the right side of eq. (11.34) must become smaller than its initial value (i.e., it becomes negative) so that the sum on the right side still equals the initial value on the left side.

(Location **b**) We interpret negative curvature as anticyclonic (clockwise turning in the N. Hemisphere). This points the jet stream equatorward.

(Location **c**) As the air approaches its starting latitude, its Coriolis parameter decreases toward its starting value. This allows the flow to become a straight line again at location c. But now the wind is blowing from the northwest, not the southwest.

(Location **d**) As the air overshoots equatorward, f_c gets smaller, requiring a positive M/R (cyclonic curvature) to maintain constant potential vorticity. This turns the jet stream back toward its starting latitude, where the cycle repeats. The flow is said to be **barotropically unstable**, because even pure, non-meandering zonal flow, if perturbed just a little bit from its starting latitude, will respond by meandering north and south.

This north-south (meridional) oscillation of the west-to-east jet stream creates the wavy flow pattern we call a **Rossby wave** or a **planetary wave**. Because the restoring force was related to the change of Coriolis parameter with latitude, it is useful to define a beta parameter as

Erroneous nesting of equation structures

where the average radius of the Earth is R_{Earth} =6371 km. For $2\cdot\Omega/REarth = 2.29x10^{-11}$ $m^{-1}\cdot s^{-1}$, one finds that β is roughly (1.5 to 2)$x10^{-11}$ $m^{-1}\cdot s^{-1}$.

The wave path in Fig. 11.50 can be approximated with a simple cosine function:

Erroneous nesting of equation structures

where the displacement distance of the Rossby wave from Y_o (its starting latitude) is y' . Let x' be the eastward distance from the start of the wave. The position of the wave crests move at phase speed c with respect to the Earth. The wavelength is λ and its amplitude is A (see Fig. 11.50). The primes indicate the deviations from a mean background state.

Barotropic Rossby waves of the jet stream have wavelengths of about λ ≈ 6000 km and amplitudes of about A ≈ 1665 km, although a wide range of both is possible. Typically 4 to 5 of these waves can fit around the earth at mid-latitudes (where the circumference

of a latitude circle is $2\pi \cdot R_{Earth} \cdot \cos\phi$, and ϕ is latitude).

The waves move through the air at **intrinsic phase speed** c_o :

Erroneous nesting of equation structures

The negative sign means that the wave crests propagate toward the west relative to the air.

However, the air in which the wave is embedded is itself moving toward the east at wind speed U_o. Thus, relative to the ground, the **phase speed** c is:

Erroneous nesting of equation structures

Given typical values for U_o, the total phase speed c relative to the ground is positive. Such movement toward the east is indeed observed on weather maps.

Barotropic Rossby waves can have a range of different wavelengths. But eq. (11.37a) says that different wavelength waves move at different intrinsic phase speeds. Thus, the different waves tend to move apart from each other, which is why eq. (11.37a) is known as a **dispersion relationship**.

Short waves have slower intrinsic phase speed toward the west, causing the background wind to blow them rapidly toward the east. **Long waves**, with their faster intrinsic phase speed westbound end up moving more slowly toward the east relative to the ground, as illustrated in Fig. 11.51. The net effect is that short waves move through long waves (see the 3rd Sample Application on the next page).

The speed that Rossby waves transport energy is called the **group velocity**, c_g :

Erroneous nesting of equation structures

which differs in sign from eq. (11.37a). This causes **teleconnections** in storminess that moves faster toward the east than the phase speed of the waves.

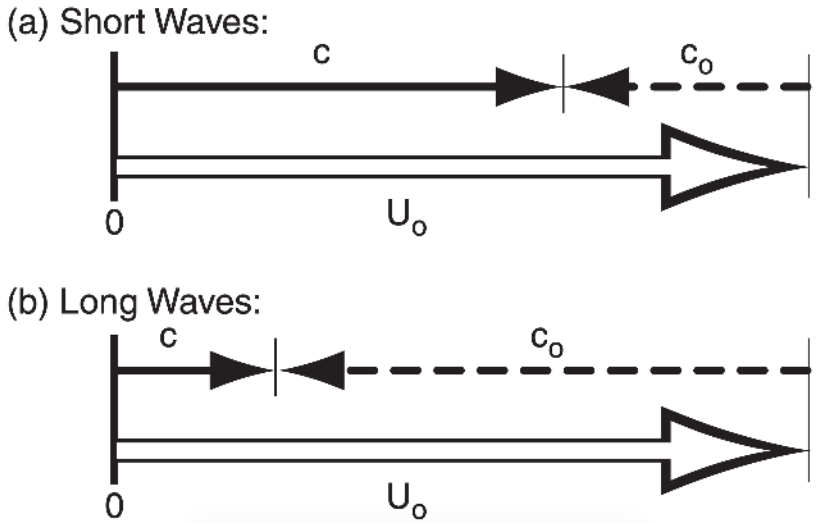


Figure 11.51 (at left) Sum of a large background west wind U_o with a smaller intrinsic Rossby-wave phase speed co from the east gives the resulting propagation speed c of Rossby waves relative to the ground.

HIGHER MATH • Planetary-wave Vorticity

Suppose the jet-stream axis oscillates north or south some distance y relative to an arbitrary reference latitude as the air flows toward the east (x):

Erroneous nesting of equation structures

where the wavelength is λ and the meridional amplitude of the wave is A. The jet speed along this wave is M, and the associated wind components (U, V) depend on the local slope s of the wave at location x. From geometry: $U^2 + V^2 = M^2$ and $s = V/U$, thus:

Erroneous nesting of equation structures

To find slope s from eq. (a), take the derivative of y:

Erroneous nesting of equation structures

Next, change eq. (11.20) from finite difference to partial derivatives:

Erroneous nesting of equation structures

But you can expand the last term as follows:

$$\zeta_r = \partial V/\partial x - (\partial U/\partial x) \cdot (\partial x/\partial y)$$

where the last factor is just one over the slope:

Erroneous nesting of equation structures

Combine eqs. (e, c, & b) to get the desired relative vorticity:

Erroneous nesting of equation structures

Fig. 11.a illustrates this for a wave with A = 1500 km, λ = 6000 km, & M = 40 m s^{-1}.

Exposition: Some calculus books give equations for the sine-wave radius of curvature R. Using that in $|\zeta_r| = 2M/R$ would give a similar equation for vorticity. The largest vorticities are concentrated near the wave crest and trough, allowing meteorologists to use positive vorticity to help find troughs.

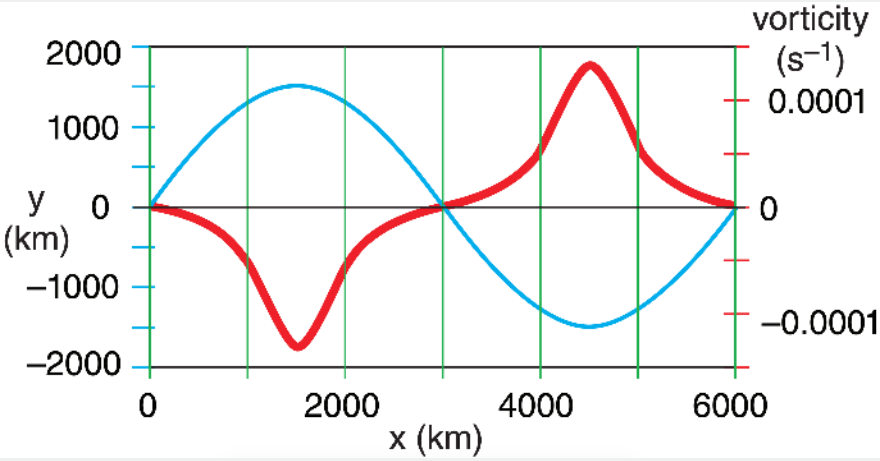


Figure 11.a. Jet-stream path (blue) and associated vorticity (thick red).

HIGHER MATH • The Beta Plane

Here is how you can get β using the definition of the Coriolis parameter f_c (eq. 10.16):

$$f_c = 2\Omega \sin\phi$$

where ϕ is latitude.

Since y is the distance along the perimeter of a circle of radius R_{Earth}, recall from geometry that

$$y = R_{\text{Earth}} \cdot \phi$$

for ϕ in radians.

Rearrange this to solve for ϕ, and then plug into the first equation to give:

$$f_c = 2\Omega \sin\left(y/R_{\text{Earth}}\right)$$

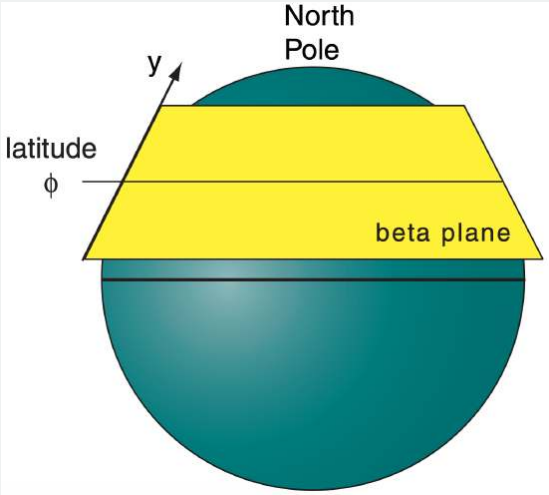
By definition of β, take the derivative to find

$$\beta = \frac{\partial f_C}{\partial y} = \frac{2 \cdot \Omega}{R_{\text{earth}}} \cdot \cos\left(\frac{y}{R_{\text{earth}}}\right)$$

Finally, use the second equation above to give:

Erroneous nesting of equation structures

For a small range of latitudes, β is nearly constant. Some theoretical derivations assume constant beta, which has the same effect as assuming that the earth is shaped like a cone. The name for this lamp-shade shaped surface is the **beta plane**.



Sample Application (§)

The jet stream meanders north and south with barotropic wavelength of 8000 km and amplitude of 1200 km relative to reference latitude 50°N. Winds in this jet are 40 m s^{-1}. Calculate the following for a barotropic wave: beta parameter, phase speeds, and wave translation distance in 12 h. Also, plot the initial and final wave streamlines y'(x') for x' of 0 to 11,000 km.

Find the Answer

Given: U_o = 40 m s^{-1}, ϕ =50°, A =1200 km, λ =8000 km, for t = 0 to 12 h

Find: β = ? $m^{-1}{\cdot}s^{-1}$, c_o = ? m s^{-1}, c = ? m s^{-1} , D = c·Δt ? km translation distance, y'(x') = ? km.

Apply eq. (11.35):

β = 2.29 x 10^{-11} ·cos(50°) = **1.47x10^{-11} $m^{-1}{\cdot}s^{-1}$**.

Next, apply eq. (11.37a):

$$c_o = -\left(1.47 \times 10^{-11}\mathrm{m^{-1}s^{-1}}\right)\left(\frac{8 \times 10^{6}\mathrm{m}}{2\pi}\right)^{2} == 23.9\mathrm{ms^{-1}}$$

Then apply eq. (11.37b):

c = (50 – 23.9) m s^{-1} = **16.1 m s^{-1}**

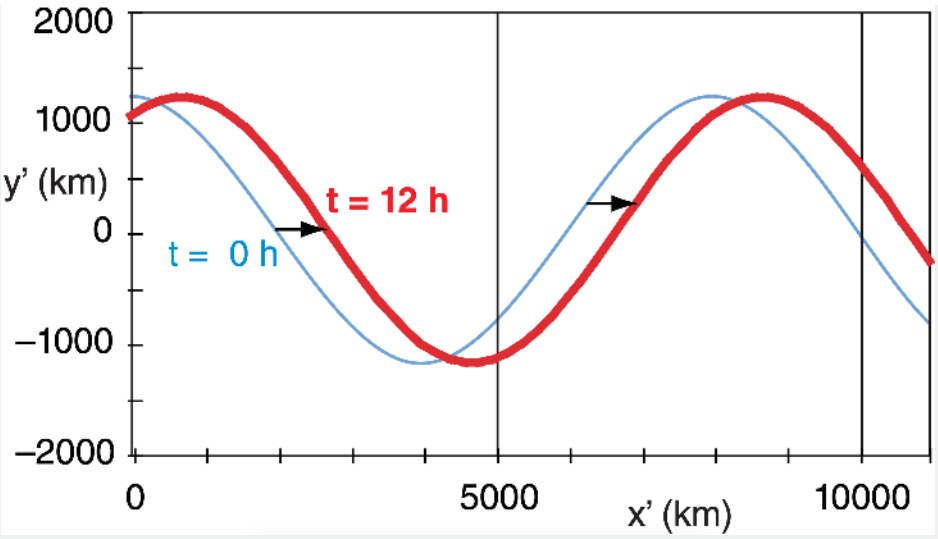
Wave crest translation distance in 12 h is

D = c·Δt = (16.1 m s^{-1})·(12h · 3600s/h) = 697.1 km

Finally solve & plot eq. (11.36) for t = 0 to 12 h:

$$y' \approx (1200\mathrm{km}) \cdot \cos\left[2\pi \cdot \left(\frac{x' - (16.1\mathrm{m/s}) \cdot t}{8 \times 10^{6}\mathrm{m}}\right)\right]$$

with conversions between m & km, and for s & h.



Check: Physics & units are reasonable.

Exposition: The thin blue streamlines plotted above show the path of the 40 m s^{-1} jet stream, but this path gradually shifts toward the east (thick red streamlines).

Although this Rossby-wave phase speed is much slower than a jet airliner, the wave does not need to land and refuel. Thus, during 24 hours, this long wave could travel about 1,400 km — roughly half the distance between San Francisco, CA & Chicago, IL, USA. Even longer waves can be stationary, and some extremely long waves can **retrograde** (move in a direction opposite to the background jet-stream flow; namely, move toward the west).

Sample Application (§)

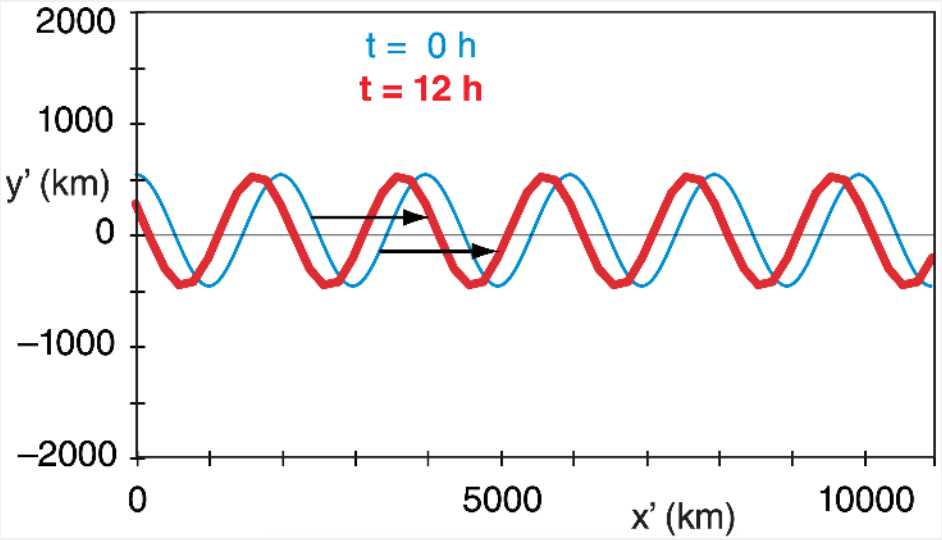
Similar to the previous Sample, but for a short wave with 500 km amplitude and 2000 km wavelength.

Find the Answer

Given: Same, but λ = 2000 km, A = 500 km.

Apply similar equations (not shown here) as before, yielding: β = **1.47x10**$^{-11}$ $m^{-1}{\cdot}s^{-1}$, c_o = –1.5 m s^{-1} , c = 38.5 m s^{-1} , and D = 1664 km.

The short-wave streamline plot:



Check: Physics & units reasonable.

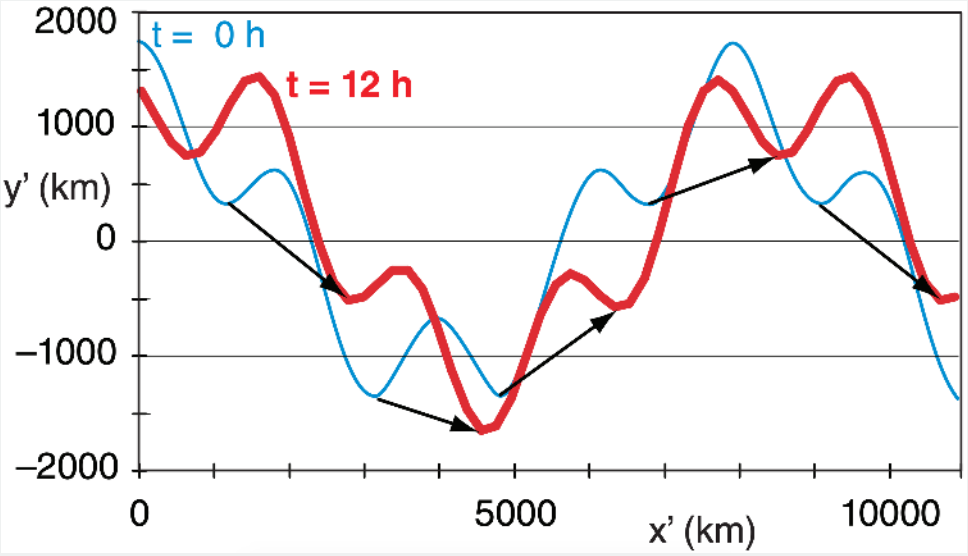
Exposition: Short waves move faster than long ones.

Sample Application (§)

Similar to the previous two Sample Applications, but superimpose the long and short waves.

Find the Answer

The combined long- and short-wave streamline plot:



Check: Physics & units reasonable.

Exposition: The short waves move rapidly along the long-wave streamline similar to trains on a track, except that this long-wave track gradually shifts east.

The short waves travel very fast, so they are in and out of any city very quickly. Thus, they cause rapid changes in the weather. For this reason, weather forecasters pay particular attention to short-wave troughs to avoid surprises. Although short waves are sometimes difficult to spot visually in a plot of geopotential height, you can see them more easily on plots of 50 kPa vorticity. Namely, each short-wave trough has a noticeable positive vorticity that can be highlighted or colorized on a forecast map.

11.11.2. Baroclinic Instability

Fig. 11.52 illustrates baroclinic instability using a toy model with thicker atmosphere near the equator and thinner atmosphere near the poles. This mimics the effect of static stability, with cold strongly stable air near the poles that restricts vertical movement of air, vs. warm weakly stable air near the equator that is less limiting in the vertical (see Fig. 5.20 again).

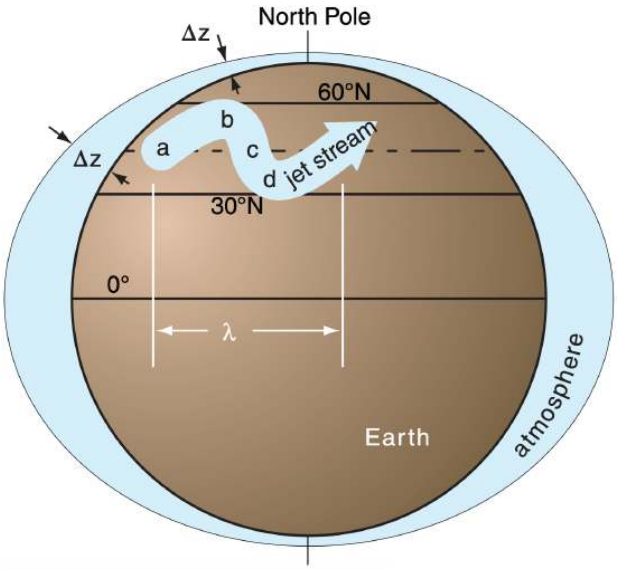


Figure 11.52 Illustration of baroclinic instability, where the effect of static stability is mimicked with an atmosphere that is thicker near the equator and thinner near the poles. The thickness change is exaggerated in this figure.

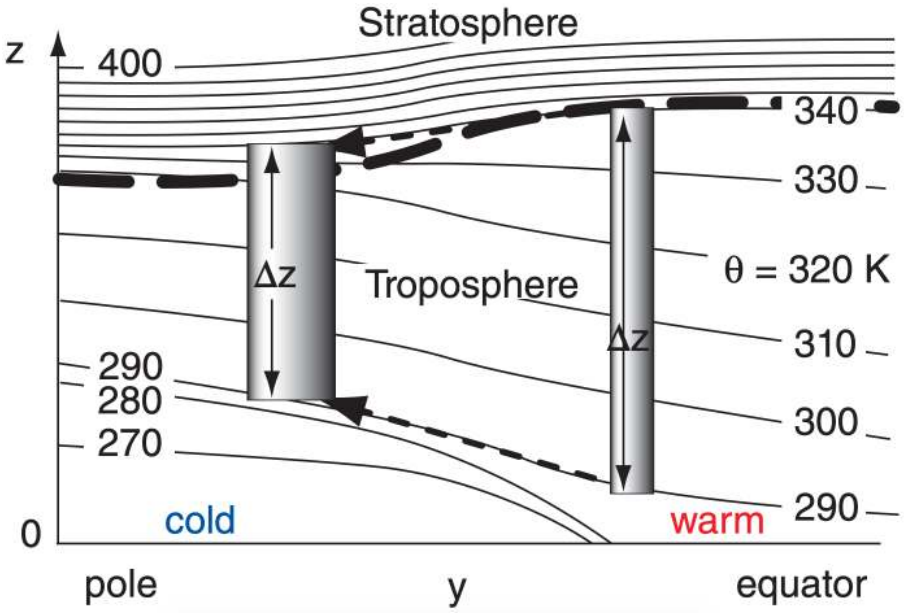


Figure 5.20 (again) Illustration of how the north-south variation of static stability can be interpreted as a change of effective depth of the atmosphere. Vertical cross section through the atmosphere, showing isentropes (lines of constant potential temperature θ). The depth Δz of column of air on the right will shrink as the column moves poleward, because air tends to follow isentropes during adiabatic processes. Thus, the same Δθ between top and bottom of the air columns spans a shorter vertical distance for the poleward column, meaning that static stability and Brunt-Väisälä frequency are greater there.

11.11.2.1. Qualitative View

As before, use potential-vorticity conservation:

Erroneous nesting of equation structures

where M is wind speed, the Coriolis parameter is f_c, and the radius of curvature is R. You must include the effective atmospheric thickness Δz because it varies south to north. This captures baroclinic effects that are intentionally neglected for barotropic instability.

For baroclinic waves, follow the jet stream as was done before for barotropic waves, from location **a** to location **d**. All the same processes happen as before, but with an important difference. As the air moves toward location **b**, not only does f_c increase, but Δz

decreases. But Δz is in the denominator, hence both f_c and Δz tend to increase the potential vorticity. Thus, the curvature M/R must be even more negative to compensate those combined effect. This means the jet stream turns more sharply.

Similarly, at location **d**, f_c is smaller and Δz in the denominator is larger, both acting to force a sharper cyclonic turn. The net result is that the combined restoring forces are stronger for baroclinic situations, causing tighter turns that create a shorter overall wavelength λ than for barotropic waves.

11.11.2.2. Quantitative View

The resulting north-south displacement y' for the baroclinic wave is:

Erroneous nesting of equation structures

where the tropospheric depth is Z_T (≈ 11 km) , the meridional amplitude is A , and where c is phase speed, λ is wavelength, x is distance East, and t is time. Notice that there is an additional cosine factor. This causes the meridional wave amplitude to be zero the middle of the troposphere, and to have opposite signs at the top and bottom.

[*ALERT: this is an oversimplification. Waves in the real atmosphere aren't always 180° out of phase between top and bottom of the troposphere. Nonetheless, this simple approach gives some insight into the workings of baroclinic waves.*]

The intrinsic phase speed co for the baroclinic wave is:

Erroneous nesting of equation structures

where eq. (11.35) gives β, wavelength is λ, and eq. (11.12) gives the internal Rossby deformation radius λ_R.

The influence of static stability is accounted for in the **Brunt-Väisälä frequency** N_{BV}, which is a factor in the equation for Rossby deformation radius λ_R (eq. 11.12). Recall from the Atmospheric Stability chapter that $N_{BV} = [\ (|g|/T_v) \cdot (\Gamma_d + \Delta T/\Delta z)\ ]^{1/2}$, where gravitation acceleration is $|g| = 9.8\ m\ s^{-2}$, absolute virtual temperature is T_v (where $T_v = T$ for dry air), dry adiabatic lapse rate is $\Gamma_d = 9.8\ °C\ km^{-1} = 0.0098\ K\ m^{-1}$, and ΔT/Δz is the change of air temperature with height.

As before, the phase speed relative to the ground is:

Erroneous nesting of equation structures

Although many different wavelengths λ are possible, the dominant wavelength (i.e., the wave for which amplitude grows the fastest) is roughly:

Erroneous nesting of equation structures

where λ_R is the internal Rossby radius of deformation (eq. 11.12). Typical values are $\lambda_d \approx 3$ to 4 Mm.

Sample Application(§)

The jet stream meanders north and south with amplitude of 900 km relative to reference latitude 50°N. Winds in this jet are 40 $m\ s^{-1}$. Temperature decreases 50°C across 12 km thick troposphere, with average temperature –20°C. Calculate the following for a baroclinic wave: Brunt-Väisälä frequency, internal Rossby radius, dominant wavelength, phase speeds, and wave translation distance in 3 h. Also, plot the initial and final wave streamlines y'(x') at both z = 0 and z = 12 km for x' ranging 0 to 5,000 km. β = **$1.47 \times 10^{-11}\ m^{-1} \cdot s^{-1}$**.

Find the Answer

Given: $U_o = 40\ m\ s^{-1}$, ϕ =50°, A =900 km, $T_{avg} = 253$ K, $\Delta T/\Delta z = -50K/(12\ km)$, $\beta = 1.47 \times 10^{-11}\ m^{-1} \cdot s^{-1}$, for z=1 & 12 km, t = 0 & 3 h

Find: N_{BV} = ? s^{-1}, λ_R = ? km, λ = ? km, co = ? $m\ s^{-1}$, c = ? $m\ s^{-1}$, y'(x') = ? km.

Get the N_{BV} eq. from the right column of this page.

$$N_{BV} = \left[\frac{(9.8 \mathrm{ms}^{-2})}{253\mathrm{K}} \left(\frac{-50\mathrm{K}}{1.2 \times 10^4 \mathrm{m}} + 0.0098 \frac{\mathrm{K}}{\mathrm{m}} \right) \right]^{1/2}$$

N_{BV} = **$0.0148\ s^{-1}$**

Apply eq. (10.16):

$f_c = \left(1.458 \times 10^{-4} s^{-1}\right) \cdot \sin\left(50^\circ\right) = 1.117 \times 10^{-4} \mathrm{s}^{-1}$

Apply eq. (11.12):

$\lambda_R = \dfrac{(0.0148 \mathrm{s}^{-1}) \cdot (12 \mathrm{km})}{1.117 \times 10^{-4} \mathrm{s}^{-1}} = 1590. \mathrm{km}$

Apply eq. (11.43):

$\lambda = \lambda_d = 2.38\ \lambda_R =$ **3784. km**

Apply eq. (11.41):

$$c_o = \frac{-\left(1.47 \times 10^{-11} \mathrm{m}^{-1} \mathrm{s}^{-1}\right)}{\pi^2 \cdot \left[\dfrac{4}{\left(3.784 \times 10^6 \mathrm{m}\right)^2} + \dfrac{1}{\left(1.59 \times 10^6 \mathrm{m}\right)^2}\right]}$$

= **–2.21 m s^{-1}**

Apply eq. (11.42):

c = (40 – 2.21)m s^{-1} = **37.8 m s^{-1}**

Use eq. (11.40) for z = 0, 12 km, and t = 0, 3 h:

$$y' \approx (900 \mathrm{km}) \cos\left(\frac{\pi \cdot z}{12 \mathrm{km}}\right) \cos\left[2\pi \left(\frac{x' - (37.8 \mathrm{m/s}) \cdot t}{3784 \mathrm{km}}\right)\right]$$

The results are plotted at right:

Check: Physics & units are reasonable.

Exposition: The Rossby wave at z = 12 km is indeed 180° out of phase from that at the surface. Namely, an upper-level ridge is above a surface trough. In the mid-troposphere (z ≈ 6 km) the baroclinic wave has zero amplitude (not shown).

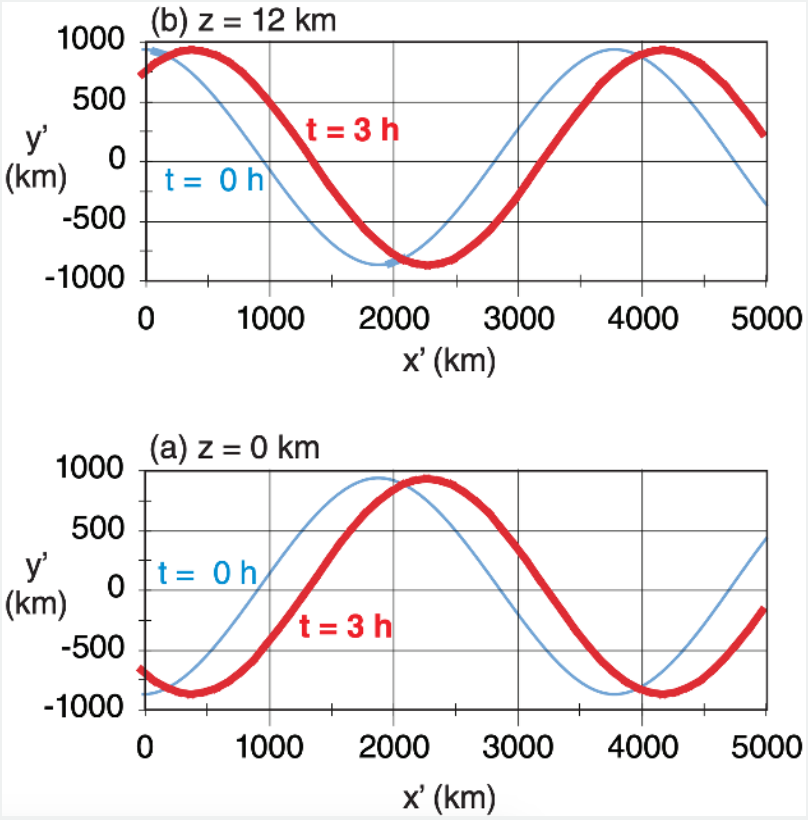


Fig. Baroclinic waves at two heights and two times, for the sample application at left.

https://geo.libretexts.org/@go/page/20937

INFO • Baroclinic Wave Characteristics

Baroclinic waves perturb many variables relative to their average background states. Represent the perturbation by e' = e – $e_{background}$ for any variable e. Variables affected include:

y' = meridional streamline displacement north,

(u', v', w') = wind components,

θ' = potential temperature,

p' = pressure, and

η' = vertical displacement.

Independent variables are time t and east-west displacement x' relative to some arbitrary location.

Let:

Erroneous nesting of equation structures

Erroneous nesting of equation structures

Thus:

Erroneous nesting of equation structures

Each of the equations above represents a wave, where the wave amplitude is indicated by the factor with the caret (^) over it. These amplitudes are always positive. For Northern Hemispheric baroclinic waves, the amplitudes are:

Erroneous nesting of equation structures

where ρ_o is average density of air at height z, and intrinsic phase speed c_o is a negative number. A (= north-south displacement) depends on the initial disturbance.

Any atmospheric variable can be reconstructed as the sum of its background and perturbation values; for example: U = $U_{background}$ + u'. Background states are defined as follows.

$U_{background} = U_g$ is the geostrophic (jet-stream) wind.

$V_{background} = W_{background} = 0$.

$P_{background}$ decreases with increasing height according to the hydrostatic equation (Chapter 1).

$\theta_{background}$ increases linearly as altitude increases as was assumed to create a constant value of N_{BV}.

$Y_{background}$ corresponds to the latitude of zeroperturbation flow, which serves as the reference latitude for calculation of f_c and β.

Background state for η is the altitude z in eq. (11.44).

All variables listed at left interact together to describe the wave. The result is sketched in Fig. 11.53. Although the equations at left look complicated, they are based on a simplified description of the atmosphere. They neglect clouds, turbulence, latent heating, meridional wave propagation, and nonlinear effects. Nonetheless, the insight gained from this simple model helps to explain the behavior of many of the synoptic weather patterns that are covered in the Extratropical Cyclones chapter.

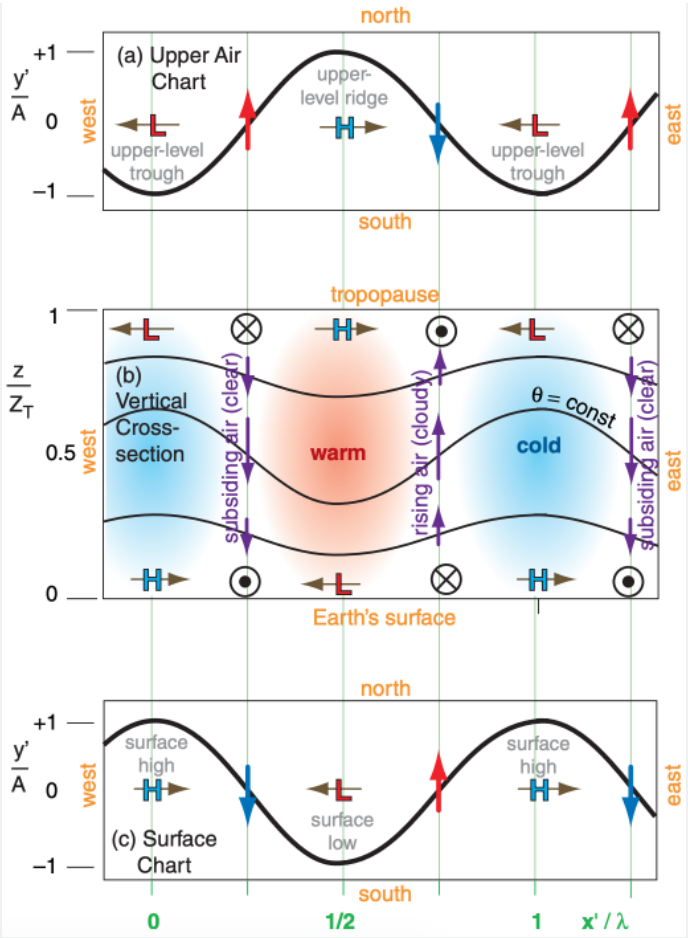


Figure 11.53 Characteristics of a N. Hemisphere baroclinic wave, based on quasi-geostrophic theory. H & L are high and low pressure. Circle-dot and circle-X are flow out-of and into the page, respectively. (after Cushman-Roisin, 1994: "Intro. to Geophysical Fluid Dynamics". Prentice Hall.)

Sample Application

Given A = 900 km, λ = 3784 km, c_o = –2.21 m s^{-1}, Z_T = 12 km, f_c = 1.117x10^{-4} s^{-1}, ρ_o = 1 kg m^{-3}, N_{BV} = 0.0148 s^{-1}, and |g|/θ_o = 0.038 m s^{-2} K^{-1}. Find the amplitudes for all baroclinic wave variables.

Find the Answer

Given: A = 900 km, λ = 3784 km, c_o = –2.21 m s^{-1}, Z_T = 12 km, f_c = 1.117x10^{-4} s^{-1}, ρ_o = 1 kg m^{-3}, N_{BV} = 0.0148 s^{-1}, and |g|/θ_o = 0.038 m s^{-2} K^{-1}.

Find: $\hat{Y}, \hat{\eta}, \hat{\theta}, \hat{P}, \hat{U}, \hat{V}, \hat{W}$

Apply eqs. (11.47). $\hat{Y} = A = 900\text{km}$

$\hat{\eta} = (900\text{km}) \cdot \pi \cdot \left(1.117 \times 10^{-4}\text{s}^{-1}\right) \cdot (2.21\text{ms}^{-1}) \,/\, \left[(12\text{km}) \cdot (0.0148\text{s}^{-1})^2\right] = 265.5\text{m}$

$\hat{\theta} = 1.53\text{K}$

$\hat{P} = 0.222\text{kPa}$

$\hat{U} = 0.109\text{ms}^{-1}$

$\hat{V} = 3.30\text{ms}^{-1}$

$\hat{W} = 0.000974\text{ms}^{-1} = 3.5\text{mh}^{-1}$

Check: Physics & units are reasonable.

Exposition: Compared to the 40 m s^{-1} jet-stream background winds, the U perturbation is small. In fact, many of these amplitudes are small. They would be larger for larger A and for shorter wavelengths and a shallower troposphere.

11.11.3. Meridional Transport by Rossby Waves

11.11.3.1. Heat Transport Meridionally

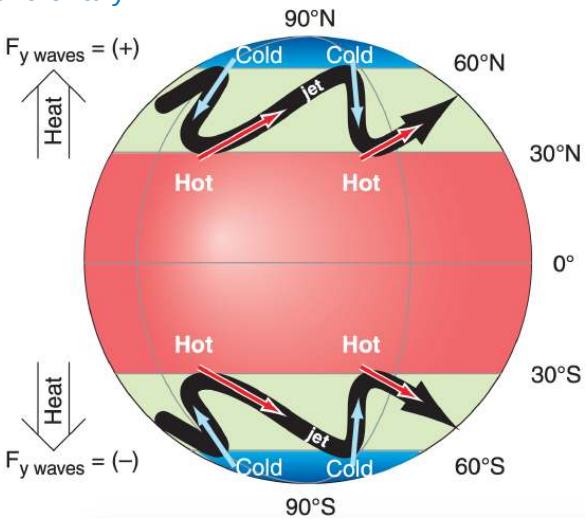


Figure 11.54 Sketch of how Rossby waves (thick black lines) transport heat poleward in midlatitudes. F_y is the meridional heat transport.

As the jet stream meanders north and south while encircling the earth with its west-to-east flow, its temperature changes in response to the regions it flows over (Fig. 11.54). For example, ridges in the planetary wave are where the jet stream is closest to the poles. This air is flowing over colder ground, and is also getting little or no direct solar heating because of the low sun angles (particularly in winter). As a result, the air becomes colder when near the poles than the average jet-stream temperature.

The opposite temperature change occurs in the troughs, where the jet stream is closest to the equator. The air near troughs is flowing over a warmer Earth surface where significant amounts of heat are moved into the jet stream via convective clouds and solar radiation. Define T' as the amount that the temperature deviates from the average value, so that positive T' means warmer (in the troughs), and negative T' means colder (in the ridges) than average.

Define v' as deviation in meridional velocity relative to the mean (V = 0). Positive v' occurs where meandering air has a component toward the north, and negative v' means a component toward the south.

Northward movement (v' = +) of warm air (T' = +) in the N. Hemisphere contributes to a positive (northward) heat flux v'T'. This is a kinematic flux, because units are K·m/s. Similarly, southward movement (v' = –) of cold air (T' = –) also contributes to a positive heat flux v'T' (because negative times negative = positive). Adding all the contributions from N different parts of the meandering jet stream gives an equation for the mean meridional heat flux F_y caused by Rossby waves:

$$F_{y\ \text{waves}} = (1/N) \cdot \Sigma\,(v' \cdot T')$$

This flux is positive (negative) in the N. (S.) Hemispheres. Thus, for both hemispheres, Rossby waves move heat from the tropics toward the poles.

The largest magnitudes of v' and T' occur at midlatitudes, where the jet stream has the largest meridional wind speeds co-located in the region of greatest meridional temperature gradient (Fig. 11.8). Fig. 11.55 shows a sketch of $F_{y\ waves}$ vs. latitude. When this line is concave to the right, it indicates positive curvature (Curv). Concave to the left is negative curvature. Curvature of $F_{y\ waves}$ will be used later.

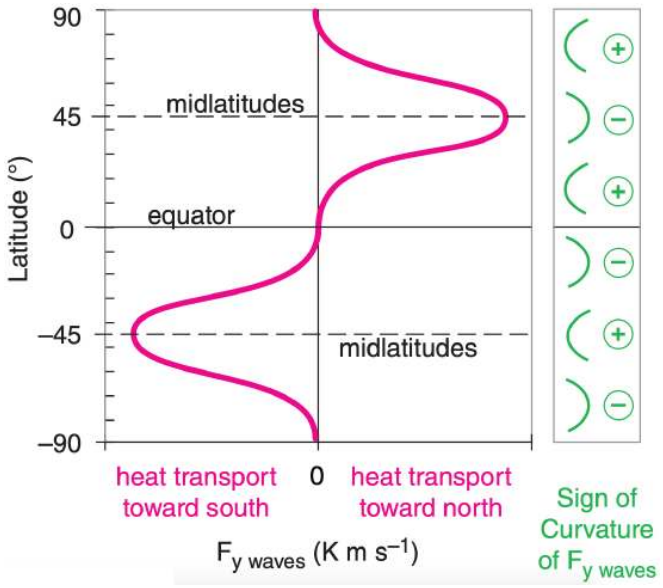


Figure 11.55 Heat flux (magenta line) toward the poles due to Rossby waves Green icons show whether the curvature of the heat-flux line is positive or negative.

11.11.3.2. Momentum Transport Meridionally

Even if the atmosphere were calm with respect to the Earth's surface, the fact that the Earth is rotating implies that the air is also moving toward the east. Latitude circles near the equator have much larger circumference than near the poles, hence air near the equator must be moving faster toward the east.

As Rossby waves move some of the tropical air poleward (positive v' in N. Hem.; negative v' in S. Hem.), conservation of U angular momentum requires the speed-up (positive U-wind perturbation: u') as the radius from Earth's axis decreases. Similarly, polar air moving equatorward must move slower (negative u' relative to the Earth's surface).

The north-south transport of U momentum is called the average kinematic momentum flux $\overline{u'v'}$:

$$\overline{u'v'} = (1/N) \cdot \Sigma\,(u' \cdot v')$$

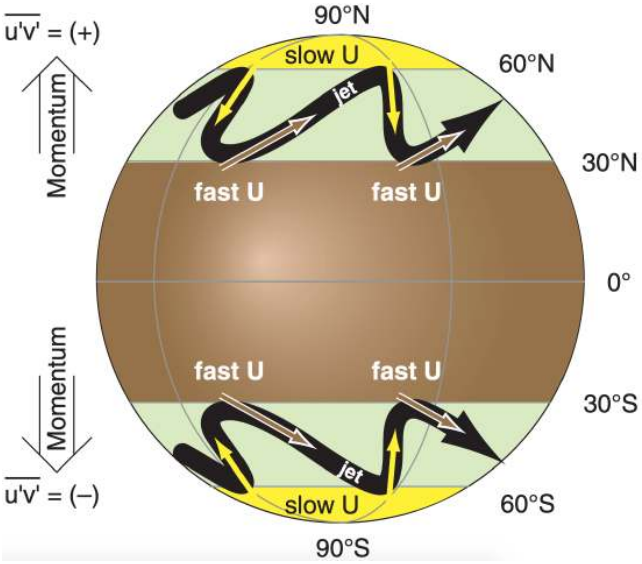


Figure 11.56 Sketch of how Rossby waves (thick black line) transport U momentum toward the poles.

For example, consider a N. Hemisphere Rossby-wave. Poleward motion (positive v' ; see brown arrows in Fig. 11.56) transports faster U winds (i.e., positive u'), causing positive $u'v'$. Similarly, equatorward motion (negative v'; see yellow arrows) transports slower U winds (i.e., negative u'), so again the product is positive $u'v'$. Averaging over all such segments of the jet stream gives

positive $\overline{u'v'}$ in the N. Hemisphere, which you can interpret as transport of zonal momentum toward the N. Pole by Rossby waves. In the Southern Hemisphere, $\overline{u'v'}$ is negative, which implies transport of U momentum toward the S. Pole.

To compensate for the larger area of fast U winds (brown in Fig. 11.56) relative to the smaller area of slower U winds (yellow in Fig. 11.56), one can multiply the angular momentum by a = cos(ϕ_s)/cos(ϕ_d):

Erroneous nesting of equation structures

where subscript s represents source location, d is destination, and $\Omega \cdot R_{Earth}$ = 463.4 m s^{-1} as before.

Sample Application

Find the weighted zonal velocity perturbation for air arriving at 50°N from 70°N.

Find the Answer

Given: ϕ_s = 70°N, ϕ_d = 50°N

Find: a·u' = ? m s^{-1}

Apply eq. (11.48): $a \cdot u' = (463.4\text{m/s}) \cdot \left[\frac{\cos^2(70^\circ)}{\cos(50^\circ)} - \cos(50^\circ)\right] \frac{\cos(70^\circ)}{\cos(50^\circ)}$

= **– 113.6 m s^{-1}**

Check: Physics reasonable, but magnitude too large.

Exposition: The negative sign means air from 70°N is moving slower from the west than any point at 50°N on the Earth's surface is moving. Thus, relative to the Earth, the wind is from the east.

For example, consider the weighted zonal velocity a·U' that reaches destination latitude 45° from other source latitudes, as sketched in Fig. 11.57. Air starting from 30° has much larger magnitude |a·U'| than does air starting from 60°. (ALERT: *Angularmomentum conservation gives unrealistically large velocities, but is qualitatively informative.*)

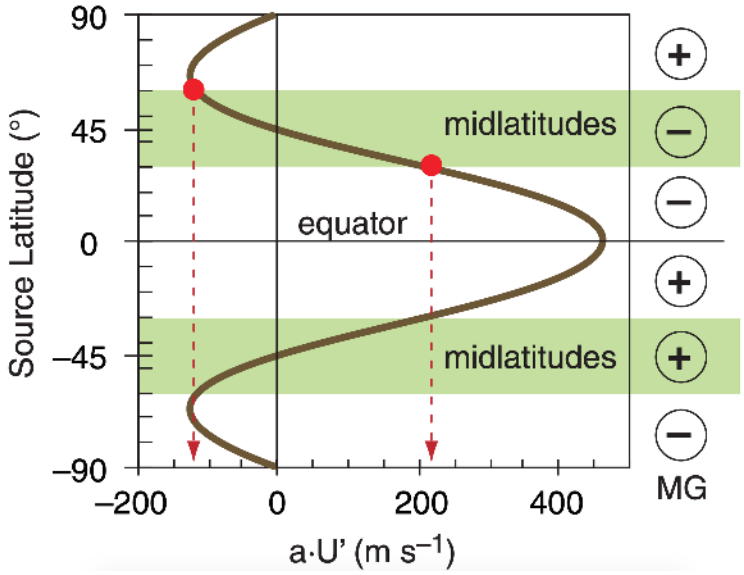


Figure 11.57 Weighted zonal velocity for air that arrives at midlatitude destination 45° from different source latitudes, where the weight is proportional to the amount of air in the source region.

The change of weighted Rossby-wave momentum flux with latitude is

Erroneous nesting of equation structures

where MG is the north-south gradient of zonal momentum. From Fig. 11.57 we infer that MG is negative (positive) in N. (S.) Hemisphere midlatitudes. This meridional gradient implies that excess zonal momentum from the tropics is being deposited at midlatitudes by the Rossby waves.

Rossby waves that transport U momentum poleward have a recognizable rounded-sawtooth shape, as sketched in Figs. 11.56 & 11.58. Specifically, the equatorward-moving portions of the jet stream are aligned more north-south (i.e., are more meridional), and sometimes even tilt backwards (toward the west as it moves toward the equator). The poleward-moving portions of the jet are more zonal (west to east).

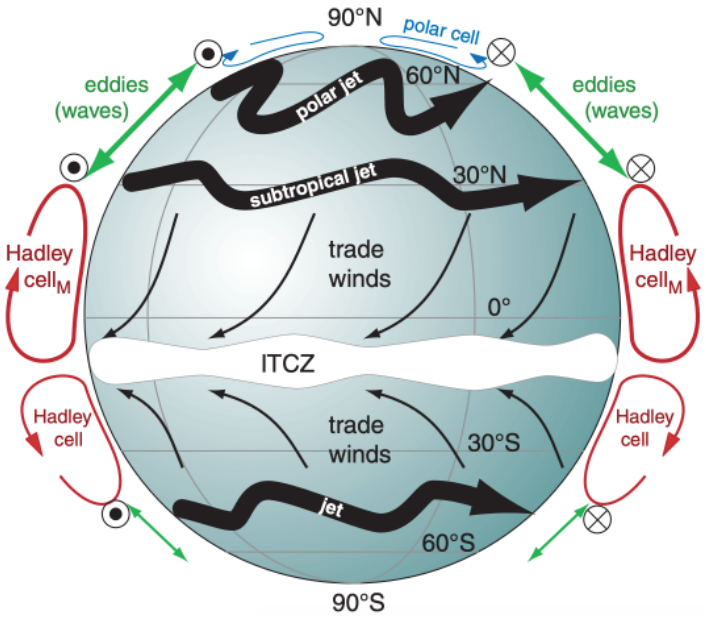


Figure 11.58 Sketch of 3-band global circulation for February (N. Hemisphere winter): 1) a Hadley cell of vertical circulation at low latitudes, 2) Rossby waves with horizontally meandering jet stream at midlatitudes, and 3) a weak polar cell of vertical circulation at high latitudes. For the jet axes, the circle-dot is wind out of the page, and circle-X is wind into the page.

This page titled 3.4.7: Extratropical Ridges and Troughs (Rossby Waves) is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.11: Extratropical Ridges and Troughs (Rossby Waves)** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.8: Three-band Global Circulation

In the chapter introduction, it was stated that Coriolis force causes the thermally-driven planetary circulation to break down into 3 latitude-bands (Fig. 11.58) in each hemisphere. These bands are: (1) a strong, **direct**, asymmetric, vertical-circulation Hadley cell in low latitudes (0° - 30°); (2) a band of mostly horizontal Rossby waves at mid-latitudes (30° - 60°); and a weak direct vertical circulation cell at high latitudes (60° to 90°). Fig. 11.58 includes more (but not all) of the details and asymmetries explained in this chapter.

The circulation bands work together to globally transport atmospheric heat (Fig. 11.14), helping undo the differential heating that was caused by solar and IR radiation. The Earth-atmosphere-ocean system is in near equilibrium thermally, with only extremely small trends over time related to global warming.

The circulation bands also work together in the meridional transport of zonal momentum. The trade winds, blowing opposite to the Earth's rotation, exert a **torque** (force times radius) that tends to slow the Earth's spin due to frictional drag against the land and ocean surface. However, in mid-latitudes, the westerlies dragging against the Earth's surface and against mountains apply an opposite torque, tending to accelerate the Earth's spin. On the long term, the opposite torques nearly cancel each other. Thus, the whole Earth-atmosphere-ocean system maintains a near-equilibrium spin rate.

INFO • Torques on the Earth

During high-wind episodes in one of the circulation bands, temporary changes in wind-drag torques are large enough to make measurable changes in the Earth's rotation rate — causing the length of a day to increase or decrease 1 - 3 μs over periods of months. In addition, external influences (lunar and solar tides, solar wind, geomagnetic effects, space dust) cause the Earth to spin ever more slowly, causing the length of a day to increase 1.4 ms/century at present.

11.12.1. A Metric for Vertical Circulation

Getting back to atmospheric circulations, one can define the strength CC of a vertical circulation cell as:

Erroneous nesting of equation structures

Direct circulation cells are ones with a vertical circulation in the direction you would expect if there were no Coriolis force. The units for circulation are s^{-1}.

Using the major Hadley cell as an example of a direct circulation (Fig. 11.59), note that w decreases as y increases; hence, $\Delta w/\Delta y$ is negative. Similarly, $\Delta V/\Delta z$ is positive. Thus, eq. (11.50) gives a positive CC value for direct circulations, and a negative value for **indirect circulations** (having an opposite rotation direction).

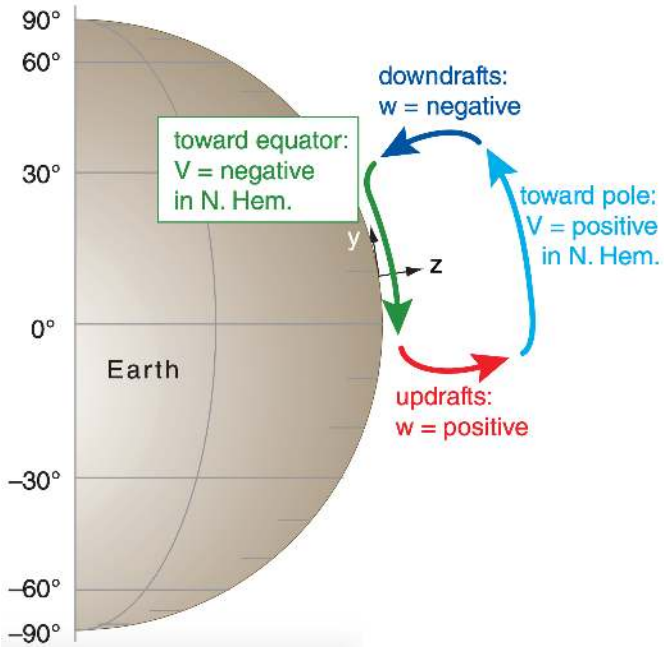


Figure 11.59 The major Hadley cell during Northern Hemisphere winter is an example of a direct circulation.

11.12.2. Effective Vertical Circulation

When the forecast equations for momentum and heat are applied to eq. (11.50), the result allows you to anticipate the value of CC for a variety of situations— even situations where vertical cells are not dominant:

Erroneous nesting of equation structures

circulation radiation wave-heat wave-momentum

In this equation are factors and terms that were discussed earlier in this chapter. For example, you can use Fig. 11.10 to estimate E_{net}, the differential heating due to radiation, and how it varies with y. The sign of the curvature (Curv) of the Rossby-wave heat flux $F_{y\ wave}$ was shown in Fig. 11.55. If we assume that the meridional gradient of zonal momentum MG ≈ 0 near the ground, then ΔMG/Δz has the same sign as MG, as was sketched in Fig. 11.57.

With this information, you can estimate the sign of CC in different latitude bands. Namely, you can anticipate direct and indirect circulations. For the Northern Hemisphere, the results are:

circulation ∝ radiation + wave-heat + wave-momentum = total

CC_{polar} ∝ positive + positive + positive = positive

CC_{midlat} ∝ positive + negative +negative = negative

$CC_{tropics}$∝ positive + positive + positive = positive

Sample Application

Suppose the Hadley cell updraft and downdraft velocities are 6 and –4 mm s^{-1}, respectively, and the meridional wind speeds are 3 m s^{-1} at the top and bottom of the cell. The major Hadley cell is about 17 km high by 3900 km wide, and is centered at about 10° latitude. Temperature in the tropical atmosphere decreases from about 25°C near the surface to –77°C at 17 km altitude. Find the vertical cell circulation.

Find the Answer

Given: Δw = –10 mm s^{-1} = –0.01 m s^{-1} , Δy = 3.9x10^3 km

ΔV = 6 m s^{-1}, Δz = 17 km, ϕ = 10°, ΔT = –102°C.

Find: CC = ? s^{-1}.

First, use eq. (10.16):

$f_c = (1.458\text{x}10^{-4}\ s^{-1})\cdot\sin(10°) = 2.53\text{x}10^{-5}\ s^{-1}$

Next, for the Brunt-Väisälä frequency, we first need:

$T_{avg} = 0.5*(25 - 77)°C = -26°C = 247\ K$

In the tropics $\Delta T/\Delta z = -6\ °C\ km^{-1} = -0.006\ K\ m^{-1}$

Then use eq. (5.4), and assume $T_v = T$:

$$N_{BV} = \sqrt{\frac{9.8\text{m/s}^2}{247\text{K}}(-0.006 + 0.0098)\frac{\text{K}}{\text{m}}} = 0.0123\text{s}^{-1}$$

Finally, use eq. (11.50):

$$CC = \left[\left(\frac{2.53 \times 10^{-5}\text{s}^{-1}}{0.0123\text{s}^{-1}}\right)^2 \cdot \frac{6\text{m/s}}{17000\text{m}}\right] - \frac{-0.01\text{m/s}}{3.9 \times 10^6\text{m}}$$

$CC = 1.493\text{x}10^{-9} + 2.564\text{x}10^{-9}\ s^{-1} = \mathbf{4.06\text{x}10^{-9}\ s^{-1}}$

Check: Physics & units are reasonable.

Exposition: Both terms contribute positively to the circulation of the major Hadley cell during Northern Hemisphere winter.

Rossby waves so efficiently transport heat and momentum at midlatitudes that the effective vertical circulation is negative. This indirect circulation is called the **Ferrel cell**.

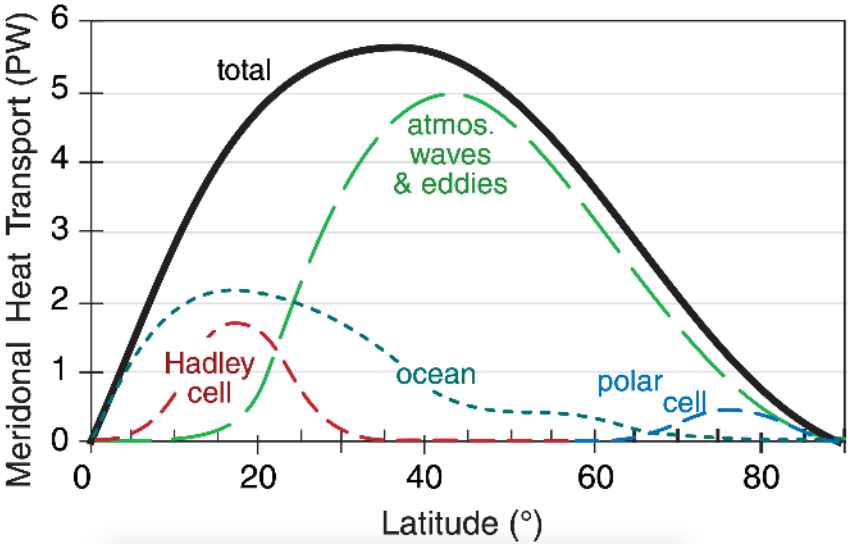


Figure 11.60 Contributions of direct circulations (medium dashed line, for Hadley cell and polar cell) and indirect circulations (long dashed line, for Rossby waves and eddies) to total meridional heat transport Tr in the N. Hemisphere

Fig. 11.60 redraws the N. Hemisphere portion of Fig. 11.14, qualitatively highlighting the relative contributions of the direct (Hadley and polar cells) and indirect (Ferrel cell/Rossby waves) atmospheric circulations to the total meridional heat transport. At mid-latitudes, the main circulation feature is the Rossby-wave meanders of the jet stream near the tropopause, and the associated low- and high-pressure centers near the surface. High- and low-latitudes are dominated by direct vertical circulations.

Ocean currents also contribute to global heat redistribution. Although ocean-circulation details are not within the scope of this book, we will introduce one ocean topic here — the Ekman spiral. This describes how wind drag can drive some ocean currents, including hurricane storm surges.

This page titled 3.4.8: Three-band Global Circulation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.12: Three-band Global Circulation** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

https://geo.libretexts.org/@go/page/20938

3.4.9: Ekman Spiral of Ocean Currents

Frictional drag between the atmosphere and ocean enables winds to drive ocean-surface currents. Coriolis force causes the surface current to be 45° to the right of the wind direction in the Northern Hemisphere. Drag between that surface-current and deeper water drives deeper currents that are slower, and which also are to the right of the current above. The result is an array of ocean-current vectors that trace a spiral (Fig. 11.61) called the **Ekman spiral**.

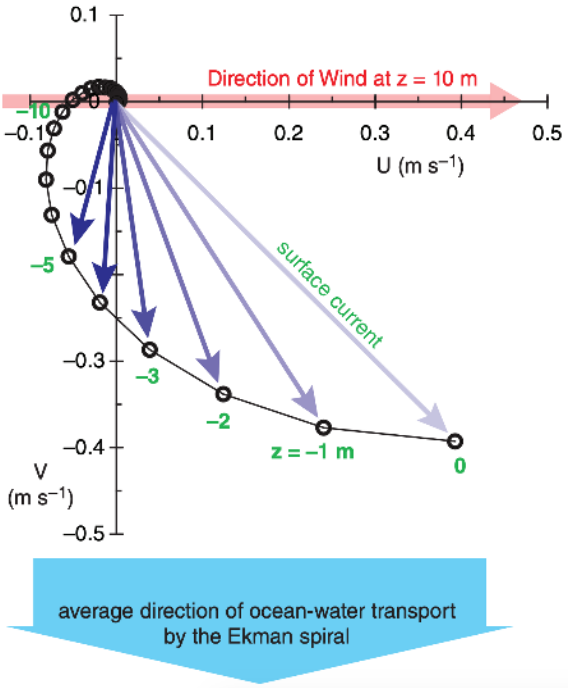


Figure 11.61 Example of an Ekman spiral in the ocean. Blue arrows show ocean currents at different depths (numbers in green). U current direction is aligned with the near-surface wind direction. Red arrow shows the wind direction, but speed is much greater. This example is based on a 10 m s^{-1} wind at 45° latitude.

The equilibrium horizontal ocean-current components (U, V, for a rotated coordinate system having U aligned with the wind direction) as a function of depth (z) are:

Erroneous nesting of equation structures

Erroneous nesting of equation structures

where z is negative below the ocean surface, f_c = Coriolis parameter, and u_{*water} is a **friction velocity** (m s^{-1}) for water. It can be found from

Erroneous nesting of equation structures

where the density ratio $\rho_{air}/\rho_{water} \approx 0.001195$ for sea water. The friction velocity (m s–1) for air can be approximated using **Charnock's relationship**:

Erroneous nesting of equation structures

where M is near-surface wind speed (at z = 10 m) in units of m s^{-1}.

The **Ekman-layer depth** scale is

Erroneous nesting of equation structures

where K is the ocean eddy viscosity (a measure of ability of ocean turbulence to mix momentum). One approximation is $K \approx 0.4\,|z|\,u_{*water}$. Although K varies with depth, for simplicity in this illustration I used constant K corresponding to its value at z = –0.2 m, which gave K ≈ 0.001 $m^2\ s^{-1}$.

The average water-mass transport by Ekman ocean processes is 90° to the right (left) of the near-surface wind in the N. (S.) Hemisphere (Fig. 11.61). This movement of water affects sea-level under hurricanes (see the Tropical Cyclone chapter).

Sample Application

For an east wind of 14 m/s at 30°N, graph the Ekman spiral.

Find the Answer

Given: M = 14 m·s^{-1}, ϕ = 30°N

Find: [U , V] (m s^{-1}) vs. depth z (m)

Using a relationship from the Forces and Winds chapter, find the Coriolis parameter: $f_c = 7.29 \times 10^{-5}$ s^{-1} .

Next, use Charnock's relationship. eq. (11.54): $u_{*air}^2 = 0.00044\ (14^{2.55}) = 0.368$ m^2 s^{-2} .

Then use eq. (11.53):

$u_{*water}^2 = 0.001195 \cdot (0.368\ \text{m}^2\ \text{s}^{-2}) = 0.00044\ \text{m}^2\ \text{s}^{-2}$

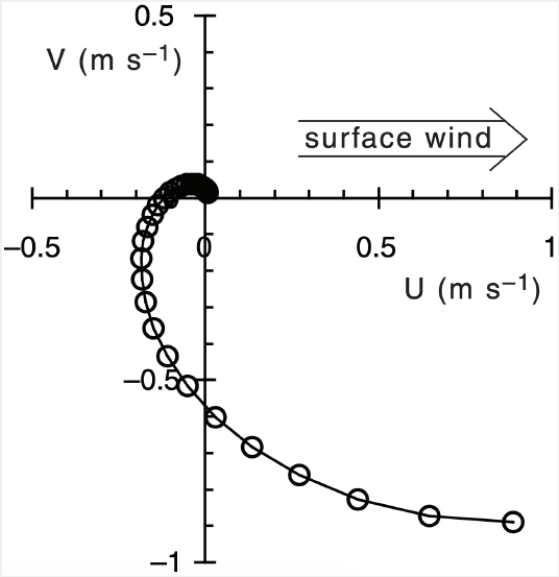
Estimate K at depth 0.2 m in the ocean from

$K \approx 0.4 \cdot (0.2\ \text{m}) \cdot [0.00044\ \text{m}^2\ \text{s}^{-2}]^{1/2} = 0.00168\ \text{m}^2\ \text{s}^{-1}$

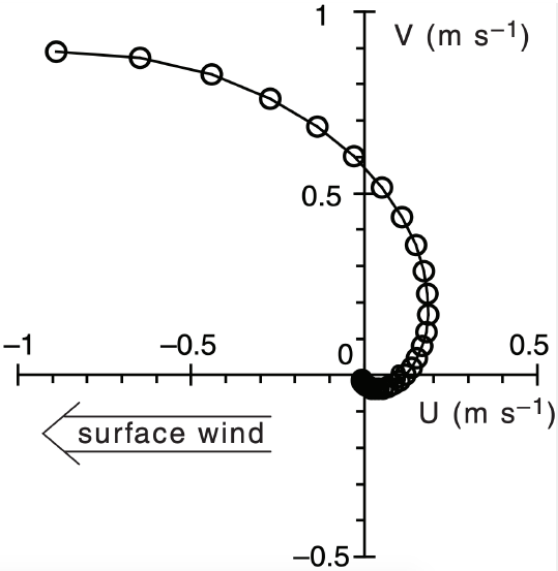
Use eq. (11.55):

$D = [2 \cdot (0.00168\ \text{m}^2\ \text{s}^{-1}) / (7.29 \times 10^{-5}\ \text{s}^{-1})]^{1/2} = 6.785$ m

Use a spreadsheet to solve for (U, V) for a range of z, using eqs. (11.52). I used z = 0, –1, –2, –3 m etc.



Finally, rotate the graph 180° because the wind is from the East.



Check: Physics & units are reasonable.

Exposition: Net transport of ocean water is toward the north for this East wind case.

This page titled 3.4.9: Ekman Spiral of Ocean Currents is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.13: Ekman Spiral of Ocean Currents** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.10: Explaining the General Circulation

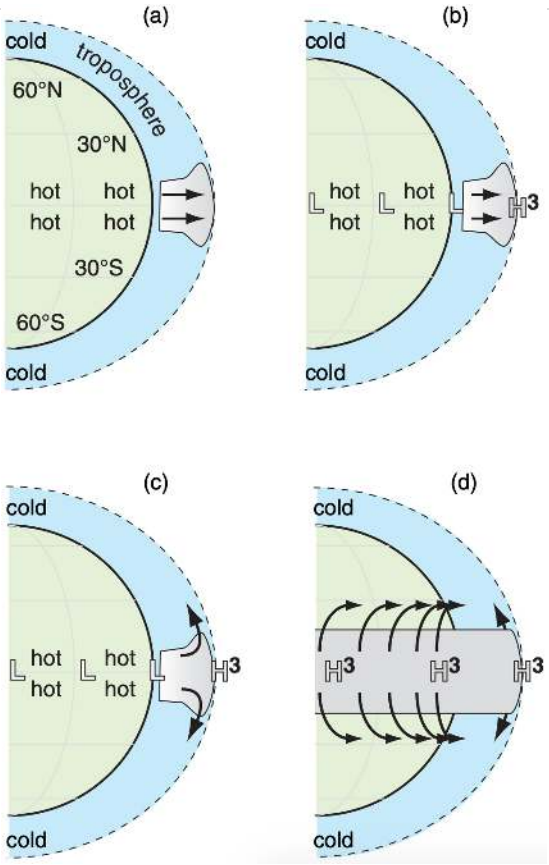


Figure 11.26 Application of physical concepts to explain the general circulation (see text). White-filled H and L indicate surface high and low pressure regions; grey-filled H and L are pressure regions near the tropopause. (continued on next pages)

11.7.1. Low Latitudes

Differential heating of the Earth's surface warms the tropics and cools the poles (Fig. 11.26a). The warm air near the equator can hold large amounts of water vapor evaporated from the oceans. Buoyancy force causes the hot humid air to rise over the equator. As the air rises, it cools and water vapor condenses, causing a belt of thick thunderstorm clouds around the equator (Fig. 11.26a) with heavy tropical precipitation.

The buoyantly forced vertical motion removes air molecules from the lower troposphere in the tropics, and deposits the air near the top of the troposphere. The result is a pressure couplet (Fig. 11.26b) of very high perturbation pressure p' (indicated with HHH or H^3 on the figures) near the tropopause, and low perturbation pressure (L in the figures) at the surface.

The belt of tropical high pressure near the tropopause forces air to diverge horizontally, forcing some air into the Northern Hemisphere and some into the Southern (Fig. 11.26c). With no Coriolis force at the equator, these winds are driven directly away from the high-pressure belt.

As these high-altitude winds move away from the equator, they are increasingly affected by Coriolis force (Fig. 11.26d). This causes winds moving into the Northern Hemisphere to turn to their right, and those moving into the Southern Hemisphere to turn to their left. Winds in both hemispheres accelerate.

But as the winds move further and further away from the equator, they are turned more and more to the east, creating the subtropical jet (Fig. 11.26e) at about 30° latitude. Coriolis force prevents these upper-level winds from getting further away from the equator than about 30° latitude (north and south), so the air accumulates and the pressure increases in those belts.

When simulations of the general circulation impose a larger Coriolis force (as if the Earth spun faster), the convergence bands occur closer to the equator. This effect is indeed found in the atmospheres of Jupiter and Saturn, as is evident by motions in their

banded structures. For weaker Coriolis force, the convergence is closer to the poles. But for our Earth, the air converges at 30° latitude.

This pressure perturbation p' is labeled as HH or H^2 in Fig. 11.26e, to show that it is a positive pressure perturbation, but not as strong as the H^3 perturbation over the equator. Namely, the horizontal pressure gradient between H^3 and H^2 drives the upperlevel winds to diverge away from the equator.

A SCIENTIFIC PERSPECTIVE • Model Sensitivity

CAUTION: Whenever you find that a model has high sensitivity (i.e., the output result varies by large magnitude for small changes in the input parameter), you should be especially wary of the results. Small errors in the parameter could cause large errors in the result. Also, if the real atmosphere does not share the same sensitivity, then this is a clue that the model is poorly designed, and perhaps a better model should be considered.

Models are used frequently in meteorology — for example: numerical weather prediction models (Chapter 20) or climate-change models (Chapter 21). Most researchers who utilize models will perform careful **sensitivity studies** (i.e., compute the output results for a wide range of parameter values) to help them gauge the potential weaknesses of the model.

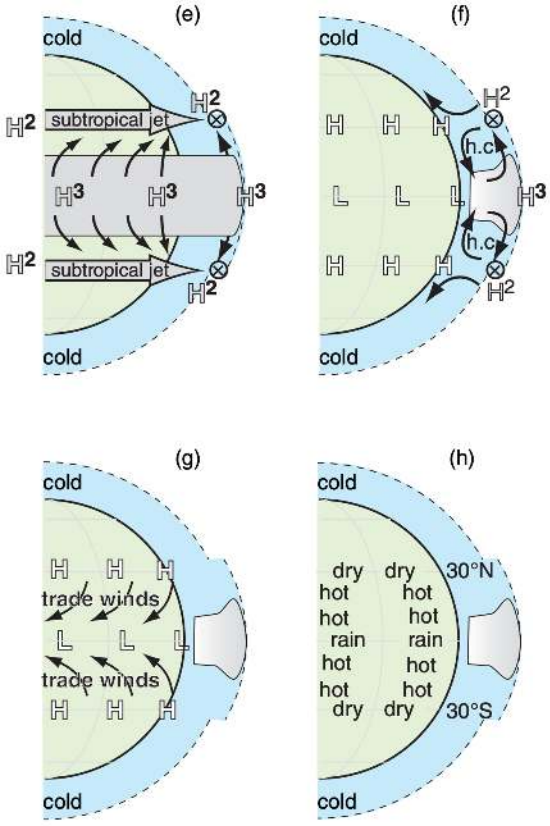


Figure 11.26 (continuation) Explanation of low-latitude portion of the global circulation. The dashed line shows the tropopause. The "x" with a circle around it (representing the tail feathers of an arrow) indicates the axis of a jet stream that goes into the page.

The excess air accumulated at 30° latitudes cannot go up into the stratosphere in the face of very strong static stability there. The air cannot go further poleward because of Coriolis force. And the air cannot move equatorward in the face of the strong upper-level winds leaving the equator. The only remaining path is downward at 30° latitude (Fig. 11.26f) as a nonhydrostatic flow. As air accumulates near the ground, it causes a high-pressure perturbation there — the belt of subtropical highs labeled with H.

The descending air at 30° latitude warms dryadiabatically, and does not contain much moisture because it was squeezed out earlier in the thunderstorm updrafts. These are the latitudes of the subtropical deserts (Fig. 11.26h), and the source of hot airmasses near the surface.

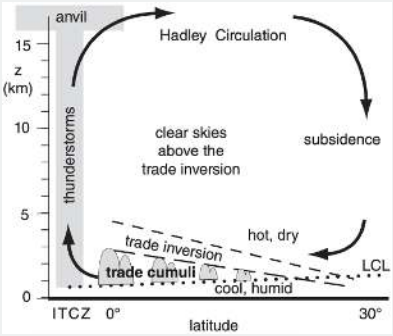
Finally, the horizontal pressure gradient between the surface subtropical highs near 30° latitude and the equatorial lows near 0° latitude drives the surface winds toward the equator. Coriolis force turns these winds toward the west in both hemispheres (Fig. 11.26g), resulting in the easterly trade winds (winds from the east) that converge at the ITCZ.

The total vertical circulation in the tropics and subtropics we recognize as the Hadley Cell (labeled h.c. in Fig. 11.26f). This vertical circulation (a **thermally-direct circulation**) is so vigorous in its vertical mixing and heat transport that it creates a deeper troposphere in the tropics than elsewhere (Fig. 11.4). Also, the vigorous circulation spreads and horizontally mixes the radiatively warmed air somewhat uniformly between ±30° latitude, as sketched by the flattened temperature curve in Fig. 11.8a.

INFO • The Trade Inversion

Descending air in the subtropical arm of the Hadley circulation is hot and dry. Air near the tropical sea surface is relatively cool and humid. Between these layers is a strong temperature inversion called the **trade inversion** or **passat inversion**. This statically stable layer (between the dashed lines in the Figure) creates a lid to the tropical convection below it. The inversion base is lowest (order of 500 m) in the subtropics, and is highest (order of 2,500 m) near the ITCZ. Fair-weather cumulus clouds (trade cumuli) between the lifting condensation level (LCL) and the trade inversion are shallowest in the subtropics and deeper closer to the ITCZ.

By capping the humid air below it, the trade inversion allows a latent-heat fuel supply to build up, which can be released in hurricanes and ITCZ thunderstorms.



11.7.2. High Latitudes

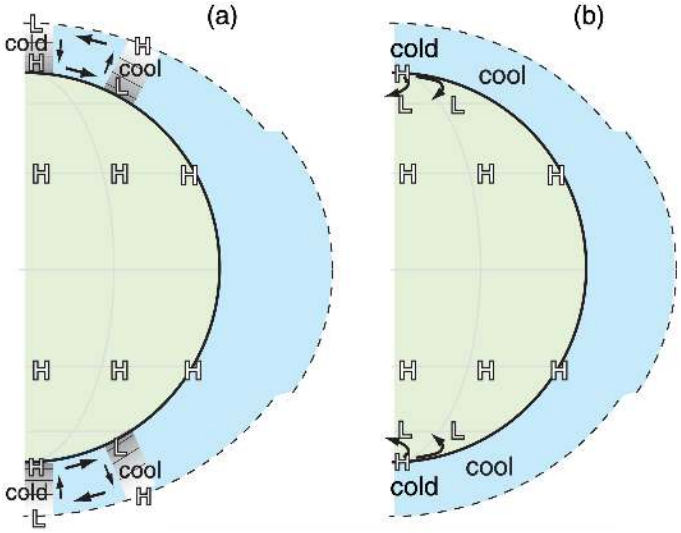


Figure 11.27 Explanation of high-latitude portion of general circulation.

As sketched in Fig. 11.8a, air temperature is very cold at the poles, and is cool at 60° latitude. The temperature difference between 60° and 90° latitudes creates opposite north-south pressure gradients and winds at the top and bottom of the troposphere, due to the thermal circulation effect (Fig. 11.17). A vertical cross section of this thermal circulation (Fig. 11.27a) shows a weak polar cell. Air generally rises near 60°N and descends near the pole.

At the poles are surface high-pressure centers, and at 60° latitudes are belts of subpolar lows at the surface. This horizontal pressure gradient drives equatorward winds, which are turned toward the west in both hemispheres due to Coriolis force (Fig. 11.27b). Namely, the winds become geostrophic, and are known as **polar easterlies**.

At the top of the shallow (6 to 8 km thick) troposphere are poleward winds that are turned toward the east by Coriolis force. These result in an upperlevel westerly flow that circulates around the upperlevel polar low (Fig. 11.3b).

11.7.3. Mid-latitudes

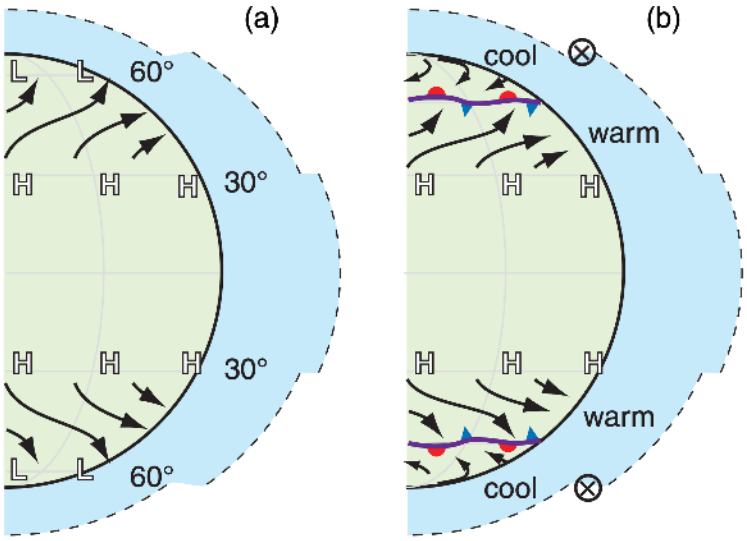


Figure 11.28 Explanation of mid-latitude flow near the Earth's surface. H and L indicate belts of high and low pressure, black arrows are average surface winds, and the polar front is shaded grey. The circle with "x" in it represents the tail feathers of the jet-stream wind vector blowing into the page.

Recall that the Hadley cell is unable to mix heat beyond about ±30° latitude. This leaves a very strong meridional temperature gradient in mid-latitudes (Fig. 11.8) throughout the depth of the troposphere. Namely, the temperature change between the equator (0°) and the poles (90°) has been compressed to a latitude band of about 30 to 60° in each hemisphere.

Between the subtropical high-pressure belt near 30° latitude and the subpolar low-pressure belt near 60° latitude is a weak meridional pressure gradient near the Earth's surface. This climatological-average pressure gradient drives weak boundary-layer winds from the west in both hemispheres (Fig. 11.28a), while the drag of the air against the surface causes the wind to turn slightly toward low pressure.

Near the subpolar belt of low pressure is a region of surface-air convergence, with easterly winds from the poles meeting westerly winds from mid-latitudes. The boundary between the warm subtropical air and the cool polar air at the Earth's surface is called the **polar front** (Fig. 11.28b) — a region of even stronger horizontal temperature gradient.

Recall from the hypsometric equation in Chapter 1 that the height difference (i.e., the thickness) between two isobaric surfaces increases with increasing temperature. As a result of the meridional temperature gradient, isobaric surfaces near the top of the troposphere in mid-latitudes are much more steeply sloped than near the ground (Figs. 11.29 & 11.32). This is related to the thermal-wind effect.

Figure 11.29 Isobaric surfaces (thin solid black lines) are spaced further apart in hot air than in cool air. Regions of steeper slope of the isobars have stronger pressure gradient, and drive faster winds. The jet-stream axis is shown with

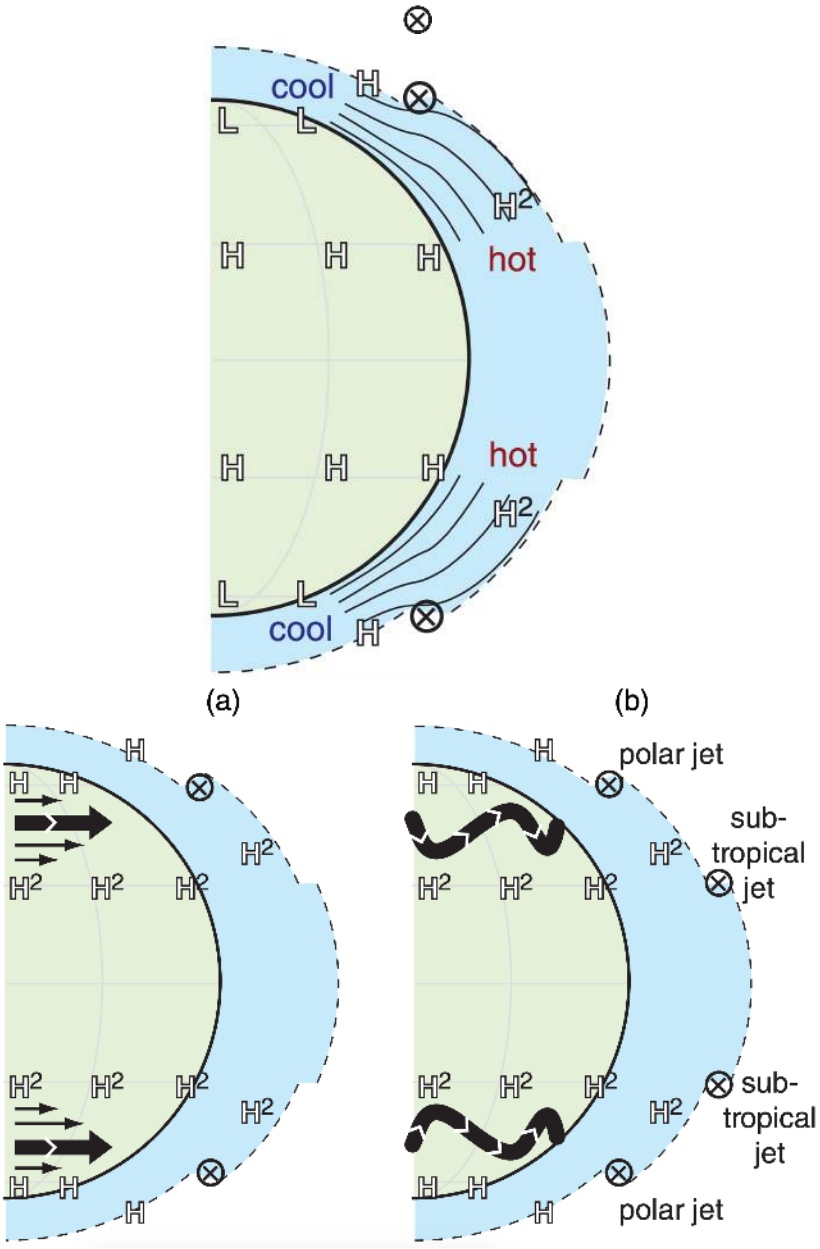


Figure 11.30 Mid-latitude flow near the top of the troposphere. The thick black arrow represents the core or axis of the jet stream: (a) average, (b) snapshot. (See caption in previous two figures for legend.)

In the Northern Hemisphere this effect is greatest in winter (Fig. 11.32), because there is the greatest temperature contrast between pole and equator. In the Southern Hemisphere, the cold Antarctic continent maintains a strong meridional temperature contrast all year.

Larger pressure gradients at higher altitudes drive stronger winds. The core of fastest westerly winds near the tropopause (where the pressure-gradient is strongest) is called the **polarjet stream**, and is also discussed in more detail later in this chapter. Thus, the climatological average winds throughout the troposphere at mid-latitudes are from the west (Fig. 11.30a) in both hemispheres.

Although the climatological average polar-jetstream winds are straight from west to east (as in Fig. 11.30a), the actual flow on any single day is unstable. Two factors cause this instability: the variation of Coriolis parameter with latitude (an effect that leads to **barotropic instability**), and the increase in static stability toward the poles (an effect that leads to **baroclinic instability**). Both of these instabilities are discussed in more detail later.

Air flow over a surface heterogeneity such as a mountain range can cause the initial meridional deviation of the jet stream. Once triggered, the jet stream continues to meander or oscillate north-south as it generally blows from the west (Fig. 11.30b). The meanders that form in this flow are called **Rossby waves**. Regions near the tropopause where the jet stream meanders equatorward are called **troughs**, because the lower pressure on the north side of the jet stream is brought equatorward there. Poleward meanders of the jet stream are called **ridges**, where higher pressure extends poleward. The locations of Rossby-wave troughs and ridges usually propagate toward the east with time, as will be explained in detail later in this chapter.

Recall that there is a subtropical jet at roughly 30° latitude associated with outflow from the Hadley Cell. Thus, in each hemisphere is a somewhat-steady subtropical jet and a meandering polar jet (Fig. 11.30b). Both of these jets are strongest in the winter hemisphere, where there is the greatest temperature gradient between cold poles and hot equator.

Troughs and ridges in the jet stream are crucial in creating and destroying cyclones and anticyclones near the Earth's surface. Namely, they cause the extremely large weather variability that is normal for mid-latitudes. The field of **synoptic meteorology** comprises the study and forecasting of these variable systems, as discussed in the Airmasses, Fronts, and Extratropical Cyclone chapters.

Figs. 11.31 and 11.32 show actual global pressure patterns at the bottom and top of the troposphere. The next section explains why the patterns over the oceans and continents differ.

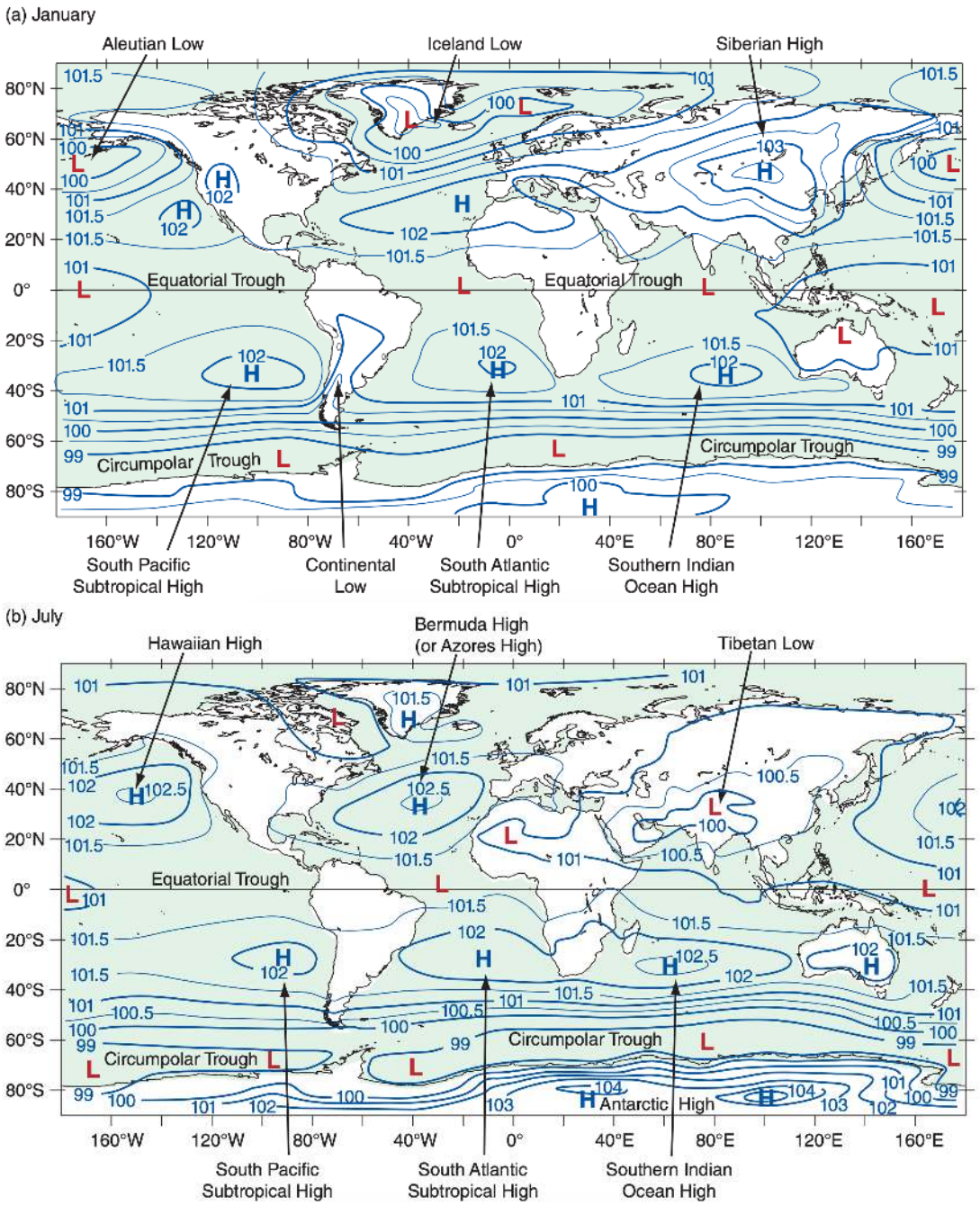


Figure 11.31 Mean sea-level pressure (kPa) for (a) January, and (b) July, averaged over 30 years: 1981 to 2010. [NCEP/NCAR Reanalysis.]

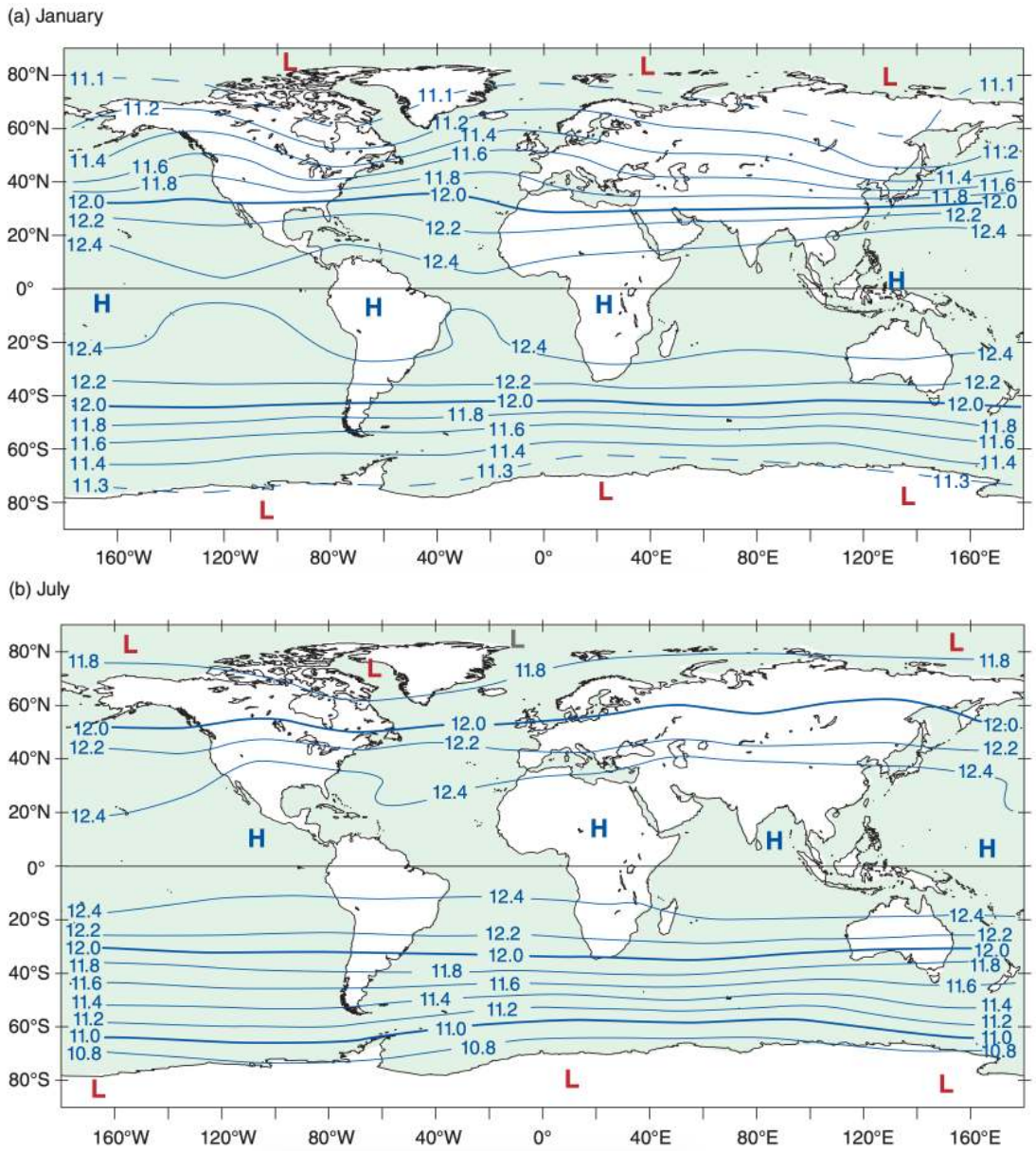


Figure 11.32 Geopotential height (km) of the 20 kPa isobaric surface (near the tropopause) for (a) January, and (b) July, averaged over 30 years: 1981 to 2010. [NCEP/NCAR Reanalysis.]

11.7.4. Monsoon

Recall from the Heat Budgets chapter that the temperature change of an object depends on the mass of material being heated or cooled, and on the material's specific heat. If you put the same amount of heat into objects of similar material but differing mass, the smaller masses will warm the most.

Rocks and soil on continents are opaque to sunlight, and are good insulators of heat. Thus, sunlight directly striking the land surface warms only a very thin top layer (mm) of molecules, causing this thin layer to get quite warm. Similarly, longwave (infrared) radiative heat loss at night causes the very thin top layer to become very cold. Namely, there is a large **diurnal** (daily) temperature contrast. Also, because there are more daylight hours in summer and more nighttime hours in winter, continental land surfaces tend to become hot in summer and cold in winter.

Water in the oceans is partially transparent and sometimes turbulent, allowing sunlight to be absorbed and spread over a thick layer (meters to tens of meters) of molecules. Also, water has a large specific heat (see the Heat Budgets chapter), hence a large input of

heat causes only a small temperature change. Thus, ocean surfaces have very small diurnal temperature changes, and have only a medium amount of seasonal temperature changes.

The net result is that during summer, continents warm faster than the oceans. During winter, continents cool faster than the oceans.

Consider a cold region next to a warm region. Over the cold surface, the near-surface air cools and develops a high-pressure center with anticyclonically rotating winds, as explained by the thermal circulation sketched in Fig. 11.17. Over warm surfaces, the thermal circulation causes near-surface air to warm and develop a low-pressure center with cyclonic winds. As already mentioned, this is called a **thermal low**.

Combining the effects from the previous two paragraphs with the strong tendency of the winds to become geostrophic (or gradient) yields the nearsurface monsoonal flows shown in Fig. 11.5. Opposite pressure patterns and circulations would occur near the top of the troposphere. The regions near surface lows tend to have rising air and abundant clouds and rain. Regions near surface highs tend to have dry fair weather with few clouds.

Monsoon circulations occur over every large continent and ocean (Fig. 11.31). Some are given names. Over the Atlantic in summer, winds on the south and west sides of the monsoonal **Bermuda High** (also called the **Azores High**) steer Atlantic hurricanes northward as they near North America. Over the North Pacific in summer is the **Hawaiian High** or **Pacific High** which provides cool northerly breezes and months of fair weather to the west coast of North America.

The summer low over northern India is called the **Tibetan Low**. It helps drive strong cross-equatorial flow (Fig. 11.33) that brings the much needed monsoon rains over India. Ghana in West Africa also receives a cross-equatorial monsoon flow.

Winter continental highs such as the **Siberian High** over Asia are formation locations for cold airmasses. Over the North Atlantic Ocean in winter is the **Icelandic Low**, with an average circulation on its south side that steers mid-latitude cyclones toward Great Britain and northern Europe. The south side of the winter **Aleutian Low** over the North Pacific brings strong onshore flow toward the west coast of North America, causing many days of clouds and rain.

The actual mean global circulation is a superposition of the zonally averaged flows and the monsoonal flows (Fig. 11.31). Also, a snapshot or satellite image of the Earth on any given day would likely deviate from the 30-year averages presented here. Other important aspects of the global circulation were not discussed, such as conversion between available potential energy and kinetic energy. Also, monsoons and the whole global circulation are modulated by El Niño / La Niña events and other oscillations, discussed in the Climate chapter.

In the previous sections, we have described characteristics of the global circulation in simple terms, looked at what drives these motions, and explained dynamically why they exist. Some of the phenomena we encountered deserve more complete analysis, including the jet stream, Rossby waves with their troughs and ridges, and some aspects of the ocean currents. The next sections give details about how these phenomena work.

This page titled 3.4.10: Explaining the General Circulation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.7: Explaining the General Circulation** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.11: Jet Streams

In the winter hemisphere there are often two strong jet streams of fast west-to-east moving air near the tropopause: the **polar jet stream** and the **subtropical jet stream** (Figs. 11.34 & 11.35).

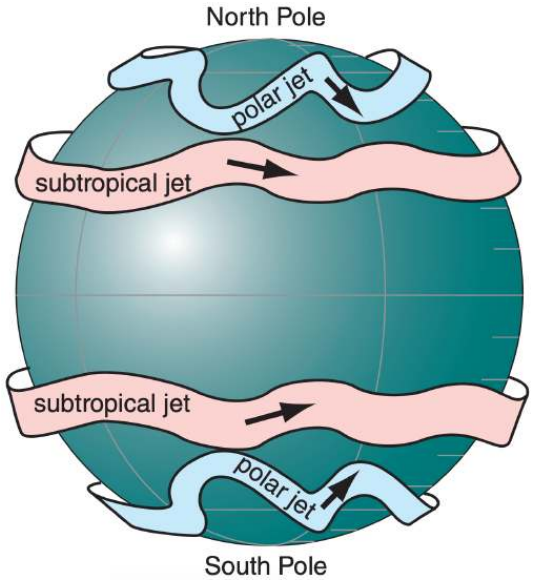


Figure 11.34 Sketch of jet streams, representing a snapshot.

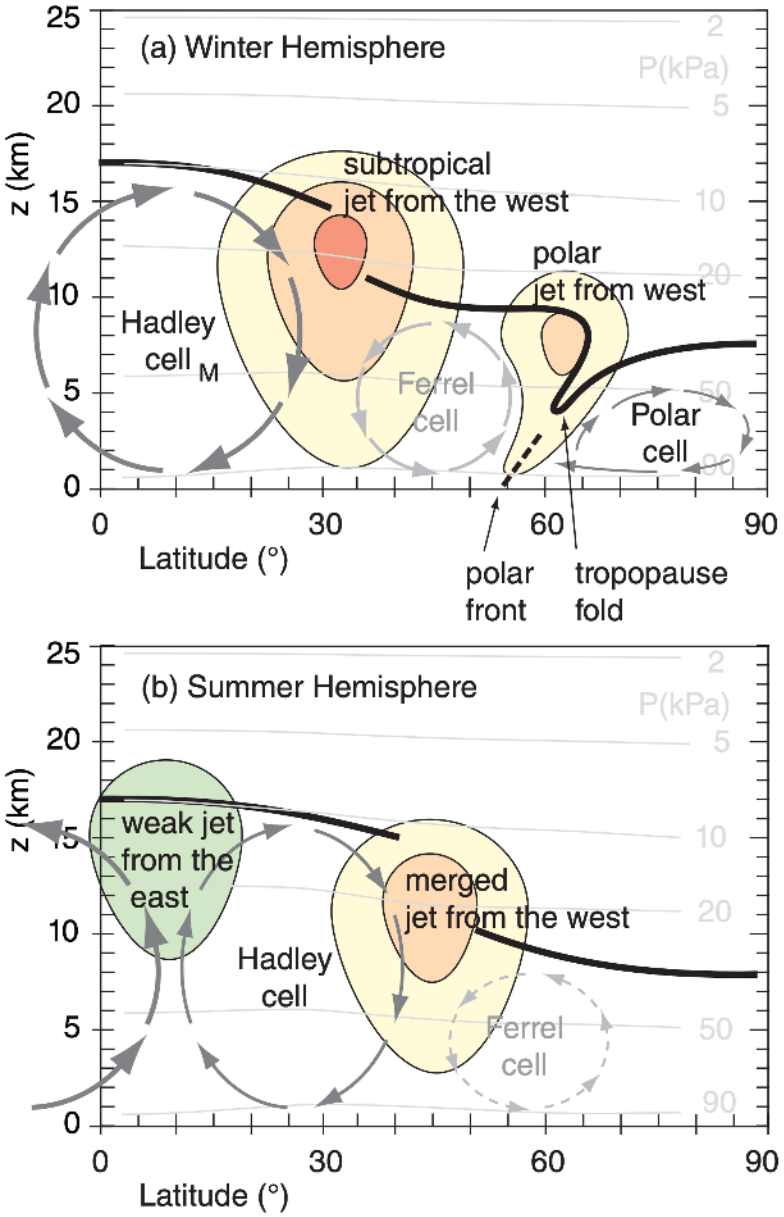


Figure 11.35 Simplified vertical cross-section. Thick solid line is the tropopause. Darker shading indicates faster winds (perpendicular to the page). This is a snapshot, not a climatological average; hence, the polar jet and polar front can be seen.

The subtropical jet is centered near 30° latitude in the winter hemisphere. This jet: (1) is very steady; (2) meanders north and south a bit; (3) is about 10° latitude wide (width ≈1,000 km); and (4) has seasonal-average speeds of about 45 m s^{-1} over the Atlantic Ocean, 55 to 65 m s^{-1} over Africa and the Indian Ocean, and 60 to 80 m s^{-1} over the western Pacific Ocean. The **core** of fast winds near its center is at 12 km altitude (Fig. 11.35). It is driven by outflow from the top of the Hadley cell, and is affected by both Coriolis force and angular-momentum conservation.

The polar jet is centered near 50 to 60° latitude in the winter hemisphere. The polar jet: (1) is extremely variable; (2) meanders extensively north and south; (3) is about 5° latitude wide; and (4) has widely varying speeds (25 to 100 m s^{-1}) driven by varying horizontal temperature gradients. The core altitude is about 9 km. This jet forms over the polar front — driven by thermal-wind effects due to the strong horizontal temperature contrast across the front.

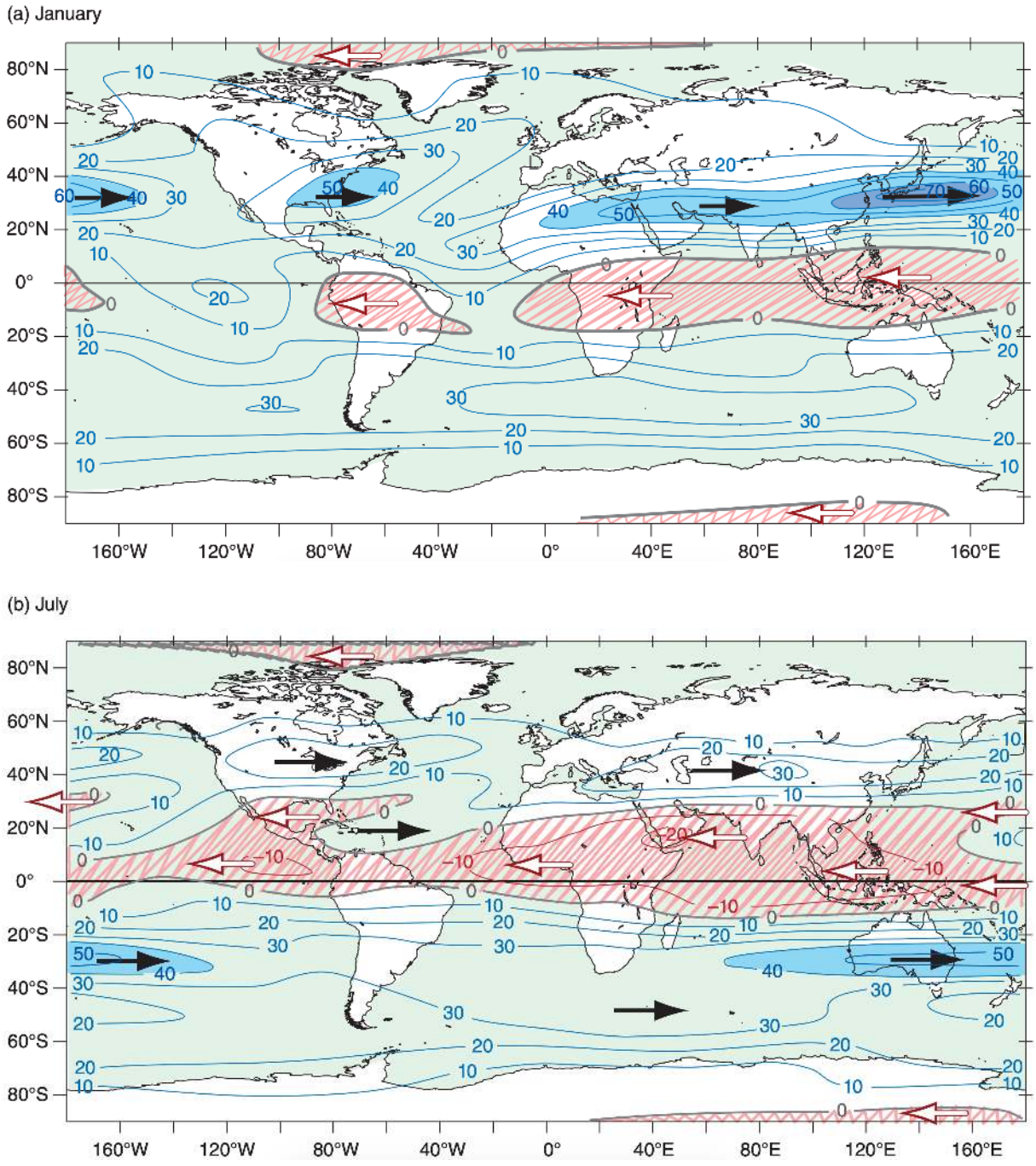
When meteorological data are averaged over 30 years, the subtropical jet shows up clearly in the data (e.g., Fig. 11.36) because it is so steady. However, the polar jet disappears because it meanders and shifts so extensively that it is washed out by the long-term average. Nonetheless, these transient meanders of the polar jet (troughs and ridges in the Rossby waves) are extremely important for mid-latitude cyclone formation and evolution (see the Extratropical Cyclone chapter).

In the summer hemisphere, the jets from the west have merged (Figs. 11.35b and 11.36), and the winds are slower because of the weaker temperature contrast between the equator and the warm pole. Core wind speeds in the jet are 0 to 10 m s^{-1} in N. Hemisphere summer, and 5 to 45 m s^{-1} in S. Hemisphere summer. This core shifts poleward to be centered near 40° to 45° latitude. A weak jet from the east is centered at roughly 10° latitude in the summer hemisphere.

INFO • Jet Stream Aspect Ratio

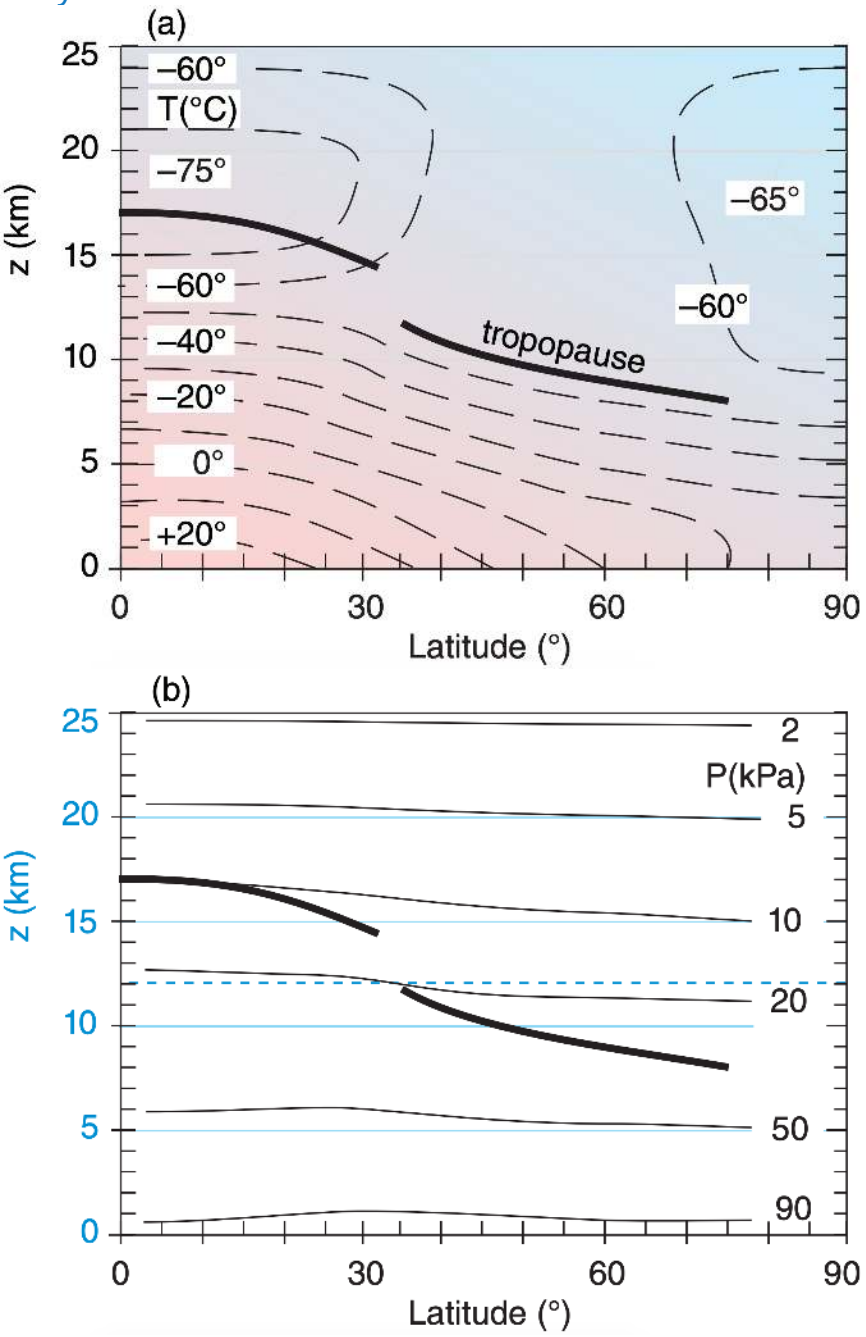
Jet streams in the real atmosphere look very much like the thin ribbons of fast-moving air, as sketched in Fig. 11.34. Jet vertical thickness (order of 5 to 10 km) is much smaller than their horizontal width (order of 1000 to 2000 km). Namely, their aspect ratio (width/ thickness) is large. Figures such as 11.35 are intentionally distorted to show vertical variations better.

Figure 11.36 Zonal (U) component of winds (m s
) at 20 kPa isobaric surface (near the
tropopause
) for (a) January, and (b) July, averaged over the 30 years of 1981 to 2010. Contour interval is 10 m s
. All negative (east) winds are lightly hatched in red, and are indicated with white arrows. Positive (west) winds are indicated with black arrows, and are shaded in blue with isotachs at 40 and 60 m s
. In polar regions, the tropopause is relatively low (at about 8 km altitude, where P ≈ 35 kPa; see Table 6-1 in Chapter 6), so 20 kPa is well above the tropopause (compare the 20 kPa winds here with tropopause winds sketched in Fig. 11.3b). [NCAR/NCEP Reanalysis.]



https://geo.libretexts.org/@go/page/20943

11.8.1. Baroclinicity & the Polar Jet



https://geo.libretexts.org/@go/page/20943

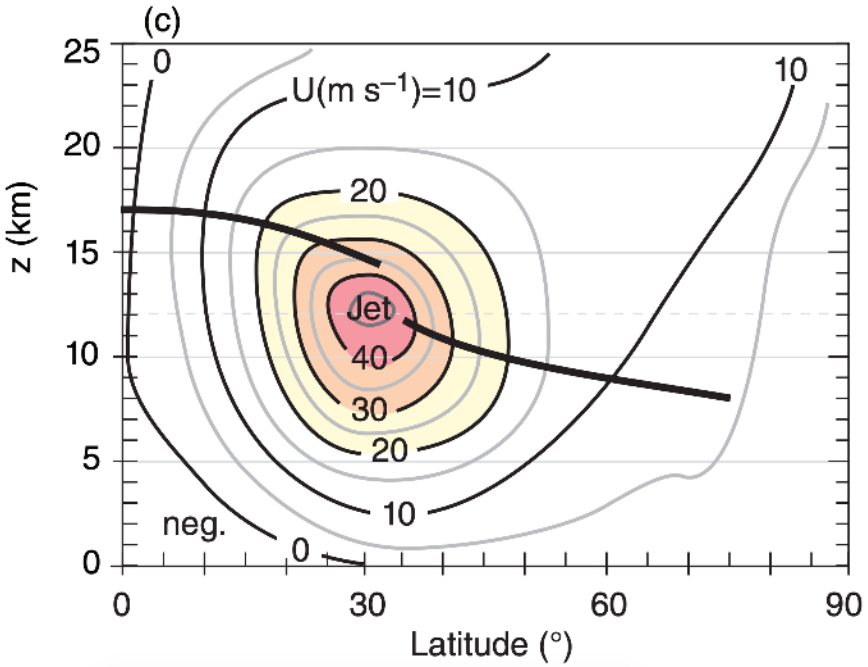


Figure 11.37 Vertical cross sections through the atmosphere on January 2003. (a) isotherms T (°C), (b) isobars P (kPa), (c) isotachs of the zonal wind U (m s^{-1}). The tropopause is indicated by a heavy solid line. For this Northern Hemisphere case, the jet-stream winds in (c) are from the west (i.e., toward the reader). In Fig. b, height contours are light blue (dashed and thin solid), and isobars are black. You can superimpose copies of these three figures to see the relationship between the **temperature field**, **mass field** (i.e., pressures), and **wind field**.

First, consider how temperature varies with height and latitude (Fig. 11.37a). At any altitude in the troposphere you will find a horizontal temperature gradient between colder poles and warmer equator. According to the hypsometric relationship, the thickness between two isobaric surfaces is smaller in the colder (polar) air and greater in the warmer (equatorial) air (Fig. 11.37b). Hence, isobaric surfaces tilt in the horizontal, which drives a geostrophic wind (Fig. 11.37c).

The greatest tilt is near 30° latitude at the tropopause, and the associated pressure gradient drives the fastest winds (**jet-stream core**) there, as expected due to the thermal wind. Because the isobars cross the isotherms (and isobars also cross the isopycnics — lines of equal density), the atmosphere is said to be **baroclinic**. It is this **baroclinicity** associated with the meridional temperature gradient that creates the west winds of the jet stream.

Notice that the troposphere is deeper near the equator than near the poles. Thus, the typical lapse rate in the troposphere, applied over the greater depth, causes colder temperatures at the tropopause over the equator than over the poles (Fig. 11.37a). In the stratosphere above the equatorial tropopause, the air is initially isothermal, but at higher altitudes the air gets warmer. Meanwhile, further north in the stratosphere, such as at latitude 60°, the air is isothermal over a very large depth. Thus, the meridional temperature gradient reverses in the bottom of the stratosphere, with warmer air over 60° latitude and colder equatorial air.

The associated north-south thickness changes between isobaric surfaces cause the meridional pressure gradient to decrease, which you can see in Fig. 11.37b as reductions in slopes of the isobars. The reduced pressure gradient in the lower stratosphere causes wind speeds to decrease with increasing altitude (Fig. 11.37c), leaving the **jet max** at the tropopause. Near the jet core is a region where the tropopause has a break or a fold, as is covered in the Fronts and Extratropical Cyclone chapters.

You can apply the concepts described above to the toy model of eq. (11.1) and Fig. 11.8. Using that model with eqs. (11.2, 11.4 & 11.13) gives

Erroneous nesting of equation structures

where ϕ is latitude, $|g| = 9.8$ m s^{-2}, U_g has been relabeled as U_{jet}, T_v is average absolute virtual temperature, $b_1 \approx 40$ K, $c = 1.18\text{x}10^{-3}$ km^{-1}, $z_T \approx 11$ km is average depth of the troposphere, and assuming $U_{jet} = 0$ at $z = 0$. The factor $2{\cdot}\Omega = 1.458\text{x}10^{-4}$ s^{-1} comes from the Coriolis-parameter definition $f_c = 2{\cdot}\Omega{\cdot}\sin\phi$.

The thermal-wind relationship tells us that the meridional variation of temperature between the cold poles and warm equator will drive westerly jet-stream winds in both hemispheres. Namely, the opposite Coriolis-parameter signs in the two hemispheres are canceled by the opposite meridional temperature-gradient signs, yielding positive values for U_g in the two hemispheres.

Although Fig. 11.8 indicates the largest meridional temperature gradients are at roughly 50° latitude (north & south) in our toy model, the meridional variation of the Coriolis parameter conspires to make the fastest jet-stream winds at 45° latitude (north & south) according to eq. (11.17). Solving eq. (11.27) for a tropopause height of z_T = 11 km with an average virtual temperature of –20°C in the troposphere gives peak jet-stream wind speeds (Figs. 11.38 & 11.39) of about 17.25 m s^{-1}.

While our simple toy model is insightful because it mimics the main features of the global circulation and allows an analytical solution that you can solve on a spreadsheet, it is too simple. Compared to Fig. 11.37, the toy jet-stream speeds are too slow, are located too far poleward, and don't give fast-enough winds near the ground. Also, actual locations in both hemispheres shift a bit to the (north, south) in northern hemisphere (summer, winter).

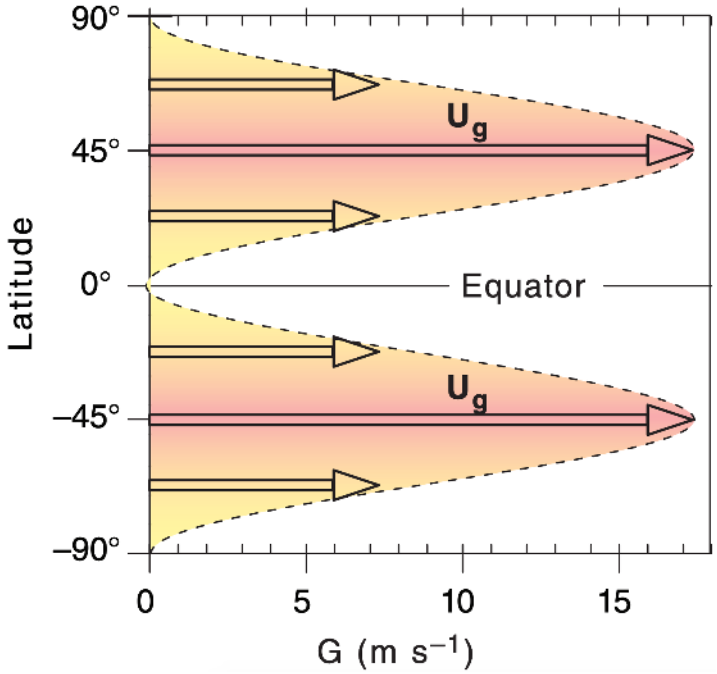


Figure 11.38 Idealized profile of 11 km altitude jet-stream winds vs. latitude.

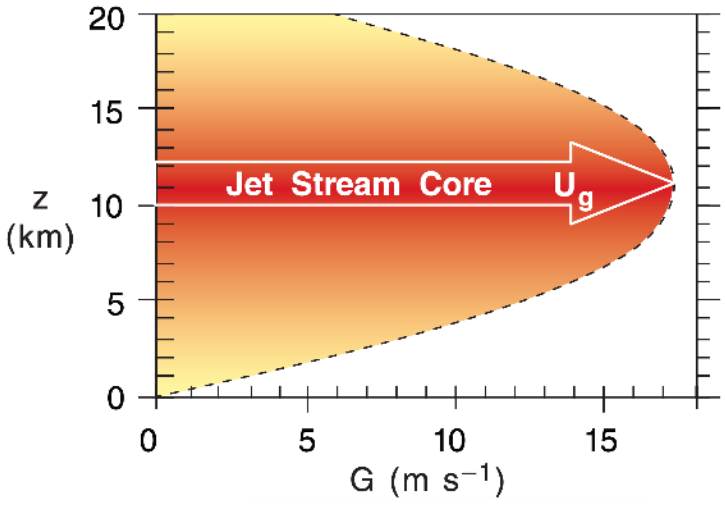


Figure 11.39 Idealized profile of 45°N latitude jet-stream winds vs. altitude.

Sample Application

Use the toy-model temperature eqs. (11.1 & 11.4) to estimate the speed of the geostrophic wind for latitude 60° at the top of the troposphere. Assume T_v = –5°C.

Find the Answer

Given: ϕ = 60°, Tv = 268 K, z = z_T.

Find: U_{jet} = ? m s^{-1}.

From Fig. 11.39 estimate troposphere top zT ≈ 9 km. Apply eq. (11.17), which already incorporates toy model eq. (11.4).

$$U_{jet} \approx \frac{(9.8\mathrm{m}\cdot\mathrm{s}^{-2})\cdot(1.18\times10^{-3}\mathrm{km}^{-1})\cdot(40\mathrm{K})}{(1.458\times10^{-4}\mathrm{s}^{-1})\cdot(268\mathrm{K})}\cdot(9\mathrm{km})\cdot(1-(1/2))\cdot\cos^2(60^\circ)\cdot\sin^2(60^\circ)$$

= **10. m s^{-1}**

Check: Physics & units are reasonable.

Exposition: Agrees with Fig. 11.37c. During winter the poles are colder relative to the equator, driving faster jet-stream winds than summer. Peak winds in the jet core can reach 100 m s^{-1} , although 3-month-average speeds are typically 40 m s^{-1}.

11.8.2. Angular Momentum & Subtropical Jet

Angular momentum can influence the subtropical jet, and is defined as mass times velocity times radius of curvature. Suppose that initially there is air moving at some zonal velocity U_s relative to the Earth's surface at some initial (source) latitude ϕ_s. Because the Earth is rotating, the Earth's surface at the source latitude is moving toward the east at velocity U_{Es}. Thus, the total eastward speed of the air parcel relative to the Earth's axis is (U_s + U_{Es}).

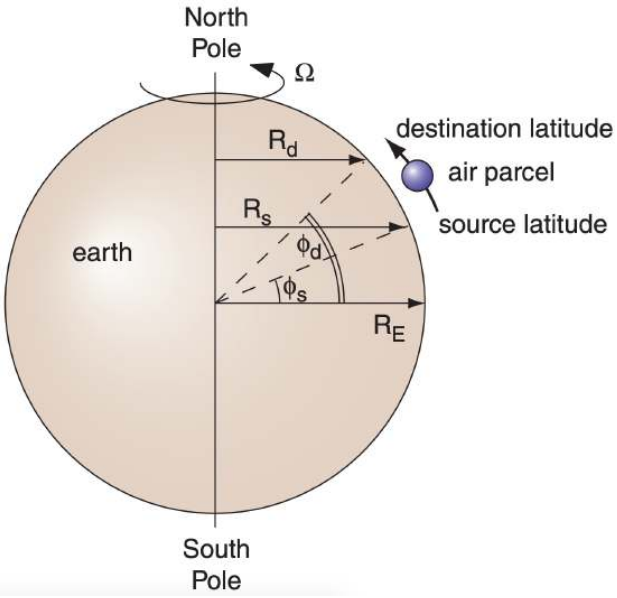


Figure 11.40 Geometry for angular-momentum calculations affecting an air parcel that moves toward the north.

As sketched in Fig. 11.40, suppose some disturbance such as a meandering jet stream moves the air to some other (destination) latitude ϕ_d, assuming that no other forces are applied. Conservation of angular momentum requires:

Erroneous nesting of equation structures

where U_d represents the new zonal air velocity relative to the Earth's surface at the destination latitude, U_{Ed} is the tangential velocity of the Earth's surface at the destination, and m is air mass.

For latitude ϕ at either the source or destination, the radius is R_s or $R_d = R_E\cdot\cos(\phi)$, where average Earth radius is R_E = 6371 km. Similarly, tangential velocities at either the source or destination are Us or $U_d = \Omega\cdot R_\phi = \Omega\cdot R_E\cdot\cos(\phi)$, for an Earth angular velocity of $\Omega = 0.729\text{x}10^{-4}$ s^{-1}.

Solving these equations for the destination air velocity Ud relative to Earth's surface gives:

Erroneous nesting of equation structures

As we already discussed, winds at the top of the Hadley cell diverge away from the equator, but cannot move beyond 30° latitude because Coriolis force turns the wind. When we use eq. (11.19) to predict the zonal wind speed for typical trade-wind air that starts at the equator with U_s = –7 m s^{-1} and ends at 30° latitude, we find unrealistically large wind speeds (125 m s^{-1}) for the subtropical jet (Fig. 11.41). Actual typical wind speeds in the subtropical jet are of order 40 to 80 m s^{-1} in the winter hemisphere, and slower in the summer hemisphere.

The discrepancy is because in the real atmosphere there is no conservation of angular momentum due to forces acting on the air. Coriolis force turns the wind, causing air to accumulate and create a pressure-gradient force to oppose poleward motion in the Hadley cell. Drag due to turbulence slows the wind a small amount. Also, the jet streams meander north and south, which helps to transport slow angular momentum southward and fast angular momentum northward. Namely, these meanders or synoptic-scale eddies cause mixing of zonal momentum.

Next, the concept of vorticity is discussed, which will be useful for explaining how troughs and ridges develop in the jet stream. Then, we will introduce a way to quantify circulation, to help understand the global wind patterns.

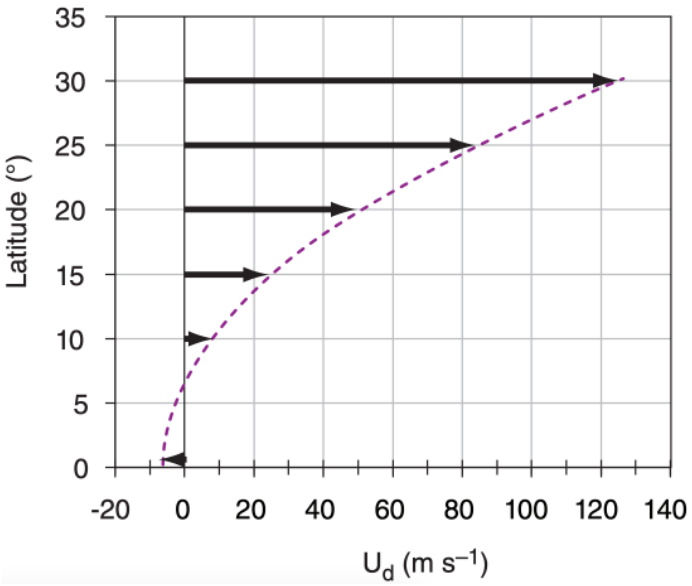


Figure 11.41 Computed zonal wind speed U_d at various destination latitudes for typical trade-wind air that starts at the equator with U_s = –7 m s^{-1}, for the unrealistic case of conservation of angular momentum. Actual zonal winds are not this fast at 30° latitude.

Sample Application

For air starting at the equatorial tropopause, what would be its zonal velocity at 20°N if angular momentum were conserved?

Find the Answer

Given: ϕ_d = 20°N, ϕ_s = 0°, $\Omega \cdot R_E$ = 463 m s^{-1}

Find: U_d = ? m s^{-1}

Assume: no turbulence; U_s = –7 m s^{-1} easterlies.

Use eq. (11.19): U_d =

$$[(463\text{m/s})\cos(0^\circ) + U_s]\frac{\cos(0^\circ)}{\cos(20^\circ)} - (463\text{m/s})\cos(20^\circ)$$

U_d = = **50.2 m s^{-1}**

Check: Physics & units are reasonable. Agrees with Fig. 11.41

Exposition: Actual winds are usually slower, because of turbulent drag and other forces..

This page titled 3.4.11: Jet Streams is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.8: Jet Streams** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.12: Types of Vorticity

11.9.1. Relative-vorticity Definition

Counterclockwise rotation about a local vertical axis defines positive vorticity. One type of vorticity ζ_r , called **relative vorticity**, is measured relative to both the location of the object and to the surface of the Earth (also see the Forces & Winds chapter).

Figure 11.42 Flower blossoms dropped into a river illustrate positive relative vorticity (counterclockwise rotation) caused by river-current shear as the blossoms translate downstream.

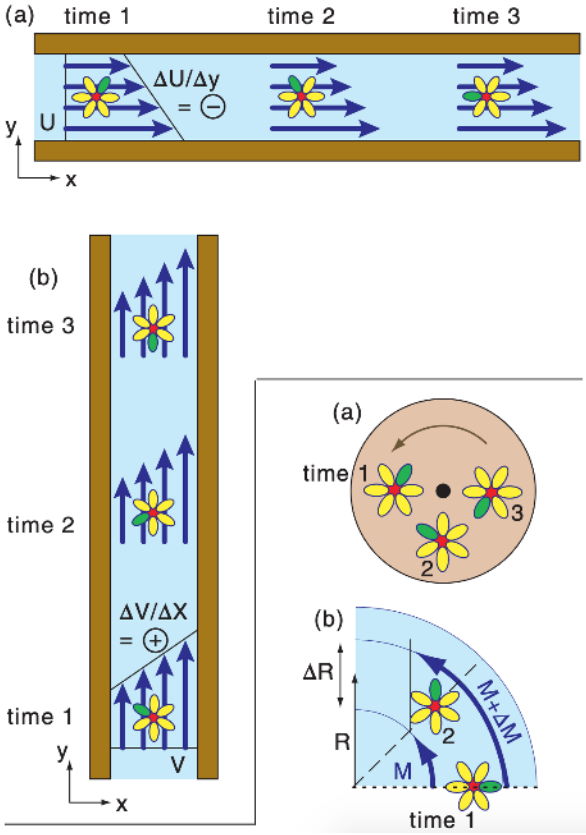


Figure 11.43 (a) Vorticity associated with solid-body rotation, illustrated with a flower blossom glued to a solid turntable at time 1. (b) Vorticity caused by radial shear of tangential velocity M.

Picture a flower blossom dropped onto a straight river, where the river current has shear (Fig. 11.42). As the floating blossom drifts (translates) downstream, it also spins due to the velocity shear of the current. Both Figs. 11.42a & b show counterclockwise rotation, giving positive relative vorticity as defined by:

Erroneous nesting of equation structures

for (U, V) positive in the local (x, y) directions.

Next, consider a flower blossom glued to a solid turntable, as sketched in Fig. 11.43a. As the table turns, so does the orientation of the flower petals relative to the center of the flower. This is shown in Fig. 11.43b. Solid-body rotation requires vectors that start on the dotted line and end on the dashed line. But in a river (or atmosphere), currents can have additional radial shear of the tangential velocity M. Thus, relative vorticity can also be defined as:

Erroneous nesting of equation structures

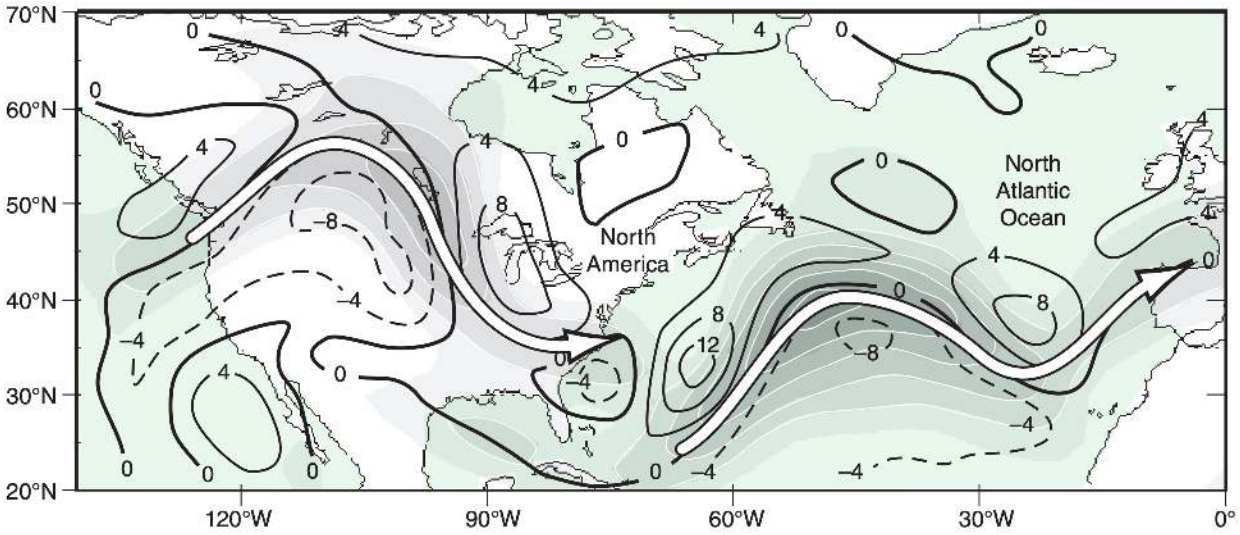


Figure 11.44 Weather map example of relative vorticity (contour lines, with units of 10^{-5} s^{-1}) near the tropopause (at the 20 kPa isobaric surface) at 12 UTC on 5 January 2001. White arrows show jet stream axis of fastest winds over North America and the Atlantic Ocean. Shading gives wind speeds every 10 m s^{-1} from 30 m s^{-1} (lightest grey) to over 80 m s^{-1} (darkest grey). (Based on NCEP/NCAR 40-year reanalysis data, utilizing the plotting tool by Christopher Godfrey, the Univ. of Oklahoma School of Meteorology.)

INFO • Solid Body Relative Vorticity

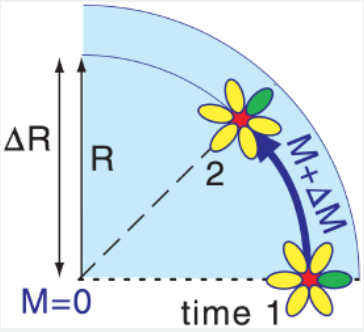
To get eq. (11.22) you can begin with eq. (11.21):

$$\zeta_r = \frac{\Delta M}{\Delta R} + \frac{M}{R}$$

For solid-body rotation, ΔM/ΔR exactly equals M/R, as is illustrated in the time 1 2 M+ΔM M=0 R ΔR figure at right. Hence:

$$\zeta_r = \frac{(M-0)}{R} + \frac{M}{R} = \frac{2M}{R}$$

which is eq. (11.22).



For pure solid-body rotation (where the tangential current- or wind-vectors do indeed start at the same dotted line and end at the same dashed line in Fig. 11.43b), eq. (11.21) can be rewritten as

Erroneous nesting of equation structures

Relative vorticity has units of s–1. A quick way to determine the sign of the vorticity is to curl the fingers of your right hand in the direction of rotation. If your thumb points up, then vorticity is positive. This is the **right-hand rule**. Fig. 11.44 shows relative vorticity of various signs, where vorticity was created by both shear and curvature.

Sample Application

At 50°N is a west wind of 100 m s^{-1}. At 46°N is a west wind of 50 m s^{-1}. Find the (a) relative, & (b) absolute vorticity

Find the Answer

Given: U_2 =100 m s^{-1}, U1 =50 m s^{-1}, $\Delta\phi$ =4°lat. V =0

Find: ζ_r = ? s^{-1} , ζ_a = ? s^{-1}

(a) Use eq. (11.20): Appendix A: 1°latitude = 111 km ζ_r = – (100 – 50 m s^{-1}) / (4.4×10^5 m) = **-1.14×10^{-4} s^{-1}**

(b) Average ϕ = 48°N. Thus, $f_c = (1.458 \times 10^{-4}\ s^{-1}) \cdot \sin(48°)$

$f_c = 1.08 \times 10^{-4}\ s^{-1}$. then, use eq. (11.23):

$\zeta_a = (-1.14 \times 10^{-4}) + (1.08 \times 10^{-4})$ = **-6×10^{-6} s^{-1}** .

Check: Physics and units are reasonable.

Exposition: Shear vorticity from a strong jet stream, but its vorticity is opposite to the Earth's rotation.

Sample Application

Given a N. Hemisphere high-pressure center with tangential winds of 5 m s^{-1} at radius 500 km. What is the value of relative vorticity?

Find the Answer

Given: |M|= 5 m s^{-1} (anti-cyclonic), R = 500,000 m.

Find: ζ_r = ? s^{-1}

In the N. Hem., anti-cyclonic winds turn clockwise, so using the right-hand rule means your thumb points down, so vorticity will be negative. Apply eq. (11.22):

ζ_r = – 2 · (5 m s^{-1}) / (5×10^5 m) = **-2×10^{-5} s^{-1}**

Check: Physics and units are reasonable.

Exposition: Anticyclones often have smaller magnitudes of relative vorticity than cyclones (Fig. 11.44).

11.9.2. Absolute-vorticity Definition

Define an absolute vorticity ζ_a as the sum of the vorticity relative to the Earth plus the rotation of the Earth relative to the so-called fixed stars:

Erroneous nesting of equation structures

where $f_c = 2\Omega \cdot \sin(\phi)$ is the Coriolis parameter, $2\Omega = 1.458 \times 10^{-4}\ s^{-1}$ is the angular velocity of the Earth relative to the fixed stars, and latitude is ϕ. Absolute vorticity has units of s^{-1}.

11.9.3. Potential-vorticity Definition

Define a potential vorticity ζ_p as absolute vorticity per unit depth Δz of the rotating air column:

Erroneous nesting of equation structures

This is the most important measure of vorticity because it is constant (i.e., is conserved) for flows having no latent or radiative heating, and no turbulent drag. Potential vorticity units are $m^{-1}\ s^{-1}$.

Sample Application

A hurricane at latitude 20°N has tangential winds of 50 m s^{-1} at radius 50 km from the center averaged over a 15 km depth. Find potential vorticity, assuming solid-body rotation for simplicity.

Find the Answer

Assume the shear is in the cyclonic direction.

Given: Δz = 15,000 m, ϕ = 20°N, M = 50 m s^{-1}, ΔR = 50,000 m.

Find: ζ_p = ? $m^{-1} \cdot s^{-1}$

Apply eqs. (11.22) with (11.24):

$$\zeta_p = \frac{\frac{2 \cdot (50 \mathrm{m/s})}{5 \times 10^4 \mathrm{m}} + \left(1.458 \times 10^{-4} \mathrm{s}^{-1}\right) \cdot \sin(20^\circ)}{15000 \mathrm{m}}$$

= $(2x10^{-3}+5x10^{-5})/15000$ m = **$\underline{1.37x10^{-7}}$ $m^{-1} \cdot s^{-1}$**

Check: Physics & units are reasonable.

Exposition: Positive sign due to cyclonic rotation.

Rewrite the potential vorticity using eqs. (11.24 & 11.21):

Erroneous nesting of equation structures

shear curvature planetary stretching

The initial values of absolute vorticity and rotatingair depth determine the value for the constant ζ_p. In order to preserve the equality in the equation above while preserving the constant value of ζ_p, any increase of depth Δz of the rotating layer of air must be associated with greater relative vorticity (air spins faster) or larger f_c (rotating air moves poleward).

11.9.4. Isentropic Potential Vorticity Definition

By definition, an isentropic surface connects locations of equal entropy. But non-changing entropy corresponds to non-changing potential temperature θ in the atmosphere (i.e., adiabatic conditions). If you calculate the absolute vorticity ($\zeta_r + f_c$) on such a surface, then it can be used to define an isentropic potential vorticity (IPV):

Erroneous nesting of equation structures

or

Erroneous nesting of equation structures

where air pressure is p, air density is ρ , gravitational acceleration magnitude is |g|, and $\Delta\theta/\Delta z$ indicates the static stability.

Sample Application

If $\zeta_a = 1.5x10^{-4}$ s^{-1}, $\rho = 0.7$ kg m^{-3} and $\Delta\theta/\Delta z = 4$ K km^{-1}, find the isentropic potential vorticity in PVU.

Find the Answer

Given: $\zeta_a = 1.5x10^{-4}$ s^{-1}, $\rho = 0.7$ kg m^{-3} , $\Delta\theta/\Delta z = 4$ K km^{-1}

Find: ζ_{IPV} = ? PVU

Apply eq. (11.26):

$$\zeta_{IPV} = \frac{\left(1.5 \times 10^{-4} \mathrm{s}^{-1}\right)}{(0.7 \mathrm{kg \cdot m}^{-3})} \cdot \left(4 \mathrm{K \cdot km}^{-1}\right) = 8.57 \times 10^{-7} \mathrm{K m}^2 \cdot \mathrm{s}^{-1} \cdot \mathrm{kg}^{-1} = \underline{0.857} \mathrm{PVU}$$

Check: Physics & units are reasonable.

Exposition: As expected for tropospheric air, the answer has magnitude less than 1.5 PVU.

Define a **potential vorticity unit** (**PVU**) such that 1 PVU = 10^{-6} $K \cdot m^2 \cdot s^{-1} \cdot kg^{-1}$. The stratosphere has lower density and greater static stability than the troposphere, hence stratospheric air has IPV values that are typically 100 times larger than for tropospheric air. Typically, ζ_{IPV} > 1.5 PVU for stratospheric air (a good atmospheric cross-section example is shown in the Extratropical Cyclone chapter).

For idealized situations where the air moves adiabatically without friction while following a constant θ surface (i.e., isentropic motion), then ζ_{IPV} is conserved, which means you can use it to track air motion. Also, stratospheric air does not instantly lose its large IPV upon being mixed downward into the troposphere.

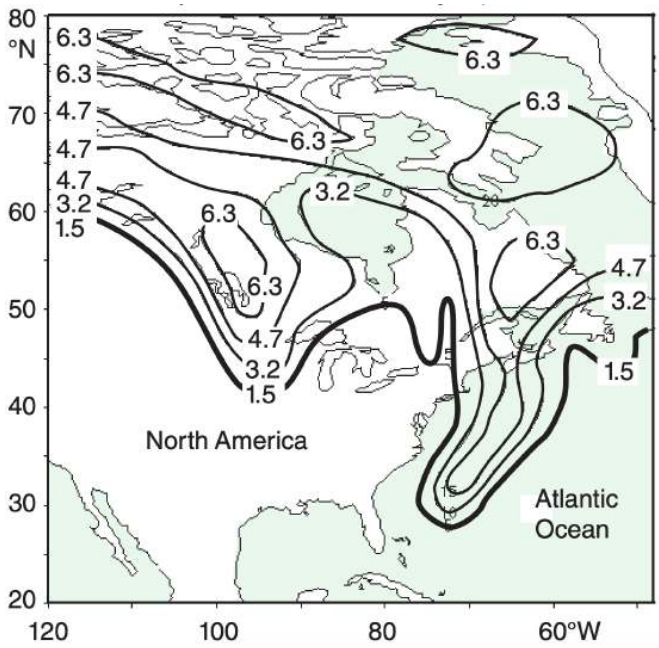


Figure 11.45 Example of isentropic potential vorticity on the 315 K isentropic surface, at 00 UTC on 5 January 2001. Units are PVU. Values greater than 1.5 are in stratospheric air. Because the tropopause is at lower altitude near the poles than at the equator, the 315 K potential temperature surface crosses the tropopause; so it is within the troposphere in the southern part of the figure and in the stratosphere in the northern part. Tropopause folds are evident by the high PVU values just west of the Great Lakes, and just east of the North American coastline. (Based on NCEP/ NCAR 40-year reanalysis data, with initial plots produced using the plotting tool by Christopher Godfrey, the University of Oklahoma School of Meteorology.)

Thus, IPV is useful for finding tropopause folds and the accompanying intrusions of stratospheric air into the troposphere (Fig. 11.45), which can bring down toward the ground the higher ozone concentrations and **radionuclides** (radioactive atoms from former atomic-bomb tests) from the stratosphere.

Because isentropic potential vorticity is conserved, if static stability ($\Delta\theta/\Delta z$) weakens, then eq. (11.26) says absolute vorticity must increase to maintain constant IPV. For example, Fig. 11.46 shows isentropes for flow from west to east across the Rocky Mountains. Where isentropes are spread far apart, static stability is low. Because air tends to follow isentropes (for frictionless adiabatic flow), a column of air between two isentropes over the crest of the Rockies will remain between the same two isentropes as the air continues eastward.

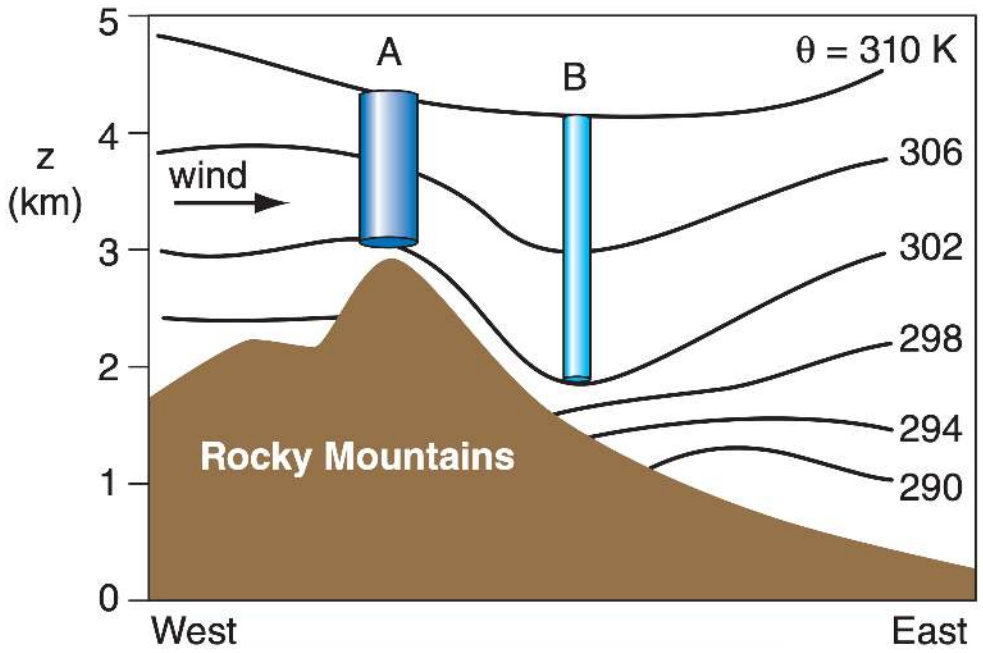


Figure 11.46 Wind blowing from west to east (A) over the Rocky Mountains creates mountain waves and downslope winds, which cause greater separation (B) between the 302 and 310 K isentropes to the lee of the mountains. This greater separation implies reduced static stability and vertical stretching.

Thus, the column of air shown in Fig. 11.46 becomes stretched on the lee side of the Rockies and its static stability decreases (same $\Delta\theta$, but spread over a larger Δz). Thus, absolute vorticity in the stretched region must increase. Such increased cyclonic vorticity encourages formation of low-pressure systems (extratropical cyclones) to the lee of the Rockies — a process called **lee cyclogenesis**.

Sample Application

Given Fig. 11.46. (a) Estimate $\Delta\theta/\Delta z$ at A and B. (b) if the initial absolute vorticity at A is 10^{-4} s^{-1}, find the absolute vorticity at B.

Find the Answer

Given: $\zeta_a = 10^{-4}$ s^{-1} at A, $\theta_{top} = 310$ K, $\theta_{bottom} = 302$ K.

Find: (a) $\Delta\theta/\Delta z$ = ? K km^{-1} at A and B.

(b) ζ_a = ? s^{-1} at B.

$\Delta\theta = 310$ K – 302 K = 8 K at A & B. Estimate the altitudes at the top and bottom of the cylinders in Fig. 11.46.

A: $z_{top} = 4.4$ km, $z_{bottom} = 3.1$ km. Thus $\Delta z_A = 1.3$ km.

B: $z_{top} = 4.2$ km, $z_{bottom} = 1.9$ km. Thus $\Delta z_B = 2.3$ km.

(a) $\Delta\theta/\Delta z$ = 8K / 1.3 km = **6.15 K km^{-1}** at A

$\Delta\theta/\Delta z$ = 8K / 2.3 km = **3.48 K km^{-1}** at B.

(b) If initial (A) and final (B) IPV are equal, then rearranging eq. (11.26) and substituting $(\zeta_r + f_c) = \zeta_a$ gives:

$\zeta_{aB} = \zeta_{aA} \cdot (\Delta z_B/\Delta z_A) = (10^{-4}\ s) \cdot [(2.3\ km)/(1.3\ km)] = 1.77 \times 10^{-4}\ s^{-1}$

Check: Sign, magnitude & units are reasonable.

Exposition: Static stability $\Delta\theta/\Delta z$ is much weaker at B than A. Thus absolute vorticity at B is much greater than at A. If the air flow directly from west to east and if the initial relative vorticity were zero, then the final relative vorticity is $\zeta_r = 0.77 \times 10^{-4}$ s^{-1}. Namely, to the lee of the mountains, cyclonic rotation forms in the air where none existed upwind. Namely, this implies cyclogenesis (birth of cyclones) to the lee (downwind) of the Rocky Mountains.

This page titled 3.4.12: Types of Vorticity is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.9: Types of Vorticity** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.13: Key Terms

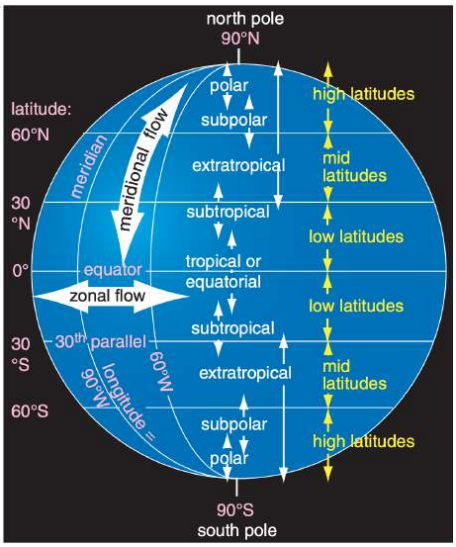


Figure 11.2 Global key terms.

Lines of constant latitude are called **parallels**, and winds parallel to the parallels are identified as **zonal flows** (Fig. 11.2). Lines of constant longitude are called **meridians**, and winds parallel to the meridians are known as **meridional flows**.

Between latitudes of 30° and 60° are the **midlatitudes**. **High latitudes** are 60° to 90°, and **low latitudes** are 0° to 30°. Each 1° of latitude = 111 km.

Tropics, **subtropics**, **subpolar**, and **polar** regions are as shown in Fig. 11.2. Regions not in the tropics are called **extratropical**; namely, poleward of about 30°N and about 30°S.

For example, **tropical cyclones** such as **hurricanes** are in the tropics. Low-pressure centers (**lows**, as indicated by L on weather maps) outside of the tropics are called **extratropical cyclones**.

In many climate studies, data from the months of June, July, and August (**JJA**) are used to represent conditions in N. Hemisphere summer (and S. Hemisphere winter). Similarly, December, January, February (**DJF**) data are used to represent N. Hemisphere winter (and S. Hemisphere summer).

This page titled 3.4.13: Key Terms is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.1: Key Terms** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.14: Review

The combination of solar radiation input and IR output from Earth causes polar cooling and tropical heating. In response, a global circulation develops due to buoyancy, pressure, and geostrophic effects, which transports heat from the tropics toward the polar regions, and which counteracts the radiative differential heating.

Near the equator, warm air rises and creates a band of thunderstorms at the ITCZ. The updrafts are a part of the Hadley-cell direct vertical circulation, which moves heat away from the equator. This cell cannot extend beyond about 30° to 35° latitude because Coriolis force turns the upper-troposphere winds toward the east, creating a subtropical jet near 30° latitude at the tropopause. Near the surface are the trade-wind return flows from the east.

A strong meridional temperature gradient remains at mid-latitudes, which drives westerly winds via the thermal-wind effect, and creates a polar jet at the tropopause. But instabilities of the jet stream cause meridional meanders called Rossby waves.

These waves are very effective at moving heat poleward, leaving only a very weak indirect-circulating Ferrel cell in midlatitudes. Near the poles is a weak direct-circulation cell. Ocean currents can be driven by the overlying winds in the atmosphere. All these ceaseless global circulations in both the atmosphere and the ocean can move enough momentum and heat to keep our planet in near equilibrium.

This page titled 3.4.14: Review is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.14: Review** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

3.4.15: Homework Exercises

11.15.1. Broaden Knowledge & Comprehension

B1. From hemispheric weather maps of winds near the tropopause (which you can access via the internet), identify locations of major global-circulation features including the jet stream, monsoon circulations, tropical cyclones and the ITCZ.

B2. Same as the previous exercise, except using water-vapor or infrared image loops from geostationary satellites to locate the features.

B3. From the web, find a rawinsonde sounding at a location in the trade-wind region, and confirm the wind reversal between low and high altitudes.

B4. Use a visible, whole-disk image from a geostationary satellite to view and quantify the cloud-cover fraction as a function of latitude. Speculate on how insolation at the Earth's surface is affected.

B5. Download a series of rawinsonde soundings for different latitudes between the equator and a pole. Find the tropopause from each sounding, and then plot the variation of tropopause height vs. latitude.

B6. Download a map of sea-surface temperature (SST), and discuss how SST varies with latitude.

B7. Most satellite images in the infrared show greys or colors that are related to brightness temperature (see the legend in whole-disk IR images that you acquired from the internet). Use these temperatures as a function of latitude to estimate the corresponding meridional variation of IR-radiation out. Hint, consider the Stefan-Boltzmann law.

B8. Download satellite-derived images that show the climatological average incoming and outgoing radiation at the top of the atmosphere. How does it relate to the idealized descriptions in this chapter?

B9. Download satellite-derived or buoy & ship-derived ocean currents for the global oceans, and discuss how they transport heat meridionally, and why the oceanic transport of heat is relatively small at mid to high latitudes in the N. Hemisphere.

B10. Use a satellite image to locate a strong portion of the ITCZ over a rawinsonde site, and then download the rawinsonde data. Plot (compute if needed) the variation of pressure with altitude, and discuss how it does or doesn't deviate from hydrostatic.

B11. Search the web for sites where you can plot "reanalysis data", such as the NCEP/NCAR reanalysis or any of the ECMWF reanalyses. Pick a month during late summer from some past year in this database, and plot the surface pressure map. Explain how this "real" result relates to a combination of the "idealized" planetary and monsoonal circulations.

B12. Same as B11, but for monthly average vertical cross sections that can be looped as movies. Display fields such as zonal wind, meridional wind, and vertical velocity, and see how they vary over a year.

B13. Capture a current map showing 85 kPa temperatures, and assume that those temperatures are surrogates for the actual average virtual temperature between 100 and 70 kPa. Compute the thermal wind magnitude and direction for a location assigned by your teacher, and see if this theoretical relationship successfully explains the wind shear between 100 and 70 kPa. Justify your reasoning.

B14 Capture a current map showing the thickness between 100 and 50 kPa, and estimate the thermal wind direction and magnitude across that layer.

B15. Use rawinsonde soundings from stations that cross the jet stream. Create your own contour plots of the jet-stream cross section for (a) heights of key isobaric surfaces; (b) potential temperature; and (c) wind magnitude. Compare your plots with idealized sketches presented in this chapter.

B16. What are the vertical and horizontal dimensions of the jet stream, based on weather maps you acquire from the internet.

B17. Acquire a 50 kPa vorticity chart, and determine if the plotted vorticity is isentropic, absolute, relative, or potential. Where are positive-vorticity maxima relative to fronts and foul weather?

B18. Calculate the values for the four types of vorticity at a location identified by your instructor, based on data for winds and temperatures. Namely, acquire the raw data used for vorticity calculations; do not use vorticity maps captured from the web.

B19. For the 20 kPa geopotential heights, use the wavy pattern of height contours and their relative packing to identify ridges and troughs in the jet stream. Between two troughs, or between two ridges, estimate the wavelength of the Rossby wave. Use that measured length as if it were the dominant wavelength to estimate the phase speed for baroclinic and barotropic waves.

B20. Confirm that the theoretical relationship between horizontal winds, temperatures, vertical velocities, and heights for baroclinic waves is consistent with the corresponding weather maps you acquire from the internet. Explain any discrepancies.

B21. Confirm the three-band nature of the global circulation using IR satellite image movie loops. In the tropics, compare the motion of low (warm) and high (cold) clouds, and relate this motion to the trade winds and Hadley circulation. In mid-latitudes, find the regions of meandering jet stream with its corresponding high and low-pressure centers. In polar regions, relate cloud motions to the polar cell.

B22. Are the ocean-surface current directions consistent with near-surface wind directions as observed in maps or animations acquired from the internet, given the dynamics describe for the Ekman spiral?

11.15.2. Apply

A1(§). For the "toy" model, make a graph of zonally-averaged temperature (°C) vs. latitude for the altitude (km) above ground level (AGL) given here:

a. 0.5	b. 1	c. 1.5	d. 2	e. 2.5	f. 3	g. 3.5
h. 4	i. 4.5	j. 5	k. 5.5	l. 6	m. 6.6	n. 7
o. 8	p. 9	q. 10	r. 11	s. 12	t. 13	u. 14

A2(§). For the "toy" model, make a graph of zonallyaveraged $\Delta T/\Delta y$ (°C km^{-1}) vs. latitude for the altitude (km AGL) given here:

a. 0.5	b. 1	c. 1.5	d. 2	e. 2.5	f. 3	g. 3.5
h. 4	i. 4.5	j. 5	k. 5.5	l. 6	m. 6.6	n. 7
o. 8	p. 9	q. 10	r. 11	s. 12	t. 13	u. 14

A3. Estimate the annual average insolation (W m^{-2}) at the following latitude:

a. 90°	b. 85°	c. 80°	d. 75°	e. 70°	f. 65°	g. 60°
h. 55°	i. 50°	j. 45°	k. 40°	l. 35°	m. 30°	n. 25°
o. 20°	p. 15°	q. 10°	r. 5°	s. equator		

A4. Estimate the annual average amount of incoming solar radiation (W m^{-2}) that is absorbed in the Earth-ocean-atmosphere system at latitude:

a. 90°	b. 85°	c. 80°	d. 75°	e. 70°	f. 65°	g. 60°
h. 55°	i. 50°	j. 45°	k. 40°	l. 35°	m. 30°	n. 25°
o. 20°	p. 15°	q. 10°	r. 5°	s. equator		

A5. Using the idealized temperature near the middle of the troposphere (at z = 5.5 km), estimate the outgoing infrared radiation (W m^{-2}) from the atmosphere at the following latitude:

a. 90°	b. 85°	c. 80°	d. 75°	e. 70°	f. 65°	g. 60°
h. 55°	i. 50°	j. 45°	k. 40°	l. 35°	m. 30°	n. 25°
o. 20°	p. 15°	q. 10°	r. 5°	s. equator		

A6. Using the results from the previous two exercises, find the net radiation magnitude (W m^{-2}) that is input to the atmosphere at latitude:

a. 90°	b. 85°	c. 80°	d. 75°	e. 70°	f. 65°	g. 60°
h. 55°	i. 50°	j. 45°	k. 40°	l. 35°	m. 30°	n. 25°
o. 20°	p. 15°	q. 10°	r. 5°	s. equator		

A7. Using the results from the previous exercise, find the latitude-compensated net radiation magnitude (W m^{-2}; i.e., the differential heating) at latitude:

a. 90°	b. 85°	c. 80°	d. 75°	e. 70°	f. 65°	g. 60°
h. 55°	i. 50°	j. 45°	k. 40°	l. 35°	m. 30°	n. 25°
o. 20°	p. 15°	q. 10°	r. 5°	s. equator		

A8. Assuming a standard atmosphere, find the internal Rossby deformation radius (km) at latitude:

a. 90°	b. 85°	c. 80°	d. 75°	e. 70°	f. 65°	g. 60°
h. 55°	i. 50°	j. 45°	k. 40°	l. 35°	m. 30°	n. 25°
o. 20°	p. 15°	q. 10°	r. 5°	s. equator		

A9. Given the following virtual temperatures at your location (20°C) and at another location, find the change of geostrophic wind with height [(m s^{-1})/ km]. Relative to your location, the other locations are:

Δx	(km)	Δy(km)	T_v(°C)	|	Δx	(km)	Δy(km)	T_v(°C)
a.	0	100	15	|	k.	100	0	15
b.	0	100	16	|	l.	100	0	16
c.	0	100	17	|	m.	100	0	17
d.	0	100	18	|	n.	100	0	18
e.	0	100	19	|	o.	100	0	19
f.	0	100	21	|	p.	100	0	21
g.	0	100	22	|	q.	100	0	22
h.	0	100	23	|	r.	100	0	23
i.	0	100	24	|	s.	100	0	24
j.	0	100	25	|	t.	100	0	25

A10. Find the thermal wind (m s^{-1}) components, given a 100 to 50 kPa thickness change of 0.1 km across the following distances:

Δx(km) =	a. 200	b. 250	c. 300	d. 350	e. 400	f. 450	g. 550	h. 600	i. 650
Δy(km) =	j. 200	k. 250	l. 300	m. 350	n. 400	o. 450	p. 550	q. 600	r. 650

A11. Find the magnitude of the thermal wind (m s^{-1}) for the following thickness gradients:

ΔTH(km) / Δx(km)	&	ΔTH(km) / Δy(km)
a. –0.2 / 600	and	– 0.1 / 400
b. –0.2 / 400	and	– 0.1 / 400
c. –0.2 / 600	and	+ 0.1 / 400
d. –0.2 / 400	and	+ 0.1 / 400
e. –0.2 / 600	and	– 0.1 / 400
f. –0.2 / 400	and	– 0.1 / 400
g. –0.2 / 600	and	+ 0.1 / 400
h. –0.2 / 400	and	+ 0.1 / 400

A12. For the toy model temperature distribution, find the wind speed (m s^{-1}) of the jet stream at the following heights (km) for latitude 30°:

a. 0.5	b. 1	c. 1.5	d. 2	e. 2.5	f. 3	g. 3.5
h. 4	i. 4.5	j. 5	k. 5.5	l. 6	m. 6.6	n. 7
o. 8	p. 9	q. 10	r. 11	s. 12	t. 13	u. 14

A13. If an air parcel from the starting latitude 5° has zero initial velocity relative to the Earth, then find its U component of velocity (m s^{-1}) relative to the Earth when it reaches the following latitude, assuming conservation of angular momentum.

a. 0°	b. 2°	c. 4°	d. 6°	e. 8°	f. 10°	g. 12°
h. 14°	i. 16°	j. 18°	k. 20°	l. 22°	m. 24°	n. 26°

A14. Find the relative vorticity (s^{-1}) for the change of (U , V) wind speed (m s^{-1}), across distances of Δx = 300 km and Δy = 600 km respectively given below.

a. 50, 50	b. 50, 20	c. 50, 0	d. 50, –20	e. 50, –50
f. 20, 50	g. 20, 20	h. 20, 0	i. 20, –20	j. 20, –50
k. 0, 50	l. 0, 20	m. 0, 0	n. 0, –20	o. 0, –50
p. –20, 50	q. –20, 20	r. –20, 0	s. –20, –20	t. –20, –50
u. –50, 50	v. –50, 20	x. –50, 0	y. –50, –20	z. –50, –50

A15. Given below a radial shear ($\Delta M/\Delta R$) in [(m s^{-1})/ km] and tangential wind speed M (m s^{-1}) around radius R (km), find relative vorticity (s^{-1}):

a. 0.1, 30, 300	b. 0.1, 20, 300	c. 0.1, 10, 300	d. 0.1, 0, 300
e. 0, 30, 300	f. 0, 20, 300	g. 0, 10, 300	h. –0.1, 30, 300
i. –0.1, 20, 300	j. –0.1, 10, 300	k. –0.1, 0, 300	

A16. If the air rotates as a solid body of radius 500 km, find the relative vorticity (s^{-1}) for tangential speeds (m s^{-1}) of:

a. 10	b. 20	c. 30	d. 40	e. 50	f. 60	g. 70
h. 80	i. 90	j. 100	k. 120	l. 140	m. 150	

A17. If the relative vorticity is $5x10^{-5}$ s^{-1}, find the absolute vorticity at the following latitude:

a. 90°	b. 85°	c. 80°	d. 75°	e. 70°	f. 65°	g. 60°
h. 55°	i. 50°	j. 45°	k. 40°	l. 35°	m. 30°	n. 25°
o. 20°	p. 15°	q. 10°	r. 5°	s. equator		

A18. If absolute vorticity is $5x10^{-5}$ s^{-1}, find the potential vorticity ($m^{-1}\cdot s^{-1}$) for a layer of thickness (km) of:

a. 0.5	b. 1	c. 1.5	d. 2	e. 2.5	f. 3	g. 3.5
h. 4	i. 4.5	j. 5	k. 5.5	l. 6	m. 6.6	n. 7
o. 8	p. 9	q. 10	r. 11	s. 12	t. 13	u. 14

A19. The potential vorticity is $1x10^{-8}$ $m^{-1}\cdot s^{-1}$ for a 10 km thick layer of air at latitude 48°N. What is the change of relative vorticity (s^{-1}) if the thickness (km) of the rotating air changes to:

a. 9.5	b. 9	c. 8.5	d. 8	e. 7.5	f. 7	g. 6.5
h. 10.5	j. 11	k. 11.5	l. 12	m. 12.5	n. 13	

A20. If the absolute vorticity is $3x10^{-5}$ s^{-1} at 12 km altitude, find the isentropic potential vorticity (PVU) for a potential temperature change of ___ °C across a height increase of 1 km.

a. 1	b. 2	c. 3	d. 4	e. 5	f. 6	g. 6.5
h. 7	i. 8	j. 9	k. 10	l. 11	m. 12	n. 13

A21. Find the horizontal circulation associated with average relative vorticity $5x10^{-5}$ s^{-1} over area (km^2):

a. 500	b. 1000	c. 2000	d. 5000	e. 10,000
f. 20,000	g. 50,000	h. 100,000	i. 200,000	

A22. For the latitude given below, what is the value of the beta parameter (m^{-1} s^{-1}):

a. 90°	b. 85°	c. 80°	d. 75°	e. 70°	f. 65°	g. 60°
h. 55°	i. 50°	j. 45°	k. 40°	l. 35°	m. 30°	n. 25°
o. 20°	p. 15°	q. 10°	r. 5°	s. equator		

A23. Suppose the average wind speed is 60 m s^{-1} from the west at the tropopause. For a barotropic Rossby wave at 50° latitude, find both the intrinsic phase speed (m s^{-1}) and the phase speed (m s^{-1}) relative to the ground for wavelength (km) of:

a. 1000	b. 1500	c. 2000	d. 2500	e. 3000
f. 3500	g. 4000	h. 4500	i. 5000	j. 5500
k. 6000	l. 6500	m. 7000	n. 7500	o. 8000

A24. Plot the barotropic wave (y' vs x') from the previous exercise, assuming amplitude 2000 km.

A25. Same as exercise A23, but for a baroclinic Rossby wave in an atmosphere where air temperature decreases with height at 4°C km^{-1}.

A26(§). Plot the baroclinic wave (y' vs x') from the previous exercise, assuming amplitude 2000 km and a height (km):

(i) 2	(ii) 4	(iii) 6	(iv) 8	(v) 10

A27. What is the fastest growing wavelength (km) for a baroclinic wave in a standard atmosphere at latitude:

a. 90°	b. 85°	c. 80°	d. 75°	e. 70°	f. 65°	g. 60°
h. 55°	i. 50°	j. 45°	k. 40°	l. 35°	m. 30°	n. 25°
o. 20°	p. 15°	q. 10°	r. 5°	s. equator		

A28. For the baroclinic Rossby wave of exercise A25 with amplitude 2000 km, find the wave amplitudes of the:

i. vertical-displacement perturbation
ii. potential-temperature perturbation
iii. pressure perturbation
iv. U-wind perturbation
v. V-wind perturbation
vi. W-wind perturbation

A29(§). For a vertical slice through the atmosphere, plot baroclinic Rossby-wave perturbation amount for conditions assigned in exercise A28.

A30. Find the latitude-weighted a·u' momentum value (m s^{-1}) for air that reaches destination latitude 50° from source latitude:

a. 80°	b. 75°	c. 70°	d. 65°	e. 60°	f. 55°	g. 45°	h. 40°	i. 35°

A31. Suppose the ____ cell upward and downward speeds are ___ and ___ mm s^{-1}, respectively, and the north-south wind speeds are 3 m s^{-1} at the top and bottom of the cell. The cell is about __ km high by ___ km wide, and is centered at about ___ latitude. Temperature in the atmosphere decreases from about 15°C near the surface to –57°C at 11 km altitude. Find the vertical circulation.

cell	W_{up} (mm s^{-1})	W_{down} (mm s^{-1})	Δz (km)	Δy (km)	ϕ (°)
a. Hadley	6	–4	17	3900	10
b. Hadley	4	–4	15	3500	10
c. Hadley	3	–3	15	3500	5
d. Ferrel	3	–3	12	3000	45
e. Ferrel	2	–2	11	3000	45
f. Ferrel	2	–2	10	3000	50
g. polar	1	–1	9	2500	75
h. polar	1	–1	8	2500	75
i. polar	0.5	–0.5	7	2500	80

A32. Find the friction velocity at the water surface if the friction velocity (m s^{-1}) in the air (at sea level for a standard atmosphere) is:

a. 0.05	b. 0.1	c. 0.15	d. 0.2	e. 0.25
f. 0.3	g. 0.35	h. 0.4	i. 0.45	j. 0.5
k. 0.55	l. 0.6	m. 0.65	n. 0.7	o. 0.75

A33. Find the Ekman-spiral depth scale at latitude 50°N for eddy viscosity (m^2 s^{-1}) of:

a. 0.0002	b. 0.0004	c. 0.0006	d. 0.0008	e. 0.001
f. 0.0012	g. 0.0014	h. 0.0016	i. 0.0018	j. 0.002
k. 0.0025	l. 0.003	m. 0.0035	n. 0.004	o. 0.005

A34(§). Create a graph of Ekman-spiral wind components (U, V) components for depths from the surface down to where the velocities are near zero, for near-surface wind speed of 8 m s^{-1} at 40°N latitude.

11.15.3. Evaluate & Analyze

E1. During months when the major Hadley cell exists, trade winds cross the equator. If there are no forces at the equator, explain why this is possible.

E2. In regions of surface high pressure, descending air in the troposphere is associated with dry (nonrainy) weather. These high-pressure belts are where deserts form. In addition to the belts at ±30° latitude, semi-permanent surface highs also exist at the poles. Are polar regions deserts? Explain.

E3. The subtropical jet stream for Earth is located at about 30° latitude. Due to Coriolis force, this is the poleward limit of outflow air from the top of the ITCZ. If the Earth were to spin faster, numerical experiments suggest that the poleward limit (and thus the jet location) would be closer to the equator. Based on the spins of the other planets (get this info from the web or a textbook) compared to Earth, at what latitudes would you expect the subtropical jets to be on Jupiter? Do your predictions agree with photos of Jupiter?

E4. Horizontal divergence of air near the surface tends to reduce or eliminate horizontal temperature gradients. Horizontal convergence does the opposite. Fronts (as you will learn in the next chapter) are regions of strong local temperature gradients. Based on the general circulation of Earth, at what latitudes would you expect fronts to frequently exist, and at what other latitudes would you expect them to rarely exist? Explain.

E5. In the global circulation, what main features can cause mixing of air between the Northern and Southern Hemispheres? Based on typical velocities and cross sectional areas of these flows, over what length of time would be needed for the portion 1/e of all the air in the N. Hemisphere to be replaced by air that arrived from the S. Hemisphere?

E6. In Fig. 11.4, the average declination of the sun was listed as 14.9° to 15° for the 4-month periods listed in those figures. Confirm that those are the correct averages, based on the equations from the Solar & Infrared Radiation chapter for solar declination angle vs. day of the year.

E7. Thunderstorms are small-diameter (15 km) columns of cloudy air from near the ground to the tropopause. They are steered by the environmental winds at an altitude of roughly 1/4 to 1/3 the troposphere depth. With that information, in what direction would you expect thunderstorms to move as a function of latitude (do this for every 10° latitude)?

E8. The average meridional wind at each pole is zero. Why? Also, does your answer apply to instantaneous winds such as on a weather map? Why?

E9. Can you detect monsoonal (monthly or seasonal average) pressure centers on a normal (instantaneous) weather map analysis or forecast? Explain.

E10. Figs. 11.3a & 11.5a showed idealized surface wind & pressure patterns. Combine these and draw a sketch of the resulting idealized global circulation including both planetary and monsoon effects.

E11. Eqs. (11.1-11.3) represent an idealized ("toy model") meridional variation of zonally averaged temperature. Critically analyze this model and discuss. Is it reasonable at the ends (boundaries) of the curve; are the units correct; is it physically justifiable; does it satisfy any budget constraints (e.g., conservation of heat, if appropriate), etc. What aspects of it are too simplified, and what aspects are OK?

E12. (a) Eq. (11.4) has the 3rd power of the sine times the 2nd power of the cosine. If you could arbitrarily change these powers, what values would lead to reasonable temperature gradients ($\Delta T/\Delta y$) at the surface and which would not (Hint: use a spreadsheet and experiment with different powers)?

(b) Of the various powers that could be reasonable, which powers would you recommend as fitting the available data the best? (Hint: consider not only the temperature gradient, but the associated meridional temperature profile and the associated jet stream.) Also, speculate on why I chose the powers that I did for this toy model.

E13. Concerning differential heating, Fig. 11.9 shows the annual average insolation vs. latitude. Instead, compute the average insolation over the two-month period of June and July, and plot vs. latitude. Use the resulting graph to explain why the jet stream and weather patterns are very weak in the summer hemisphere, and strong in the winter hemisphere.

E14. At mid- and high-latitudes, Fig. 11.9 shows that each hemisphere has one full cycle of insolation annually (i.e., there is one maximum and one minimum each year).

But look at Fig. 11.9 near the equator.

a. Based on the data in this graph (or even better, based on the eqs. from the Solar & Infrared Radiation chapter), plot insolation vs. relative Julian day for the equator.
b. How many insolation cycles are there each year at the equator?
c. At the equator, speculate on when would be the hottest and coldest "seasons".
d. Within what range of latitudes near the equator is this behavior observed?

E15. Just before idealized eq. (11.6), I mentioned my surprise that E_2 was approximately constant with latitude. I had estimated E_2 by subtracting my toy-model values for E_{insol} from the actual observed values of E_{in}. Speculate about what physical processes could cause E_2 to be constant with latitude all the way from the equator to the poles.

E16. How sensitive is the toy model for E_{out} (i.e., eq. 11.7) to the choice of average emission altitude z_m? Recall that z_m, when used as the altitude z in eqs. (11.1-11.3), affects T_m. Hint: for your sensitivity analysis, use a spreadsheet to experiment with different z_m and see how the resulting plots of E_{out} vs. latitude change. (See the "A SCIENTIFIC PERSPECTIVE" box about model sensitivity.)

E17(§). Solve the equations to reproduce the curves in figure:

a. 11.10	b. 11.11	c. 11.12	d. 11.13

E18. We recognize the global circulation as a response of the atmosphere to the instability caused by differential heating, as suggested by **LeChatelier's Principle**. But the circulation does not totally undo the instability; namely, the tropics remain slightly warmer than the poles. Comment on why this remaining, unremoved instability is required to exist, for the global circulation to work.

E19. In Fig. 11.12, what would happen if the surplus area exceeded the deficit area? How would the global circulation change, and what would be the end result for Fig. 11.12?

E20. Check to see if the data in Fig. 11.12 does give zero net radiation when averaged from pole to pole.

E21. The observation data that was used in Fig. 11.14 was based on satellite-measured radiation and differential heating to get the total needed heat transport, and on estimates of heat transport by the oceans. The published "observations" for net atmospheric heat transport were, in fact, estimated as the difference (i.e., residual) between the total and the ocean curves. What could be some errors in this atmosphere curve? (Hint: see the A SCIENTIFIC PERSPECTIVE box about Residuals.)

E22. Use the total heat-transport curve from Fig. 11.60. At what latitude is the max transport? For that latitude, convert the total meridional heat-flux value to horsepower.

E23. For Fig. 11.15, explain why it is p' vs. z that drive vertical winds, and not P_{column} vs. z.

E24.

a. Redraw Figs. 11.16 for downdraft situations.
b. Figs. 11.16 both show updraft situations, but they have opposite pressure couplets. As you already found from part (a) both pressure couplets can be associated with downdrafts. What external information (in addition to the pressure-couplet sign) do you always need to decide whether a pressure couplet causes an updraft or a downdraft? Why?

E25.

a. For the thermal circulation of Fig. 11.17(iv), what needs to happen for this circulation to be maintained? Namely, what prevents it from dying out?
b. For what real-atmosphere situations can thermal circulations be maintained for several days?

E26.

a. Study Fig. 11.18 closely, and explain why the wind vectors to/from the low- and high-pressure centers at the equator differ from the winds near pressure centers at mid-latitudes.
b. Redraw Fig. 11.5a, but with continents and oceans at the equator. Discuss what monsoonal pressures and winds might occur during winter and summer, and why.

E27.

a. Redraw Fig. 11.19, but for the case of geostrophic wind decreasing from its initial equilibrium value. Discuss the resulting evolution of wind and pressure fields during this geostrophic adjustment.
b. Redraw Fig. 11.19, but for flow around a lowpressure center (i.e., look at gradient winds instead of geostrophic winds). Discuss how the wind and pressure fields adjust when the geostrophic wind is increased above its initial equilibrium value.

E28. How would the vertical potential temperature gradient need to vary with latitude for the "internal Rossby radius of deformation" to be invariant? Assume constant troposphere depth.

E29. In the Regional Winds chapter, gap winds and coastally-trapped jets are explained. Discuss how these flows relate to geostrophic adjustment.

E30. At the top of **hurricanes** (see the Tropical Cyclones chapter), so much air is being continuously pumped to the top of the troposphere that a highpressure center is formed over the hurricane core there. This high is so intense and localized that it violates the conditions for gradient winds; namely, the pressure gradient around this high is too steep (see the Forces & Winds chapter).

Discuss the winds and pressure at the top of a hurricane, using what you know about geostrophic adjustment. Namely, what happens to the winds and air mass if the wind field is not in geostrophic or gradient balance with the pressure field?

E31. In the thermal-wind relationship (eqs. 11.13), which factors on the right side are constant or vary by only a small amount compared to their magnitude, and which factors vary more (and are thus more important in the equations)?

E32. In Fig. 11.20, how would it change if the bottom isobaric surface were tilted; namely, if there were already a horizontal pressure gradient at the bottom?

E33. Draw a sketch similar to Fig. 11.20 for the thermalwind relationship for the Southern Hemisphere.

E34. In maps such as Fig. 11.21, explain why thickness is related to average temperature.

E35. Redraw Fig. 11.22 for the case cold air in the west and warm air in the east. Assume no change to the bottom isobaric surface.

E36. Copy Fig. 11.24.

a. On your copy, draw the G_1 and G_2 vectors, and the M_{TH} vector at point B. Confirm that the thermal wind relationship is qualitatively satisfied via vector addition. Discuss why point B is an example of veering or backing.
b. Same as (a) but calculate the actual magnitude of each vector at point B based on the spacing between isobars, thickness contours, or height contours. Again, confirm that the thermal wind relationship is satisfied. (1° latitude = 111 km)

E37. Using a spreadsheet, start with an air parcel at rest at the tropopause over the equator. Assume a realistic pressure gradient between the equator and 30° latitude. Use dynamics to solve for acceleration of the parcel over a short time step, and then iterate over many time steps to find parcel speed and position. How does the path of this parcel compare to the idealized paths drawn in Fig. 11.26d? Discuss.

E38. In the thunderstorms at the ITCZ, copious amounts of water vapor condense and release latent heat. Discuss how this condensation affects the average lapse rate in the tropics, the distribution of heat, and the strength of the equatorial high-pressure belt at the tropopause.

E39. Summarize in a list or an outline all the general-circulation factors that make the mid-latitude weather different from tropical weather.

E40. Explain the surface pressure patterns in Figs. 11.31 in terms of a combination of idealized monsoon and planetary circulations.

E41. Figs. 11.31 show mid-summer and mid-winter conditions in each hemisphere. Speculate on what the circulation would look like in April or October.

E42. Compare Figs. 11.32 with the idealized planetary and monsoon circulations, and discuss similarities and differences.

E43. Based on Figs. 11.32, which hemisphere would you expect to have strong subtropical jets in both summer and winter, and which would not. What factors might be responsible for this difference?

E44. For the Indian monsoon sketched in Fig. 11.33, where are the updraft and downdraft portions of the major Hadley cell for that month? Also, what is the relationship between the trade winds at that time, and the Indian monsoon winds?

E45. What are the dominant characteristics you see in Fig. 11.34, regarding jet streams in the Earth's atmosphere? Where don't jet streams go?

E46. In Figs. 11.35, indicate if the jet-stream winds would be coming out of the page or into the page, for the:

a. N. Hemisphere,
b. S. Hemisphere.

E47. Although Figs. 11.36 are for different months than Figs. 11.32, they are close enough in months to still both describe summer and winter flows.

a. Do the near-tropopause winds in Figs. 11.36 agree with the pressure gradients (or height gradients) in Figs. 11.32?
b. Why are there easterly winds at the tropopause over/near the equator, even though there is negligible pressure gradient there?

E48. Describe the mechanism that drives the polar jet, and explain how it differs from the mechanism that drives the subtropical jet.

E49. In Fig. 11.37b, we see a very strong pressure gradient in the vertical (indicated by the different isobars), but only small pressure gradients in the horizontal (indicated by the slope of any one isobar). Yet the strongest average winds are horizontal, not vertical. Why?

E50. Why does the jet stream wind speed decrease with increasing height above the tropopause?

E51. a. Knowing the temperature field given by the toy model earlier in this chapter, show the steps needed to create eq. (11.17) by utilizing eqs. (11.2, 11.4 and 11.13). b. For what situations might this jetwind-speed equation not be valid? c. Explain what each term in eq. (11.17) represents physically.

E52. Why does an air parcel at rest (i.e., calm winds) near the equator possess large angular momentum?

What about for air parcels that move from the east at typical trade wind speeds?

E53. At the equator, air at the bottom of the troposphere has a smaller radius of curvature about the Earth's axis than at the top of the troposphere. How significant is this difference? Can we neglect it?

E54. Suppose that air at 30° latitude has no eastwest velocity relative to the Earth's surface. If that air moves equatorward while preserving its angular momentum, which direction would it move relative to the Earth's surface? Why? Does it agree with real winds in the general circulation? Elaborate.

E55. Picture a circular hot tub of 2 m diameter with a drain in the middle. Water is initially 1.2 m deep, and you made rotate one revolution each 10 s. Next, you pull the plug, allowing the water depth to stretch to 2.4 m as it flows down the drain. Calculate the new angular velocity of the water, neglecting frictional drag. Show your steps.

E56. In eq. (11.20), why is there a negative sign on the last term? Hint: How does the rotation direction implied by the last term without a negative sign compare to the rotation direction of the first term?

E57. In the Thunderstorm chapters, you will learn that the winds in a portion of the tornado can be irrotational. This is surprising, because the winds are traveling so quickly around a very tight vortex. Explain what wind field is needed to gave **irrotational winds** (i.e., no relative vorticity) in air that is rotating around the tornado. Hint: Into the wall of a tornado, imagine dropping a neutrally-buoyant small paddle wheel the size of a flower. As this flower is translated around the perimeter of the tornado funnel, what must the local wind shear be at the flower to cause it to not spin relative to the ground? Redraw Fig. 11.43 to show what you propose.

E58. Eq. (11.25) gives names for the different terms that can contribute toward vorticity. For simplicity, assume Δz is constant (i.e., assume no stretching). On a copy of Fig. 11.44, write these names at appropriate locations to identify the dominant factors affecting the vorticity max and min centers.

E59. If you were standing at the equator, you would be rotating with the Earth about its axis. However, you would have zero vorticity about your vertical axis. Explain how that is possible.

E60. Eq. (11.26) looks like it has the absolute vorticity in the numerator, yet that is an equation for a form of potential vorticity. What other aspects of that equation make it like a potential vorticity?

E61. Compare the expression of horizontal circulation C with that for vertical circulation CC.

E62. Relate Kelvin's circulation theorem to the conservation of potential vorticity. Hint: Consider a constant Volume = $A \cdot \Delta z$.

E63. The jet stream sketched in Fig. 11.49 separates cold polar air near the pole from warmer air near the equator. What prevents the cold air from extending further away from the poles toward the equator?

E64. If the Coriolis force didn't vary with latitude, could there be Rossby waves? Discuss.

E65. Are baroclinic or barotropic Rossby waves faster relative to Earth's surface at midlatitudes? Why?

E66. Compare how many Rossby waves would exist around the Earth under barotropic vs. baroclinic conditions. Assume an isothermal troposphere at 50°N.

E67. Once a Rossby wave is triggered, what mechanisms do you think could cause it to diminish (i.e., to reduce the waviness, and leave straight zonal flow).

E68. In Fig. 11.50 at point (4) in the jet stream, why doesn't the air just continue turning clockwise around toward points (2) and (3), instead of starting to turn the other way?

E69. Pretend you are a newspaper reporter writing for a general audience. Write a short article describing how baroclinic Rossby waves work, and why they differ from barotropic waves.

E70. What conditions are needed so that Rossby waves have zero phase speed relative to the ground? Can such conditions occur in the real atmosphere?

E71. Will Rossby waves move faster or slower with respect to the Earth's surface if the tropospheric static stability increases? Why?

E72. For a baroclinic wave that is meandering north and south, consider the northern-most point as the wave crest. Plot the variation of this crest longitude vs. altitude (i.e., x vs. z). Hint: consider eq. (11.40).

E73. Use tropopause-level Rossby-wave troughaxes and ridge-axes as landmarks. Relative to those landmarks, where east or west is: (a) vertical velocity the greatest; (b) potential-temperature deviation the greatest; and (c) vertical displacement the greatest?

E74. In Figs. 11.51 and 11.53 in the jet stream, there is just as much air going northward as there is air going southward across any latitude line, as required by mass conservation. If there is no net mass transport, how can there be heat or momentum transport?

E75. For the Southern Hemisphere: (a) would a direct circulation cell have positive or negative CC? (b) for each term of eq. (11.51), what are their signs?

E76. Compare definitions of circulation from this chapter with the previous chapter, and speculate on the relevance of the static stability and Earth's rotation in one or both of those definitions.

E77. Consider a cyclonic air circulation over an ocean in your hemisphere. Knowing the relationship between ocean currents and surface winds, would you anticipate that the near-surface wind-driven ocean currents are diverging away from the center of the cyclone, or converging toward the center? Explain, and use drawings. Note: Due to mass conservations, horizontally diverging ocean surface waters cause upwelling of nutrient-rich water toward the surface, which can support ocean plants and animals, while downwelling does the opposite.

11.15.4. Synthesize

S1. Describe the equilibrium general circulation for a non-rotating Earth.

S2. Circulations are said to **spin-down** as they lose energy. Describe general-circulation spin-down if Earth suddenly stopped spinning on its axis.

S3. Describe the equilibrium general circulation for an Earth that spins three times faster than now.

S4. Describe the **spin-up** (increasing energy) as the general circulation evolves on an initially non-rotating Earth that suddenly started spinning.

S5. Describe the equilibrium general circulation on an Earth with no differential radiative heating.

S6. Describe the equilibrium general circulation on an Earth with cold equator and hot poles.

S7. Suppose that the sun caused radiative cooling of Earth, while IR radiation from space caused warming of Earth. How would the weather and climate be different, if at all?

S8. Describe the equilibrium general circulation for an Earth with polar ice caps that extend to 30° latitude.

S9. About 250 million years ago, all of the continents had moved together to become one big continent called **Pangaea**, before further plate tectonic movement caused the continents to drift apart. Pangaea spanned roughly 120° of longitude (1/3 of Earth's circumference) and extended roughly from pole to pole. Also, at that time, the Earth was spinning faster, with the solar day being

only about 23 of our present-day hours long. Assuming no other changes to insolation, etc, how would the global circulation have differed compared to the current circulation?

S10. If the Earth was dry and no clouds could form, how would the global circulation differ, if at all? Would the tropopause height be different? Why?

S11. Describe the equilibrium general circulation for an Earth with tropopause that is 5 km high.

S12. Describe the equilibrium general circulation for an Earth where potential vorticity isn't conserved.

S13. Describe the equilibrium general circulation for an Earth having a zonal wind speed halfway between the phase speeds of short and long barotropic Rossby waves.

S14. Describe the equilibrium general circulation for an Earth having long barotropic Rossby waves that had slower intrinsic phase speed than short waves.

S15. Describe the nature of baroclinic Rossby waves for an Earth with statically unstable troposphere.

S16. Describe the equilibrium general circulation for an Earth where Rossby waves had no north-south net transport of heat, momentum, or moisture.

S17. Describe the equilibrium general circulation for an Earth where no heat was transported meridionally by ocean currents.

S18. Describe the equilibrium ocean currents for an Earth with no drag between atmosphere and ocean.

S19. Suppose there was an isolated small continent that was hot relative to the surrounding cooler ocean. Sketch a vertical cross section in the atmosphere across that continent, and use thickness concepts to draw the isobaric surfaces. Next, draw a planview map of heights of one of the mid-troposphere isobaric surfaces, and use thermal-wind effects to sketch wind vectors on this same map. Discuss how this approach does or doesn't explain some aspects of monsoon circulations.

S20. If the Rossby wave of Fig. 11.50 was displaced so that it is centered on the equator (i.e., point (1) starts at the equator), would it still oscillate as shown in that figure, or would the trough of the wave (which is now in the S. Hem.) behave differently? Discuss.

S21. If the Earth were shaped like a cylinder with its axis of rotation aligned with the axis of the real Earth, could Rossby waves exist? How would the global circulation be different, if at all?

S22. In the subtropics, low altitude winds are from the east, but high altitude winds are from the west. In mid-latitudes, winds at all altitudes are from the west. Why are the winds in these latitude bands different?

S23. What if the Earth did not rotate? How would the Ekman spiral in the ocean be different, if at all?

This page titled 3.4.15: Homework Exercises is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Roland Stull via source content that was edited to the style and standards of the LibreTexts platform.

- **11.15: Homework Exercises** by Roland Stull is licensed CC BY-NC-SA 4.0. Original source: https://www.eoas.ubc.ca/books/Practical_Meteorology.

SECTION OVERVIEW

3.5: Atmospheric Composition

Learning Objectives

By the end of this chapter, you should be able to:

- explain the role that each atmospheric constituent plays in atmospheric structure and weather
- identify changes in minor and trace gas amounts and the impacts these changes have on the atmosphere
- explain how the atmosphere cleanses itself using methane as an example
- use chemical equations to show how ozone is formed in the stratosphere and the troposphere and how they differ
- diagram the lifecycle of aerosol particles with an emphasis on their role in weather

The atmosphere consists mostly of dry air - mostly molecular nitrogen (78%), molecular oxygen (21%), and Argon (0.9%) - and highly variable amounts of water vapor (from parts per million in air to a few percent). Now we will consider gases and particles in the atmosphere at trace levels. The most abundant of the trace gases in the global atmosphere is carbon dioxide (~400 parts per million, or 400 x 10^{-6}), but there are thousands of trace gases with fractions much less than a few parts per million. Some, particularly the reactive hydroxyl (OH) radical, are important even though their abundance is less than 1 part per trillion (10^{-12}). The atmosphere also contains small particles with sizes from nanometers (10^{-9} m) to microns (10^{-6} m) coming from many sources. These trace gases and particles are as important to atmospheric structure and weather as are nitrogen, oxygen, and water vapor and they also play a huge role in human and ecological health and global climate. In this chapter, we will examine the atmosphere's composition and its changes over time. The atmosphere is continually inundated with surface emissions of gases and particles (and some from space) but it has chemical mechanisms to clean itself. We will see how two atmospheric pollutants - ozone and small particles - are produced. In later lessons, we will see that without these chemical processes and particles, there would be no clouds and, thus, no real weather.

3.5.1: Atmospheric Composition

3.5.2: Changes in Atmospheric Composition

3.5.3: Other Trace Gases

3.5.4: Stratospheric Ozone Formation

3.5.5: The Story of the Atmosphere's PAC-MAN

3.5.6: Where do Cloud Condensation Nuclei (CCN) come from?

3.5.7: Summary and Final Tasks

This page titled 3.5: Atmospheric Composition is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by William Brune (John A. Dutton: e-Education Institute) via source content that was edited to the style and standards of the LibreTexts platform.

3.5.1: Atmospheric Composition

The major gases that comprise today's atmosphere are in Table 3.5.1.1. The mixing ratio of a gas X is defined as the fraction of total moles that are the moles of gas X. For instance, 78 moles of every 100 total moles of air is nitrogen, so nitrogen's mixing ratio is 0.78. Note that in atmospheric composition, the mixing ratio is the *moles* of the gas divided by the *total moles of air*. In contrast, the water vapor mixing ratio is the *mass* of water vapor divided by the *mass of dry air*.

Table 3.5.1.1: Major Constituents in Earth's Present Atmosphere

Constituent	Molecular Mass (g/mol)	Mixing Ratio (mol mol^{-1})	Role in the Atmosphere
nitrogen (N_2)	28.013	0.7808	transparent; provides heat capacity and momentum; exchanged with biomass; decomposed in combustion
oxygen (O_2)	31.998	0.2095	transparent except for in extreme ultraviolet; provides some heat capacity and momentum; exchanged with life; source of important reactive gases like ozone
argon (Ar)	39.948	0.0093	no role
carbon dioxide (CO_2)	44.010	0.000385 (385 ppmv)	transparent in visible; absorbs infrared light (i.e., contributes to global warming); exchanged with life; product of combustion
neon (Ne)	20.183	0.0000182	no role, but makes colorful glowing signs
water vapor (H_2O)	18.015	$2x10^{-6}$ to 0.05	gas transparent in visible; absorbs infrared light (i.e., contributes to global warming); exists as vapor, liquid, and solid; exchanged with life; product of combustion
aerosol particles	varies	0-500 ug m^{-3} (note different units)	essential for cloud formation; interact with visible and infrared light; exchanged with surfaces and life
methane (CH_4)	16.04	0.00000182 (1820 ppbv)	transparent in visible; absorbs in infrared (i.e. contributes to global warming); exchanged with life; source of CO_2 and H_2O
ozone (O_3)	48.00	0.01 – 10 ppm	transparent in visible; absorbs in UV and infrared; reactive and source of more reactive gases
particles	varies	0-100's µg m^{-3} of air	absorbs and scatters light; acts as CCN and IN (see below)

Key features of the gases include their compressibility (i.e., ability to expand or shrink in volume), their transparency in the visible, their momentum, and their heat capacity. Water vapor has the additional important feature of existing in the vapor, liquid, and solid phases in the atmosphere and on Earth's surface. The most important properties of small particles include their ability to dissolve in water in order to be **Cloud Condensation Nuclei (CCN)** or to maintain a lattice structure similar to ice in order to be Ice Nuclei (IN), as well as their ability to absorb and scatter sunlight. These properties depend completely on the particle size and composition. Most atmospheric gases participate in the atmosphere's chemistry, which is initiated by sunlight, as you will soon see.

Units used when quantifying atmospheric composition

Three different units are typically used when specifying the amounts of gases. One is the mass mixing ratio, which is the mass of a chemical species divided by the total mass of air. You have already encountered this with the specific humidity of water vapor. A second is the volume mixing ratio, which is just the number of molecules of a chemical species in a unit volume divided by the total number of all molecules in a unit volume. For gases with relatively large fractions like nitrogen, oxygen, and argon, we use percent to indicate this fraction. For minor gases like carbon dioxide and ozone, we use parts per million (10^{-6}) ppmv or parts per billion (10^{-9}) ppbv by volume (meaning by number not mass). Lastly we need to use the concentration, or number per unit volume, to calculate reaction rates and lifetimes.

To convert between volume mixing ratios and concentrations, use the following procedure. For a species X, to convert from a mixing ratio, notated χ_X, to a concentration, notated [X], use the Ideal Gas Law to find the number of total molecules in a cm^3 and then multiply by χ_X, expressed as a fraction. Suppose p = 960 hPa (or mb) and T = 296K, and X = 60 ppbv, then

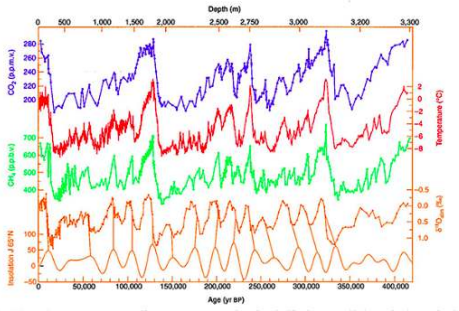
$$[X] = \frac{p}{kT}\chi_X = \frac{96000Pa}{\left(1.38 \times 10^{-23} JK^{-1}\ \text{molecule}^{-1}\right)(296K)}\left(\frac{1m^3}{10^6 cm^3}\right)60x10^{-9} \qquad (3.5.1.1)$$
$$= 1.4 \times 10^{12}\, molecules cm^{-3}$$

This page titled 3.5.1: Atmospheric Composition is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by William Brune (John A. Dutton: e-Education Institute) via source content that was edited to the style and standards of the LibreTexts platform.

3.5.2: Changes in Atmospheric Composition

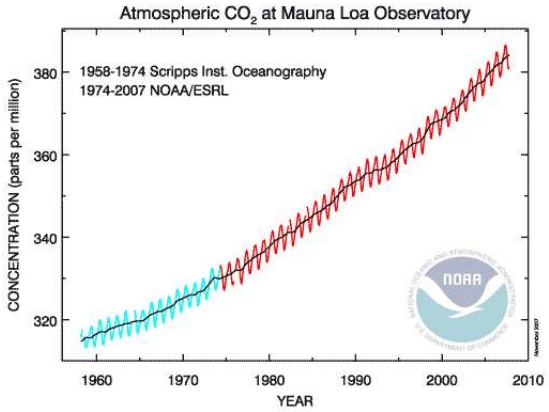
Since the rise of oxygen, 2 billion years ago, the nitrogen and oxygen fractions in the atmosphere have been stable. Water vapor is highly variable but, on average, appears to also have been fairly stable. Recent data from satellites and sondes indicate that perceptible water (the total amount of water that is in a column from the surface to space) has increased 1.3 ± 0.3% per decade over the oceans in the past 25 years (Trenberth et al., Climate Dynamics, 2005).

Historical (up to 500 years before the present) changes in CO_2 and CH_4 show large, rapid variations. Note that the historical range for CO_2 is 200-300 parts per million (10^{-6}), and for methane is 350-400 parts per billion (ppbv)**.** These changes in gas amounts have been driven by changes in Earth's temperature, which come from changes in Earth's orbit, the axis of Earth's rotation, and volcanoes. Until recently, changes in Earth's temperature caused changes in these gas amounts, which then reinforced the warming. In the past century, changes in the gas amounts have been driving the observed change in Earth's temperature.



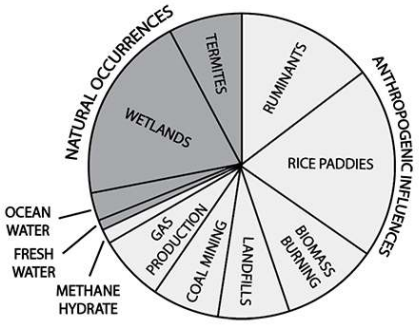
Changes in carbon dioxide, methane, and temperature over the past half million years. As measured from gases trapped in an ice core at Vostok, Antarctica. Credit: "Vostok 420ky curves insolation". Licensed under Public Domain via Wikimedia Commons

The recent changes in carbon dioxide show a fairly constant increase over the past 50 years. There is a smaller seasonal cycle imposed on this trend. This seasonal behavior occurs because CO_2 is taken up by plants in the northern hemisphere in summer, since most of the plants are in the northern hemisphere. Note that the current increase to above 400 ppm now extends well above any other time in the past half million years. Much of this CO_2increase can be linked to fossil fuel combustion. We will examine the scientific consequences of these CO_2 levels in the lesson on applications of radiation.

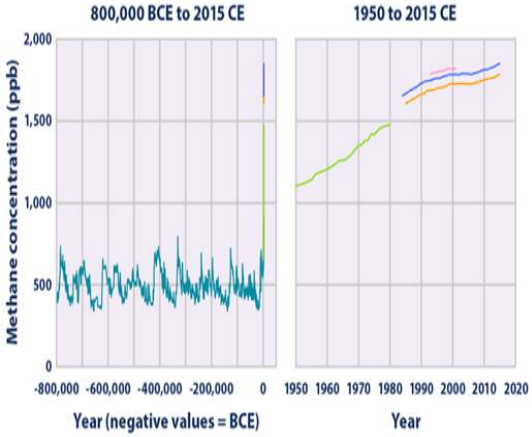


Atmospheric CO_2 mixing ratios measured at Mauna Loa, Hawaii for more than 50 years. Credit: NOAA

Another important trace gas is methane (CH_4), which is often called natural gas when it is used to produce power and heat. Methane has many sources, some of them natural and some of them anthropogenic (meaning human-made). Its atmospheric mixing ratio is now greater than 1800 parts per billion (ppb, or a fraction of 10^{-9} of air). We see from the figures below that almost 2/3 of atmospheric methane sources are anthropogenic and that methane measured at Hawaii, like all other locations, has been increasing except for the early to mid 2000s. This slowdown is not understood, but since the late 2000s, methane has been increasing again.



Sources of atmospheric methane: Freshwater, ocean water, wetlands, termites, ruminants, rice paddies, biomass burning, landfills, coal mining, gas production, methane hydrate. Ruminants are cattle, sheep, goats, etc. 2/3 of the total is due to human activities.



Changes in methane. Left, since 800,000 years ago. Right, since 1950. Note the tripling in the past 100 years, well above any levels since 800,000 years ago. The decade-long slow-down in the methane increase from 2000 to 2010 is not well understood. Credit: EPA

There are trends and variations in many of the other trace gases as well. Some others, like nitrous oxide (N_2O), are increasing, while others, like human-made chlorofluorocarbons (CCl_2F_2, CCl_3F) are decreasing. There are other trace gases that increase as the sun rises and decrease as it sets and are heavily involved in atmospheric chemistry. We will talk about these gases next.

This page titled 3.5.2: Changes in Atmospheric Composition is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by William Brune (John A. Dutton: e-Education Institute) via source content that was edited to the style and standards of the LibreTexts platform.

3.5.3: Other Trace Gases

Hundreds of different trace gases have been measured in the atmosphere and perhaps thousands more have yet to be measured. Many of these are **volatile organic compounds (VOCs)**. *Volatile* means that the compound may exist in the liquid or solid phase but that it easily evaporates. *Organic* means that the compound contains carbon but is not carbon dioxide, carbon monoxide, or carbides and carbonates found in rocks. There are also other chemicals like the nitrogen oxides (e.g., nitric oxide (NO), nitrogen dioxide (NO_2), nitric acid (HNO_3)), sulfur compounds (e.g., sulfur dioxide (SO_2), sulfuric acid (H_2SO_4)) and halogen compounds (e.g., natural methyl chloride (CH_3Cl), human-made chlorofluorocarbons (CCl_2F_2)). If we pay attention, we can often smell and identify many of these chemicals, even at trace levels, although some, like methane, carbon monoxide (CO), and chlorofluorocarbons are odorless. We enjoy smelling the VOCs emitted by trees in a forest – aah, that fresh pine smell – but we hold our nose to escape the smells of a stagnant swamp.

In addition to these thousands of chemicals that are emitted into the atmosphere every day, there are also some very reactive compounds that are created by atmospheric chemistry and play the important role of cleaning the atmosphere of many gases. The most important reactive gases are ozone (O_3) and hydroxyl (OH). We will focus the discussion of atmospheric chemistry on these two.

The Atmosphere's Oxidizing Capacity

Earth's atmosphere is an *oxidizing environment*. This term means what you think it would: gases that are emitted into the atmosphere react in a way that increases their oxygen content. Gases that contain oxygen tend to be "stickier" on surfaces and more water soluble, which means that they stick when they hit a surface or they can be readily taken up in clouds and rain drops and be deposited on Earth's surface. We call gases hitting the surface and sticking "dry deposition" and gases being taken up in precipitation and rained out "wet deposition."

Let's consider a natural gas that is very important in our lives – methane (a.k.a., natural gas). More and more methane is being extracted from below Earth's surface and used to run our electrical power plants, heat our homes, cook our food, and, increasingly, to run our transportation vehicles. Methane is a simple molecule – CH_4 – in which each of carbon's four bonds is made with a hydrogen atom. Energy comes from heating methane to high enough temperatures that cause it to react, giving off energy as more stable molecules are formed. In complete combustion, each methane molecule is converted into CO_2 and two H_2O. In the process, four oxygen atoms or two oxygen molecules are consumed.

This same process occurs in the atmosphere, but at much lower temperatures and at a much slower rate. In both cases, the first step in the methane oxidation sequence is the reaction with the hydroxyl radical (OH). In water, hydroxyl loses an electron and is ionized (OH^-), but in the atmosphere, hydroxyl is not ionized. We call OH a free radical because it has an odd number of electrons (eight for oxygen and one for hydrogen). Any gas with an odd number of electrons is reactive because the electrons want to be paired up in molecules because that makes them more stable.

Often, combustion is inefficient, resulting in the formation of carbon monoxide (CO). Examples include forest fires, humans burning fields to clear them for planting, poorly tuned vehicles, inefficient industrial processes, and other human-caused processes. The primary way that CO is removed from the atmosphere is by reacting with atmospheric OH. It takes a while for CO to be removed from the atmosphere by the reaction with OH, so that satellite instruments can track CO plumes as they emerge from their sources and flow around the world.

Where does OH come from?

Before we tackle this question, let's first look at where ozone (O_3) comes from. We will start with the stratosphere (a.k.a, good ozone because it blocks solar UV that harms humans, other animals, agriculture, and ecosystems) and then eventually we will consider tropospheric ozone (a.k.a., bad ozone, which is the ozone that hurts our health when we breathe it and that damages plants and their fruit).

Discussion Activity: Trace Gases

I would like you to think about which trace gas is the most important and why. By trace gas I mean a gas with a mixing ratio of less than 20 ppm in the atmosphere. Defend your choice. Use information from this lesson as well as other sources (credit them, please!) to describe the qualities of this gas that make you think that it is the most important trace gas. Then read the choices of your classmates and respond to their choices and follow-up with further questions and/or analysis

This page titled 3.5.3: Other Trace Gases is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by William Brune (John A. Dutton: e-Education Institute) via source content that was edited to the style and standards of the LibreTexts platform.

3.5.4: Stratospheric Ozone Formation

Ozone is ozone no matter where it is in the atmosphere. Good ozone is good only because it is in the stratosphere where we cannot breathe it (Figure 3.5.4.1). Bad ozone also absorbs solar ultraviolet light, but it is down near Earth's surface where we can breathe it. For UV protection, we are interested in the total number of ozone molecules between us and the Sun. 90% of ozone molecules are in the stratosphere and 10% are in the troposphere - some down near Earth's surface where we can breathe them. There are important issues affecting human and ecological health for both good ozone and bad ozone. For good ozone, the most important issues are the reduction of ozone globally, the **Antarctic Ozone Hole**, and Arctic ozone loss that is caused by chlorofluorocarbons. Reduced ozone means more solar UV gets to the ground causing more skin cancer. For bad ozone, the most important issues include the production of too much ozone in cities and nearby regions that is caused by too many pollutants from traffic, industrial processes, power generation, and other human activities. Increased ozone means more people have respiratory and heart problems. Let's look at both the good and the bad, starting with the stratospheric ozone.

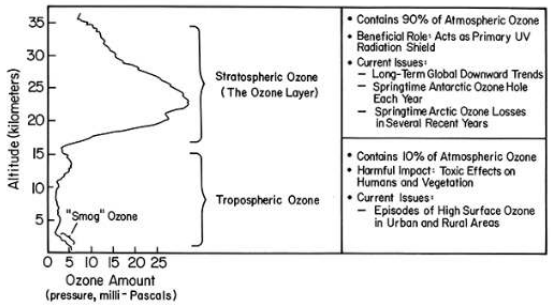


Figure 3.5.4.1: A cross-section of the typical vertical ozone profile for the tropics. Credit: World Meteorological Organization

Click for a text description of the Cross Section image.

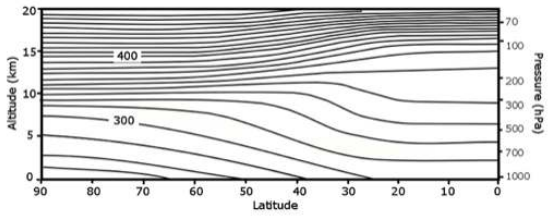
Explanation diagram of the typical vertical ozone profile for the tropics

Stratospheric Ozone (the ozone layer, ~15-35 km) contains 90% of atmospheric ozone beneficial role: acts as primary UV radiation shield Current issues

- long-term global downward trends
- springtime Antarctic ozone hole each year
- springtime arctic ozone losses in several recent years
- episodes of high surface ozone in urban and rural areas

Tropospheric Ozone (0-15 km) contains 10% of atmospheric ozone harmful impact: toxic effects on humans and vegetation current issues

To get the total amount of ozone between us and the Sun, we simply add up the ozone amount starting at the surface and going up to the top of the ozone layer. Note how much more ozone there is in the stratosphere. At higher latitudes, the bottom of the stratospheric ozone layer is at approximately 10-12 km. Recall the following image from Lesson 2:



Potential temperature (solid lines, K) as a function of latitude and altitude. Note that the decrease in potential temperature with height is small in the troposphere and large in the stratosphere. Credit: W. Brune, after Andews, Holton, and Leovy

The process of stratospheric ozone formation starts with ozone (O_3), which is made by ultraviolet sunlight in the stratosphere (but not the troposphere, as we shall see). The two reactions are:

$$O_2 + \text{hard}\, UV \rightarrow O + O \tag{3.5.4.1}$$

$$O + O_2 + N_2 \rightarrow O_3 + N_2 \tag{3.5.4.2}$$

Note that N_2 doesn't really react in this last chemical equation, but instead, simply bumps into the O_3 molecule as it is being formed and stabilizes it by removing some of the energy from O_3. We call O_3 an oxidant because it can react with some compounds and oxidize them.

This O_3 can be broken apart by ultraviolet light to make O_2 and O. Usually O combines with O_2 to form O_3 in this way: $O+O_2+N_2 \rightarrow O_3+N_2$, so nothing really happens, except that the solar energy that breaks apart the O_2 ends up as extra energy for the O_3 and for the colliding N_2 and, as a result, ends up warming the air. Sometimes O collides with O_3 and reacts: $O+O_3 \rightarrow O_2+O_2$. Putting all of the reactions together, we can see the chemical lifecycle of ozone in the stratosphere. This set of reactions was proposed in the 1930s by Chapman:

Stratospheric Ozone and Atomic Oxygen Production, Cycling, and Loss

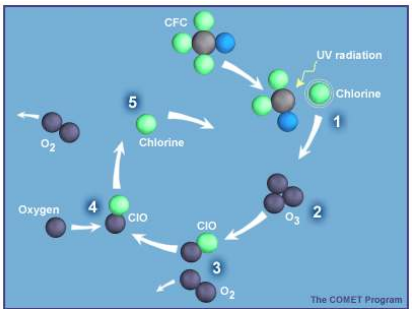
O_2 + hard UV → O + O	production
2(O + O_2 + N_2 → O_3 + N_2)	cycling
O_3 + UV → O_2 + O	cycling
O + O_3 → O_2 + O_2	loss
Net: UV → higher T	

These four reactions could produce the basic characteristics of the ozone layer as it was in the 1940s through the 1970s. However, this theory produced peak ozone levels that were 50 milli-Pascals, not the 25-30 milli-Pascals seen in the first figure above. Thus, the measured levels of stratospheric ozone were about half of those predicted by Chapman's theory - it was a real puzzle. However, in the 1970s, scientists proposed new sets of reactions by other gases that accomplished the same results as the loss reaction shown above. A famous example involved chlorine, which comes mostly from human-made chlorofluorocarbons (CFCs):

The Stratospheric Chlorine Catalytic Cycles That Destroy Ozone

CFCs + UV → product + Cl	production
Cl + O_3 → ClO + O_2	cycling
ClO + O → Cl + O_2	cycling
Cl + CH_4 → HCl + CH_3	loss
Net O_3+O: O_3 + O → O_2 + O_2	

During the cycle, chlorine (Cl) and chlorine monoxide (ClO) aren’t destroyed but instead are just recycled into each other. With each cycle, two ozone molecules are lost (one directly and a second because O almost always reacts with O_2 to form O_3). This cycle can run for hundreds of thousands of times before Cl gets tied up in HCl. So ClO and Cl levels of tens of parts per trillion of air (10^{-12}) are able to destroy several percent of the few parts per million of O_3. Sherry Rowland and Mario Molina figured this cycle out and wrote a paper about it in 1974. They received a Nobel Prize in Chemistry in 1995 for this work. When catalytic cycles involving chlorine, nitrogen oxides, and OH are included with the theory, the agreement between the theory and the measurements gets much better.



The chlorine catalytic cycle that destroys ozone. Credit: UCAR

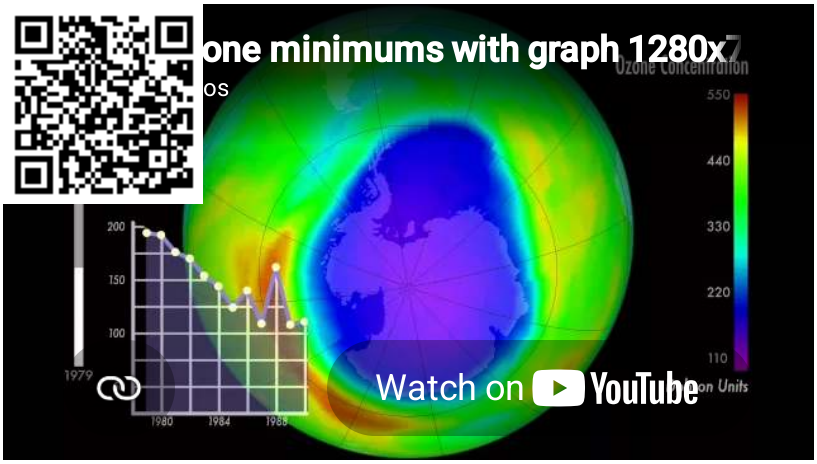
? Exercise

Note that the total ozone amount at midlatitudes is greater than the amount in the tropics. This should seem strange to you because the solar UV that is part of the Chapman mechanism is strongest in the tropics. Why do you think that total ozone is distributed this way?

Click for answer.

Answer: The ozone distribution is due to the motion of air through the stratosphere. Air comes from the troposphere into the stratosphere mostly in the tropics and, as it rapidly moves from west to east, it slowly moves from the tropics to near the poles, where it re-enters the troposphere. So most of the ozone is made in the tropics at higher altitudes and then some of this ozone is destroyed by chemical reactions as the ozone is transported poleward and downward. So, while the ozone mixing ratio decreases slightly from tropics to high latitudes, the ozone concentration (moles per volume) increases as it is transported to lower altitudes where the pressure and number of moles is greater, more than twice as large (see the Lesson 2 image above).

The low ozone over Antarctica above is the Antarctic Ozone Hole; the video below (:31) entitled "Ozone Minimums With Graph" (from NASA) shows changes in ozone concentration between 1979 and 2013.



Satellite map of total ozone from September 6, 2004. Note that the total ozone amount is greater at midlatitudes than it is in the tropics, even though the solar UV is most intense in the tropics. The extremely low ozone over Antarctica is the Antarctic Ozone Hole. Credit: NASA GSFC

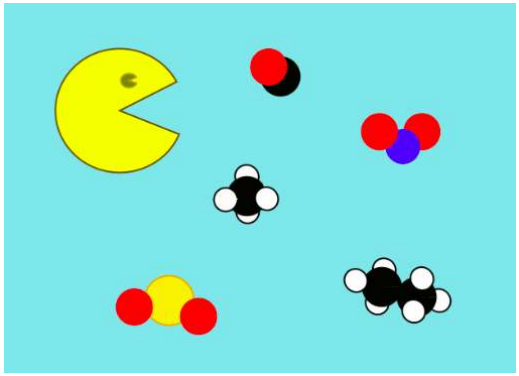
The Antarctic ozone hole is an extreme example of the destructive power of chlorine catalytic cycles. Different catalytic cycles dominate the ozone destruction over Antarctica and, to a lesser extent, the Arctic. But, when aided by chemistry on the surfaces of naturally occurring polar stratospheric clouds, all the Cl in the form of HCl is liberated so that the polar catalytic cycles are able to destroy a few percent of the ozone per day in a plug the size of Antarctica from an altitude of 12 km all the way up to 20 km.

Fortunately, the amount of chlorine being injected into the stratosphere is decreasing due to the Montreal Protocol, the world's first international global environmental treaty.

This page titled 3.5.4: Stratospheric Ozone Formation is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by William Brune (John A. Dutton: e-Education Institute) via source content that was edited to the style and standards of the LibreTexts platform.

3.5.5: The Story of the Atmosphere's PAC-MAN

The atmosphere's oxidation capacity is its ability to clean itself of all of the gases that are emitted into it. What does stratospheric ozone have to do with the atmosphere's oxidation capacity, which mostly occurs in the troposphere and mostly by the atmosphere's PAC-MAN, hydroxyl (OH)? It turns out that natural dynamic processes actually pull air down from the stratosphere and mix it into the troposphere, eventually mixing some of this ozone to Earth's surface. This naturally occurring surface ozone provides a baseline value for near-surface ozone, but ozone pollution is more than ten times greater than this baseline in cities. Ozone is both sticky on surfaces and fairly reactive in the atmosphere. It is lost both by depositing on surfaces and through being chemically destroyed by reactions in the atmosphere.



Hydroxyl (OH) is the PAC-MAN™ of the atmosphere. It reacts with thousands of different molecules, including carbon monoxide (CO), nitrogen dioxide (NO_2), methane (CH_4), sulfur dioxide (SO_2), and ethane (C_2H_6). Credit: W. Brune

The following chemical sequences are the humble beginnings of the atmosphere's PAC-MAN. OH is generated throughout the stratosphere and troposphere by a two-step reaction sequence. The first step is:

$$O_3 + UV \rightarrow O_2 + O^* \tag{3.5.5.1}$$

where O* is an excited-state oxygen atom that has extra chemical energy. O* can lose this extra energy by colliding with N_2 and O_2, but it can also collide with a water molecule to make two OH molecules:

$$O^* + H_2O \rightarrow OH + OH \tag{3.5.5.2}$$

OH is very reactive. You can think of OH as being water that has had a hydrogen taken away and wants it back. There are other sources for OH, but this one is the most important globally. OH reacts with many other atmospheric constituents. In fact, it is so reactive, that its lifetime in the atmosphere is less than a second.

Another important oxidant is nitric oxide (NO). It comes from combustion (power plants, internal combustion engines, fires) or lightning. In cities, the NO mixing ratio is tens of ppb (10^{-9}, by moles) during morning rush hour and a bit smaller during evening rush hour, but there is typically about a ppb around during the day. In very remote areas, the levels of NO are a hundred times less. NO can react with many other chemicals, but it reacts with O_3:

$$NO + O_3 \rightarrow NO_2 + O_2 \tag{3.5.5.3}$$

which forms nitrogen dioxide, NO_2. NO_2 is not very stable:

$$NO_2 + \text{nearUV} \rightarrow NO + O \tag{3.5.5.4}$$

ut the O reacts immediately with O_2 to form ozone:

$$O + O_2 + N_2 \rightarrow O_3 + N_2 \tag{3.5.5.5}$$

If a NO_2 molecule is produced, then an O_3 molecule will be produced during the day when the sun is out. *Note that if we think of these three reactions as a cycle, no ozone was either created or destroyed because it is destroyed in [4.5] and created in [4.2].*

What happens to all of the methane emitted into the atmosphere?

Methane is a volatile organic compound (VOC). Methane oxidation is a good model for what happens to all of the volatile organic compounds that you smell every day and all the ones that you can't smell. I am not going to show you the entire reaction sequence. Instead, here are just a few steps.

The first step is the reaction between methane and hydroxyl:

$$CH_4 + OH \rightarrow CH_3 + H_2O \tag{3.5.5.6}$$

Note that water vapor is made and CH_3 is a radical because it has 12+3 = 15 protons and, therefore, electrons. Just as for most other VOCs, and some other trace emissions, the reaction with OH is the main way methane is removed from the atmosphere. Otherwise, it would build up to high abundance.

CH_3 is very reactive. It adds an O_2:

$$\mathrm{CH_3 + O_2 + N_2 \rightarrow CH_3O_2 + N_2} \tag{3.5.5.7}$$

If there is any NO around, the following reaction happens:

$$\mathrm{CH_3O_2 + NO \rightarrow CH_3O + NO_2} \tag{3.5.5.8}$$

followed by:

$$CH_3O + O_2 \rightarrow CH_2O + HO_2 \tag{3.5.5.9}$$

and:

$$HO_2 + NO \rightarrow OH + NO_2 \tag{3.5.5.10}$$

The chemical CH_2O is formaldehyde. Some of you may have encountered it in high school chemistry or biology and so may be familiar with the smell. You also see that we got the OH molecule back.

Ultimately, formaldehyde gets broken down to CO and the net reaction of methane oxidation is:

$$CH_4 + 2NO + 3O_2 \rightarrow\rightarrow\rightarrow\rightarrow CO_2 + 2H_2O + 2NO_2 \tag{3.5.5.11}$$

Remember that NO_2 is easily broken apart by the UV sunlight that reaches Earth's surface, so we can take this reaction sequence a step further and show that in the presence of sunlight, reactions [4.6] and [4.2] give:

$$NO_2 + \mathrm{nearUV} + O_2 \rightarrow NO + O_3 \tag{3.5.5.12}$$

or

$$CH_4 + 5O_2 \rightarrow\rightarrow\rightarrow\rightarrow CO_2 + 2H_2O + 2O_3 \tag{3.5.5.13}$$

In this final chemical equation, we do not see OH, HO_2, NO, or NO_2, yet they are essential to the formation of ozone. They are catalytic, which means that they are neither created nor destroyed in the reaction sequence, but instead are simply recycled between OH and HO_2 and between NO and NO_2. There are other reactions that destroy these reactive chemicals by producing other chemicals that are much less reactive and sticky, a main one being:

$$OH + NO_2 + N_2 \rightarrow HNO_3 + N_2 \tag{3.5.5.14}$$

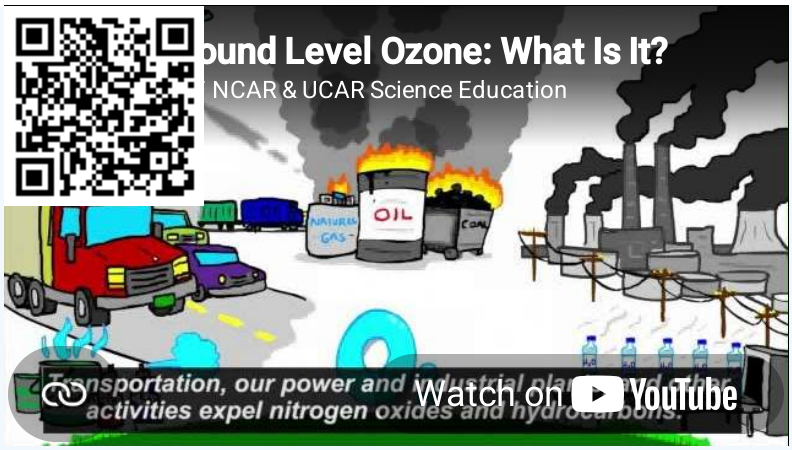
where HNO_3 is nitric acid, a very sticky and water soluble chemical. However, each OH that is produced can typically oxidize more than ten methane molecules before it reacts with NO_2 to form nitric acid. And as reaction [4.13] shows, each time methane is completely oxidized, two O_3 molecules are produced. That's a lot of chemical steps to remember, but I don't want you to necessarily remember them. *I want you to see that the process started with a reaction of OH with a volatile organic compound (in this case methane) and that in the subsequent reactions, the product molecules had more and more oxygens attached to them. This process is why we say that the atmosphere is an oxidizing environment.*

Where does ozone pollution come from?

Ozone is a different sort of pollutant from others because it is not directly emitted by a factory or power plant or vehicle but instead is produced by atmospheric chemistry.

Three ingredients are needed to make ozone pollution: *volatile organic compounds (VOCs)* (like methane); *nitric oxide (NO from combustion)*; and *sunlight*. When we say this, we assume that we already have some ozone and water to provide the OH to get the reactions started. Every VOC goes through an oxidation process that is similar to the methane oxidation reaction sequence. In the methane oxidation sequence, steps [4.9] and [4.11] make NO_2, which in the presence of sunlight makes ozone through step [4.6] followed by step [4.2]. Voila! Ozone is formed from methane oxidation in the presence of nitrogen oxides and sunlight. Now imagine the thousands of volatile organic compounds in the atmosphere and realize that all of them - both anthropogenic and natural - can participate in the production of ozone pollution. Now you have seen the sequence of chemical reactions that produce tropospheric ozone.

Let's look at a video (3:14) entitled "Ground Level Ozone: What Is It?" that explains ozone production without getting into the gory details of the chemistry.



Ground Level Ozone: What is it? Credit: UCARConnect

Click here for transcript of the Ground Level Ozone video.

We're all pretty familiar with what O2 is. I hope so. You need to breathe it to live. Yes, O2 is oxygen, that life-giving gas, but what is O3? O3 is another gas essential to our survival but it's definitely not for breathing. O3 is ozone high up in the stratosphere. It's made naturally and absorbs harmful ultraviolet rays from the Sun. Without it life as we know it wouldn't, couldn't exist. We need the ozone layer in the stratosphere. We want it, we rely on it. But don't get too used to singing ozone's praises. High ozone levels at lower altitudes, what we call the troposphere, where we live and breathe or anything but natural and beneficial. In fact, down here it turns out to be a toxic atmospheric pollutants. Yep, you heard me right. ground-level ozone primarily exists due to human activities that burn fossil fuels. Transportation, our power and industrial plants, and other activities expel nitrogen oxides and hydrocarbons. When those compounds interact with sunlight, voila, ozone is created a contributor to smog. that's why I ozone levels increase during the summer months when sunlight is abundant. Yes, smog love summer just like many of us. We run, bike, hike, fish, play, stroll, oh yeah, and breathe. Yes, the fact that more people are outside when it's warmer makes us particularly vulnerable to Ozone's harmful impacts. Ozone is a harmful oxidant when we inhale it it's like getting a sunburn inside your lungs and it can be particularly serious for the young, old, active, and those with respiratory conditions at any age. And it's not just humans that are vulnerable ozone harms plants, crops, and agricultural yield interfering with pretty important processes like well, photosynthesis and even our economy. To make matters worse ozone production increases with higher temperatures which are occurring more frequently with climate change. The EPA sets national ambient air quality standards for several pollutants in the United States including ground level ozone. When a county is out of compliance they need to know what can be done to improve air quality. and let's not forget that air pollution is a global comments. air pollution is shared from surrounding cities states and also country's halfway around the world. What can we do, what are we willing to do to improve current levels? Drive less, carpool, avoid car idling, set your home's thermostat higher in the summer and lower in the winter, avoid gas powered lawn & garden tools on severe ozone days. There's a lot to do and lots to know about air quality knowing more about the sources and contributors to ozone and other atmospheric pollutants will help us chart our course.

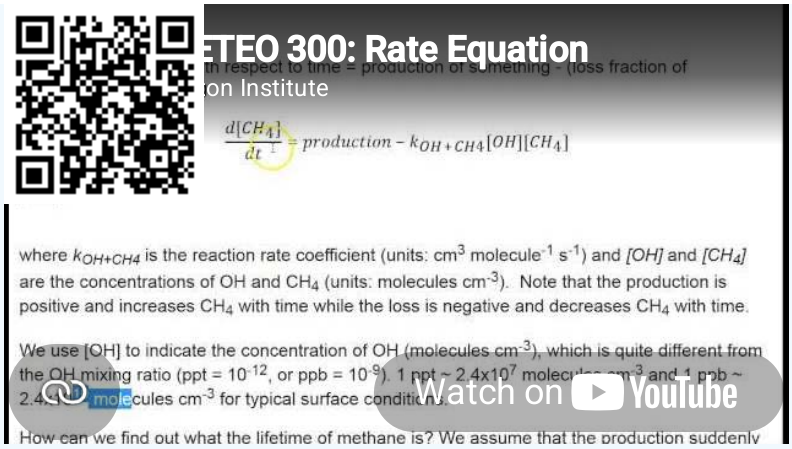
Ozone pollution is bad for the health of people, crops, and forests. Ozone can react with some types of VOCs, including types that make up our lungs, and breathing it can cause serious health problems and even death. Ozone reacts with the VOCs that make up plants and stunts their growth and damages their fruit. The Clean Air Act from the 1970s has dramatically decreased the levels of air pollution in the United States, including ozone. The EPA can take the credit for much of the progress against air pollution in the United States. But there is still a ways to go and the progress may be reversed due to effects of climate change. Since ozone pollution increases at higher temperatures, the increases in global temperatures could actually reverse the steady progress in ozone reduction and ozone pollution could once again increase, unless volatile organic compounds and nitrogen oxides are reduced even more.

Now you can see why OH is called the PAC-MAN of the atmosphere. But how can we tell how long it will take for OH to remove from the atmosphere some trace gas like methane? Let's look at an equation for the budget of methane. It is produced in the atmosphere by all the emissions from cows and wetlands. It is removed from the atmosphere by reactions with OH [4.7]. The rate of removal, that is the change in the methane concentration, is always proportional to the amount of the two reactants, in this case, CH_4 and OH. So, the change in methane is given by the balance between the production and the loss by reaction with OH:

$$\frac{d\,[CH_4]}{dt} = \)production\(- k_{OH+CH4}[OH]\,[CH_4] \qquad (3.5.5.15)$$

where k_{OH+CH4} is the reaction rate coefficient (units: cm^3 $molecule^{-1}$ s^{-1}) and *[OH]* and *[CH_4]* are the concentrations of OH and CH_4 (units: molecules cm^{-3}). Note that the production is positive and increases CH_4 with time while the loss is negative and decreases CH_4 with time.

We use [OH] to indicate the concentration of OH (molecules cm^{-3}), which is quite different from the OH mixing ratio (ppt = 10^{-12}, or ppb = 10^{-9}). 1 ppt ~ $2.4x10^7$ molecules cm^{-3} and 1 ppb ~ $2.4x10^{10}$ molecules cm^{-3} for typical surface conditions. See the video below (1:47) entitled "Rate Equation" for further explanation:



Rate Equation

Click here for transcript of the Rate Equation video

Let me explain equation 4.15, which is a rate equation for methane. A rate equation is just a differential equation. The change of something with respect to time equals the production rate of something, minus the fraction of something that is lost each unit of time, multiplied by the amount of something. Note that the loss rate of something is always proportional to something. That something can be anything. It does not have to be a chemical concentration. It could be the amount of milk in your refrigerator, or the number of socks in your drawer, both of which tend to disappear over time. And equation 4.15 is the methane concentration, which has units of molecules per centimeter cubed. The production rate is in units of molecules per centimeter cubed per second. Remember, each term of the equation must have the same units. The last term is the loss rate. The reaction rate coefficient has units of centimeter cubed per molecule per second, but when we multiply it by the OH concentration, we get a product that has units of per second, which is a frequency. Now, OH varies from almost 0 at night, to a peak value at midday. However, we can take an average OH to find the average loss rate of methane. Note that if we assume that the production rate suddenly goes to 0, then we find a very simple equation, which has an exponential

solution. We designate the time that it takes the exponential factor to go to minus 1 as a lifetime, which is just the inverse of a loss frequency.

How can we find out what the lifetime of methane is? We assume that the production suddenly stops and equals 0. Then [4.15] becomes:

$$\frac{d\left[CH_4\right]}{dt} = -k_{OH+CH4}[OH]\left[CH_4\right] \tag{3.5.5.16}$$

$$\frac{d\left[CH_4\right]}{\left[CH_A\right]} = -k_{OH+CH4}[OH]dt \tag{3.5.5.17}$$

k_{OH+CH4} is the reaction rate coefficient for this reaction. Assume that OH is constant. Because OH is generated mostly from sunlight, it follows the sunshine and is greatest near midday and is very small at night. However, we assume that the OH concentration is the average over the day and night in order to assign it a constant value. Now integrate both sides of the equation:

$$\int_{[CH_4]_0}^{[CH_4]} \frac{d\left[CH_4\right]}{\left[CH_4\right]} = \int_0^t -k_{OH+CH_4}[OH]dt \tag{3.5.5.18}$$

$$\ln\left(\left[CH_4\right]\right) - \ln\left(\left[CH_4\right]_0\right) = -k_{OH+CH_4}\overline{[OH]}t \tag{3.5.5.19}$$

$$\ln\left(\frac{\left[CH_4\right]}{\left[CH_4\right]_0}\right) = -k_{OH+CH_4}\overline{[OH]}t \tag{3.5.5.20}$$

take exponential of both sides

$$\frac{\left[CH_4\right]}{\left[CH_4\right]_0} = e^{-k_{OH+CH_4}\overline{[OH]}t} \tag{3.5.5.21}$$

$$\left[CH_4\right] = \left[CH_4\right]_0 e^{-k_{OH+CH_4}\overline{[OH]}t} \tag{3.5.5.22}$$

So we see that *methane decreases exponentially with time*.

The atmospheric lifetime is defined as the time it takes something to decrease to e^{-1} = 0.37 of its initial value. So to find the lifetime of methane in the atmosphere, we see when k_{OH+CH4}[OH] t = 1, or:

$$\tau = \frac{1}{k_{OH+CH4}[OH]} \tag{3.5.5.23}$$

where τ indicates the lifetime. k_{OH+CH4} =3x10^{-15} cm^3 $molecule^{-1}$ s^{-1} and [OH] ~ 10^6 molecules cm^{-3}, so:

$$\tau = \frac{1}{3x10^{-15}10^6} = 3 \times 10^8 \text{seconds} \sim 10 \tag{3.5.5.24}$$

years

This reaction rate coefficient is fairly slow. Other VOCs have reaction rate coefficients that are typically hundreds to hundreds of thousands of times faster, so the lifetimes of most VOCs is hours to days.

The atmospheric lifetime of a gas is very important for determining how far a gas can travel from its source. Some trace gases have lifetimes of hours, so unless they are made by atmospheric chemistry, they can't travel more than a few tens of kilometers from their sources. Other gases have much longer lifetimes - methane is a good example with its 10-year lifetime. In 10 years, it can travel from its sources to most anywhere around the globe, even to the stratosphere. NASA measures the amounts of several gases from space. An excellent NASA website for accessing these satellite data and having it plotted as global maps is the Center for Trace Gas Data & Information Website at the NASA Goddard Space Flight Center's Earth Sciences Distributed Active Archive Center (GES DISC).

This concept of atmospheric lifetime is pretty important. For instance, what if an industry is spewing a chemical into the atmosphere that is toxic at a certain concentration in the atmosphere? Then it is important to know if that chemical is removed in less time than it takes to become toxic or if it is going to continue to build up at toxic levels and not leave the atmosphere for a

long, long time. If its atmospheric lifetime is hundreds to thousands of years, then maybe we shouldn't let that industry dump that chemical into the air.

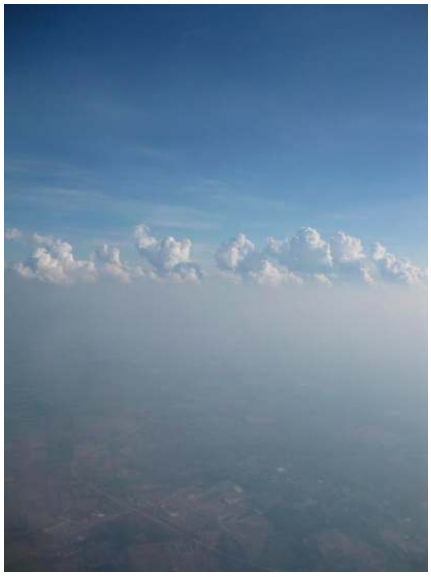
This page titled 3.5.5: The Story of the Atmosphere's PAC-MAN is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by William Brune (John A. Dutton: e-Education Institute) via source content that was edited to the style and standards of the LibreTexts platform.

3.5.6: Where do Cloud Condensation Nuclei (CCN) come from?

Now that you know everything about the atmosphere's gas-phase composition, it is time to look at its particle composition. We are interested in atmospheric particles for several reasons:

- smaller ones can get into the lungs and cause serious health problems;
- smaller ones can absorb or scatter sunlight, thus affecting climate;
- some of them are good cloud condensation nuclei, which are essential for the formation of clouds.

Atmospheric aerosol is most obvious to us on warm muggy summer days. Under these conditions, there are lots of aerosol particles and they absorb water and swell up to a size that is quite efficient at scattering sunlight. The following picture was taken over Maryland on a flight between Washington Dulles airport and State College airport. Above the fair weather cumulus clouds is blue sky in the free troposphere. Below the clouds is the atmospheric boundary layer, which is filled with aerosol that has been well mixed by warm, moist air parcels rising and stirring the boundary layer air. The haze is so thick that it is a little hard to see the ground.



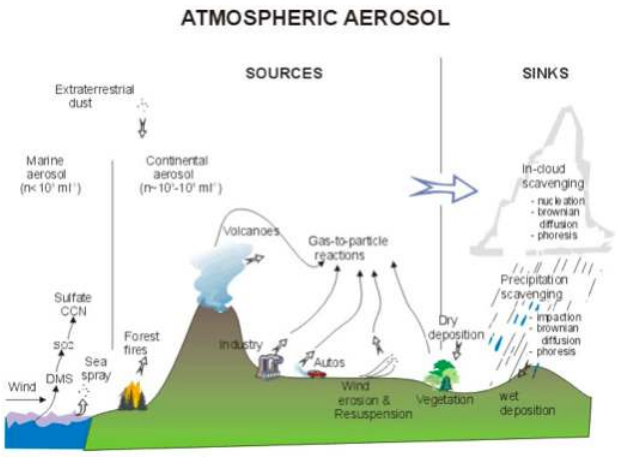
Summertime haze over Maryland. Credit: W. Brune

Atmospheric particles come from many different sources. Good cloud condensation nuclei (CCN) must be *small particles*, so that they do not settle too fast, and must be *hydrophilic*, which means that water can stick. They can be *either soluble* (i.e., dissolvable in water), *or insoluble*, but most are soluble.

Most particles originate from emissions from Earth's surface. **Primary aerosols** are emitted directly from the source, although the smaller ones start off as hot gases that rapidly condense to form particles even before they leave the smokestack or tailpipe. **Secondary aerosols** are gaseous emissions that are converted to aerosol particles by chemical reactions in the atmosphere. Some of these become CCN. This process is often called gas-to-particle conversion. Most CCN are secondary aerosols.

The sources are both natural and anthropogenic (human-made). Seaspray, volcanoes, forests, and forest fires, as well as gas-to-particle conversion of naturally occurring gases such as sulfur dioxide (SO_2) and some naturally occurring VOCs, such as α-pinene (which gives the pine smell) are important natural particle sources. Industry, power plants, using fires to clear cropland, transportation, and gas-to-particle conversion of anthropogenic SO_2 and numerous other gas emissions are important anthropogenic particle sources.

Note that we must pay attention not only to the aerosol sources but also the aerosol sinks, as shown in the diagram below.



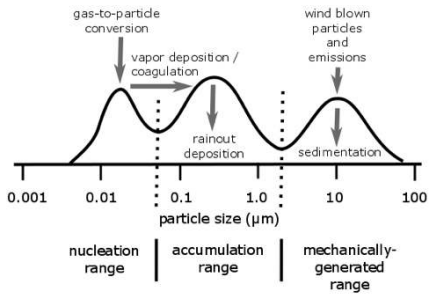
Atmospheric Aerosol Sources and Sinks Sources of Atmospheric Aerosol: forest fires, extra terrestrial dust, continental aerosol, volcanoes, industry, autos, wind erosion and resuspension, gas-to-particle reactions Sources of Atmospheric inks: In-cloud scavenging (nucleation, bornian diffusion, phoresis), precipitation scavenging (impaction, brownian diffusion, phoresis) *Credit: NOAA*

The different sources make particles of different sizes. The typical size distribution (i.e. number of particles in a volume of air, plotted as a function of size) has bumps in it, with more particles at some sizes than at others, as seen in the diagram below. Reading these bumps tells us a lot about how the particles were made.

The **nucleation mode** (there are other designations for this) includes particles that are made by gas-to-particle conversion. A low-volatility vapor is one that will condense onto particles or other surfaces when its vapor pressure exceeds its low saturation vapor pressure. This situation is analogous to water.

Coarse mode includes particles made by mechanical processes. The hydrophilic coarse particles can be CCN, but they settle out pretty fast.

Accumulation mode particles are usually made when nucleation particles collide and stick (called coagulation) or when gases accumulate on a nucleation mode particle. They neither settle fast nor coagulate, so they tend to hang around in the atmosphere for a few weeks. They make pretty good CCN.



Typical aerosol size distributions and their sources and sinks. Nucleation range particles quickly grow into accumulation range particles, which fall slowly to the ground and stay a week or two in the atmosphere. The larger particles, called coarse particles, fall to the ground within hours or sometimes days. Gas-to-particle conversion means that the particles start as gases but are converted by reactions to sticky chemicals that form particles. Credit: W. Brune

PM2.5 - Secondary Particles from Gas-to-Particle Formation

PM2.5 is a particle size designation that means “Particle matter smaller than 2.5 µm in diameter”. Another common term is PM10, which is "particle matter smaller than 10 µm in diameter". PM2.5 particles are the ones that are most important for human health and climate, and, in many cases, cloud formation because of their longer lifetime in the atmosphere.

Secondary particles start with the emission of VOCs or sulfur compounds, which react mainly with OH to start a sequence of reactions. These reactions tend to add oxygen to the molecules, which chemically makes them stickier (with a lower saturation vapor pressure) and more water soluble, which is just what is needed to make them better cloud condensation nuclei.

For particles that start as gaseous sulfur compounds, such as sulfur dioxide (SO_2), the reaction sequence starts with OH and the reaction product is sulfuric acid, a compound that has a very low vapor pressure and is very sticky.

$$SO_2 + OH \rightarrow\rightarrow H_2SO_4 \tag{3.5.6.1}$$

Sulfuric acid is easily taken up into cloud drops and raindrops and then can be deposited on Earth's surface when it rains. The good news is that the rain cleans the atmosphere. The bad news is that the rain is very acidic and has earned the name "acid rain" because of its harmful effects on forests and on buildings, memorials, and statues.

If sulfur sources are upwind of an area, the particles in that area will contain some sulfur. But almost all atmospheric particles also contain some organic compounds and sometimes particles are mainly made up of carbon-containing organic compounds. Some of these organic particles are primary, but most of the small ones are made by gas-to-particle conversion, which is just a simple way to say the volatile organic compounds react in the atmosphere with OH or O_3 to form less volatile organic compounds that become aerosol particles. The chemicals in these particles can continue to oxidize, thus making them even better CCN.

We can demonstrate gas-to-particle conversion of a VOC that is often emitted into the atmosphere by trees. This compound is limonene and also comes from oranges. In the video (4:47) below entitled "Demonstration of Gas to Particle Conversion," I will use orange peel to demonstrate this effect.



Gas to Particle Conversion Demonstration

Click here for transcript of the Gas to Particle Conversion video.

Today I'm going to show you how particles can come from gases. So what I have here is I have an artificial atmosphere right here, a glass jar. I have a light source in there, which will act like the sun, so here's the sun right here. I'll set the sun. Sun's off right now. And then I have a source of gases, an orange. So let me first peel the orange. Let's peel the orange here. A little bit of orange. Take some peel off the orange. OK, that's enough. So now here we have some orange peel. That smells really good. So what I'm going to do is take some orange peel and squeeze it a little bit, and I'm going to drop it in the atmosphere. Squeeze some more and drop it in the atmosphere. Now I'll put the lid on the atmosphere. OK. There we go. Now you see I have orange peel here and so there's some gases that smell really good, these volatile organic compounds. So I can smell, and they smell great. And I'm going to show you that even though I put that in there, I have a light source. Here's a little laser, just a laser pointer, and you see I'm shining it in there and you don't really see any particles at all in here. So look in here and there are no particles. See that? No particles. OK, now I'm going to turn on the sun, which is this nice ultraviolet light here. So I'll turn on the sun, and we're just going to wait a little bit. So the sun is on. You can see it glowing here and here. And so what that's doing is that's producing lots and lots and lots of OH, the hydroxyl radical. And it's also producing lots of ozone, and so it's doing what the atmosphere does, which is oxidizing the volatile organic compounds that came from the orange peel. Those volatile organic compounds, one of the main ones is called limonene, which is of course also in lemons and limes. And so we are just going to let that cook a little bit and let the day go on a little bit. And then what we can do is we can see if we're making any particles. So let me shake it up a little bit like there's a little bit of a wind. There we go. So now you see we have a very bright light from the lamp. It's really gotten bright. And we can see that it's doing a lot. Heat it up a little bit more. Now we'll just take a quick look and see if we see anything. So the laser

light here we can't see anything in the atmosphere outside of the chamber, so I'm shining it. Right now this is hitting my hand, as you see here. And there's no sort of scattering. But I've turned off the light now. Now we shine this in here, and we see a tremendous beam. And so all this is small particles that we made which are now scattering that light. See you can see that very strong beam. And all those particles came from the volatile organic compounds, limonene and others, that came out of the orange peel and then they were oxidized in the atmosphere to make less volatile compounds, one that had lower vapor pressure, and then those all stuck and made these nice little particles that you see here. So there's a demonstration of gas to particle conversion.

This page titled 3.5.6: Where do Cloud Condensation Nuclei (CCN) come from? is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by William Brune (John A. Dutton: e-Education Institute) via source content that was edited to the style and standards of the LibreTexts platform.

3.5.7: Summary and Final Tasks

Atmospheric composition, even of trace gases, has a huge influence on weather and climate. Carbon dioxide is the most abundant trace gas – its mixing ratio is 400 ppm and growing, but other trace gases are also emitted into the atmosphere. The atmosphere cleans itself of these gases by atmospheric chemistry, which oxidizes the gas emissions and produces new chemicals that contain oxygen and so are stickier and more water soluble.

These new chemicals can be removed from the atmosphere either by hitting surfaces and sticking or by being taken up in clouds or rain drops and precipitated to the ground. The main oxidant is hydroxyl (OH), which is made with ozone, UV sunlight, and water vapor and starts the removal sequence by reacting with gas emissions. In these reaction sequences ozone pollution is produced if the pollutant nitric oxide (NO) is also present.

This *pollutant ozone* is nearby and is harmful to human health and agriculture. *Stratospheric ozone*, on the other hand, shields Earth from harmful UV, and is made a completely different way – by the breaking apart of O_2 to produce O, which reacts readily with O_2 to form O_3. Some of this stratospheric ozone is then transported to Earth, but at levels much lower than pollutant levels. Methane oxidation is an example of the VOC reactions that produce ozone and particles.

An important concept is the atmospheric lifetime of gases and particles. This can be determined by solving a simple linear differential equation. The methane lifetime was shown to be about 10 years.

Particles have many natural and anthropogenic sources; some are emitted directly from the sources (primary particles) and some are produced by atmospheric chemistry (secondary particles). Particles affect human health, visibility, scattering and absorption of light, and are essential for cloud formation, as will be seen in the next lesson on cloud physics.

Reminder - Complete all of the Lesson 4 tasks!

You have reached the end of Lesson 4! Double-check that you have completed all of the activities before you begin Lesson 5.

3.5.7: Summary and Final Tasks is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.6: Biogeochemical reactions in the troposphere

The planetary boundary layer, PBL, is the initial recipient of substances from ground sources. Because of the dynamic character of the atmosphere, material from the PBL will mix into the free troposphere and can be transported horizontally over great distances. Vertical transport across the PBL is significantly enhanced in the presence of deep convective cloud systems. These dynamic processes provide a transition from high mixing ratios of trace gases near source regions to low mixing ratios in the remote troposphere. These changes in mixing ratios result in changes in chemical reaction paths, thus justifying a chemical classification of the troposphere into the background troposphere, the source region, and a "transition" regime between the two. The troposphere is governed by heterogeneous chemistry far more so than the stratosphere. Heterogeneous processes of interest here involve scavenging of trace gases by aerosols, cloud and precipitation elements leading to aqueous phase chemical reactions and to temporary and permanent removal of material from the gas phase. They also include exchange of material with the earth's surface, such as dry deposition. Dry deposition is a major removal process for ozone, as well as for other gases of importance in tropospheric photochemistry (e.g., NO_2).

The Chemistry of Ozone

Ozone is formed by the association reaction of ground state O atoms with O_2 :

$$O\left(^3P\right) + O_2 + M \rightarrow O_3 + M \tag{3.6.1}$$

and is removed by photolysis

$$\begin{aligned} O_3 + h\nu &\rightarrow O\left(^1D\right) + O_2 \\ O_3 + h\nu &\rightarrow O\left(^3P\right) + O_2. \end{aligned} \tag{3.6.2}$$

The metastable $O\left('D\right)$ is quenched by O_2 and N_2

$$O\left(^1D\right) + M \rightarrow O\left(^3P\right) + M \tag{3.6.3}$$

though it can also react with H_2O to form the hydroxyl radical HO

$$O\left(^1D\right) + H_2O \rightarrow HO + HO \tag{3.6.4}$$

Nitric oxide reacts with ozone to form NO_2

$$NO + O_3 \rightarrow NO_2 + O_2 \tag{3.6.5}$$

with NO_2 removed by photolysis

$$NO_2 + h\nu \rightarrow O\left(^3P\right) + NO \tag{3.6.6}$$

Ozone removed by (6) is reconstituted by (7) followed by (1). It is convenient, because of the cyclic nature of the chemistry, to define a family of species undergoing rapid reactions leading to formation or removal of ozone. We define this family, odd oxygen, as the sum of $O_3, O\left(^1D\right), O\left(^3P\right)$ and NO_2.

Odd oxygen is conserved in reactions (1)-(4), (6) and (7). It is formed by reactions of NO with HO_2, CH_3O_2, and RO_2.

$$HO_2 + NO \rightarrow HO + NO_2 \tag{3.6.7}$$

$$\begin{aligned} CH_3O_2 + NO &\rightarrow CH_3O + NO_2 \\ RO_2 + NO &- RO + NO_2 \end{aligned} \tag{3.6.8}$$

The species RO_2 represents a variety of complex organic peroxy radicals. These reactions occur during the photooxidation of CO, CH_4 and hydrocarbons, for example by

$$\begin{gathered} HO + CO + O_2 \rightarrow HO_2 + CO_2 \\ HO_2 + NO \rightarrow NO_2 + HO \\ NO_2 + h\nu \rightarrow NO + O \\ O + O_2 + M \rightarrow O_3 + M \\ \text{NET} : CO + 2O_2 \rightarrow CO_2 + O_3 \end{gathered}$$

and by

$$\begin{gathered} HO + CH_4 \rightarrow CH_3 + H_2O \\ CH_3 + O_2 + M \rightarrow CH_3O_2 + M \\ CH_3O_2 + NO \rightarrow CH_3O + NO_2 \\ CH_3O + O_2 \rightarrow CH_2O + HO_2 \\ HO_2 + NO \rightarrow NO_2 + HO \\ 2 \times \left(NO_2 + h\nu \rightarrow NO + O\left(^3P\right)\right) \quad \lambda < 420 \text{ nm} \\ 2 \times \left(O\left(^3P\right) + O_2 + M \rightarrow O_3 + M\right) \\ \text{NET} : CH_4 + 4O_2 \rightarrow CH_2O + H_2O + 2O_3 \end{gathered}$$

Odd oxygen is removed by reaction of $O\left(^1D\right)$ with H_2O, by reaction of O_3 with HO_2 and HO,

$$\begin{gathered} HO_2 + O_3 \rightarrow OH + 2O_2 \\ HO + O_3 \rightarrow HO_2 + O_2 \end{gathered} \tag{3.6.9}$$

by formation of nitric acid from NO_2,

$$HO + NO_2 + M \rightarrow HNO_3 \tag{3.6.10}$$

and by heterogeneous reactions of O_3 and NO_2 at the earth's surface.

Oxidation of CO, CH_4 and hydrocarbons leads to net production of odd oxygen in the presence of adequate NO_x. The rate for production of ozone is roughly proportional to the concentration of NO, while the rate of loss is almost independent of $NO_x\ (NO + NO_2)$ for concentrations below 200 ppt (Fishman et al., 1979). Loss of odd oxygen, primarily by reactions (5) and (11) is balanced by production in reactions (8) and (9) for concentrations of NO near 30 ppt (Fishman et al., 1979; Logan et al., 1981; Crutzen, 1983). Hence, regions of the globe characterized by extremely low concentrations of NO, such as the remote Pacific, are likely to provide a net photochemical sink for odd oxygen (Liu et al., 1983), while the continental boundary layer at midlatitudes, characterized by higher concentrations of NO, is likely to provide a net source. Measurements of NO_x in the troposphere are few, and the lack of data for NO_x contributes significantly to uncertainties in global estimates for the photochemical source of ozone.
Production of ozone in the troposphere is limited ultimately by supply of CO, CH_4, and hydrocarbons, if NO_x is available. One molecule of ozone may be formed for each molecule of CO (Crutzen, 1973b), while the yield of ozone from oxidation of CH_4 could be as large as 3.5 (e.g., Logan et al., 1981). Production of ozone from nonmethane hydrocarbons is discussed in more detail later in this chapter.

The Chemistry of HO

Reaction with HO in the troposphere is the primary removal mechanism for many trace gases which influence the chemical composition and radiative balance of the stratosphere. Consequently, tropospheric and stratospheric chemistry are inextricably linked through the HO radical.

The hydroxyl radical is formed by reaction (2) followed by (5). It is removed by reaction with carbon monoxide and methane,

$$\begin{aligned} HO + CO + O_2 &\rightarrow HO_2 + CO_2 \\ HO + CH_4 + O_2 &\rightarrow CH_3O_2 + H_2O \end{aligned} \tag{3.6.11}$$

with reaction (14) dominant over (15). HO_2 is removed by reaction (8), and (16)

$$HO_2 + O_3 \rightarrow HO + 2O_2 \tag{3.6.12}$$

leading to regeneration of HO, or by

$$HO_2 + HO_2 \rightarrow H_2O_2 + O_2 \tag{3.6.13}$$

Hydrogen peroxide is photolysed,

$$H_2O_2 + h\nu \rightarrow HO + HO \tag{3.6.14}$$

it can react with HO,

$$HO + H_2O_2 \rightarrow H_2O + HO_2 \tag{3.6.15}$$

or may be removed by heterogeneous processes such as precipitation scavenging, as discussed below.

$$H_2O_2 + \text{ rain } \rightarrow \text{ removal.} \tag{3.6.16}$$

It is convenient, because of the fairly rapid interconversion of H, HO, HO_2 and H_2O_2 to consider these species as a family, odd hydrogen $(= H + HO + HO_2 + 2 \times H_2O_2)$. Reaction (5) provides the dominant source for odd hydrogen, while reactions (19) and (20) provide important sinks for odd hydrogen. Reactions (8), (14) and (16)-(18) do not influence the concentration of odd hydrogen.

Atmospheric oxidation of methane (Figure 2-1) may also provide sources and sinks for odd hydrogen, in addition to providing a source for odd oxygen. CH_3O_2, formed in (15), is removed primarily by reaction with NO or HO_2.

$$\begin{aligned} CH_3O_2 + NO &\rightarrow CH_3O + NO_2 \\ CH_3O_2 + HO_2 &\rightarrow CH_3OOH + O_2 \end{aligned} \tag{3.6.17}$$

Subsequent reaction of methyl hydroperoxide with HO leads to loss of odd hydrogen.

$$CH_3OOH + HO \rightarrow CH_3O_2 + H_2O \tag{3.6.18}$$

$$HO_2 + HO \rightarrow H_2O + O_2. \tag{3.6.19}$$

Recent measurements of a relatively fast rate for reaction (22) (Niki et al., 1983) indicate that this reaction may provide an important sink for odd hydrogen at low concentrations of NO_x (Logan, 1985, private communication).

The net source of odd hydrogen is determined first by competition between reaction (9) and reaction (21), and second by competition between reactions (22)-(23) which remove odd hydrogen,

$$CH_3OOH + \text{ rain } \rightarrow \text{ removal} \tag{3.6.20}$$

and reaction (24)

$$CH_3OOH + h\nu \rightarrow CH_3O + HO \tag{3.6.21}$$

which recycles odd hydrogen. Reaction of the methoxy radical with oxygen yields formaldehyde.

$$CH_3O + O_2 \rightarrow CH_2O + HO_2 \tag{3.6.22}$$

Photolysis of CH_2O

$$\begin{aligned} CH_2O + h\nu &\rightarrow H + HCO \\ CH_2O + h\nu &\rightarrow H_2 + CO \end{aligned} \tag{3.6.23}$$

via path (26a) provides an important source for odd hydrogen in the upper troposphere, while path (26b) and reaction of HO with CH_2O have no net effect on odd hydrogen.

The major pathways in the methane oxidation chain are thought to be understood fairly well. Recent kinetic data for some of the intermediate species have changed the net effect of the cycle on the budget of odd hydrogen. It should be emphasized that the influence of CH_4 chemistry on the budgets of both odd hydrogen and odd oxygen depends critically on ambient concentrations of NO_x, because of competition between reactions (9) and (21). A discussion of the chemistry of HO and associated uncertainties may be found in Chameides and Tan (1981), Logan et al. (1981) and NRC (1984). Logan et al. (1981) present a detailed discussion of the potential influence of CH_4 chemistry on odd hydrogen.

The Chemistry of Oxides of Nitrogen

Nitrogen oxides act as catalysts in the photochemical production of ozone. Any reaction which converts NO to NO_2, other than the reaction of NO with ozone, provides a photochemical source of ozone. Present measurements of NO_x are inadequate for definition of its distribution. Preliminary data indicate that ambient concentrations are highly variable in space and time, as discussed in Chapter 3. The lack of data for NO_x precludes accurate quantification of the net global chemical source for ozone.

Nitrogen oxides are produced in the troposphere primarily in the form of NO. Nitric oxide and nitrogen dioxide are rapidly interconverted by reactions (6), (7), (8) and (9) on a time scale of minutes. Nitrogen oxides are removed from the atmosphere by conversion to nitric acid,

$$HO + NO_2 + M \rightarrow HNO_3 \tag{3.6.24}$$

followed by heterogeneous processes, i.e., rainout or surface deposition of HNO_3. Surface deposition of NO_2 could provide another important tropospheric sink. There have been few field studies of the rate of uptake of NO_x by surfaces, as discussed later in this chapter, and removal rates of NO_x are not well defined at present.

Current models suggest that NO_x is converted to HNO_3 by reaction (13) within 1-2 days in summer. Nitric acid may be converted to aerosol nitrate, for example by reaction with sea-salt aerosol (Savoie and Prospero, 1982) or by reaction with ammonia (Tang, 1980). Nitric acid and aerosol nitrate are likely to be removed from the atmosphere by precipitation and surface deposition, with mean lifetimes of a few days (Junge, 1963; Levine and Schwartz, 1982). Nitric acid is converted back to NO_x by reactions (20) and (21) more slowly, with a time scale of 2-4 weeks.

$$\begin{aligned} HNO_3 + HO &\rightarrow H_2O + NO_3 \\ HNO_3 + h\nu &\rightarrow HO + NO_2 \end{aligned} \tag{3.6.25}$$

These processes are significant primarily in the upper troposphere.
Nitric oxide and NO_2 may be converted to $HNO_2, HO_2NO_2, NO_3, N_2O_5$ and organic nitrates in addition to HNO_3. Most of these molecules decompose thermally or photolytically and therefore provide temporary reservoirs for NO_x. Peroxynitric acid is formed by reaction of HO_2 with NO_2

$$HO_2 + NO_2 + M \rightarrow HO_2NO_2 + M. \tag{3.6.26}$$

It decomposes rapidly in the lower troposphere,

$$HO_2NO_2 \rightarrow HO_2 + NO_2 \tag{3.6.27}$$

but it is thermally stable in the colder upper troposphere, where it is removed by photolysis and by reaction with HO. Current models suggest that a significant fraction of acidic nitrate may be present in the form of HO_2NO_2 in the upper troposphere, but there are no observational data at present.

Peroxyacetyl nitrate (PAN) is formed during the degradation of hydrocarbons (see below). PAN is more stable than HO_2NO_2 with respect to thermal decomposition, with a lifetime of about a day at $275^{\circ}K$ and several years at temperatures characteristic of the upper troposphere (Cox and Coffey, 1977; Hendry and Kenley, 1979). Recent kinetic data indicate that the lifetime of PAN towards photolysis is about four months (Senum et al., 1984), with a similar value for attack by HO (Wallington *et* al., 1984). Recent measurements of PAN show that it is an important reservoir for NO_x in clean marine air (Singh and Salas, 1983a, b) and in rural air in North America (Bottenheim et al., 1984; Spicer et al., 1983) and in Europe (Brice et al., 1984).
The nitrate radical (NO_3) is formed by

$$NO_2 + O_3 \rightarrow NO_3 + O_2 \tag{3.6.28}$$

and is photolysed rapidly during the day. At night a steady state should be established between NO_3 and N_2O_5.

$$NO_3 + NO_2 + M \leftrightarrows NO_2O_5 + M. \tag{3.6.29}$$

The nitrate radical has been observed in a variety of "clean-air" locations including Hawaii and rural Colorado. In most cases the concentration of NO_3 at a given level of NO_2 was lower than predicted by model calculations (Noxon et al., 1980; Platt et al., 1980b, 1981, 1984; Platt and Perner, 1980; Noxon, 1983). Some processes, as yet unidentified, appear to be removing NO_3 and/or N_2O_5 from the atmosphere at night. If these processes lead to removal of NO_x, then nighttime removal of NO_x by reactions involving NO_3 could be comparable to daytime by reaction (13). Platt *et* al. $(1981, 1984)$ speculate that reactions of NO_3, or more likely N_2O_5, on wet aerosols may account for rapid removal of NO_3 under conditions of high humidity, and may provide another sink for NO_x in clouds and fog. A detailed discussion of NO_3 scavenging by cloud droplets, and its possible conversion to $NO_3^-$by chemical reactions in the liquid phase is given in the section on heterogeneous chemistry. The rate of the homogeneous gas-phase reaction between N_2O_5 and water vapor,

$$N_2O_5 + H_2O \rightarrow HNO_3 + HNO_3 \tag{3.6.30}$$

is extremely slow, but this reaction could provide an important source for HNO_3 if it proceeds at the upper limit $\left(k < 1.3 \times 10^{-21}\ cm^3\ \text{molecule}^{-1}\text{sec}^{-1}\right)$ given by Tuazon et al. (1983). The role of (33) in nighttime chemistry will remain uncertain until its rate constant is better defined.

CHEMISTRY OF OZONE FORMATION IN THE POLLUTED TROPOSPHERE

The discussion on clean tropospheric chemistry indicated that the oxidation of CO, CH_4, and possibly other biogenic hydrocarbons leads to a net production of odd oxygen, if sufficient NO_x is present (i.e., $NO_x > 30ppt$). In the polluted troposphere, unlike the clean troposphere, the rate of ozone formation is not necessarily proportional to the concentration of NO_x, but varies in a complex way that is dependent upon the ratios and concentrations of the hydrocarbons and NO_x as well as the chemical composition of the hydrocarbons themselves. The chemistry of ozone in the polluted troposphere has been studied for over thirty years. The research emphasis has been predominantly in elucidating the processes for oxidant/ozone formation in urban areas. The focus of interest in urban areas is due mainly to the high emission densities of ozone precursors (hydrocarbons and NO_x) which result typically in very high ozone concentrations under summertime meteorological conditions. In many cases the concentration levels of ozone in urban areas exceed ambient air quality standards which have been set to protect public health and welfare.
Several reviews of the chemistry of polluted atmospheres are available (Leighton, 1961; Stern, 1977; Seinfeld, 1975; Heiklen, 1976), as are detailed discussions of reaction mechanisms (Demerjian et al., 1974; Carter et al., 1979; Baldwin et al., 1977; Falls

and Seinfeld, 1978; Whitten et al., 1980) and reaction rate constant reviews (Baulch et al., 1980; Demerjian et al., 1980; Baulch et al., 1982). The discussions which follows provides an overview of the chemistry of the polluted troposphere.

The chemistry that occurs in sunlight-irradiated polluted atmospheres involves the interaction of a host of chemical species. These include: hydrocarbons such as alkanes, alkenes, and aromatics; other organics such as aldehydes and ketones; nitric oxide (NO); nitrogen dioxide (NO_2); ozone (O_3); peroxyacetyl nitrate (PAN); nitric acid (HNO_3); atomic oxygen (O^3P) and its first electronic excited state $O\,(^1D)$; hydroxy radical (HO); hydroperoxyl radical (HO_2); alkylperoxyl radicals (RO_2); acylperoxyl radicals $(R(O)O_2)$; nitrogen trioxide (NO_3); and dinitrogen pentoxide (N_2O_5).

This chemistry explains the rapid conversion of NO to NO_2 observed in the ambient polluted atmosphere (Leighton, 1961; Altshuller and Bufalini, 1965; Demerjian et al., 1974). The key lies in a sequence of reactions involving the same free radical species thought to be important in the chemistry of the clean troposphere and a number of organic radicals derived from the host of more complex hydrocarbon and organic molecules present additionally in the polluted atmosphere.

The most important atmospheric reactions governing the decomposition of alkanes, alkenes, and aromatics involve their reaction with hydroxyl radical. Reaction sequences describing the pathways for their oxidation as a result of HO attack have been studied extensively (Demerjian et al., 1974; Carter et al., 1979; Niki, 1978; Falls and Seinfeld, 1978; Perry et al., 1977; Atkinson et al., 1985; Grosjean, 1984). In a very simplified form this may be represented as follows:

$$HO + \text{ Hydrocarbon } \rightarrow R + H_2O \tag{3.6.31}$$

The alkyl radical (R) produced as a result of the HO attack on the hydrocarbon reacts with an oxygen molecule to form an alkylperoxyl radical (RO_2)

$$R + O_2 \rightarrow RO_2 \tag{3.6.32}$$

The alkylperoxyl radical reacts with NO to form NO_2 and an alkoxy radical (RO)

$$RO_2 + NO \rightarrow RO + NO_2 \tag{3.6.33}$$

Hydrogen abstraction from the alkoxy radical by molecular oxygen will produce a hydroperoxyl radical (HO_2) and a carbonyl compound $(R(C=O)H)$:

$$RO + O_2 \rightarrow R(C=O)H + HO_2 \tag{3.6.34}$$

The aldehydes formed in the RO oxidation, or emitted as such from combustion sources, react with HO or photolyze, introducing another important source of radicals to the atmosphere.

$$\begin{aligned} R(C=O)H + HO &\rightarrow R(C=O) + H_2O \\ R(C=O)H + h\nu(\lambda < 400 \text{ nm}) &\rightarrow R + H(C=O) \end{aligned} \tag{3.6.35}$$

The hydroperoxyl radical can then react with NO to form NO_2 and another hydroxyl radical, which reenters the cycle.

$$HO_2 + NO \rightarrow \quad HO + NO_2 \tag{3.6.36}$$

Reactions (8) and (10) taken together provide an efficient mechanism for the rapid conversion of NO to NO_2. This increase in the concentration of "odd oxygen" (previous section) results in an increase of ozone through the perturbation of the chemical steady state relationship (39) derived from the following reactions:

$$\begin{aligned} NO_2 + h\nu &\rightarrow NO + O\,(^3P) \\ O\,(^3P) + O_2 + M &\rightarrow O_3 + M \\ O_3 + NO &\rightarrow NO_2 + O_2 \end{aligned} \tag{3.6.37}$$

The chemical steady state relationship (39) has been shown to be a valid approximation over a considerable range of atmospheric

pollutant conditions (Stedman and Jackson, 1975; Calvert, 1976).

$$[O_3] = \frac{J_7\,[NO_2]}{k_6[NO]} \tag{3.6.38}$$

Ratios of NO_2 and NO emitted into the atmosphere by man's activities are typically less than one and would lead to very low ozone steady state concentrations under typical atmospheric solar irradiation conditions. The impact of reactions (8) and (10) is to drive the ratio of NO_2 to NO up by oxidizing the NO to NO_2, in direct competition with the ozone-nitric oxide reaction and thereby allowing ozone to build up in concentration.

Typical observations made in a polluted atmosphere of this phenomenon are shown in Figures 4-1a and *b*. The diurnal pattern begins with the emission of hydrocarbons, carbon monoxide and NO_x from motor vehicles in the early morning. The NO is converted to NO_2 and ozone accumulation begins when

most of the NO has been oxidized. The ozone concentration maximizes and then declines either as a result of reaction with additional emissions of NO through the day, dilution due to meteorology, or in the case of extended periods of time and transport, by interaction with the ground surface, which is a major sink for ozone in the atmosphere as will be discussed later in this chapter.

The interaction of organic free radicals produced by hydrocarbon oxidation with NO and NO_2 represents an important aspect of the chemistry of the oxides of nitrogen in the polluted atmosphere. They represent key processes in the conversion of NO to NO_2 and the formation of organic nitrates. It has recently been found that the more complex peroxyalkyl radicals formed during alkane photooxidation can add NO to form alkyl nitrates in non-negligible yields, in competition with reaction (10) (Atkinson et al., 1985):

$$RO_2 + NO \rightarrow RONO_2 \tag{3.6.39}$$

Also the hydroperoxyl radical formed from hydrocarbon oxidation in reaction (36) supplements that formed in reactions (12) and (14) to create hydrogen peroxide via reaction (17), and aldehydes react with the hydroxy radical to produce peroxyacyl radicals and ultimately peroxyacyl nitrates by reaction with NO_2

$$\begin{gathered} RCHO + OH \rightarrow RCO + H_2O \\ RCO + O_2 \rightarrow R(C=O)O_2 \\ R(C=O)O_2 + NO_2 \rightarrow R(C=O)O_2NO_2 \end{gathered} \tag{3.6.40}$$

A thermal equilibrium exists between the peroxyacyl nitrate and its components

$$R(C=O)O_2NO_2 \leftrightarrows R(C=O)O_2 + NO_2 \tag{3.6.41}$$

and the peroxyacyl radical is removed from the atmosphere by reaction with NO (Cox and Coffey 1977)

$$R(C=O)O_2 + NO \rightarrow R(C=O)O + NO_2 \tag{3.6.42}$$

The most commonly occurring peroxyacyl nitrate is peroxyacetyl nitrate or PAN and there is a great similarity between the behavior of PAN and ozone in polluted atmospheres. Their diurnal variations at ground based sites are virtually identical on many occasions (Garland and Penkett 1976) primarily because their formation and removal mechanisms are very similar. PAN is thus an excellent tracer of the photochemical reactivity of the troposphere and since it has no large stratospheric source it can be used in studies designed to investigate the seasonal behavior of tropospheric photochemical activity (Brice et al., 1984).

4.2.2 Night-time Chemistry of the Source Region

The nitrate radical NO_3 can be formed by reaction of NO_2 with ozone

$$NO_2 + O_3 \rightarrow NO_3 + O_2 \tag{3.6.43}$$

During the day NO_3 is rapidly photolyzed back to its precursors but during the night an equilibrium with N_2O_5 can be established

$$NO_3 + NO_2 = N_2O_5 \tag{3.6.44}$$

Night-time concentrations of NO_3 as high as $355 pptv$ have been recorded in the Los Angeles basin (Platt et al., 1980b, 1984). Some NO_3 is converted to HNO_3 by reactions with hydrocarbons and aldehydes,

$$\begin{aligned} NO_3 + RH &\rightarrow R + HNO_3 \\ NO_3 + R(C=O)H &\rightarrow R(C=O) + HNO_3 \end{aligned} \tag{3.6.45}$$

but in the absence of NO no chain reactions can be initiated by the peroxy radicals, which will be rapidly lost. The N_2O_5 which is formed will be stable until sunrise or it can be removed either by hydrolysis in deliquescent aerosols or by uptake on ground surfaces.

It is quite possible that heterogeneous chemistry is important in the overall NO_x cycle particularly in the formation of nitrous acid, HONO, which can act as an efficient source of HO radicals in conditions of low light intensity, but this is an area where much more research is required before definite statements can be made.

Excerpted from:

V.A. Mohnen, W. Chameides K.L. Demerjian D. H. Lenschow J.A. Logan R.J. McNeal, S.A. Penkett, U. Platt, U. Schurath, P. da Silva Dias. Tropospheric Chemistry, in *Atmospheric Ozone 1985: Assessment of Our Understanding of the Processes Controlling Its Present Distribution and Change, World Meteorological Organization.* Accessed November 2023 at https://csl.noaa.gov/assessments/ozone/1985/report.html

3.6: Biogeochemical reactions in the troposphere is shared under a Public Domain license and was authored, remixed, and/or curated by NOAA Tropospheric Chemistry https://csl.noaa.gov/assessments/ozone/1985/vol1/chapter4.pdf.

3.7: Atmospheric deposition

Heterogeneous chemical reactions are normally defined as ones which occur at the interface between two phases, e.g., gas-liquid, gas-solid, or liquid-solid. In the atmosphere processes are considered to be heterogeneous as long as the overall reaction involves two phases even though specific steps in the reaction may be homogeneous. An example of this is the oxidation of SO_2 in cloudwater, i.e. the dissolution of gaseous SO_2 into a cloud droplet followed by the oxidation of aquated SO_2 by homogeneous aqueousphase reactions. Because SO_2 is transferred from the gas to the aqueous phase, this process is considered to be heterogeneous even though the actual oxidative reaction is homogeneous.

Interactions with Aerosols and Particulates

The presence of particles in the atmosphere raises the possibility that atmospheric gaseous species interact with these particles heterogeneously. In most cases this interaction results in a gas-to-particle conversion; i.e. the transfer of a chemical species from the gas phase to an aerosol or liquid droplet or the formation of a new particle from a gaseous species. Typically gas-to-particle conversion processes are classified into three categories (Kiang et al., 1973; Schryer, 1982):
1) Homogeneous, homomolecular nucleation (the formation of a new stable liquid or solid ultrafine particle from a gas involving one gaseous species only);
2) Homogeneous, heteromolecular nucleation (the formation of a new particle from two or more gaseous species);
3) Heterogeneous, hetermolecular condensation (the growth of pre-existing particles due to deposition of molecules from the gas phase).

Of the above processes, the third, sometimes referred to as "aerosol scavenging", probably has the greatest impact upon the gas-phase levels of nitrogen oxides and hydrogen oxides and thus on the atmospheric abundances of ozone and HO (Turco et al., 1982). Scavenging reactions that are most relevant in this regard are those involving HNO_3, NO_3, N_2O_5, organic nitrates $(RONO_2), H_2O_2, HO_2$, and organic peroxides $(ROOH)$. These reactions can be represented by

(Rla) $HNO_3 \rightarrow$ aerosol
(Rlb) $NO_3 \rightarrow$ aerosol
(R1c) $N_2O_5 \rightarrow$ aerosol
(R1d) $RONO_2 \rightarrow$ aerosol
(Rle) $H_2O_2 \rightarrow$ aerosol
(R1f) $HO_2 \rightarrow$ aerosol
(R1g) $ROOH \rightarrow$ aerosol

A complete understanding of scavenging reactions such as those listed above requires a thorough description of all the processes occurring at the surface interface and thus a knowledge of the aerosol surface structure and composition and its interaction with the relevant reactants and products. At the present time we have not yet developed this understanding; to do so will require the continued design and application of sophisticated experimental techniques used in conjunction with advanced theoretical studies of surface interactions. As a result present-day descriptions of heterogeneous reactions such as aerosol scavenging are quite rudimentary and involve the use of simple parameterizations to treat many of the complex molecular processes that occur but are not yet understood. One such simple description is described below.
The rate of scavenging of a gaseous species J is normally assumed to be proportional to the species ambient concentration, n_J, so that

\[\text{Rate of aerosol scavenging} =\mathrm{n}_{\mathrm{J}} \mathrm{k}_{\text {part }}\]

where the proportionality constant, k_{part}, has units of s^{-1}. The parameter k_{part} can be represented by (Chameides and Davis, 1982; Heikes and Thompson, 1983),

$$k_{part} = \int \phi_J(r) n_p(r) dr \tag{3.7.1}$$

where $n_p(r)$ is the concentration of aerosol particles having radii between r and $r + dr$ and $\phi_J(r)$ is the rate at which species **J** diffuses and sticks to an aerosol particle of radius r.$\phi_J(r)$ can be represented by (Schwartz, 1983)

$$\phi_J(r) = \frac{4}{3}\pi \ell r V_j \left(1 + \frac{4\ell}{3r\alpha}\right) \frac{[n_J - n_J^o]}{n_J} \tag{3.7.2}$$

where ℓ is the mean free path, V_J is the species thermal velocity, α is the appropriate sticking or accommodation coefficient for species J impinging upon the aerosol surface, and n_J^o is the species' concentration at the surface of the aerosol. Typically, n_J^o is assumed to be zero for solid particles and $(a_J/H_J RT)$ for wet particles, where a_J is the activity of J in solution, H_J is the species solubility constant in the appropriate aerosol solution, R is the gas constant, and T is the temperature.

While appearing to be quite straightforward, the application of Equations (1), (2), (3) to determine the rate of aerosol scavenging of a given species J is quite difficult. For one the value of α is highly uncertain. The little data that does exist suggest that α can vary widely depending upon the aerosol composition (i.e. basic or acidic), relative humidity (i.e. for hygroscopic aerosols, solid particles at low humidity and solution droplets at high humidity), and the nature of the impinging gas. For instance while sticking coefficients for species such as HO_2, H_2O_2, NO_3 impinging upon water solutions with pH 's above about 5 may approach unity, the α for these species impinging upon H_2SO_4 particles or similar highly acidic dry particles are likely of the order of 10^{-6} to 10^{-4} (cf. Chameides and Davis, 1982).

Another complication arises for hygroscopic aerosol particles which, provided the humidity is high enough, exist as small suspended solution droplets of relatively high ionic strength (i.e. $\mu > 1$). At these ionic strengths, the particle can no longer be treated as an ideal solution as one does for lower ionic strengths. The appropriate values for a_J and H_J are generally not known and can only be accurately determined by detailed experimental studies such as those carried out by Pytkowicz (1984) for seawater.

In spite of these uncertainties, Equation (1), (2), and (3) can be used to obtain a preliminary indication of how important these scavenging reactions may be for a range of possible values for uncertain parameters such as α. As an example, values for k_{part} for NO_3 and N_2O_5 as a function of α and the total number density of aerosol particles (assuming a log-normal distribution) are illustrated in Figure 4-3 for the case when $n_J^O = 0$. (Very similar scavenging coefficients would be obtained for other species such as HNO_3, HO_2, and H_2O_2). The results of Figure 4-3 suggest that for species with sufficiently long lifetimes (i.e. of the order of days or more), aerosol scavenging would represent a significant sink in regions of modestto-large aerosol loadings if $\alpha \geq 10^{-2}$. Indeed the large levels of NO_3^-found in aerosols in the marine boundary layer would appear to confirm that scavenging of gaseous HNO_3 by sea salt aerosol represents a major sink for this species in the marine atmosphere (Huebert and Lazrus, 1980; Liu et al., 1983). At night when NO_3 photochemical lifetimes become large, scavenging by aerosols could also be a significant sink for NO_3; in fact this mechanism has been proposed to explain the low levels of NO_3 observed at night (Noxon et al., 1978; Platt et al., 1980b; Heickes and Thompson, 1983). It is also interesting to note that in urban locations where aerosol number densities of $10^4 - 10^5$ cm^{-3} are not uncommon, aerosol scavenging could also be an important sink for HO_2, whose photochemical lifetime is about 100 s, if $\alpha =$ values near 1 are appropriate.

Thus these calculations imply that aerosol scavenging can be an important sink for HNO_3 and NO_3, as well as H_2O_2 and possibly HO_2. As such this process would represent a major heterogeneous sink for nitrogen oxides and hydrogen oxides. However, many uncertainties are associated with this assessment. Before a more definitive assessment of aerosol scavenging can be made, experimental studies establishing accurate values for α as well as field measurements characterizing the chemical composition and structure of aerosol surfaces will be needed.

In addition to aerosol scavenging another potentially important heterogeneous sink of nitrogen oxides that needs to be considered is the formation of NH_4NO_3 from gaseous NH_3 and HNO_3, i.e.

$$(R2) \quad (NH_3)_g + (HNO_3)_g - (NH_4NO_3)_g \tag{3.7.3}$$

In regions of high NH_3 levels this process can represent a significant heterogeneous sink for nitrogen oxides. Studies by Stelson and Seinfeld (1982) have indicated that the levels of NH_4NO_3 in the particulate phase relative to that of gas-phase NH_3 and HNO_3 can be accurately described in terms of the thermodynamic equilibrium between the phases. While the condensation of NH_4NO_3 can occur via heteromolecular, homogeneous condensation as well as hetermolecular, heterogeneous condensation, observations indicate that the later process is the dominant one in the atmosphere.

Scavenging

Similar to the interactions of gases with aerosol particles, scavenging or heterogeneous removal is probably the most common process by which hydrometeors affect gas-phase species. This can occur via rainout (removal of gases by cloud droplets) as well as washout (removal of gases by raindrops). Of these two it is generally believed that rainout is more important than washout simply because of the longer lifetime and greater surface area afforded by cloud drops when compared to that of raindrops. With regard to nitrogen oxides and hydrogen oxides, removal of HNO_3, NO_3, H_2O_2, and HO_2 are most relevant because of their high solubility or reactivity in clouds.

The rate of incorporation of a gaseous species J into a cloud droplet can be treated in much the same manner as that for aerosol particles described in the previous section. Thus
Rate of cloud droplet scavenging $= \mathrm{n_J k_{cloud}}$
where

$$\mathrm{k_{cloud}} = \int \phi_J(\mathrm{r}) \mathrm{n_d}(\mathrm{r}) \mathrm{dr} \tag{3.7.4}$$

and n_d is the number of cloud droplets with radii between r and $r + dr.\phi_J j(r)$ in Equation (5) is essentially the same as that given by Equation (3), except that now $n_J{}^\circ = [J]/HRT$, where $[J]$ is the concentration of J in solution in units of moles (liter) $)^{-1}$. (Because ionic strengths in cloudwater are small, the solution can be treated as ideal so that $\mathrm{a_J} = [\mathrm{J}]$ and $\mathrm{H_J}$ is given by the Henry Law constant for J in a pure water solution.) Values for α for cloud drops, while not accurately known, are probably not nearly as uncertain or variable as those for aerosol particles since the water droplet surface is better defined. (On the other hand, the reader should note that the presence of organic films on droplets could further complicate the issue as discussed by Gill et al., 1983). Sticking coefficients for most species of interest impinging on water are probably of the order of 10^{-2} or higher although values as low as 10^{-4} cannot be ruled out at this time (cf., Chameides and Davis, 1982).

Within a cloud, gaseous species are transferred to droplets until equilibrium is attained. Calculations simulating the transfer of gases to droplets indicates that, for reasonable values of α, equilibrium between the two phases is rapidly attained. This fact is illustrated in Table 4-2 where equilibration times, τ_{eq}, are listed for several species of interest as a function of α and r. Because equilibrium does apply the concentration of a species J in solution can be related to its ambient gas-phase concentration in cloud-free regions via

$$[\mathrm{J}] \cong [\mathrm{J}]^{\mathrm{eq}} = \mathrm{n_J} / \left(\mathrm{ALx10^{-9}} + (\mathrm{HRT})^{-1}\right) \tag{3.7.5}$$

where $[\mathrm{J}]^{\mathrm{eq}}$ is the concentration of J in the cloudwater at equilibrium, A is Avogadro's number, and L is the liquid water content in units of $\mathrm{gm^{-3}}$.

Given Equation (6) as well as statistics for the average cloudiness of the atmosphere, the rate of precipitation from clouds, and the liquid water content of clouds, an expression for the rate of removal of a species via rainout can be obtained. The results of such an analysis are found in Figure 4-4a, where the rainout

lifetimes, τ_{J}, are illustrated as a function of z for different values of H and varying types of storm cycles. For species of low solubility, τ_{J} is linearly related to H but independent of the length of time between storms. For species of high solubility which are virtually completely removed from the atmosphere during each storm, τ_{J} is independent of H but linearly related to the time between storms; in this case $\tau_{\mathrm{J}} = \mathrm{T}$ where T is the storm period.

Thus in regions of long dry periods between storms, the rainout lifetimes for highly soluble species can become significantly longer than the average 10 day lifetime for water vapor that has been previously assumed to be representative of the lifetime of highly soluble species. It can be seen in Figure 4-4b that for nitrogen oxide species strikingly different profiles can be obtained in model calculations using this parameter for different assumed storm periods even though identical source strengths were assumed in each case. Thus, given the large seasonal and latitudinal variations in the water vapor-cycle, this result would seem to imply that in addition to $\mathrm{NO_y}$ sources, $\mathrm{NO_y}$ removal rates can also lead to large temporal and spatial variations in $\mathrm{NO_y}$ concentrations.

In addition to providing a liquid phase where soluble gases can be dissolved, clouds also offer an active chemical medium where aqueous-phase chemical reactions can occur which affect the removal rate of atmospheric species in general and nitrogen oxide and

hydrogen oxide species in particular. For instance while the distribution of gaseous HNO_3

$$(R3)\ \ (HNO_3)_g - (H^+)_{aq} + (NO_3^-)_{aq} \tag{3.7.6}$$

provides a major rainout sink for atmospheric nitrogen oxides and source of dissolved NO_3- in rainwater, another important sink can arise in regions where $NO_x\,(= NO + NO_2)$ is large. This sink occurs via the production of gaseous NO_3 radicals in clouds at night. Because NO_3 radicals have long chemical lifetimes at night and can be rapidly scavenged by cloud droplets, their night time production in a cloud is followed by their incorporation into cloudwater. Once in the aqueous-phase, NO_3 is rapidly converted to $NO_3{}^-$. A typical reaction sequence leading to NO_3^-production in cloudwater is

$$(R4)\ \ (NO_2)_g + (O_3)_g \rightarrow (NO_3)_g + (O_2)_g \tag{3.7.7}$$

(R5) $(NO_3)_g \rightarrow (NO_3)_{aq}$

$$(R6)\ \ (NO_3)_{aq} + (Cl^-)_{aq} \rightarrow (NO_3^-)_{aq} + (Cl)_{aq} \tag{3.7.8}$$

Figure 4-5 taken from a numerical simulation of the coupled gas- and aqueous-phase chemistry of a cloud, illustrates the enhanced levels of $NO_3{}^-$that can result from this process at night.

SURFACE EXCHANGE AND VERTICAL REDISTRIBUTION

Surface Exchange

Vertical transport in the boundary layer (i.e., the layer of air that is coupled directly to the surface by turbulent exchange processes on a time scale of about an hour or less), and between the boundary layer and the overlying free atmosphere, plays an important role in determining the fate of trace atmospheric species. At the surface, trace species may be both deposited and emitted, so that the flux at the surface is a net balance between the two processes. Deposition at the surface requires transport of the trace species to the surface by turbulent eddies.

Redistribution between the boundary layer and the overlying free atmosphere is mostly a result of highly intermittent cloud processes. Cycling of trace species through clouds can result in chemical transformations and loss through wet deposition (see section 4). The latent heat released by clouds can result in several kilometer ascents of boundary layer air into the free atmosphere within a few minutes. Compensating downward motions can transport air from the free atmosphere into the boundary layer. Eddy diffusivity models are not adequate for describing mixing of trace species over such large vertical displacements. Although some progress has been made in parameterizing these processes, much still remains to be done.
The subsequent sections discuss our current understanding of surface exchange of trace species important in estimating the tropospheric ozone budget, including both dry deposition and surface emission, as well as the exchange of trace species between the boundary layer and the overlying free atmosphere. Both are important in estimating the ozone budget throughout the troposphere.

Dry deposition is the transfer of an atmospheric constituent from the air directly to the earth's surface, regardless of whether the surface is wet or dry, or whether the loss is at the ground or in elements that are attached to, or touching the ground, such as a forest canopy or a snow cover. Because of the complexity of the earth's surface, and the variety of mechanisms for capturing species, precise physical descriptions of how this process takes place are difficult. Instead, the rate of dry deposition is commonly specified with a gross parameter such as a deposition velocity. The deposition velocity is defined as the ratio of the downward flux F_S of a species s to its mean concentration at some reference level s

$$v_d = F_s/\overline{s} \tag{3.7.9}$$

Trace constituents are transferred from the atmospheric boundary layer to within a centimeter of the surface (or surface protrusions) by turbulent eddies. Below this, the transfer is predominantly by molecular diffusion. We can, therefore, define the downward flux through a particular level as the averaged product of instantaneous departures of the constituent concentration s' and vertical

velocity w' from their means, measured at that level; i.e., $F_s = -\overline{w's'}$, where the overbar denotes an average over a distance or time long enough to ensure a statistically reliable result.

Dry deposition is determined by the efficiency of the turbulence transport process, and the properties of both the constituent and the surface. The constituent may be either adsorbed or absorbed at the surface; that is, the constituent may either be stored on the surface without changing its identity, or it may combine chemically with the surface material. The deposition rate for constituents which are efficiently absorbed at the surface is controlled mainly by the ability of the turbulent eddies to transport the constituents to impinging constituents stick to the surface. As discussed by Wesely (1983) and Wesely and Hicks (1977), a convenient way to express this mathematically is to separate the effects of turbulence transport from constituent and surface properties by considering the reciprocal of the deposition velocity as a "resistance" to deposition, and equating it to the sum of the individual resistances contributed by (1) the turbulence transport process, commonly called the aerodynamic resistance r_a; (2) that contributed by molecular transport through the quasilaminar layer within about a centimeter of the surface and surface elements r_s; and (3) that contributed by how well the constituent sticks to the particular surface r_c :

$$v_d^{-1} = r_a + r_s + r_c \tag{3.7.10}$$

In this way, properties of the constituent and surface can, in principle, be separated from aerodynamic properties.

Many species are not only deposited, but are also emitted at the surface. Therefore, in these cases, the surface flux is the net result of both deposition and emission, and the concept of a surface resistance is no longer applicable for estimating surface flux, except when emission is negligible compared to deposition. Nitrogen oxides $(NO_x = NO + NO_2)$ for example, are emitted from vegetated surfaces. Depending upon the situation, either deposition or emission may dominate; thus the surface flux may be either positive or negative. Their emissions are mostly the result of the activity of various soil microbes, whose rates of production are governed mainly by soil composition, temperature, moisture and pH, available oxygen, and fertilization practices. Over water, nitrite photolysis is believed to be a source of NO emission (Zafiriou and McFarland, 1981). Surface emissions of O_3 and HNO_3 are considered to be negligible, so that the resistance concept can be used to estimate their surface flux.

The actual process of deposition depends upon many factors. Over vegetated surfaces, the condition of the vegetation can be important, particularly for reactive gas species. Ozone deposition, for example, is much larger when the leaf stomata are open than when they are closed. Thus, ozone deposition is maximized over vegetated areas in the daytime when plants are actively growing. Observations over the ocean indicate a deposition rate several times that predicted on the basis of laboratory studies over pure water. This indicates that trace constituents in the water and on the water surface are important in determining the actual deposition rate.

Excerpted from:

V.A. Mohnen, W. Chameides K.L. Demerjian D. H. Lenschow J.A. Logan R.J. McNeal, S.A. Penkett, U. Platt, U. Schurath, P. da Silva Dias. Tropospheric Chemistry, in *Atmospheric Ozone 1985: Assessment of Our Understanding of the Processes Controlling Its Present Distribution and Change*, *World Meteodrological Organization*. Accessed November 2023 at https://csl.noaa.gov/assessments/ozone/1985/report.html

3.7: Atmospheric deposition is shared under a Public Domain license and was authored, remixed, and/or curated by NOAA https://csl.noaa.gov/assessments/ozone/1985/report.html.

3.7.1: Acid Deposition

Acid deposition is a term referring to a mixture of wet and dry deposition (deposited material) from the atmosphere containing higher than normal amounts of nitric and sulfuric acids. The precursors, or chemical forerunners, of acid deposition formation result from both natural sources, such as volcanoes and decaying vegetation, and man-made sources, primarily emissions of sulfur dioxide (SO_2) and nitrogen oxides (NO_x) resulting from fossil fuel combustion (figure 3.7.1.1). Acid deposition occurs when these gases react in the atmosphere with water, oxygen, and other chemicals to form various acidic compounds. The result is a mild solution of sulfuric acid and nitric acid. When sulfur dioxide and nitrogen oxides are released from power plants and other sources, prevailing winds blow these compounds across state and national borders, sometimes over hundreds of miles.

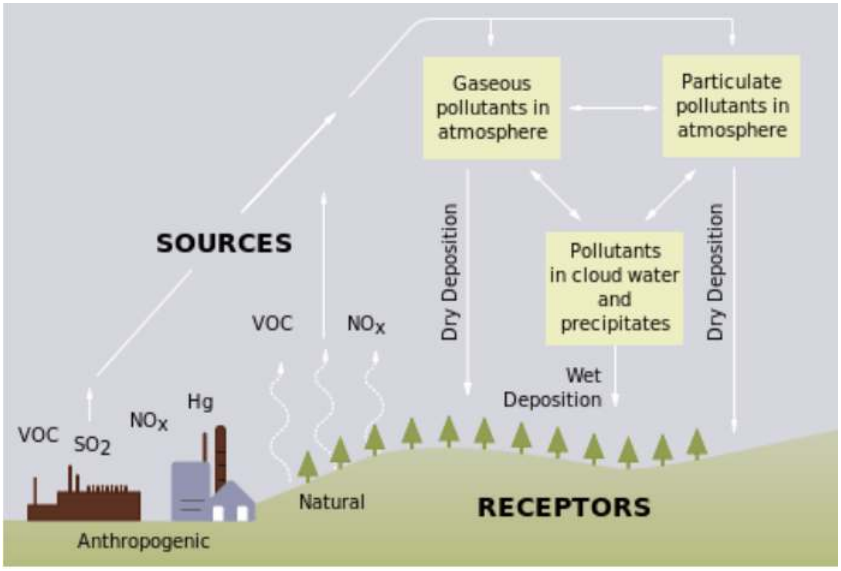


Figure 3.7.1.1**:** This diagram shows both anthropogenic and natural sources of acid deposition. Volatile organic compounds (VOCs), sulfur dioxide (SO_2), nitrogen oxides (NO_X), and mercury (Hg) are released, but only SO_2 and NO_X react to form acids. Three categories of pollutants interact with each other. These are gaseous pollutants in the atmosphere, particulate pollutants in the atmosphere, and pollutants in cloud water and precipitates. The former two categories form dry deposition, and the latter forms wet deposition. Acid deposition (both dry and wet deposition) reaches natural areas, which could be terrestrial (as pictured) or aquatic.

Acid deposition is measured using a scale called “pH.” The lower a substance’s pH, the more acidic it is. Pure water has a pH of 7.0. However, normal rain is slightly acidic because carbon dioxide (CO_2) dissolves into it forming weak carbonic acid, giving the resulting mixture a pH of approximately 5.6 at typical atmospheric concentrations of CO_2. The average pH of precipitation in the United States in 1980 was 4.6, but it has been gradually increasing due to reductions in sulfur dioxide and nitrogen oxide emissions (figure 3.7.1.2).

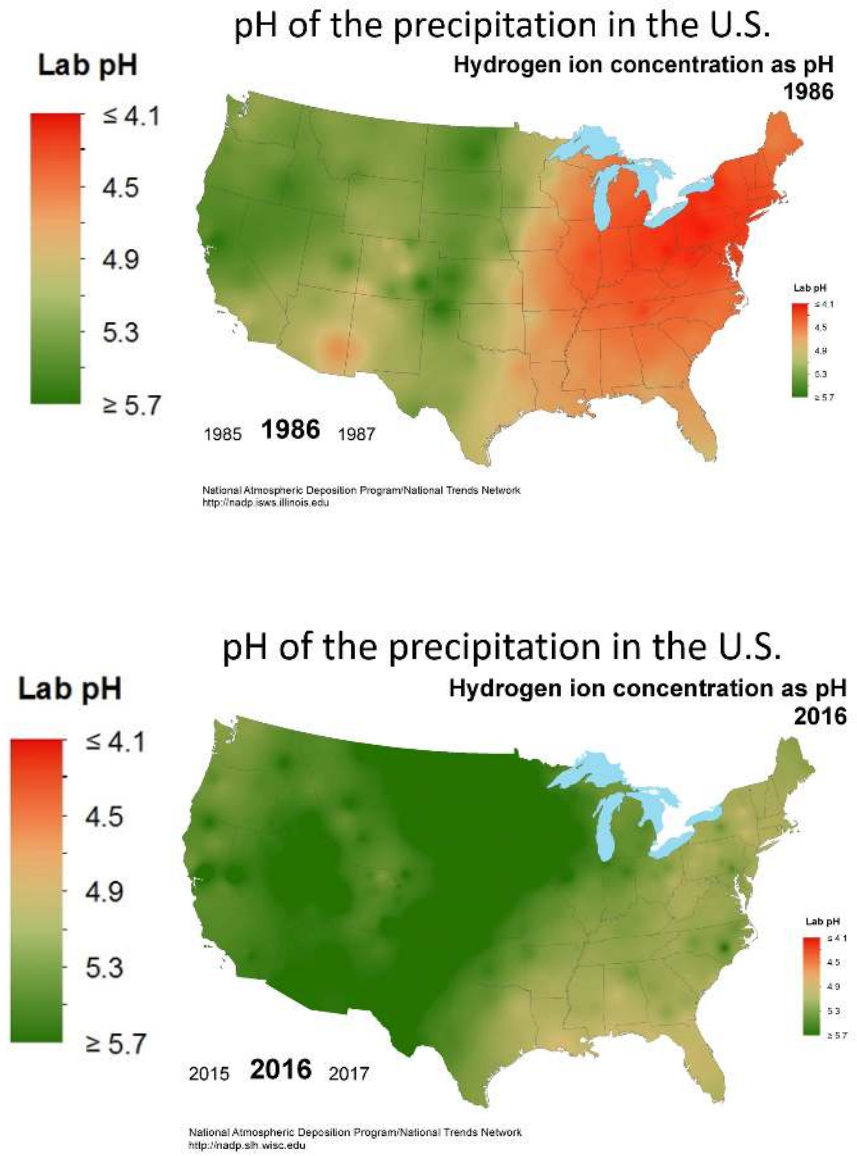


Figure 3.7.1.2**:** pH measures how acidic or basic a solution is, which depends on the concentration of hydrogen ions. Precipitation is normally slightly acidic, with a pH of around 5.6. (Anything below pH = 7 is considered acidic.) pH of precipitation was even more acidic in much of the East Coast of the U.S. in 1986 (acid deposition) with parts of Michigan, Ohio, Pennsylvania, and New York state reaching a pH of less than or equal to 4.1. The pH of the western United States in 1986 was closer to 5.0. Due to efforts to reduce air pollution, the pH of precipitation has moved closer to normal. By 2016, the pH of the eastern United States had increased to around 4.9, and the pH of the western United States had mostly returned to normal, with many regions having a pH of greater than or equal to 5.7. Images by National Atmospheric Deposition Program (public domain).

Effects of Acid Deposition

Acid deposition causes **acidification** of lakes and streams and contributes to the damage of trees at high elevations (for example, red spruce trees above 2,000 feet) and many sensitive forest soils. In addition, acid deposition accelerates the decay of building materials and paints, including irreplaceable buildings, statues, and sculptures that are part of our nation's cultural heritage. Prior to falling to the earth, sulfur dioxide (SO_2) and nitrogen oxide (NO_x) gases and their particulate matter derivatives—sulfates and nitrates—contribute to visibility degradation and harm public health.

The **ecological** effects of acid deposition are most clearly seen in the aquatic, or water, environments, such as streams, lakes, and marshes. Most lakes and streams have a pH between 6 and 8, although some lakes are naturally acidic even without the effects of acid deposition. Acid deposition primarily affects sensitive bodies of water, which are located in watersheds whose soils have a limited ability to neutralize acidic compounds (called “buffering capacity”). Lakes and streams become acidic (i.e., the pH value goes down) when the water itself and its surrounding soil cannot buffer the acid deposition enough to neutralize it. In areas where buffering capacity is low, acid deposition releases aluminum from soils into lakes and streams; aluminum is highly toxic to many species of aquatic organisms. Acid deposition causes slower growth, injury, or death of **forests**. Of course, acid deposition is not the only cause of such conditions. Other factors contribute to the overall stress of these areas, including air pollutants, insects, disease, drought, or very cold weather. In most cases, in fact, the impacts of acid deposition on trees are due to the combined effects of acid deposition and these other environmental stressors.

Acid deposition contributes to the corrosion of **metals** (such as bronze) and the deterioration of paint and stone (such as marble and limestone). These effects significantly reduce the societal value of buildings, bridges, cultural objects (such as statues, monuments, and tombstones), and cars (figure 3.7.1.3).



Figure 3.7.1.3**:** A gargoyle that has been damaged by acid deposition.

Sulfates and nitrates that form in the atmosphere from sulfur dioxide (SO_2) and nitrogen oxides (NO_x) emissions contribute to **visibility impairment,** meaning we cannot see as far or as clearly through the air. The pollutants that cause acid deposition—sulfur dioxide (SO_2) and nitrogen oxides (NO_x)—damage **human health**. These gases interact in the atmosphere to form fine sulfate and nitrate particles that can be transported long distances by winds and inhaled deep into people’s lungs. Fine particles can also penetrate indoors. Many scientific studies have identified a relationship between elevated levels of fine particles and increased illness and premature death from heart and lung disorders, such as asthma and bronchitis.

Contributors and Attribution

Modified by Kyle Whittinghill and Melissa Ha from the following sources

- Acid Rain from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.

This page titled 3.7.1: Acid Deposition is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **10.3: Acid Rain** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

3.8: Biogeochemical reactions in the stratosphere

Ozone is present in the earth's atmosphere at all altitudes from the surface up to at least $100\ \mathrm{km}$. The bulk of the ozone resides in the stratosphere with a maximum ozone concentration of 5×10^{12} molecule cm^{-3} at about $25\ \mathrm{km}$. In the mesosphere $(>60\ \mathrm{km})\mathrm{O}_3$ densities are quite low and are not discussed in the present report. Although O_3 concentrations in the troposphere are also less than in the stratosphere, ozone plays a vital role in the atmospheric chemistry in this region and also affects the thermal radiation balance in the lower atmosphere.

Atmospheric ozone is formed by combination of atomic and molecular oxygen.

$$\mathrm{O} + \mathrm{O}_2 + \mathrm{M} \rightarrow \mathrm{O}_3 + \mathrm{M} \tag{3.8.1}$$

where M is a third body required to carry away the energy released in the combination reaction. At altitudes above approximately $20\ \mathrm{km}$ production of O atoms results almost exclusively from photodissociation of molecular O_2 by short wavelength ultraviolet radiation $(\lambda < 243\ \mathrm{nm})$:

$$\mathrm{O}_2 + \mathrm{h}\nu \rightarrow \mathrm{O} + \mathrm{O} \tag{3.8.2}$$

At lower altitudes and particularly in the troposphere, O atom formation from the photodissociation of nitrogen dioxide by long wavelength ultraviolet radiation is more important:

$$\mathrm{NO}_2 + \mathrm{h}\nu \rightarrow \mathrm{NO} + \mathrm{O} \tag{3.8.3}$$

Ozone itself is photodissociated by both UV and visible light:

$$\mathrm{O}_3 + \mathrm{h}\nu \rightarrow \mathrm{O}_2 + \mathrm{O} \tag{3.8.4}$$

but this reaction together with the combination reaction (1) only serves to partition the 'odd oxygen' species between $\mathrm{O}_{\text{and}}\ \mathrm{O}_3$. The production processes (2) and (3) are balanced by chemical and physical loss processes. Until the $1950\ \mathrm{s}$, chemical loss of odd oxygen was attributed only to the reaction:

$$\mathrm{O} + \mathrm{O}_3 \rightarrow \mathrm{O}_2 + \mathrm{O}_2 \tag{3.8.5}$$

originally proposed by S. Chapman (1930). It is now known that ozone in the stratosphere is removed predominantly by catalytic cycles involving homogeneous gas phase reactions of active free radical species in the $\mathrm{HO}_x, \mathrm{NO}_x, \mathrm{ClO}_x$ and BrO_x families:

$$\begin{array}{l} \mathrm{X} + \mathrm{O}_3 \rightarrow \mathrm{XO} + \mathrm{O}_2 \\ \mathrm{XO} + \mathrm{O} \rightarrow \mathrm{X} + \mathrm{O}_2 \\ \hline \text{net:} \quad \mathrm{O} + \mathrm{O}_3 \rightarrow 2\mathrm{O}_2 \end{array} \tag{3.8.6}$$

where the catalyst $\mathrm{X} = \mathrm{H}, \mathrm{OH}, \mathrm{NO}, \mathrm{Cl}$ and Br. Thus these species can, with varying degrees of efficiency, control the abundance and distribution of ozone in the stratosphere. Assignment of the relative importance and the prediction of the future impact of these catalytic species is dependent on a detailed understanding of the chemical reactions which form, remove and interconvert the active components of each family.

Physical loss of ozone from the stratosphere is mainly by dynamical transport to the troposphere where further photochemically driven sources and sinks modify the ozone concentration field. Ozone is destroyed at the surface of the earth and so there is an overall downward flux in the lower part of the atmosphere. Physical removal of ozone and other trace gaseous components can also occur in the precipitation elements and on the surface of atmospheric aerosols. Since most of the precursor and sink molecules for the species catalytically active in ozone removal in the stratosphere are derived from or removed in the troposphere, global tropospheric chemistry is a significant feature of overall atmospheric ozone behavior.

HO_x Chemistry

There have been relatively few changes recently in the kinetics data base for HO_x. The principal catalytic cycle for odd-oxygen destruction within the HO_x family is:

(I)

$$\text{net:} \quad \frac{O + HO_2 \rightarrow OH + O_2}{O + O_3 \rightarrow 2O_2} \tag{3.8.7}$$

$$\begin{matrix} OH + O_3 \rightarrow HO_2 + O_2 \\ \frac{O+HO_2 \rightarrow OH+O_2}{O+O_3 \rightarrow 2O_2} \end{matrix} \tag{3.8.8}$$

Depending on altitude the following cycles may also become important:

(II)

$$\begin{matrix} HO_2 + O_3 \rightarrow OH + 2O_2 \\ \text{net:} \\ \text{III} \\ 2O_3 \rightarrow 3O_2 \\ O + OH \rightarrow H + O_2 \\ H + O_2 + M \rightarrow HO_2 + M \\ O + HO_2 \rightarrow OH + O_2 \\ \text{net:} \\ 2O \rightarrow O_2 \end{matrix} \tag{3.8.9}$$

2.1.3 NO_x Chemistry

Odd nitrogen species are important in the stratosphere because they are involved in catalytic cycles which directly destroy O_3,

$$\begin{matrix} NO + O_3 \rightarrow NO_2 + O_2 \\ NO_2 + O \rightarrow NO + O_2 \\ \hline O + O_3 \rightarrow 2O_2 \end{matrix} \tag{3.8.10}$$

and

$$\begin{matrix} NO + O_3 \rightarrow NO_2 + O_2 \\ NO_2 + O_3 \rightarrow NO_3 + O_2 \\ \text{net: } NO_3 + h\nu \rightarrow NO + O_2 \\ \hline O_3 \rightarrow 3O_2 \end{matrix} \tag{3.8.11}$$

The first of these two cycles is much more important than the second. Even though all the above reactions have been studied in the laboratory, there exist some uncertainties in the values for the rate coefficient for reaction (22) at stratospheric temperatures and the quantum yield for NO_{in} NO_3 photolysis. In addition to their involvement in direct O_3 destruction, NO_x species play crucial roles in the partitioning of odd hydrogen and odd chlorine into various forms. The rates of conversion of HO_2 to OH and ClO to Cl are determined by the reactions involving NO,

$$\begin{matrix} HO_2 + NO \rightarrow OH + NO_2 \\ ClO + NO \rightarrow Cl + NO_2 \end{matrix} \tag{3.8.12}$$

and those involving $O\,(^3P)$, i.e., $O + HO_2 \rightarrow OH + O_2$ and $O + ClO \rightarrow Cl + O_2$. Thus, these reactions in conjunction with reactions

of OH and Cl with O_3, control the ratios $[HO_2]/[OH]$ and $[ClO]/[Cl]$. Both reactions 26 and 27 are well characterized. NO_x species are also involved in sequestering HO_x species in temporary reservoirs e.g.:

and those involving $O(^3P)$, i.e., $O + HO_2 \rightarrow OH + O_2$ and $O + ClO \rightarrow Cl + O_2$. Thus, these reactions in conjunction with reactions of OH and Cl with O_3, control the ratios $[HO_2]/[OH]$ and $[ClO]/[Cl]$. Both reactions 26 and 27 are well characterized. NO_x species are also involved in sequestering HO_x species in temporary reservoirs e.g.:

$$\begin{aligned} OH + NO_2 &\overset{M}{=} HNO_3 \\ HO_2 + NO_2 &\overset{M}{\rightarrow} HO_2NO_2 \end{aligned} \tag{3.8.13}$$

The above processes have been thoroughly investigated and their rate coefficients are quite well established. The photolysis of NO_2, reaction (3), serves as the major source of odd oxygen in the troposphere. The absorption cross section for NO_2 and the quantum yield for O atom production are still somewhat uncertain, as are their temperature dependences.

In addition to the above mentioned reactions, the majority of reactions involving NO_x that are important in understanding stratospheric chemistry are well characterized. In the following section, we will discuss only the problem areas and areas where significant new data have been recently reported.

N_2O is the major source of NO_x in the stratosphere. The predominant path for N_2O destruction is photolysis. Its reaction with $O(^1D)$ contributes only 2% to N_2O destruction but is currently assumed to be the main NO_x production mechanism. Therefore, the possibility of N_2O photolysis to give $NO + N$ needs to be very carefully assessed. Even if such a pathway constitutes only 1% of the total N_2O photolysis rate, it could be equal to the $O(^1D) + N_2O$ source [for each N_2O photolyzed to give $NO + N$, one more molecule of NO is produced due to the reaction of N with O_2 or O_3].

The majority of $O(^1D)$ produced by O_3 is physically deactivated to $O(^3P)$. The thermal rate coefficients for the reaction/deactivation of $O(^1D)$ by atmospheric gases $N_2, O_2, O_3, CO_2, Ar, N_2O, H_2O$ and CH_4 are well defined (NASA evaluation). However, the yield of NO due to the $O(^1D) + N_2O$ reaction in the stratosphere is uncertain by as much as 30%. This uncertainty is partly due to the combined errors in the measured values of all the rate coefficients for $O(^1D)$ removal reactions, and is partly due to the possibility that the branching ratio of $O(^1D) + N_2O$ reaction to yield NO (as opposed to N_2 and O_2) changes with the kinetic energy of $O(^1D)$. Since $O(^1D)$ produced by ozone photolysis is translationally hot, and since the $O(^1D) + N_2$ and $O(^1D) + O_2$ quenching rates are temperature dependent, the uncertainty of the atmospheric rate of $O(^1D) + N_2O \rightarrow 2NO$ reaction branch is further enhanced if translationally hot $O(^1D)$ reacts differently than thermal $O(^1D)$. Therefore, experiments designed to measure NO production under stratospheric conditions which do not rely on the accuracy of the individual reaction rates need to be carried out.

Currently, all stratospheric N_2O is assumed to be produced at the ground level and transported into the stratosphcre. However, local production of N_2O due to reactions such as $N_2(A^3\Sigma) + O_2$ and $OH(A^2\Pi) + N_2$ cannot be ruled out (Zipf, Prasad 1982). If such reactions occur in the mesosphere they could influence the stratospheric NO_x budget by downward transport of N_2O.

The main known process which removes NO_x from the stratosphere is transport of long lived species such as HNO_3, but a small amount of NO_x loss occurs through the $N + NO$ and $N + NO_2$ reactions in the upper stratosphere. The latter reaction may produce N_2O as a major product. Kinetic data for reaction of N with NO are reasonably well established, but the rate constant for reaction with NO_2 is only reliable to within a factor of 3 at room temperature and its temperature dependence has not been established.

Removal of odd-hydrogen in the lower stratosphere occurs mainly by the reaction of OH with nitric acid and peroxynitric acid:

$$\begin{aligned} OH + HNO_3 &\rightarrow H_2O + NO_3 \\ OH + HO_2NO_2 &\rightarrow H_2O + NO_2 + O_2 \end{aligned} \tag{3.8.14}$$

Changes in the recommended rate coefficients for these reactions have previously resulted in significant revisions of the calculated ozone column. The existence of a negative temperature dependence for the $OH + HNO_3$ reaction is now well established and confirmation of the small pressure dependence may help explain some of the divergence between results of the kinetics studies in different laboratory systems (NASA, 1985). The equally important $OH + HO_2NO_2$ reaction is not as well characterized, either with regard to the temperature dependence or the reaction products. New data have been reported recently for the temperature and pressure dependence of the HO_2NO_2 formation reaction (Sander and Peterson, 1984):

$$HO_2 + NO_2 + M \rightarrow HO_2NO_2 + M \tag{3.8.15}$$

The rate constant for stratospheric conditions is about 40% lower than previously recommended. The products and temperature dependence of the photodissociation of HO_2NO_2 are still not established and the equilibrium constant for HO_2NO_2 formation is not reliably known. These gaps in the data base lead to some uncertainty in the description of peroxynitric acid behavior in the lower stratosphere and the troposphere.

The possibility of formation of an isomer of nitric acid in the recombination reaction of OH with NO_2 reaction (28) has also been considered. Such an isomer, if more reactive than $HONO_2$, would serve to reduce the effective rate of nitric acid formation. To date no firm evidence has been found for a complication of this kind in the kinetics of the $OH + NO_2$ reaction. NO_2 and NO have been made. The reliability of the data base for these reactions is now greatly improved.

$$\begin{aligned} NO_3 + NO &\rightarrow 2NO_2 \\ NO_3 + NO_2 + M &\rightarrow N_2O_5 + M \end{aligned} \tag{3.8.16}$$

The equilibrium constant for the formation in the latter reaction of the important temporary reservoir species N_2O_5 has also been measured directly in several studies, but there remains some uncertainty in this quantity.

2.1.4 ClO_x Chemistry

The principal odd oxygen destruction cycle involving ClO_x is:

$$\begin{array}{r} Cl + O_3 \rightarrow ClO + O_2 \\ \underline{O + ClO \rightarrow Cl + O_2} \\ O + O_3 \rightarrow 2O_2 \end{array} \tag{3.8.17}$$

However, in large parts of the stratosphere, the conversion of ClO to Cl occurs mainly by coupling with NO_x :

$$NO + ClO \rightarrow NO_2 + Cl \tag{3.8.18}$$

In this case the sequence: reaction (34) followed by reaction (27) does not destroy odd oxygen, because NO_2 is rapidly photolyzed, reaction (3).

The main sink of active chlorine species is the reaction

$$Cl + CH_4 \rightarrow HCl + CH_3 \tag{3.8.19}$$

Excerpted from

V.A. Mohnen, W. Chameides K.L. Demerjian D. H. Lenschow J.A. Logan R.J. McNeal, S.A. Penkett, U. Platt, U. Schurath, P. da Silva Dias. Tropospheric Chemistry, in *Atmospheric Ozone 1985: Assessment of Our Understanding of the Processes Controlling Its Present Distribution and Change*, *World Meteorological Organization*. Accessed November 2023 at https://csl.noaa.gov/assessments/ozone/1985/report.html

3.8: Biogeochemical reactions in the stratosphere is shared under a Public Domain license and was authored, remixed, and/or curated by NOAA https://csl.noaa.gov/assessments/ozone/1985/report.html.

3.9: Models

Climate models are tools used in climate research. They are attempts to synthesize our theoretical and empirical knowledge of the climate system in a computer code. This chapter describes how climate models are constructed, and how they are evaluated, and it discusses some applications.

a) Construction

Climate models solve budget equations numerically on a computer. The equations are based on the **conservation of energy, momentum, and mass** (air, water, carbon, and other relevant elements, substances, and tracers). Typically they are solved in separate **boxes** that represent specific regions of Earth's climate system components (Figure 3.9.1). Along their boundaries the boxes exchange energy, momentum and mass. Exchange with the flow of water or air from one box to another is called advection. Prognostic variables such as temperature, specific humidity in the atmosphere, or salinity in the ocean, and three velocity components (zonal, meridional, and vertical) are calculated in each box. The momentum equations, which are used to calculate the velocities, are based on Newton's laws of motion and they include effects of the rotating Earth such as the Coriolis force. The temperature equations are based on the laws of thermodynamics. Thus, climate models represent the fundamental laws of physics as applied to Earth's climate system.

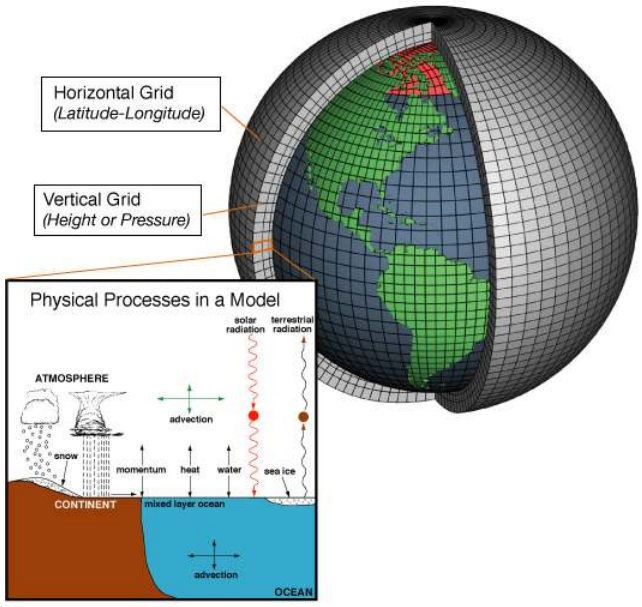


Figure 3.9.1: Schematic of a three-dimensional coupled climate model. From Wikipedia.org.

The evolution of the prognostic variables in the interior boxes are solved one **time step** at a time (see equation B4.1.4). After the prognostic variables have been updated the fluxes between boxes (I and O) are calculated, which are used for the next time step. Then the prognostic variables are updated again using the fluxes, and so on. This procedure is called forward modeling because all model variables at the next time step are calculated only from the model variables at the previous time step and the **boundary conditions** without the use of observations. Boundary conditions such as the incident solar radiation at the top-of-the-atmosphere or concentrations of greenhouse gases are usually required as input to the calculations. These are also called radiative forcings. Other boundary conditions need to be applied at the lower boundary: the topography (mountains) and bathymetry (sea floor). To start a forward model simulation **initial conditions** need also to be provided. Those can be taken from observations or idealized distributions.

Climate models range from the simplest zero-dimensional (0D) Energy Balance Model (EBM) discussed in the theory chapter to the most complex three-dimensional General Circulation Models (GCMs). The range of models ordered with respect to complexity is called the **hierarchy** of climate models. The 0D-EBM can be expanded by solving the energy budget equation separately at different latitudinal bands. This is called the one-dimensional EBM (1D-EBM). The 1D-EBM is still vertically averaged but it includes energy exchange between latitudinal bands as discussed in the discussion of processes. Meridional energy transport in

1D-EBMs is typically treated as a diffusive process proportional to the temperature gradient, such that heat flows from warm to cold regions. 1D-EBMs typically treat the ocean and atmosphere as one box, so that one cannot distinguish between heat transport in the ocean and atmosphere.

One-dimensional models are also used for vertical energy transfer in the atmosphere. These are called radiative-convective models and they work similar to the models used to produce Fig. 4.6 in chapter 4. Radiative-convective models themselves range in complexity from line-by-line models, which calculate radiative transfer in the atmosphere at individual wavelengths, to models that average over a range of frequencies (band models). Line-by-line models are computationally expensive and cannot be used in three-dimensional GCMs, which use band models. Band models are calibrated and tested by comparison to line-by-line models. As discussed in Chapter 4, radiative fluxes alone would cause a much warmer surface and much colder upper troposphere than observed. Therefore, radiative-convective models include convection, mostly by limiting the lapse rate to the observed or moist adiabatic rate.

Intermediate complexity models consist of 2D-EBMs, which are still vertically averaged but include zonal transport of energy and moisture (in this case they are called Energy-Moisture-Balance Models or EMBMs), zonally-averaged ocean models coupled to a 1D-EBM, and zonally averaged dynamical atmospheric models (resolving the Hadley circulation). Intermediate complexity models also often include biogeochemistry and land ice components. They are computationally relatively inexpensive and can be run for many thousands or even millions of years.

The first climate models developed in the 1960s were 1D-EBMs and simple GCMs of the **atmosphere** and **ocean** at very coarse resolution. Initially ocean and atmospheric models were developed separately and only later they were coupled. Coupling involves the exchange of heat and water fluxes and momentum at the surface. Current state-of-the-science coupled, three-dimensional GCMs also include **sea ice** and **land surface** processes such as snow cover, soil moisture and runoff of water through river drainage basins into the ocean. Many models also include dynamic **vegetation** with separate plant functional types such as trees and grasses. However, most current coupled GCMs that are used for future projections do not include interactive **ice sheet** components. This is because ice sheets have long equilibration (response) times of tens of thousands of years and therefore they need to be run for a much longer time than the other climate system components, which is currently not possible for most climate models.

The deep ocean has an equilibration time of about a thousand years. For future projections people are mostly interested in the next hundred or perhaps a few hundreds of years. For the most reliable and detailed projections on these timescales global climate modeling groups try to configure their models at the finest possible **resolution**. Currently the typical resolution is a few degrees (~200 km) in the horizontal directions and 20 to 30 vertical levels each in the atmosphere and ocean components. Finer resolution global models are being developed at various climate modeling centers around the world, but currently most climate projections of centennial and longer time scales are based on coarser resolution models (Fig. 3.9.2).



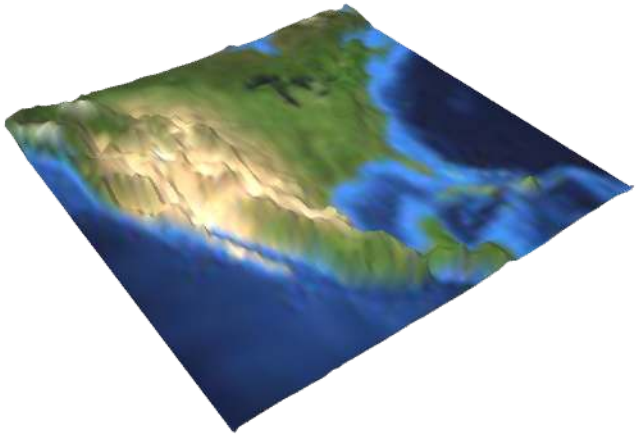


Figure 3.9.2: Illustrations of land surfaces (green and brown colors) and sea floor (blue) at fine (top; ~10 km) and coarse (botton; ~100 km) resolution. Note that the elevations and depths are strongly exaggerated with respect to horizontal distances. From ucar.edu.

Ice sheet models have finer spatial resolution (10s of kilometers) in order to resolve the narrow and steep ice sheet margin, but they have much larger time steps (1 year) than atmospheric (seconds) and ocean (minutes to hours) models because of the slow ice velocities (10-100 m/yr) compared to velocities of ocean currents (1-10 cm/s) or winds (1-10 m/s). The time step in a model depends on the velocity of the fluid and the grid-box size. The higher the velocity and the smaller the grid-box size the smaller the time step has to be to guarantee numerical stability. Finer resolution models therefore have to use smaller time steps, which is an additional burden on the computational resources. Another obstacle to move to higher resolution is the amount of data that accumulate. The highest resolution ocean model simulations currently require petabytes (10^{15} bytes = 1000 terrabytes) of storage for all the model output. Processing these huge amounts of data is a challenge.

One way to avoid increasing computational resources at higher resolution is to construct **regional climate models** for a specific region of interest, e.g. North America. However, the disadvantage of regional climate models is that boundary conditions at the margins of the model domain have to be prescribed. The resulting solution in the interior depends strongly on those boundary conditions, which often are taken from global climate models. Therefore, any bias from the global climate model at the boundary would be propagated by the regional model into the interior of the model domain. However, in the interior the regional climate model can account for details e.g. of the topography, that a global model cannot. Thus, although not a silver bullet, regional climate models are useful for simulating climate in more spatial detail than possible with global models.

Typically the resolution and grids of the atmospheric and ocean components are different. Therefore, surface fluxes and variables needed to calculate the fluxes need to be mapped from one grid to the other. This is often accomplished by a coupler, which is software that does interpolation, extrapolation, and conservative mapping. All calculations need to be numerically sound such that energy, water, and other properties are conserved. However, numerical schemes, e.g. for the transport from one box to the next, are associated with errors and artifacts.

Models that include **biogeochemistry**, such as the **carbon cycle**, and/or ice sheets are called **Earth System Models**. Earth System Models calculate atmospheric CO_2 concentrations interactively based on changes in land and ocean carbon stocks. They can be forced directly with emissions of anthropogenic carbon, whereas models without carbon cycles need to be forced with prescribed atmospheric CO_2 concentrations.

Due to the limited resolution of the models, processes at spatial scales below the grid box size cannot be directly simulated. For example, individual clouds or convective updrafts in the atmosphere are often only a few tens or hundred meters in size and therefore cannot be resolved in global atmospheric models. Similarly, turbulence and eddies in the ocean, which are important for

the transport of heat and other properties, cannot be resolved by global ocean models. These processes must be parameterized. A **parameterization** is a mathematical description of the process that depends on the resolved variables, e.g. the mean temperature in the grid box, and one or more parameters. A simple example of a parameterization is the meridional heat flux $F_m = -K\partial T/\partial y$ in a 1D-EBM, which can be parameterized as a diffusive process, where K>0 is the diffusivity, $\partial T/\partial y$ is the meridional temperature gradient and y represents latitude. This parameterization transports heat down-gradient (note the minus sign), which means from warmer to colder regions. In this case the parameter K can be determined from observations of meridional heat flux and $\partial T/\partial y$. Parameterizations can be derived from empirical relationships based on detailed measurements or high-resolution model results. The parameter values are usually not precisely known but they influence the results of the climate model. Therefore, parameterizations are a source of error and uncertainty in climate models. Parameters in the cloud parameterization of a model, for example, will impact its cloud feedback and therefore its climate sensitivity.

b) Evaluation

Climate models are evaluated by comparing their output to observations. Figure 3.9.3 shows the multi model mean (the average of all models) from the most recent IPCC report (Flato et al., 2013). The model simulated surface temperature distribution is similar to the observations with warm (20-30°C) temperatures in the tropics and cold (<0°C) temperatures near the poles and at high altitudes (Himalayas). The models also reproduce some of the observed zonal gradients such as the cooler temperatures in the eastern equatorial Pacific compared to the western Pacific warm pool and the warmer temperatures in the northeast Atlantic compared to the northwest Atlantic, which are caused by the upper ocean circulation. However, the models are not perfect as indicated by biases such as too cold temperatures in the northern North Atlantic and too warm temperatures in the southeast Atlantic and Pacific. The warm biases in the southeast Atlantic and Pacific are most likely caused by the coarse resolution ocean models that do not resolve well the narrow upwelling in these regions. A similar bias is seen in the California current in the northeast Pacific. Despite these biases the multi model mean agrees with the observed temperatures to within plus/minus one degree Celsius in most regions. Even the larger regional biases such as those mentioned above are relatively small compared to the ~60°C range of temperature variations on Earth. This indicates that the models reproduce observed surface temperatures relatively well.

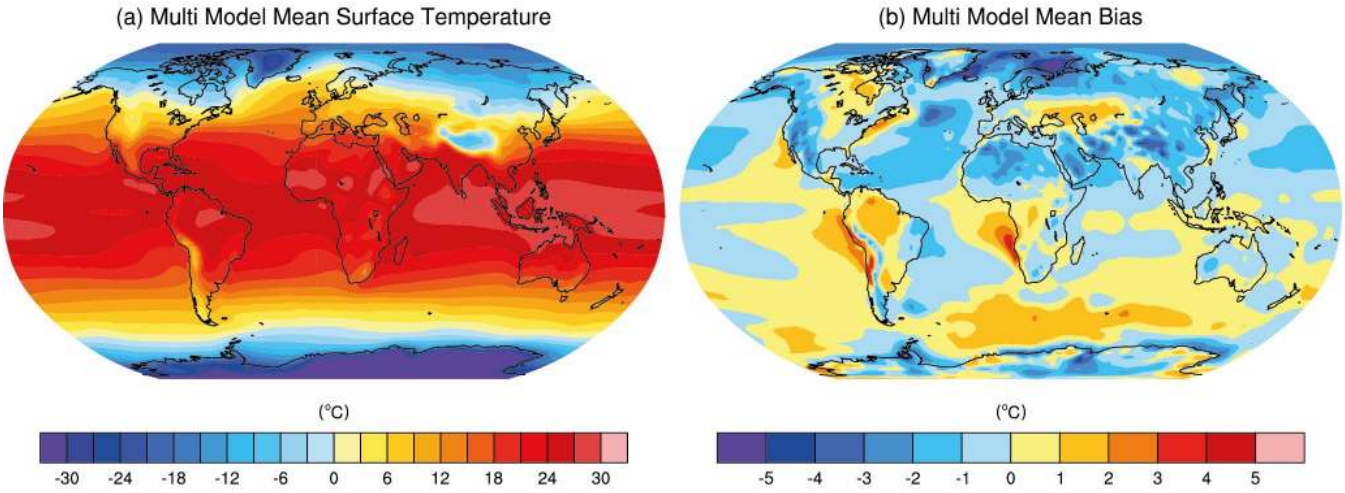


Figure 3.9.3: Annual mean surface air temperature distribution from 1980-2005 as simulated by climate models (a) and their bias (b). The bias $b = T_m - T_o$ is the difference between the model simulated temperature T_m and that estimated from observations T_o. From ipcc.ch.

Fig. 3.9.4 shows the multi model mean precipitation. The models reproduce the general pattern of the observations (Chapter 6, Fig. 6.8) such as more precipitation in the tropics and at mid-latitudes compared with less precipitation in the subtropics and at the poles. They also reproduce some of the observed zonal differences such as dryer conditions over the eastern parts of the subtropical ocean basins compared with wetter conditions further west. However, the models also display systematic biases such as the double Intertropical Convergence Zone (ITCZ) over the East Pacific and too dry conditions over the Amazon. The relative errors in precipitation are generally larger than those for temperature. This indicates that the models are better in simulating temperature than

precipitation. This may not be surprising given that the simulation of precipitation depends strongly on parameterized processes such as convection and clouds

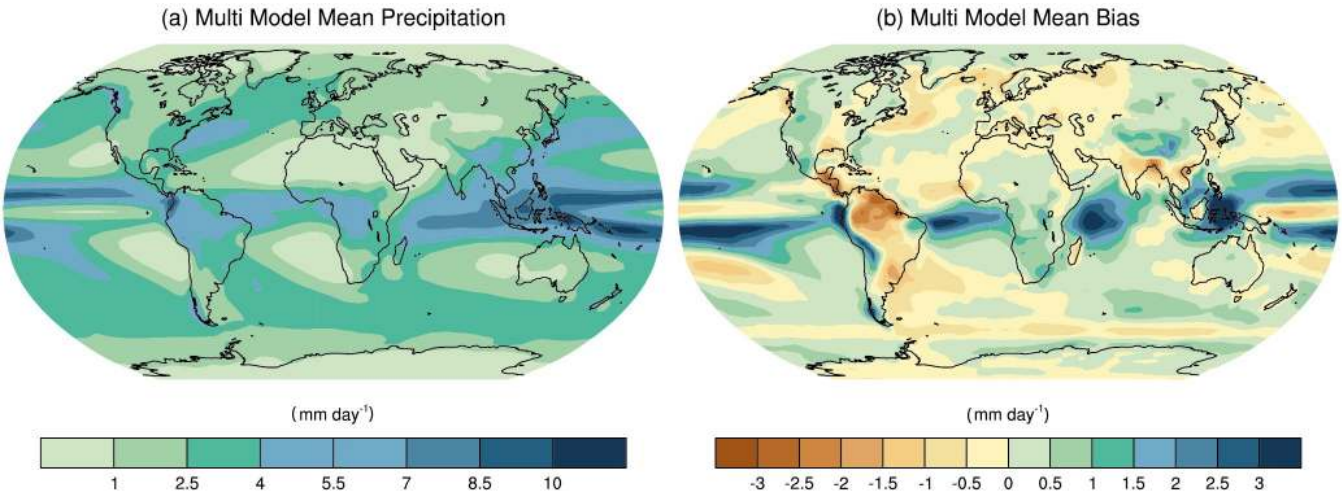


Figure 3.9.4: As Figure 3.9.3 but for annual mean precipitation.

Fig. 3.9.5 compares correlation coefficients of different variables. It confirms our previous conclusion that the models are better in simulating temperature than precipitation. It also shows that the models have very good skill in simulating Emitted Terrestrial Radiation at the top-of-the-atmosphere, whereas they are less good at simulating clouds. The current generation of climate models (CMIP5) are improved compared with the previous generation (CMIP3) particularly for precipitation. Another interesting feature also apparent in Fig. 3.9.5 is that the the multi model mean is almost always in better agreement with the observations than any one particular model. A similar phenomenon, which has been called the wisdom of the crowd, has been noted by Sir Francis Galton (1907), who analyzed villager's guesses for the weight of an ox at an English livestock fair. He found that many guesses where too high or too low, but the mean of all guesses was almost exactly the correct weight of the animal.

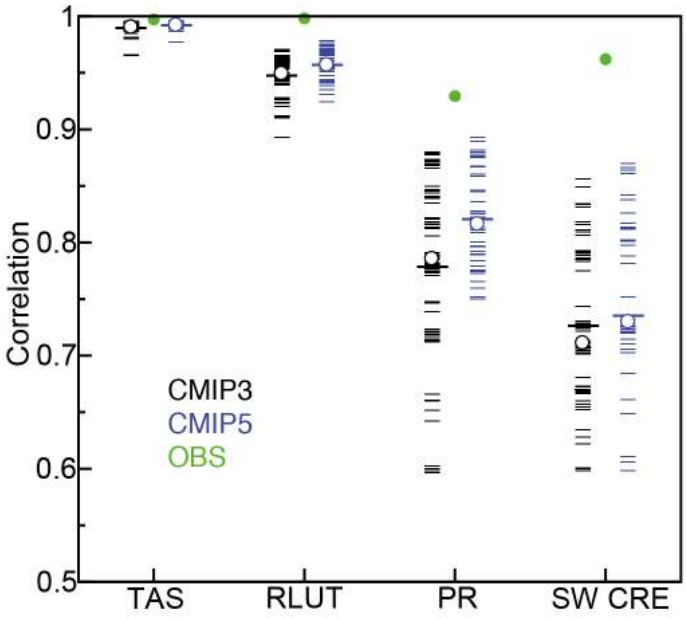


Figure 3.9.5: Correlation coefficient for the annual mean surface air temperature (TAS), outgoing longwave radiation (RLUT; in chapter 6 we called this variable emitted terrestrial radiation from the atmosphere: ETR_a), precipitation (PR), and the shortwave cloud radiative effect (SW CRE). The correlation coefficient $-1 \leq r \leq 1$ is a measure of pattern agreement. See this Wikipedia.org page for a detailed definition of r. If two variables are unrelated r = 0, whereas if they are perfectly in linear agreement r = 1. Each short horizontal line is one model. The longer line is the multi model mean and the circles are the median. Results are shown for the current generation of climate models from the Climate Modeling Intercomparison Project 5 (CMIP5) and the previous generation (CMIP3). Green dots show correlation coefficients for alternative observational datasets indicating uncertainties in the observations. They represent the best the models could do given uncertainties in the observations. From ipcc.ch.

Fig. 3.9.6 shows that most models overestimate temperatures in the thermocline by about 1°C, presumably due to too much vertical diffusion. Again these errors are relatively small given the large range (~20°C) of deep ocean temperature variations. Model errors in salinity are larger near the surface than in the deep ocean. In the southern hemisphere subtropics near surface waters are too fresh, perhaps related to the double ITCZ bias and associated too wet conditions in the atmosphere there (Fig. 3.9.5).

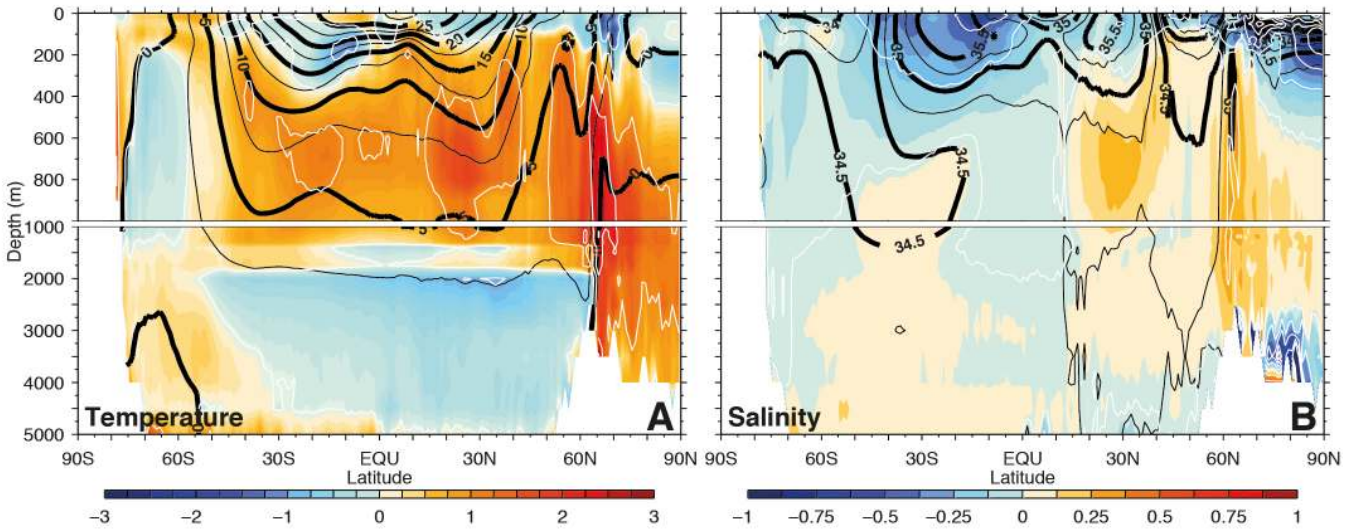


Figure 3.9.6: Zonally averaged differences in temperature (A) and salinity (B) between the multi model mean and observations (color contours and white lines). Black lines indicate observed temperature and salinity from the World Ocean Atlas 2009. From ipcc.ch.

All climate models simulate an increasing ocean heat content over the last 40 years, consistent with observations. Some models simulate more and others less heat uptake. The multi model mean, however, is in good agreement with the observations. This indicates that the models are skillful not only in simulating the mean state of ocean temperatures but also its recent temporal evolution.

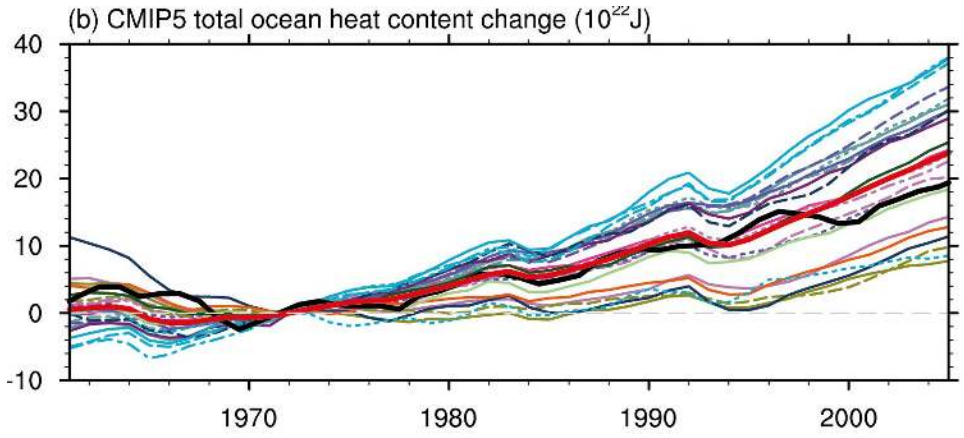


Figure 3.9.7: Changes in ocean heat content simulated by the CMIP5 models (thin lines), and their multi model mean (red thick line), compared to observations (black thick line). From ipcc.ch.

Historical and paleoclimate variations are also used to test and evaluate climate models as will be discussed next.

c) Applications

Some of the main applications of climate models are paleoclimate studies, detection and attribution studies, and future projections. Future projections will be discussed in the next chapter.

Paleoclimate model studies are not only useful for a better understanding of past climate changes and their impacts but they can also be used to evaluate the models. For example, model simulations of the Last Glacial Maximum (LGM) are broadly consistent with temperature reconstructions (see Chapter 3) that show global cooling of 4-5°C, polar amplification and larger changes over land than over the ocean. The models reproduce these basic features of the reconstructions, which indicates that they have skill

simulating climates different from the present (Masson-Delmotte et al., 2013). On the other hand, there are also differences between model results and observations, e.g. in deep ocean circulation (Muglia et al., 2015), which indicates that model's skills in those aspects are questionable.

Detection and attribution studies attempt to determine which observed climate changes are unusual (detection) and what are its causes (attribution). Climate models driven with only natural forcings show variations from one year to the next due to internal climate variability (e.g. El Niño) and short term cooling in response to large volcanic eruptions (Fig. 3.9.8). However, they do not show a long-term warming trend over the past century, in contrast to the observations. This suggests that the global warming observed since about the 1970's is highly unusual and cannot be explained by internal climate variability (as represented in the models) nor by natural drivers. However, if anthropogenic forcings are included the models reproduce very well the observed long term trend. The multi model mean also reproduces well the observed short term coolings associated with large volcanic eruptions of the past 50 years.

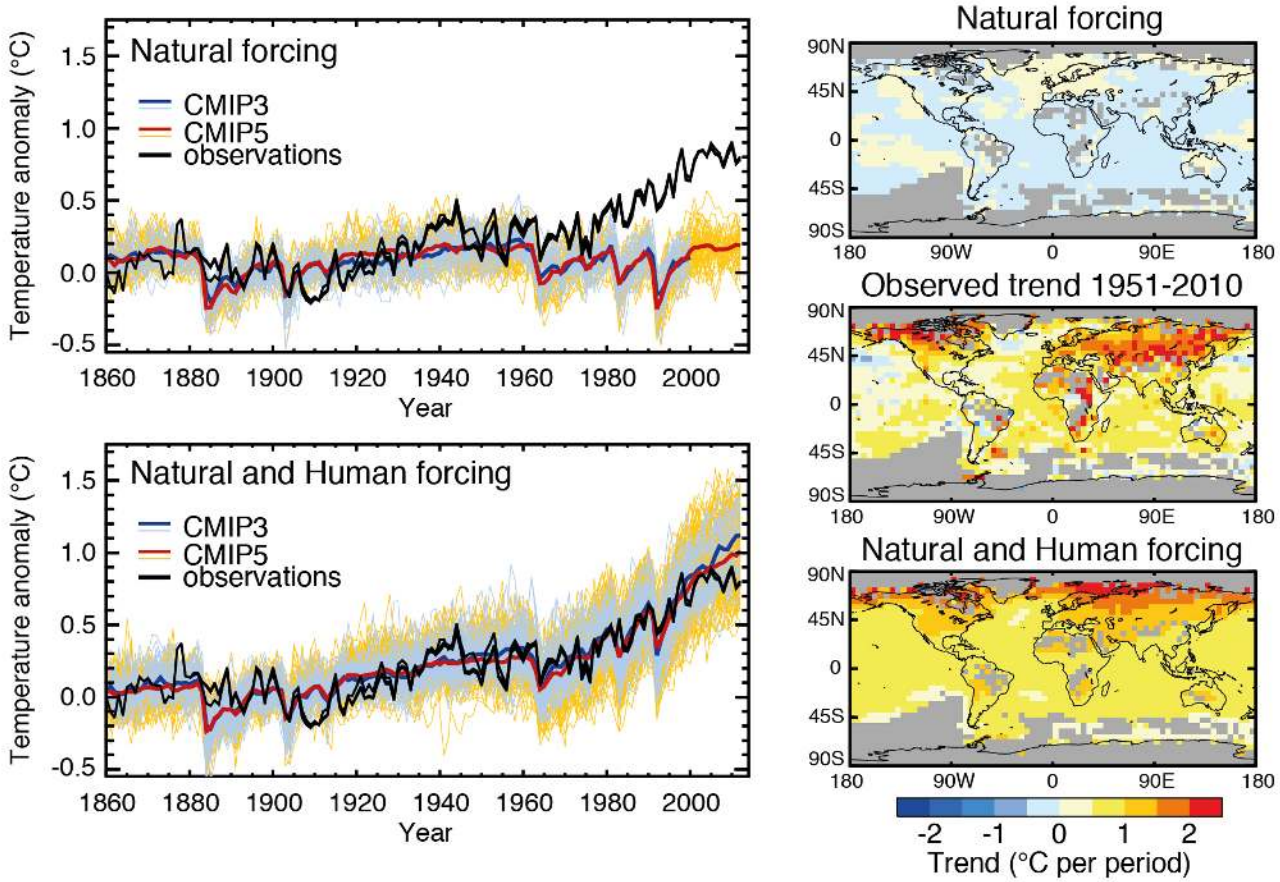


Figure 3.9.8: Comparison of climate model simulations with observed surface temperature changes. Left panels show timeseries of globally averaged surface temperatures from the current generation of climate models (CMIP5; yellow thin lines show individual models, red thick line shows the multi model mean) and the previous generation (CMIP3; blue lines, mostly obscured by the overlain yellow lines). Two sets of simulations are show. Top panels show models forced with natural forcings only, bottom panels show models forced with both natural and human forcings. From ipcc.ch. This is a key figure.

Models driven with both natural and anthropogenic forcings reproduce well not only the observed global mean temperature changes but also its spatial distribution such as larger warming at high northern latitudes (polar amplification) and over land (land-sea contrast). These results represent evidence that human activities are the main reason for the observed warming during the past 50-60 years. The IPCC's AR5 concludes that "it is extremely likely that human activities caused more than half of the observed increase in GMST from 1951 to 2010" (Bindhoff et al., 2013).

Questions

Answer the following questions:

- How do climate models work? Describe the equations that are solved.
- When were the first climate models developed?
- What is the hierarchy of climate models?
- List three different types of climate models of different complexity.
- Which components of the climate system are currently included interactively in most climate models?
- Which components are currently not included in most models?
- What is an Earth System Model?
- What is the typical resolution of a General Circulation Model (include units)?
- How many layers does a typical three-dimensional climate model have in the ocean?
- How many does it have in the atmosphere?
- Are observations used in forward climate model simulations?
- What are some inputs (boundary conditions) required for a climate model simulation? List two.
- How are climate models evaluated?
- Which variable to climate models simulate better: temperature or precipitation?
- What are climate models used for? List two applications.
- How are climate models used to attribute recent climate change?
- What is the result of those attribution studies?

Videos

Recorded Lecture: Climate Models

References

Bindoff, N.L., P.A. Stott, K.M. AchutaRao, M.R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain, I.I. Mokhov, J. Overland, J. Perlwitz, R. Sebbari and X. Zhang, 2013: Detection and Attribution of Climate Change: from Global to Regional. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Flato, G., J. Marotzke, B. Abiodun, P. Braconnot, S.C. Chou, W. Collins, P. Cox, F. Driouech, S. Emori, V. Eyring, C. Forest, P. Gleckler, E. Guilyardi, C. Jakob, V. Kattsov, C. Reason and M. Rummukainen, 2013: Evaluation of Climate Models. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Galton, F. (1097) Vox Populi, *Nature*, 75, 450-451. pdf

Masson-Delmotte, V., M. Schulz, A. Abe-Ouchi, J. Beer, A. Ganopolski, J.F. González Rouco, E. Jansen, K. Lambeck, J. Luterbacher, T. Naish, T. Osborn, B. Otto-Bliesner, T. Quinn, R. Ramesh, M. Rojas, X. Shao and A. Timmermann, 2013: Information from Paleoclimate Archives. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Muglia, J., and Schmittner, A. (2015) Glacial Atlantic overturning increased by wind stress in climate models, *Geophysical Research Letters*, 42, doi:10.1002/2015GL064583.

This page titled 3.9: Models is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Andreas Schmittner (Oregon State University, Open Educational Resources Unit) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.6: Models** by Andreas Schmittner is licensed CC BY-NC 4.0.

3.10: Impacts

Climate change impacts not only physical but also biological and human systems. Possible impacts of future climate change are important considerations for current decision making, but they are associated with uncertainties. In this chapter we will discuss future projections with climate models, the main sources of uncertainty for those projections, and some global scale impacts.

a) Projections

Prediction is difficult, especially about the future[1]. This is also true for projections of future climate change. They rely on assumptions about future radiative forcings such as anthropogenic emissions of greenhouse gases and aerosols, which are unknown. For this reason they are called projections and not predictions.

Climate scientists believe that the best projections consider results from state-of-the-science climate models because they are syntheses of theoretical and empirical knowledge. However, as we have learned in the previous chapters, climate models are imperfect. Uncertainties in future projections will therefore arise from both, assumptions about future greenhouse gas emissions and climate model errors.

The most recent IPCC assessment report (AR5) uses future scenarios that specify anthropogenic radiative forcing (Fig. 3.10.1; Collins et al., 2013; Stocker et al., 2013). They are called Representative Concentration Pathways (RCPs) and are followed by a number that indicates the radiative forcing around year 2100. The goal is to cover a range of possible futures. Note that the radiative forcing slowly decreases after the year 2050 for scenario RCP2.6. In this case the maximum forcing of ~2.5 Wm^{-2} is reached during the middle of this century. Scenarios RCP4.5 and RCP6.0 continue to increase the radiative forcing during this century and stabilize close to their maximum around year 2100, whereas for scenario RCP8.5 the forcing continues to increase until about year 2200 after which is stabilizes at 12 Wm^{-2}.

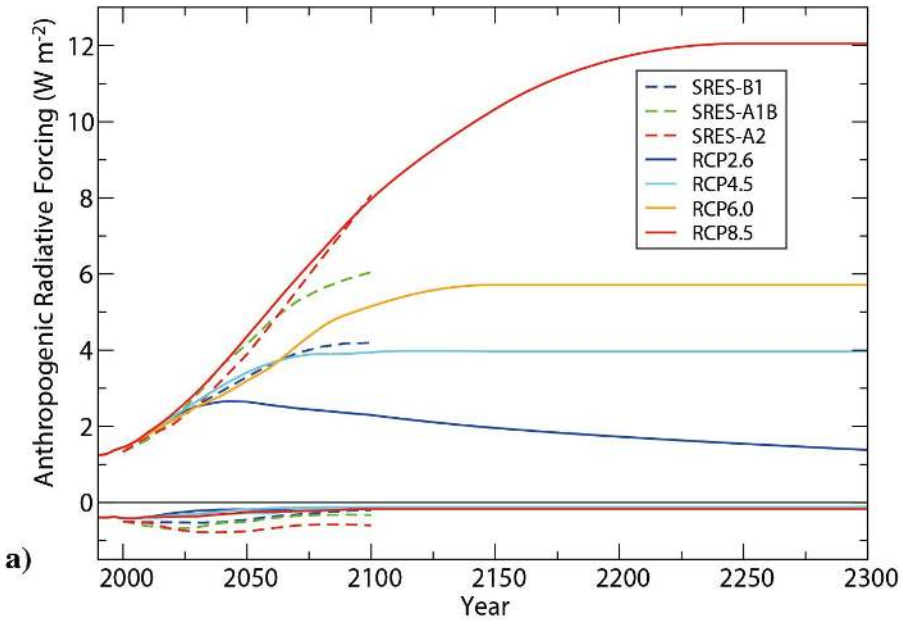


Figure 3.10.1: Anthropogenic radiative forcing as a function of time for the current (RCP) and previous (SRES) IPCC scenarios. From ipcc.ch.

The prescribed radiative forcing is then used to calculate corresponding CO_2 concentrations (Fig. 3.10.2), e.g. by inverting Eq. (4.2) in Chapter 4. Scenario RCP2.6 leads to a stabilization of CO2 concentrations below 450 ppm during the 21st century, whereas in scenarios RCP4.5 and RCP6.0 CO2 concentrations continue to increase to about 550 ppm and 650 ppm, respectively, at year 2100. In scenario RCP8.5 CO2 concentrations increase to about 900 ppm at year 2100. For the upcoming AR6 different scenarios, called Shared Socioeconomic Pathways or SSPs, have been developed.

Carbon cycle models are used to calculate carbon emissions consistent with those CO_2 concentration pathways. Due to uncertainties in the models a range of carbon emission scenarios emerges (shaded bands in Fig. 3.10.2). Scenario RCP2.6 implies

strong reductions in carbon emissions after about year 2020, whereas the business-as-usual scenario RCP8.5 assumes continued increases in emissions until the year 2100 (Fig. 3.10.2). The two intermediate scenarios RCP4.5 and RCP6.0 assume continued increases in emissions until about 2050 or 2080 followed by emissions reductions.

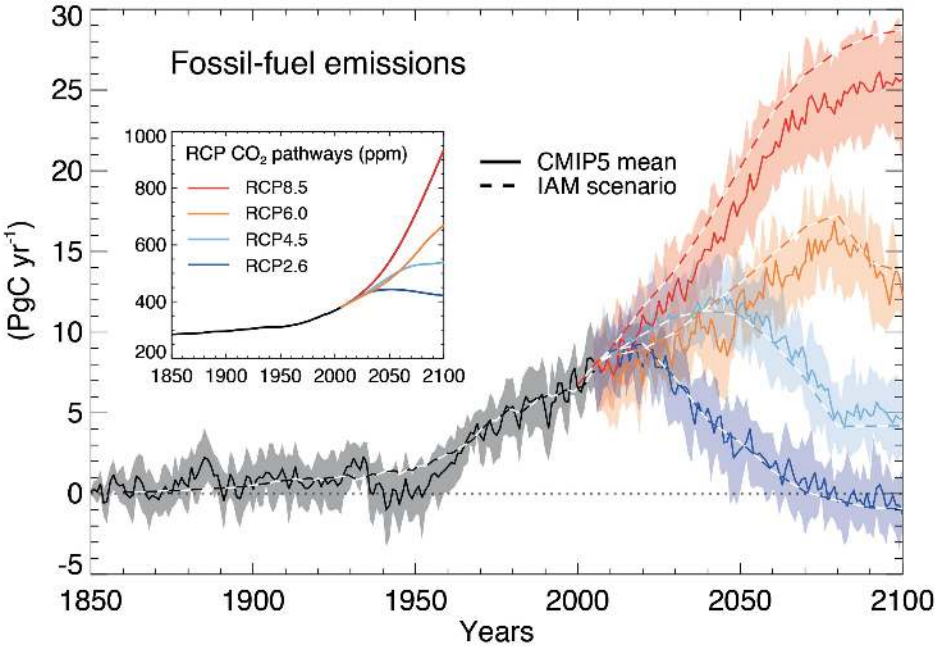


Figure 3.10.2: Evolution of CO_2 emissions and concentrations (inset) for RCP scenarios. Emissions are calculated from simple carbon cycle box models given the CO_2 concentrations. Uncertainties in these calculations are indicated by the shading and the two different lines labeled CMIP5 mean and IAM scenario. From ipcc.ch.

The CO2 concentration pathways are used as input to comprehensive climate models, which project a range of global temperature responses (Fig. 3.10.3). For scenario RCP2.6 the models project further warming of less than 1°C above current levels by year 2050 and subsequent slow cooling. Scenarios RCP4.5 and RCP6.0 result in additional warming of about 1.5 to 2°C until year 2100, whereas for the business-as-usual scenario RCP8.5 temperatures increase by 4°C by the year 2100 and by 8°C by year 2300. The latter corresponds to twice the temperature difference between the Last Glacial Maximum and the pre-industrial (see Chapter 3). Note that the uncertainty as indicated by the shading is much larger for the high emission scenario RCP8.5 compared with the intermediate and low scenarios.

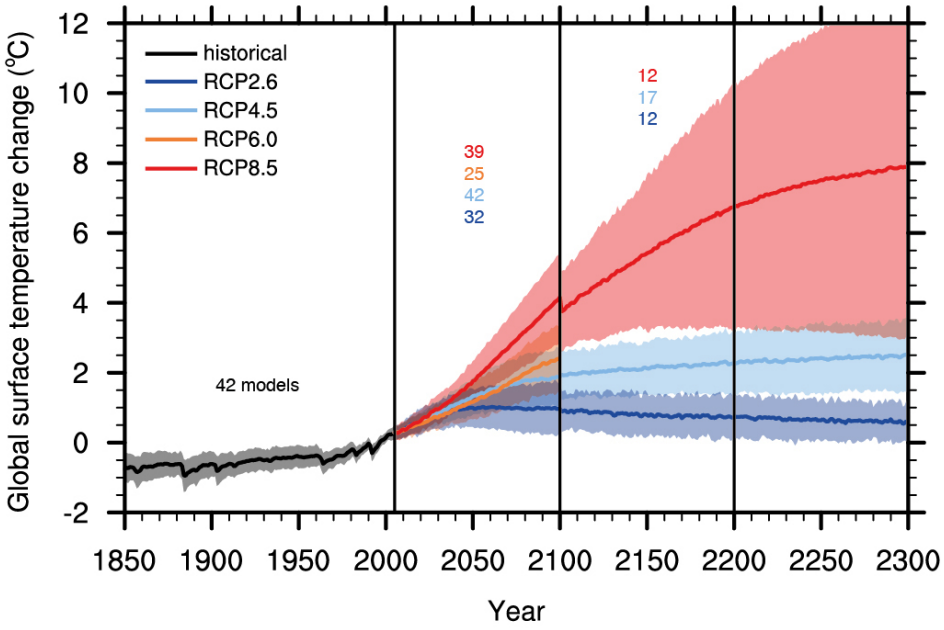


Figure 3.10.3. Global average surface temperature projections relative to 1986-2005. Shading indicates the 5 and 95% range and the line is the multi model mean. The black line with the grey shading represents historical simulations including variations in volcanic forcing. Since volcanic forcing is impossible to predict for the future, changes in natural radiative forcings are not considered in the projections. The number of climate models used is indicated. Not all models have simulated the full period until year 2300, which leads to a jump in the red line at year 2100. From ipcc.ch.

The climate models project more warming over land than over the ocean and more warming at high latitudes, especially in the Arctic (Fig. 3.10.4). This pattern is similar for all scenarios and time periods. However, the absolute numbers are smaller for lower emission scenarios than for the business-as-usual scenario RCP8.5 shown in Fig. 3.10.4). We have encountered these patterns of land-sea contrast and polar amplification before in the observations of historical temperature changes discussed in Chapter 2, in reconstructed temperatures from the Last Glacial Maximum discussed in Chapter 3, and in our discussions of detection and attribution studies in Chapter 7. We conclude that these patterns are robust features of the climate system that are relatively well understood and reproduced in models. Warming in the Arctic is more than twice as much as the global average, which is dominated by the oceans. The IPCC concludes that, on average, warming over land will be 1.4 to 1.7 times larger than warming over the ocean (Collins et al., 2013).

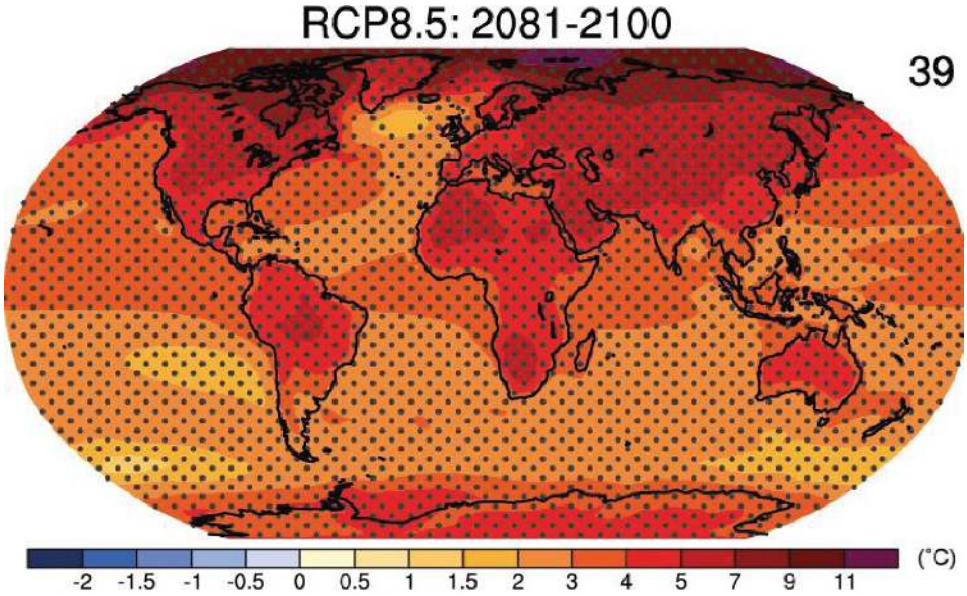


Figure 3.10.4: Surface temperature changes projected for the business-as-usual scenario RCP8.5 at the end of the 21st century. Stippling indicates regions of highly significant changes, where the multi model mean is larger than two standard deviations of modeled internal variability and where at least 90% of models agree on the sign of the change. Click on the map to see projections for other scenarios and times. From ipcc.ch.

Explore Projections

Explore projections of future climate using this website:

- Select a country, variable and model of your choice.
- How does the model simulation of the seasonal cycle (lower left plot) for the 1980-2004 period compare with the observations?
- Try a different model and then the model mean. Which agrees better with the observations?
- Compare simulations of temperature and precipitation with observations. Which fits better?
- Now compare the model simulation of the future (2050-2074) with the recent past (1980-2004). Describe your observations.

Another good website with climate model projections but also other data is climexp.knmi.nl.

Fig. 3.10.5) shows that models project more warming in the upper tropical troposphere than at the surface. Increased water vapor content in the warmer atmosphere (see discussions in Chapter 4 of the Clausius-Clapeyron relation and the lapse rate feedback, and in Chapter 6 of the hydrological cycle) causes more upward moisture transport in the ascending part of the Hadley cell. Enhanced condensation and latent heat release leads to intensified warming there.

Models also project significant cooling in the stratosphere. Since the stratosphere is stably stratified (there is no convection) its energy balance is determined by local heating from the aabsorption of solar radiation by ozone and local cooling from the emission of infrared radiation by CO_2 (Manabe and Strickler, 1964). Thus, higher CO_2 concentrations in the stratosphere will lead to more emission of infrared radiation and cooling there. Stratospheric cooling is observed in contemporary measurements, consistent with the cooling effect of CO_2. However, the current cooling is also caused by decreases in stratospheric ozone caused by human emissions of Chlorofluorocarbons (CFCs). Nevertheless, the projected stratospheric cooling for the future will be dominated by CO_2 increases since CFC emissions have decreased and the ozone hole is expected to heal over the next decades.

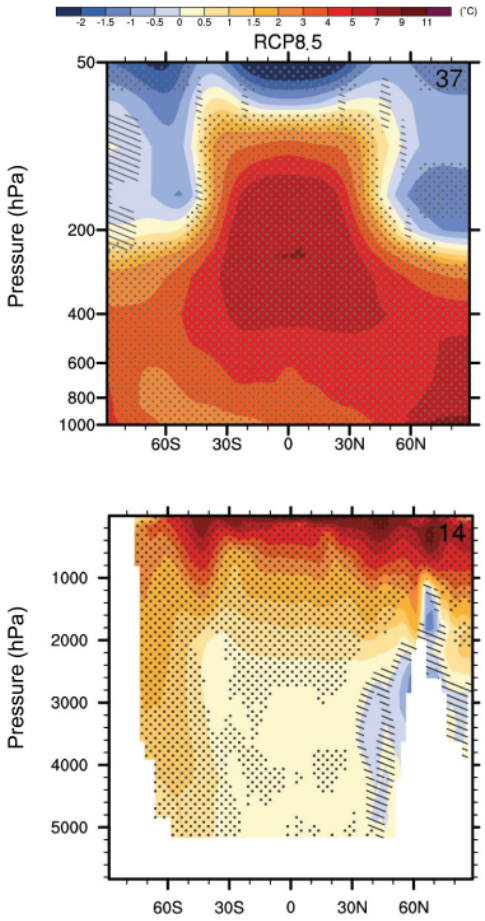


Figure 3.10.5). Hatching indicates where changes are not significant with respect to 2 sigma variations of internal variability. Click on the ocean panel to see results for other scenarios. From ipcc.ch.

The observed warming of the upper ocean is projected to continue and penetrate deeper into the ocean interior particularly in regions of deep water formation such as the Southern Ocean (Fig. 3.10.5)). More warming of the surface will increase the stratification of the ocean and reduce vertical mixing, with implications for the ocean's meridional overturning circulation and nutrient delivery to the euphotic zone, both of which will likely continue to decrease.

Warming of the surface is projected to increase evaporation everywhere except in the northern North Atlantic and Southern Ocean where the warming is smallest and over some land areas where precipitation is projected to decrease (Fig. 3.10.6)). Precipitation is generally projected to increase in regions that are already wet like the tropics and at mid- to high-latitudes, and decrease in regions that are already dry like the subtropics. Relative humidity is generally projected to decrease over land areas and increase over the oceans. The net surface freshwater loss, which is determined by evaporation minus precipitation (E − P), is projected to increase in the subtropics, whereas the tropics and mid- to high latitudes will gain water. Runoff is projected to increase in most regions except around the Mediterranean, the Southwest US and Mexico, southwest Africa, and parts of South America. Soil moisture, on the other hand, is projected to decrease almost everywhere except in parts of North Africa, Asia, and South America, where precipitation increases substantially.

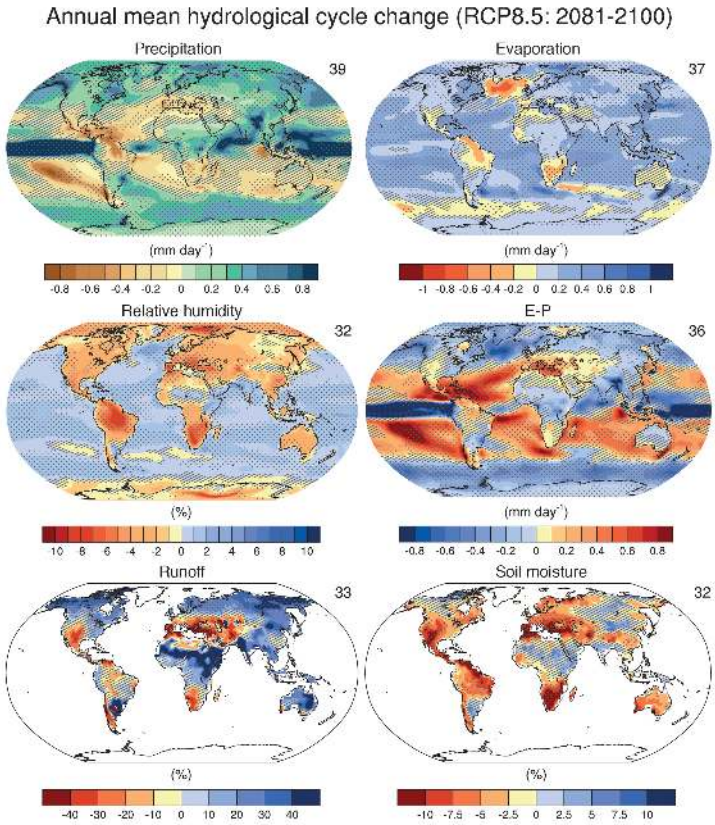


Figure 3.10.6): As Figure 3.10.4), but for changes in the hydrological cycle. From ipcc.ch.

Projections suggest that Arctic sea ice will continue to decline in the future particularly at the end of summer (Fig. 3.10.7)). However, much larger declines are projected for the high emission scenario (RCP8.5), for which all models project an almost completely ice free Arctic ocean, compared to the low emission scenario (RCP2.6), for which all models project a substantial remaining sea ice cover.

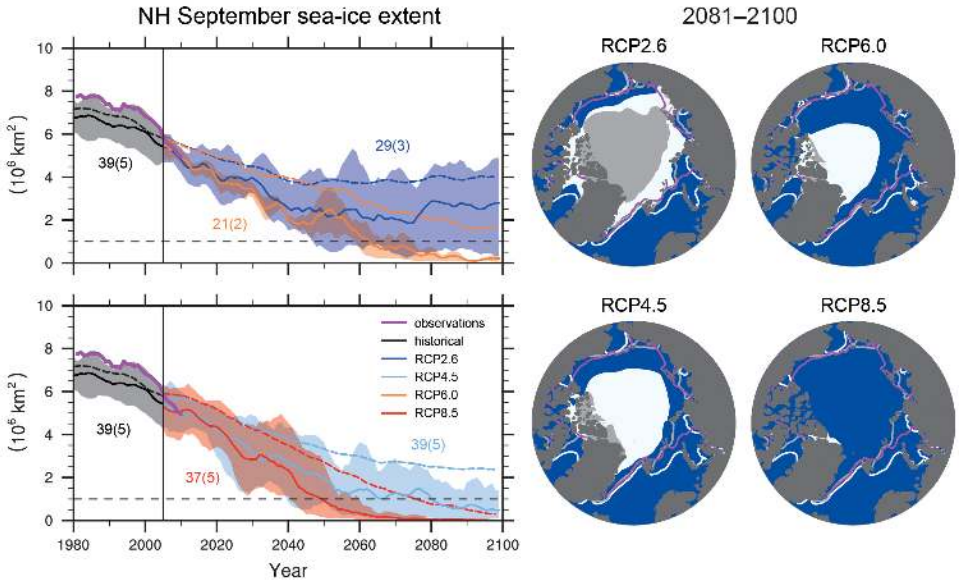


Figure 3.10.7): Projections of Arctic sea ice in September. Results from all models are shown as dashed lines in the time series panels on the left and as white shading in the maps on the right. Results from a subset of models that agree best with observed sea ice (numbers in brackets) are shown as solid lines and grey shading. From ipcc.ch.

Mountain glaciers are also projected to continue to melt, as well as the Greenland and Antarctic ice sheets. Due to this input of mass together with the expansion of the ocean from warming **sea levels** are projected to increase by 40 to 70 cm by the year 2100 (Fig. 3.10.8)). Note that sea level projections are very similar for all scenarios until mid-century. This indicates the **commitment** to the near future sea level rise from past carbon emissions. In other words, considerable climate change and the associated impacts are already baked into the system and cannot be avoided. It implies the **necessity to adapt** to these unavoidable impacts.

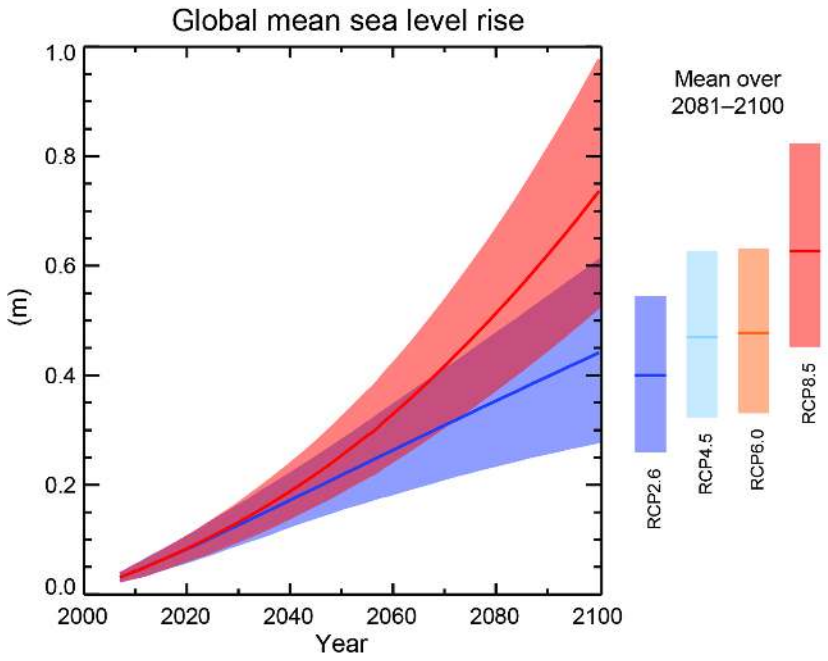


Figure 3.10.8): Global sea level rise projected for the 21st century. The blue and red curves show projections for scenarios RCP2.6 and RCP8.5, respectively. The shading indicates the likely (66-100%) uncertainty range. Bars on the right show the 2081-2000 averages for four scenarios. From ipcc.ch.

Increasing stratification from warming and freshening of North Atlantic surface waters will lead to a slowdown of the Atlantic meridional overturning circulation, by about 10% for each degree Celsius global mean warming (Bakker et al., 2016). Many other impacts also depend approximately linearly on global mean temperature and global mean temperature can be well approximated by cumulative carbon emissions (Matthews et al., 2009) such that 500 Gt lead to approximately 1°C of warming (Figure 3.10.9)). Since we've already emitted about 500 Gt, this means that if we want to stay below 2°C, the goal of the Paris Agreement, we have only another 500 Gt of carbon to emit.

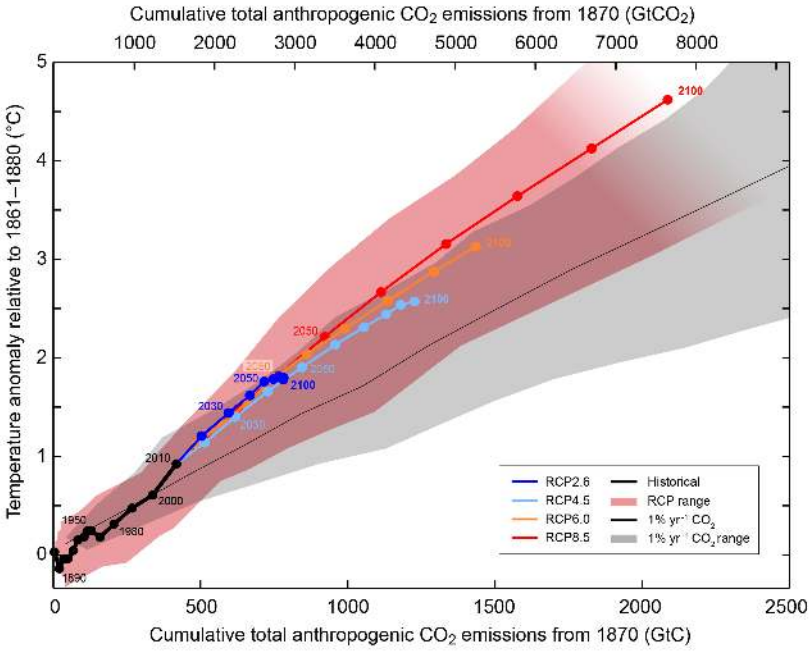


Figure 3.10.9). Colored lines show multi-model means from the RCP scenarios and colored plume indicates uncertainties. Results from exponential CO_2 increase experiments at a rate of 1% per year are shown as the black line and grey shading. RCP scenarios lead to somewhat larger warming because they include additional non-CO_2 forcings. From ipcc.ch.

Tipping Points

A tipping point is a threshold that when crossed will lead to an irreversible transition into a different state. Examples for tipping points are

- melting of the Greenland ice sheet,
- melting of the West Antarctic ice sheet,
- collapse of the Atlantic meridional overturning circulation,
- ecosystem shifts,
- species extinctions.

Ice sheets can have tipping points due to positive feedbacks, such as the *ice sheet elevation – mass balance feedback (Weertman 1961)*. This feedback can be understood by considering an idealized ice sheet on a flat bed as illustrated in Fig. B 3.10.1) below. An ice sheet gains mass through accumulation of snow in the upper central parts, whereas it looses ice through melting along the lower-lying margins. For simplicity we assume in the following that the accumulation and melt rates are spatially constant. Due to gravity and the resulting horizontal pressure gradient the ice flows slowly from the center towards the margins. Moving upward on a glacier or ice sheet one starts at the margin to walk on ice. At some point the ice turns into snow, which covers the interior of the sheet all the way to the summit. This is the so-called snow line or equilibrium line. It indicates the transition from the melt-zone to the accumulation zone. If the mean height of the ice sheet $\overline{H}$ is larger than the height of the equilibrium line H_E the ice sheet is in a stable equilibrium. However, if warming temperatures rise the equilibrium line to $\overline{H}$ the ablation area increases and the ice sheet looses more mass than it gains. It shrinks and its height drops, which leads to even more melting. This feedback can lead to a rapid and irreversible melting of an entire ice sheet.

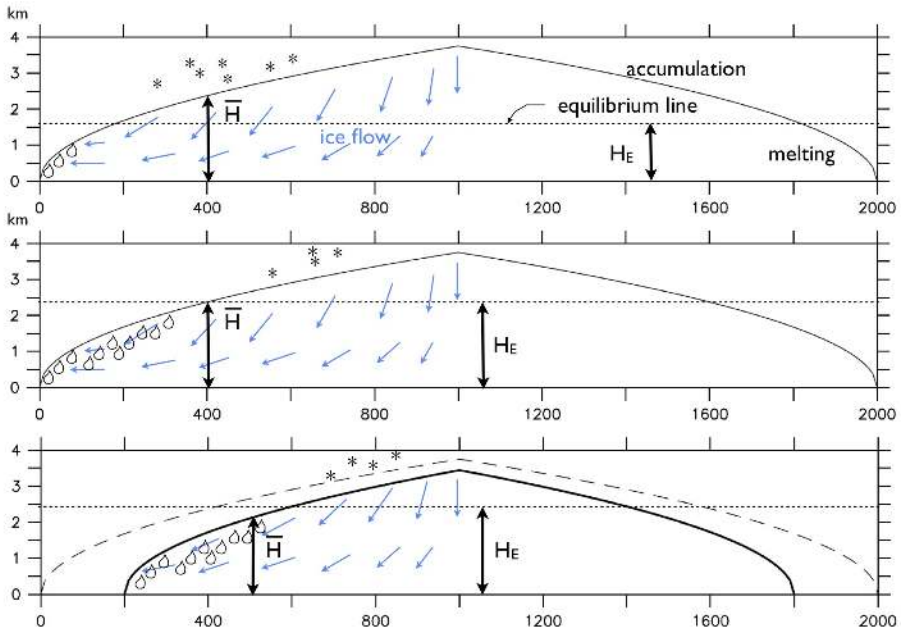


Figure B3.10.1): Schematic section across an idealized ice sheet. The horizontal axis represents the distance in km from its left margin. The vertical axis represents elevation above the flat base.

As a result of this feedback ice sheets exhibit hysteresis behavior as illustrated schematically in Fig. B3.10.2). If the height of the equilibrium line is between zero and $\overline{H}$ two states are possible. One without ice and one with an ice sheet. Assume we're in a state without ice on the lower branch of the hysteresis curve. If the climate cools and the equilibrium line drops below zero ice will grow and we will switch to the state with an ice sheet. If the climate now warms again the ice sheet will stay in place until the height of the equilibrium line raises above $\overline{H}$. At this point the ice sheet will melt. So, for equilibrium line heights between zero and $\overline{H}$ two states are possible and which state the system is in depends on its history. This is a typical property of hysteresis as observed e.g. in magnetism. It also can lead to rapid changes in state as illustrated by the vertical arrows. Due to its latitude being further poleward and its height being higher it is thought that the East Antarctic ice sheet is more stable and further away from the melting threshold than the Greenland ice sheet.

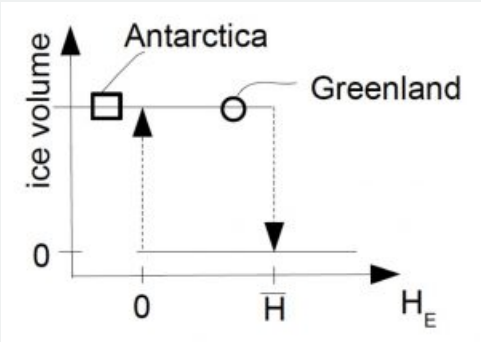


Figure B3.10.2): Schematic hysteresis curve for an ice sheet. The vertical and horizontal axes represent ice volume and height of the equilibrium line, respectively.

More complex ice sheet models exhibit similar threshold behavior. Current estimates suggest that the **Greenland ice sheet** may disappear almost completely for global warming of 0.8-3.2°C above pre-industrial temperatures (Figure B3.10.3). However, even though this threshold may be exceeded in the near future due to continued anthropogenic carbon emissions, the ice sheet will not disintegrate immediately. It may take hundreds to thousands of years to melt completely. The larger the warming the more rapid the ice melt will be. Melting of the entire Greenland ice sheet would raise global sea level by about 7 m.

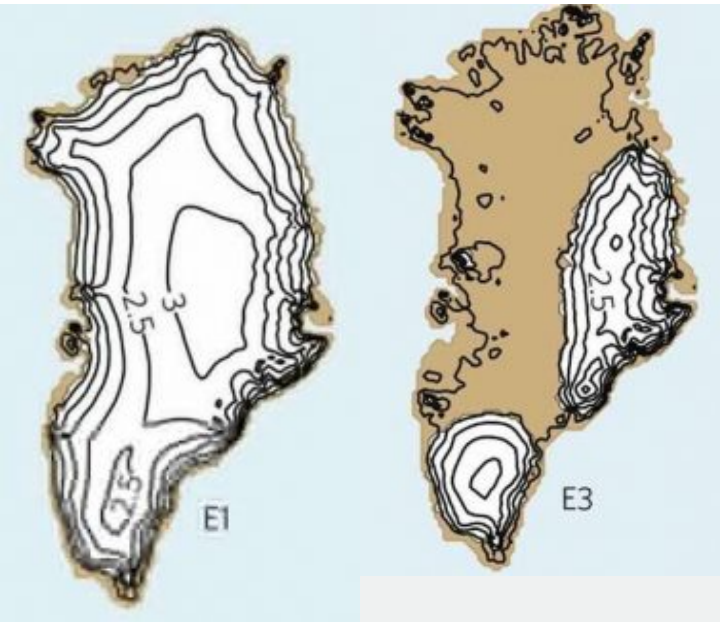


Figure B3.10.3): Greenland fully glaciated (left) and mostly deglaciated (right). Contour lines denote ice sheet thickness in km. From Robinson et al. (2012) at nature.com.

The **West Antarctic ice sheet** is also vulnerable to warming. Most of it is below sea level and it is fringed with ice shelves that float on the ocean (Fig. B3.10.4). The grounding line separates the ice shelf from the grounded inland ice. Ice shelves often provide a buttressing force that keeps the interior ice sheet from flowing faster into the ocean. Warm Circumpolar Deep Water (CDW) intrusion can melt the bottom of the ice shelf, increase calving of ice bergs and lead to retreat of the grounding line. Because the deepening of the bed towards the interior retreat of the grounding line and reduced buttressing can cause an increase in the outward ice flow and initiate an instability that leads to the collapse of the entire ice sheet. Recent models indicate that the West Antarctic ice sheet could disintegrate within hundreds of years for high and intermediate carbon emission scenarios, whereas it is stable for low emission scenarios (DeConto and Pollard, 2016). If the West Antarctic ice sheet collapsed global sea level would rise by about 5 m.

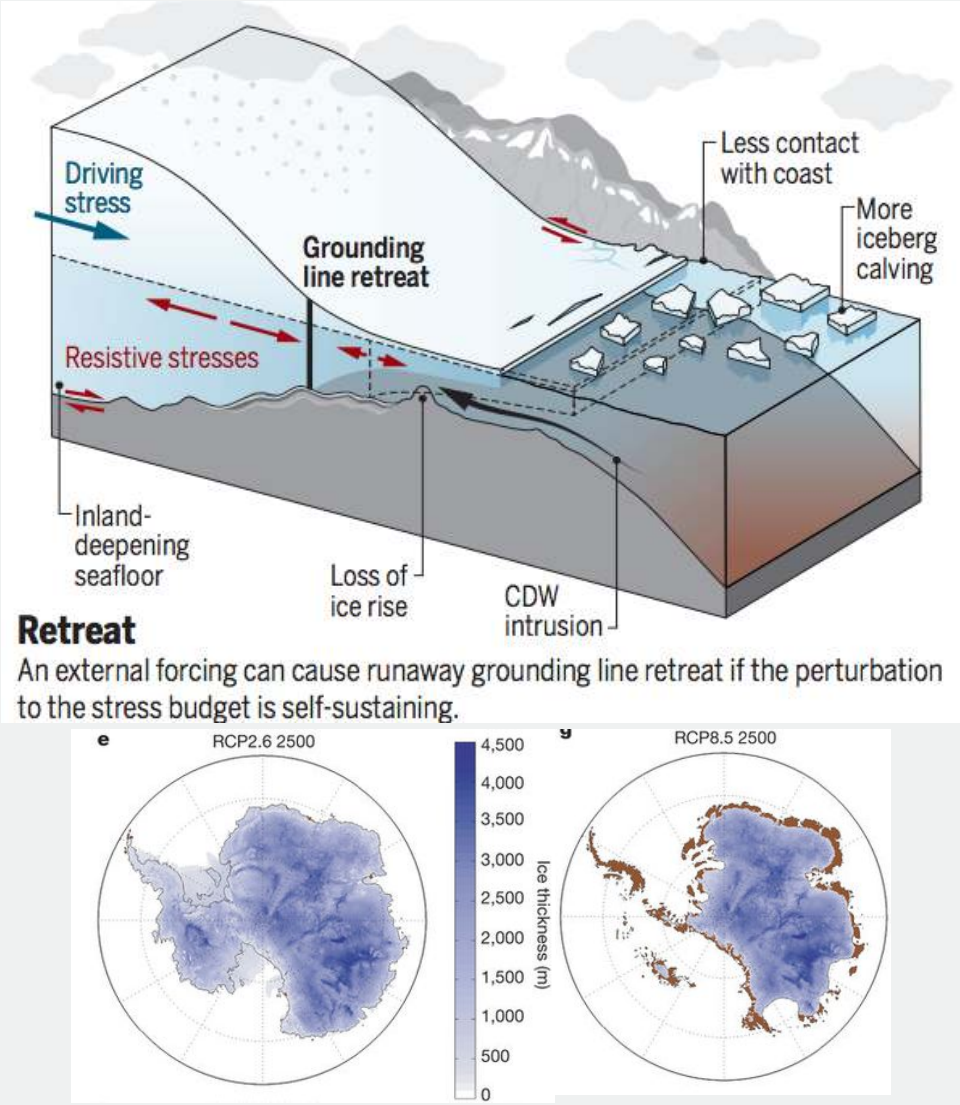


Figure B3.10.4: Antarctic ice sheet vulnerabilities. Top: Schematic illustration of the West Antarctic ice sheet from carbonbrief.org. Bottom: Projection of the Antarctic ice sheet at year 2500 for emission scenarios RCP2.6 (left) and RCP8.5 (right) from DeConto and Pollard, 2016).

The **Atlantic Meridional Overturning Circulation** (AMOC) also has the potential to switch into a collapsed state once a threshold is exceeded. Paleoclimate evidence indicates that such shifts happened repeatedly in the past. The AMOC exhibits hysteresis behavior because of a positive feedback between the circulation and salinities in the North Atlantic. The circulation transports high saline waters from the subtropics towards the North Atlantic. However, a reduction of the flow and thus the northward transport of saline waters will lead to decreasing salinities in the North Atlantic regions of deep water formation and this will reduce the flow even further. Currently it is thought that an AMOC collapse is unlikely for low emission scenarios but the probability increases for high emission scenarios (Bakker et al., 2016). A collapse or large reduction of the AMOC would impact not only the North Atlantic due to reduced northward heat transport, it would also affect the tropics by shifting the Intertropical Convergence Zone towards the south, which would affect precipitation patterns, vegetation, ecosystems and humans there and it would affect ocean biogeochemical cycles (e.g. of nutrients and carbon) and ecosystems.

Ecosystems will continue their poleward and upward shifts. Present day tundra, for example, will be replaced by taiga (Figure B3.10.5). Species that used to live at high latitudes or on the upper elevations of mountains may be replaced by species moving in from further south or from lower altitudes. This can lead to **species extinctions**. Especially vulnerable are species that live

near the top of mountains such as the cute little pika, or near disappearing sea ice such as polar bears, walrus, and narwhales (Larsen et al., 2014).



Figure B3.10.5: Tundra (left) is a biome at far northern latitudes characterized by grasses and shrub vegetation and permafrost beneath the seasonally thawed soil. Musk oxen, geese, caribou (reindeer), Arctic fox, Arctic hare, snow owl, and lemmings are some of the animals that call the tundra home. Warming leads to taiga (a.k.a. boreal forest, right) with coniferous trees replacing the tundra.

b) Ecosystems

Vegetation models project further poleward shifts in biomes (Fig. 3.10.10) due to warming consistent with the currently observed ongoing trends. Potential vegetation is the natural vegetation that would grow in a certain area without human land use changes such as agriculture or forestry. The biome shifts are projected with the most confidence in those regions close to the biome boundaries towards the migration direction of the invading biome. E.g. close to the tundra/taiga boundary there will be almost certainly a shift from tundra to taiga. For individual species such as aspen in North America large shifts are possible.

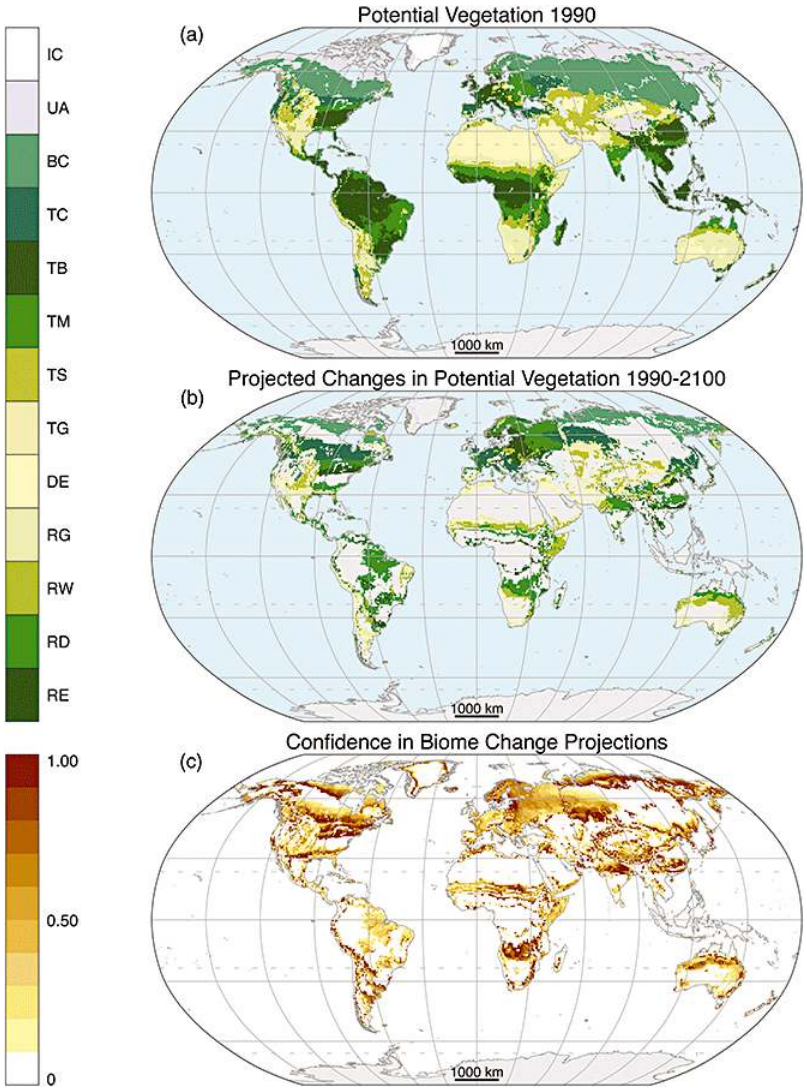


Figure 3.10.10: Modeled potential vegetation in 1990 (a), projected changes by the end of the century (b) and the confidence in the projections (c). Biomes from poles to equator are ice (IC), tundra and alpine (UA), boreal conifer forest (taiga, BC), temperate conifer forest (TC), temperate broadleaf forest (TB), temperate mixed forest (TM), temperate shrubland (TS), temperate grassland (TG), desert (DE), tropical grassland (RG), tropical woodland (RW), tropical deciduous broadleaf forest (RD), tropical evergreen broadleaf forest (RE). From Gonzalez et al. (2010).

Increasing CO_2 concentrations in the atmosphere will increase the water-use efficiency of plants. They can grow more with the same water use because they don't need to open their stomata as much for the same amount of CO_2 to flux in. They can also grow in regions that were previously too dry for them under lower CO_2. This greening effect is currently observed e.g. from satellites.

Changes in precipitation will also impact vegetation distribution. The effects on vegetation of reduced precipitation in the subtropics in a warmer climate may be compensated to some degree by the increased water-use efficiency under higher CO_2. Longer growing season, higher CO_2, and warmer temperatures will increase net plant production but also respiration, e.g. in soils.

Wildfires are difficult to project but in most regions they are likely to get worse due to climate change (Fig. 3.10.11. Fire needs fuel, which in the case of wildfire is burnable biomass (dry wood or grass). It also needs the right climatic or weather conditions. Hot and dry summers increase fire hazards and strong winds can lead to rapidly spreading fires. In many regions, e.g. North America, Australia and Russia, fire suppression by humans have caused a build up of fuel, which, once ignited, can lead to larger fires. In many parts of the tropics, on the other hand, humans burn to clear agricultural fields. Thus, changes in wildfires are not only caused by climate change. However, hotter drier summers, projected e.g. for the western United States, will lead to increased fire probability in the future. This is consistent with scientific evidence that fires have already increased there in recent decades due

to anthropogenic climate change (Abatzoglou and Williams, 2016). In other regions such as the high northern latitudes, where woody taiga vegetation will replace tundra the resulting increase in available fuel may lead to more or larger fires. Note that the changes in projected fire frequency are highly uncertain as indicated by different models and methods yielding different results.

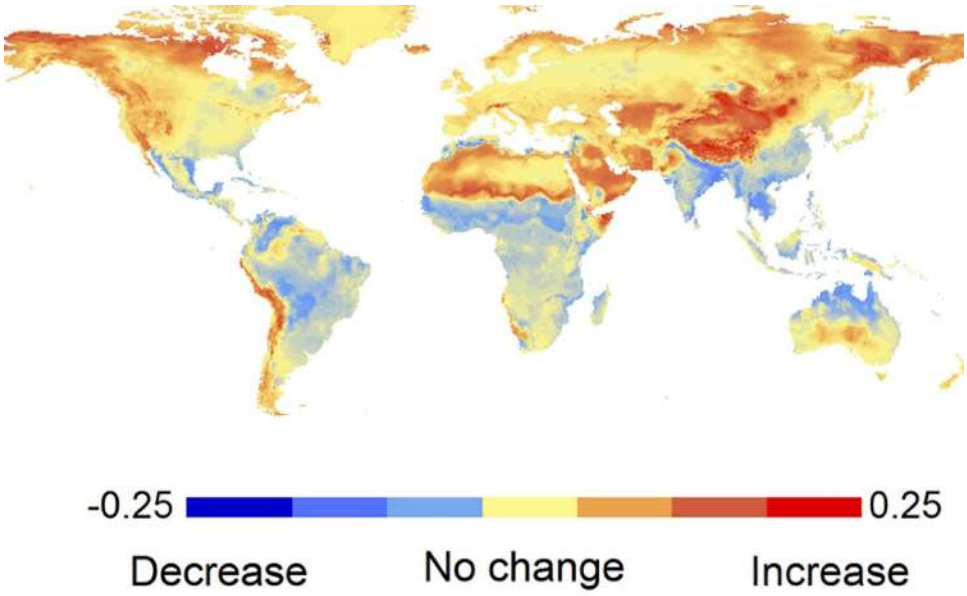


Figure 3.10.11: Projected change in fire probability from 1971-2000 to 2070-2099 for the high SRES-A2 emissions scenario using a statistical method and climate models. From Moritz et al. (2012).

One of the most sensitive **ocean ecosystems** are probably coral reefs. Corals are threatened not only by bleaching, which refers to the expulsion of their symbiotic algae resulting from warmer waters, but also by ocean acidification, which inhibits their calcium carbonate production. The presently observed shift of fish species towards higher latitudes will most likely continue.



Figure 3.10.12: Healthy (left) and bleached (right) coral.

https://geo.libretexts.org/@go/page/24038

In the Arctic narwhales may be replaced by invading killer whales (Figure 3.10.13). Similarly, for many other species from phytoplankton to whales, some will benefit and others will suffer from climate change. Calcifiers such as coccolithophores will presumably be among the losers due to ocean acidification, whereas cyanobacteria, which is another group of phytoplankton, may be among the winners (Dutkiewitcz et al., 2015). Because of this complex response, many consequences for ecosystems cannot be predicted. It is therefore likely that there will be surprises that scientists could not foresee.



Figure 3.10.13: Narwhal (left) and killer whales (right).

Generally, impacts of climate change on Earth's ecosystems will be larger for high emission scenarios and smaller for low emission scenarios. Earth's ecosystems are resilient. They have experienced large climate changes in the past, such as during the ice age cycles of the Pleistocene. However, current rates of change and those expected for the future are larger for some climate variables than past rates (Fig. 3.10.14), which has many scientists concerned about the adaptability of ecosystems.

c) Long-Term Changes

Due to the long lifetime of carbon in the Earth system current human activities will affect many future generations. The ultimate fate of anthropogenic carbon is burial in ocean sediments. This is a slow process that takes tens of thousands of years to remove all of the extra carbon. For this reason much of the carbon we're putting into the atmosphere today will impact climate and Earth's physical, biological and human systems for a long time. The total amount of available fossil fuels still in the ground is uncertain (~10,000 PgC, GEA, 2012), but it is clear that enough exists to melt all major ice sheets, which would raise sea level by about 65 m. Fig. 3.10.14 shows scenarios of up to about 5,000 GtC, which leads to complete melting of the Greenland and West Antarctic ice sheets and most of the East Antarctic ice sheet with a sea level rise of about 50 m. Even the relatively low (1,280 GtC) emission scenario would lead to a long term sea level rise of about 25 m. However, the rate of change would be much slower for the lower emission scenario (0.5 m/century) compared with the higher scenario (3 m/century). Such rates are unprecedented in more than 8,000 years. High emission scenarios would melt essentially all mountain glaciers on Earth, whereas low emission scenarios would melt about 70% of current glaciers, mostly within this and the next century (Marzeion et al., 2012). Global warming for all but the low emission scenarios will be similar to or even exceed that from the Last Glacial Maximum to the early Holocene (~4ºC). Transformations of Earth's ecosystems similar in magnitude to those documented for the last deglaciation can therefore be expected for intermediate and high emission scenarios. Because of the prominence of the human influence the period since the industrial revolution has been called the Anthropocene.

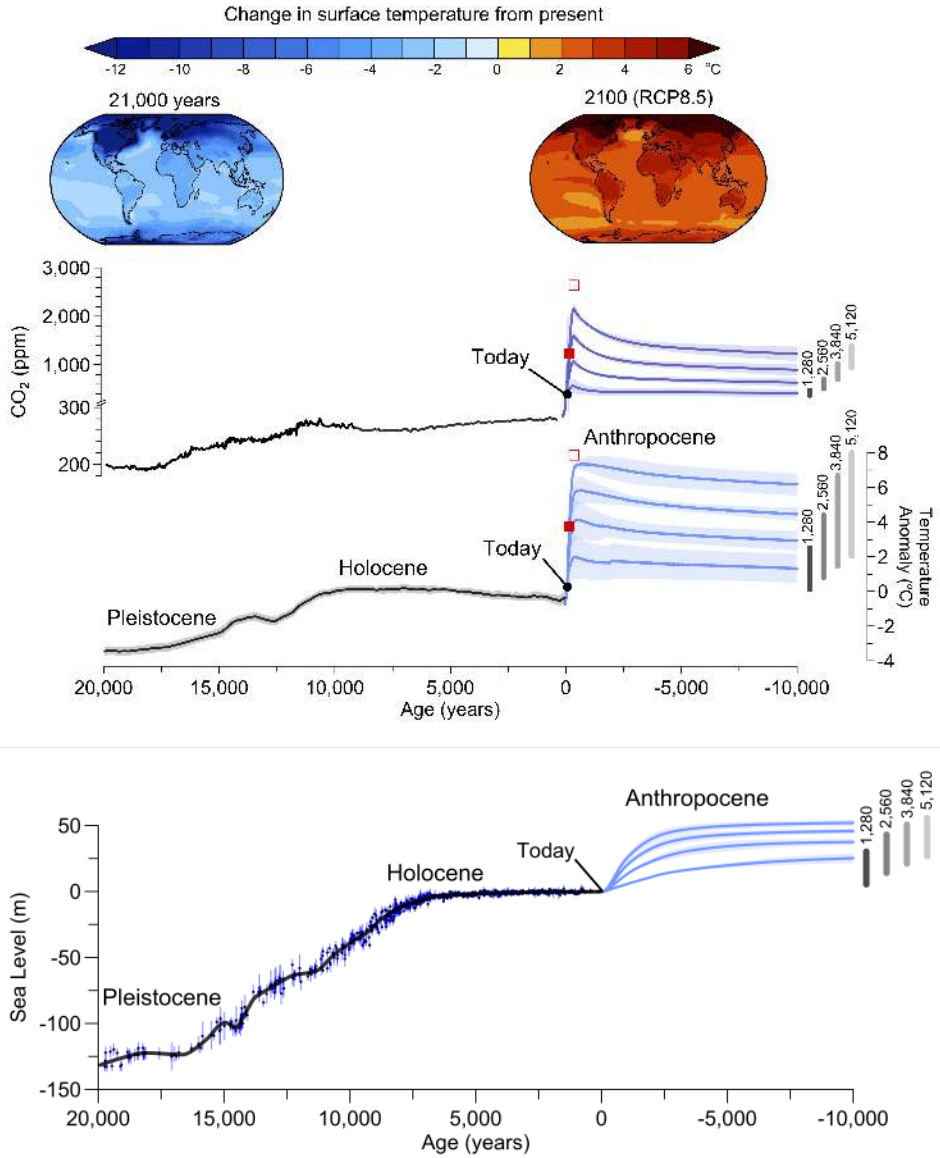


Figure 3.10.14: Long-term projected future in the perspective of past changes in atmospheric CO_2, surface temperature and sea level. Top: maps show model simulated temperature anomalies (with respect to the preindustrial) for the Last Glacial Maximum (21,000 years ago at the end of the Pleistocene) and projections for the year 2100 based on the RCP8.5 emission scenario. Center and bottom: changes in CO2, global surface temperature and sea level from paleo data and model projections. Intermediate complexity models were used for the future projections assuming a total of 1,280, 2,560, 3840, and 5,120 PgC emissions shown in the blue lines with shading indicating uncertainty. Red squares indicate results from years 2100 (solid) and 2300 (open) for the RCP8.5 scenario for comparison. Vertical gray bars show the range of uncertainty based on a range for climate sensitivity from 1.5 to 4.5ºC. Modified from Clark et al. (2016). Figure courtesy of Shaun Marcott and Peter Clark.

Sea level rise will have tremendous effects on people living in coastal regions. Clark et al. (2016) estimate that their low emission scenario (1,280 PgC) would submerge an area where currently 1.3 billion people live (19% of the global population) including 25 megacities such as Calcutta, New York, Tokyo, Shanghai, and Cairo. Due to the time lag associated with sea level rise we are already committed to future sea level rise from past emissions. By the year 2000, e.g. humans had emitted ~470 PgC and were committed to a sea level rise of about 2 m. Releasing another 470 PgC would commit us to another 9 m of long-term sea level rise.

https://geo.libretexts.org/@go/page/24038

d) Regional Changes

Projected climate impacts vary strongly depending on the region. Small islands and coastal regions will be affected by sea level rise. Polar and alpine regions will be affected by snow and ice loss. Dry regions such as the subtropics will get drier. Detailed regional projections exist for some regions such as North America and Europe. A good resource for regional impacts is part B of this IPCC (2014) report. Here we will only look at one example, which is the western United States of America. Fig. 3.10.13 shows that the snowpack there has decreased dramatically during the past 60 years. On average the April 1st snowpack has declined by 15-30% or 25-30 km^3, which is similar in volume to the region's largest man-made reservoir, Lake Mead. Further decreases in snowpack in this region can be expected due to continued warming with impacts on summer stream flows, agriculture and other human water use downstream.

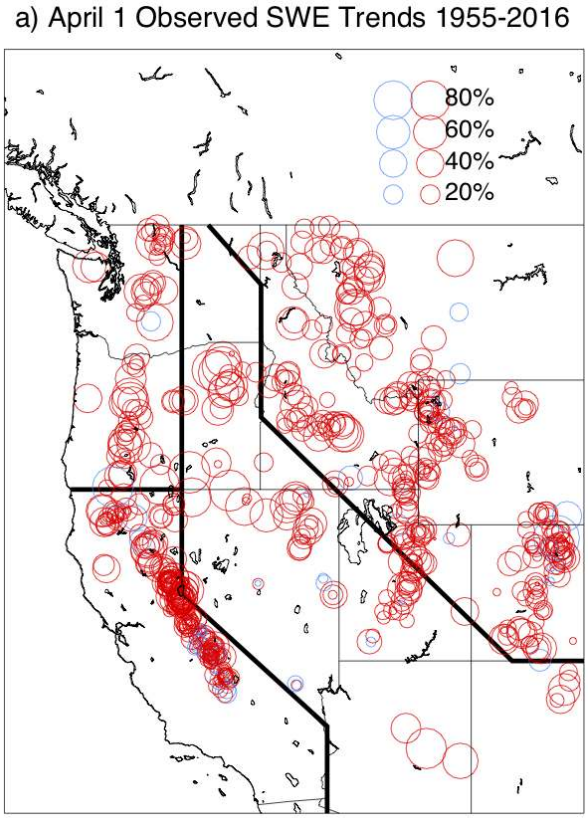


Figure 3.10.15: Linear trends in April 1st Snow Water Equivalent (SWE) observed for the period 1955-2016. Red circles indicate decreased snowpack, blue increased. From Mote et al. (2018).

Figure 3.10.14 illustrates that vegetation west of the Cascades mountain range in Oregon and Washington is projected to shift from conifer to more drought-tolerant mixed forest, whereas some of the grassland that occupies the eastern portions of those states as well as parts of Idaho and Montana is projected to be replaced by shrubs and conifer forest. Fires are projected to increase in frequency from 37 to 9 years on average in the full region. Modeled fires for the past are infrequent west of the Cascades (~80 year recurrence time on average), but they are projected to become much more common for high emission scenarios (~27 year recurrence time for RCP8.5) due to warmer and dryer summers. Intermediate emission scenarios lead to less dramatic changes in fire and vegetation.

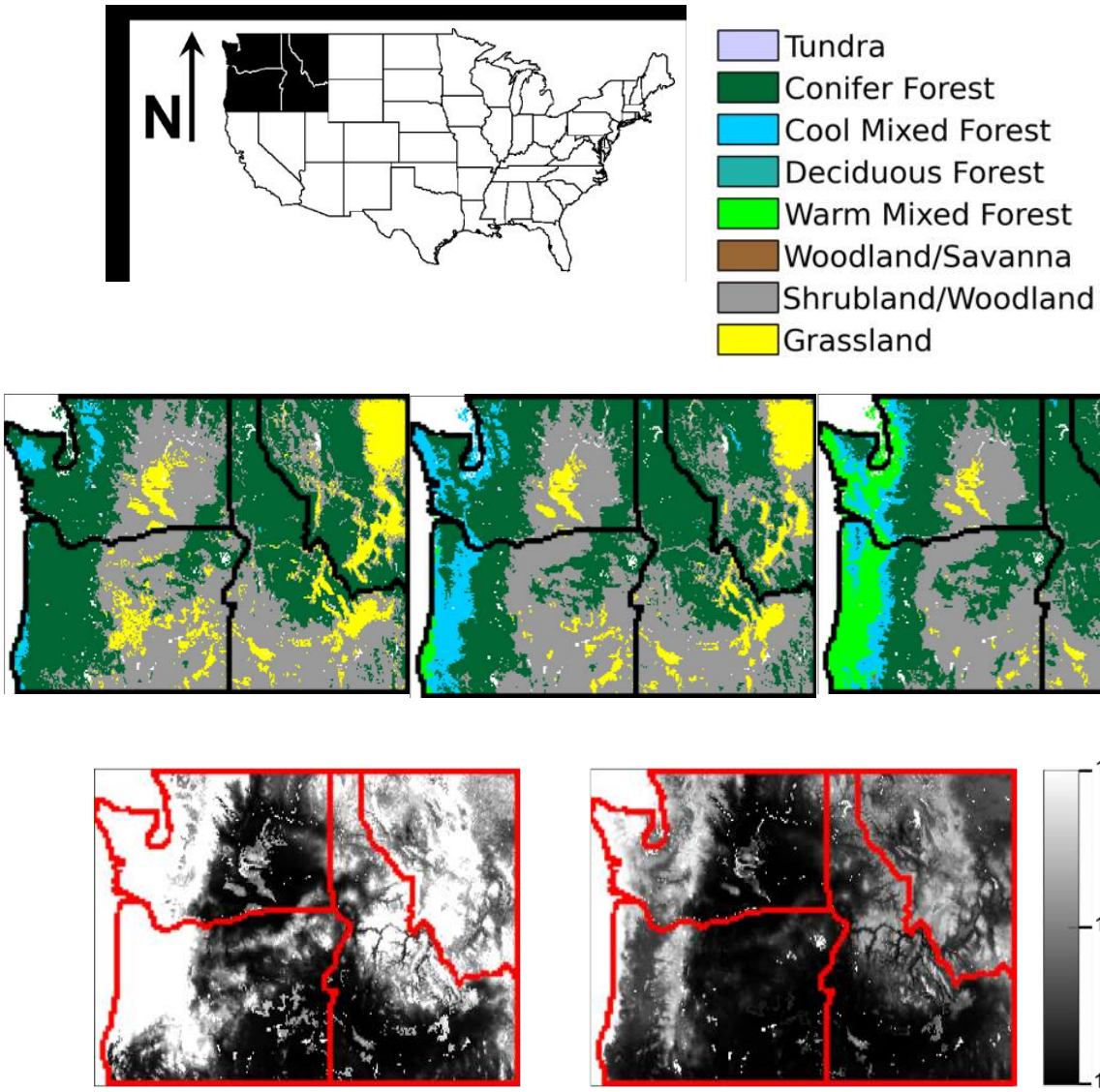


Figure 3.10.16: Projected changes in vegetation (center row with legend in top right) and fire frequency (bottom) in the Pacific Northwest (filled area in the top left map of the USA). Vegetation is shown for years 1971-2000 (left), 2036-2065 (center) and 2071-2100 (right). Fire frequency in years is shown for the 20th century (left) and the 21st century (right). Modified from Sheehan et al. (2015). Simulations without fire suppression are shown. Figure courtesy of Tim Sheehan.

e) Extremes

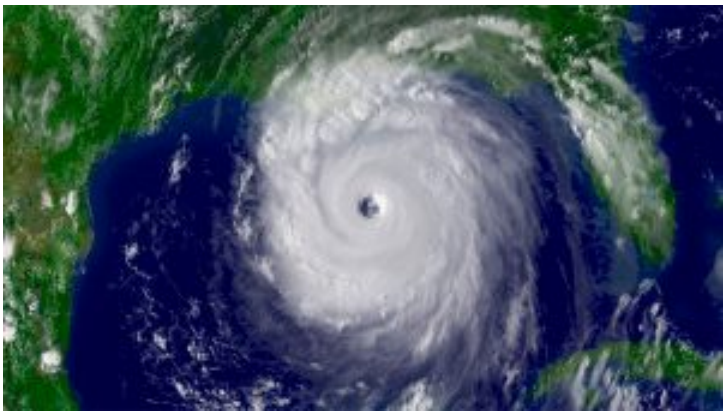


Figure 3.10.17: Satellite Image of Hurricane Katrina.

A warming climate implies a shift in the probability distribution such that hot extremes become more frequent and cold extremes become less frequent. This is what is currently observed and we can expect this trend to continue into the future. Due to the intensification of the hydrological cycle we can also expect more droughts and more floods. Changes in other extreme events are less well understood. Occurrences of total hurricanes and typhoons (tropical cyclones) are projected to decrease but the strongest hurricanes are projected to become more frequent (Knutson et al., 2010). This is a concern because those are the ones that inflict the most damage. Hurricane development depends on warm ocean water as an energy source. Latent heat release is also an important fuel for hurricanes and other storms. Thus warmer sea surface temperatures and more latent heat release due to more water vapor in the warmer air will strengthen hurricanes, consistent with observations of increases in the destructiveness of tropical storms in the past decades (Emanuel, 2005). Another important factor for hurricane development, particularly in its early stages, is wind shear. That is how fast winds increase with elevation. Storms require low shear to develop into a coherent vortex, which may become less likely in the future. The end result of changes in wind shear and changes in temperatures is that in a warmer climate the total number of hurricanes will decrease, but strong hurricanes will become more frequent and they will cause more rainfall.

f) Impacts on Humans

Climate change is already affecting people and more changes can be expected for the future. The decline of sea ice in the Arctic, for example, already affects native people there such as the Inuit. Their life depends on sea ice for hunting and travel and will be greatly affected by the expected future sea ice loss (Watt-Cloutier, 2015; watch her talk here).



Sea ice also dampens waves and their effect on erosion of shorelines. Some villages are already threatened by increased erosion due to sea ice loss. On the other hand, a seasonally ice free Arctic will allow for ships to take a shortcut through the Northwest Passage and reduce travel time between the North Atlantic and North Pacific.

Similarly, some mining companies are already planning operations in Greenland in areas that were previously inaccessible due to ice cover. But this opportunity for resource extraction comes of course at the price of sea level rise, which will affect millions of people who live close to shore. Many people will also be affected by the loss of mountain glaciers and their summer water supply. Many current ski areas may no longer be viable in the future due to less snow cover and a shorter season, which will result in job losses there.

Climate change can also lead to conflict. The ongoing civil war in Syria, for example, has been partly caused by a drought, which has been attributed to anthropogenic climate change (Kelley et al., 2015). This attribution is in line with climate model projections that indicate dryer conditions around the Mediterranean in a warmer world (Fig. 3.10.6: and paleoclimate data (Cook et al., 2016). Violent conflict about freshwater resources is not new in the history of this and other regions but the current conflict may be the first that is partly caused by anthropogenic climate change. The resulting migrants, many of which fled to central Europe, may be one of the first climate refugees. Conflict and climate refugees may be also be expected from people displaced by sea level rise, such as in Bangladesh or low laying islands. Sea level rise affects people not only through flooding but also by saltwater seeping into the freshwater lens that exists below the ground and provides islanders often with the only source of freshwater. This can also be a problem in other coastal regions.

Questions

- What is the difference between prediction and projection?
- List two uncertainties in climate projections.
- Which tool do climate scientists use for climate projections?
- Are climate models perfect?
- Describe scenarios RCP2.6 and RCP8.5 in terms of carbon emissions and climate change. For which scenario will the impacts be greater?
- How does the global average surface temperature change projected for the end of the 22nd century compare with that between the Last Glacial Maximum and the preindustrial?
- Is projected surface temperature change larger on land or on the ocean? What are the reasons for these differences?
- Where is projected surface temperature change larger, in the tropics or at the poles? What are the reasons for these differences?
- Where are atmospheric temperatures projected to increase most in the tropics? At the surface or at high elevations? What is the reason for this?
- Where is precipitation projected to increase, where to decrease?
- How is evaporation projected to change?
- How is Arctic sea ice projected to change?
- List five different possible impacts of future climate change.
- How much carbon have humans emitted so far, cumulatively?
- How much carbon can we emit in the future if we want to stay below 2ºC above preindustrial global mean temperature?
- What is a tipping point?
- Which ice sheet is closer to a tipping point, Greenland or East Antarctica?
- Why is the West Antarctic ice sheet vulnerable to a warming climate?
- How is the tundra ecosystem projected to change in the future?
- What will be some of the long-term global impacts of past and future anthropogenic carbon emissions?
- What will be some of the regional impacts in the Pacific Northwest?
- How would you expect the frequency of heat waves and cold spells change in the future?
- How are hurricanes expected to change?
- How do you expect future climate change to impacts humans?

Explore Impacts

Explore regional impacts of possible future climate change using this websites:

Videos

Lecture: Projections

Lecture: Impacts

Lecture: Tipping Points

References

Abatzoglou, J. T., and A. P. Williams (2016), Impact of anthropogenic climate change on wildfire across western US forests, Proceedings of the National Academy of Sciences, 113(42), 11770. doi: 10.1073/pnas.1607171113.

Bakker, P., A. Schmittner, J. T. M. Lenaerts, A. Abe-Ouchi, D. Bi, M. R. van den Broeke, W. L. Chan, A. Hu, R. L. Beadling, S. J. Marsland, S. H. Mernild, O. A. Saenko, D. Swingedouw, A. Sullivan, and J. Yin (2016), Fate of the Atlantic Meridional Overturning Circulation: Strong decline under continued warming and Greenland melting, Geophys Res Lett, 43(23), 12,252-212,260, doi: 10.1002/2016GL070457.

Clark, P. U., J. D. Shakun, S. A. Marcott, A. C. Mix, M. Eby, S. Kulp, A. Levermann, G. A. Milne, P. L. Pfister, B. D. Santer, D. P. Schrag, S. Solomon, T. F. Stocker, B. H. Strauss, A. J. Weaver, R. Winkelmann, D. Archer, E. Bard, A. Goldner, K. Lambeck, R. T. Pierrehumbert, and G.-K. Plattner (2016), Consequences of twenty-first-century policy for multi-millennial climate and sea-level change, Nature Clim. Change, 6, 360-369, doi: 10.1038/nclimate2923.

Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver and M. Wehner, 2013: Long-term Climate Change: Projections, Commitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. pdf

Cook, B. I., K. J. Anchukaitis, R. Touchan, D. M. Meko, and E. R. Cook (2016), Spatiotemporal drought variability in the Mediterranean over the last 900 years, Journal of Geophysical Research: Atmospheres, 121(5), 2060-2074, doi: 10.1002/2015JD023929.

DeConto, R. M., and D. Pollard (2016), Contribution of Antarctica to past and future sea-level rise, Nature, 531(7596), 591-597, doi: 10.1038/nature17145.

Dutkiewicz, S., J. J. Morris, M. J. Follows, J. Scott, O. Levitan, S. T. Dyhrman, and I. Berman-Frank (2015), Impact of ocean acidification on the structure of future phytoplankton communities, Nature Climate Change, 5, 1002, doi: 10.1038/nclimate2722.

Emanuel, K. (2005), Increasing destructiveness of tropical cyclones over the past 30 years, Nature, 436(7051), 686-688, doi: 10.1038/nature03906.

GEA (2012) Global Energy Assessment – Toward a Sustainable Future, Cambridge University Press, Cambridge, UK and New York, NY, USA and the International Institute for Applied Systems Analysis, Laxenburg, Austria. link

Gonzalez, P., R. P. Neilson, J. M. Lenihan, and R. J. Drapek (2010), Global patterns in the vulnerability of ecosystems to vegetation shifts due to climate change, Global Ecology and Biogeography, 19(6), 755-768, doi: 10.1111/j.1466-8238.2010.00558.x.

IPCC (2014) Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 688. link

Kelley, C. P., S. Mohtadi, M. A. Cane, R. Seager, and Y. Kushnir (2015), Climate change in the Fertile Crescent and implications of the recent Syrian drought, Proceedings of the National Academy of Sciences, 112(11), 3241. link

Knutson, T. R., J. L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J. P. Kossin, A. K. Srivastava, and M. Sugi (2010), Tropical cyclones and climate change, Nat Geosci, 3, 157, doi: 10.1038/ngeo779.

Larsen, J.N., O.A. Anisimov, A. Constable, A.B. Hollowed, N. Maynard, P. Prestrud, T.D. Prowse, and J.M.R. Stone, 2014: Polar regions. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1567-1612. link

Manabe, S., and R. F. Strickler (1964), Thermal Equilibrium of the Atmosphere with a Convective Adjustment, J. Atmos. Sci., 21, 361-385. pdf

Marzeion, B., A. H. Jarosch, and M. Hofer (2012), Past and future sea-level change from the surface mass balance of glaciers, The Cryosphere, 6(6), 1295-1322, doi: 10.5194/tc-6-1295-2012.

Matthews, H. D., N. P. Gillett, P. A. Stott, and K. Zickfeld (2009), The proportionality of global warming to cumulative carbon emissions, Nature, 459(7248), 829-U823, doi: 10.1038/Nature08047.

Moritz, M. A., M.-A. Parisien, E. Batllori, M. A. Krawchuk, J. Van Dorn, D. J. Ganz, and K. Hayhoe (2012), Climate change and disruptions to global fire activity, Ecosphere, 3(6), 1-22, doi: 10.1890/ES11-00345.1.

Mote, P. W., S. Li, D. P. Lettenmaier, M. Xiao, and R. Engel (2018), Dramatic declines in snowpack in the western US, npj Climate and Atmospheric Science, 1(1), 2, doi: 10.1038/s41612-018-0012-1.

Robinson, A., R. Calov, and A. Ganopolski (2012), Multistability and critical thresholds of the Greenland ice sheet, Nature Climate Change, 2, 429, doi: 10.1038/nclimate1449.

Sheehan, T., D. Bachelet, and K. Ferschweiler (2015), Projected major fire and vegetation changes in the Pacific Northwest of the conterminous United States under selected CMIP5 climate futures, Ecological Modelling, 317, 16-29, doi: 10.1016/j.ecolmodel.2015.08.023.

Stocker, T.F., D. Qin, G.-K. Plattner, L.V. Alexander, S.K. Allen, N.L. Bindoff, F.-M. Bréon, J.A. Church, U. Cubasch, S. Emori, P. Forster, P. Friedlingstein, N. Gillett, J.M. Gregory, D.L. Hartmann, E. Jansen, B. Kirtman, R. Knutti, K. Krishna Kumar, P. Lemke, J. Marotzke, V. Masson-Delmotte, G.A. Meehl, I.I. Mokhov, S. Piao, V. Ramaswamy, D. Randall, M. Rhein, M. Rojas, C. Sabine, D. Shindell, L.D. Talley, D.G. Vaughan and S.-P. Xie, 2013: Technical Summary. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. link

Watt-Cloutier, S. (2015) The Right to Be Cold, Penguin Canada Books, Ontario, Toronto. linkyoutube

Weertman, J. (1961) Stability of Ice-Age Ice Sheets, J. Geophys. Res. 66 (11), 3783-3792, doi: 10.1029/JZ066i011p03783.

1. This quote has been attributed to various people including the Danish physicist Niels Bohr and American writer Mark Twain. ↵

This page titled 3.10: Impacts is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Andreas Schmittner (Oregon State University, Open Educational Resources Unit) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.7: Impacts** by Andreas Schmittner is licensed CC BY-NC 4.0.

3.11: Summary



Under construction



3.11: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

4: The Lithosphere

- 4.1: The Earth - The Ground Beneath Your Feet
- 4.2: Minerals
- 4.3: Rocks
- 4.4: Bedrock
- 4.5: Weathering
- 4.6: The Mineral Particles of the Earth's Surface Materials
- 4.7: Regolith
- 4.8: Sediment
- 4.9: Soils
- 4.10: Soil Chemistry
- 4.11: Soil Biodiversity and Ecology
- 4.12: Soil Nutrient Cycling
- 4.13: Weathering, Erosion, and Mass Movement
  - 4.13.1: Weathering
  - 4.13.2: Mass Movement
  - 4.13.3: Water Erosion
  - 4.13.4: Review and Additional Resources
- 4.14: Summary

4: The Lithosphere is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.1: The Earth - The Ground Beneath Your Feet

To a great extent in this course, we will be dealing with processes that act on the solid materials at and near the Earth's surface. This chapter might better be called "the ground beneath your feet". This is the place to deal with the nature of the Earth's surface materials, which in later sections of the chapter I will be calling regolith, sediment, and soil.

I purposely did not specify any previous knowledge of geology as a prerequisite for this course, so it is important, here in the first part of this chapter, for me to provide you with some background on Earth materials.

We will be dealing almost exclusively with the Earth's continental surfaces. There are profound geological differences between the continents and the ocean basins, in terms of origin, age, history, and composition. Here I'll present, very briefly, some basic things about geology. (For more depth on such matters you would need to take a course like "The Earth: What It Is, How It Works", given in the Harvard Extension program in the fall semester of 2005– 2006 and likely to be offered again in the not-too-distant future.)

In a gross sense, the Earth is a layered body (Figure 2-1). To a first approximation, it consists of concentric shells: the core, the mantle, and the crust.

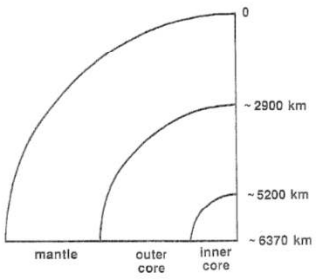


Figure 2-1: Schematic cross-section through the Earth

The core: The core consists mostly of iron, alloyed with a small percentage of certain other chemical elements. The outer part of the core is liquid, and the inner part is solid. There is no need here to deal with how that's known—or indeed why there is an iron core in the first place.

The mantle: The mantle, which constitutes the greater part of the mass of the Earth, has its lower boundary, with the core, at a depth of about 2900 km.
(The radius of the Earth is about 6400 km.) The mantle consists almost entirely of solid rock, except for certain places in the uppermost mantle where magma (= melted rock) is formed at certain times. (There is a deplorably widespread and thoroughly mistaken belief that the upper mantle is everywhere molten.) Saying that the mantle is solid rock, true as that statement is, is somewhat misleading, though, because the solid rock of the mantle can flow plastically in the same way that a very viscous liquid, like honey or corn syrup, flows. You will see the same kind of behavior in glacier ice, in a later chapter. "Silly putty" is another good example of such a material. The uppermost part of the mantle, however, is sufficiently cool that it behaves not as a very stiff liquid but as a rigid solid. This situation, whereby the uppermost rigid part of the mantle (which, together with the overlying crust, is called the ***lithosphere***) rides on the plastically flowing part of the mantle beneath (called the ***asthenosphere***), is the basis for what is called plate tectonics (Figure 1-2).

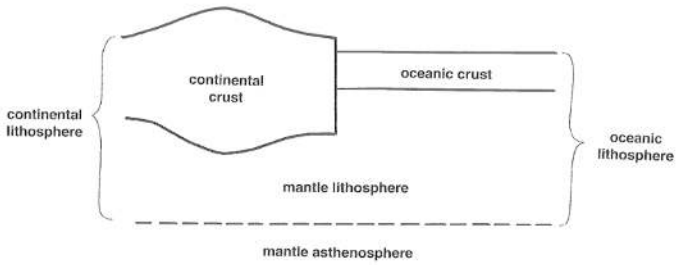


Figure 2-2. Schematic cross-section through the outer part of the Earth

The crust: The uppermost skin of the Earth, above the mantle, is called the crust (Figure 2-2). There are two basic kinds of crust, very different from one another in properties and origin: oceanic crust and continental crust. Oceanic crust is relatively thin, seldom more than seven to eight kilometers. Continental crust, on the other hand, is relatively thick, mostly thirty to fifty kilometers. The features of the upper surface of the continental crust, and the great variety of physical and chemical (and biological) processes and phenomena associated with that surface, are the focus of this course.

You probably know something about, or at least have heard about, plate tectonics. The lithosphere, the outermost part of the Earth, comprising the crust and the uppermost, rigid part of the mantle (Figure 2-2), is in the form of a number of segments, called ***lithospheric plates*** (or just *plates*), which are in contact with one another along what are called plate boundaries. The movements of these plates, and the consequences of those movements, especially in terms of what happens at plate boundaries, are collectively termed ***plate tectonics***.

We could get along fairly well in this course without reference to the core or the mantle, or, for that matter, even to the lower part of the continental crust. Deep subsurface processes, in the crust and in the underlying upper mantle, have a great indirect importance to this course, however: such processes (the nature of which is beyond the scope of this course!) cause *uplift and subsidence of the land surface of the continents*, both sharp and localized, in the case of active mountain belts like the western Cordillera of North America, and broad and gentle, affecting large areas of the stable continent of North America east of the Cordillera. Uplift and subsidence, in turn, are instrumental in erosion, transportation, and deposition of sedimentary materials and in shaping of the surface topography of the continent. We'll return to such matters in the final chapter.

This page titled 4.1: The Earth - The Ground Beneath Your Feet is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.1: The Earth - The Ground Beneath Your Feet** by John Southard is licensed CC BY-NC-SA 4.0. Original source: https://ocw.mit.edu/courses/12-090-the-environment-of-the-earths-surface-spring-2007/.

4.2: Minerals

The technical definition of a mineral is elegant: *a naturally occurring crystalline solid*. That definition calls for a bit of commentary, though. By “crystalline” is meant that the atoms of the mineral are arranged in a regular three-dimensional array, called its ***crystal structure***. The qualification “naturally occurring” excludes the multitude of crystalline solids that can be synthesized in the laboratory but are not found in the natural environment. Also, there are naturally occurring solids, like glass or amber, that do not have a crystal structure and therefore technically are not minerals.

There are thousands of named minerals—but, fortunately for us in this course, only a few are common rock-forming minerals, and even fewer figure prominently in the unconsolidated material that mantles the continents.

Most of the common minerals are of a class called ***silicate minerals***. Silicate minerals have as their basic building blocks a polyatomic (five-atom) unit, called a ***silica tetrahedron***, that consists of one atom of silicon, relatively small, surrounded by four atoms of oxygen, relatively large, to give the shape of a tetrahedron (Figure 2-3). The five atoms are bonded very strongly together. The reason why silicate minerals are so common is that, in terms of abundances of the chemical elements in the crust, oxygen is the big number one and silicon is second.

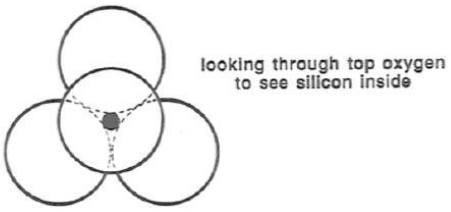


Figure 2-3. Schematic view of the silica tetrahedron, looking straight down on one of the oxygens. The small black dot represents the silica atom.

The atoms in mineral crystals are held together by strong forces called ***bonds***. An understanding of the nature of bonds necessitates some familiarity with atomic structure. All atoms consist of a positively charged nucleus around which a number of negatively charged electrons are in orbit. The fundamental nature of the electron orbits is complex, but for the purposes of understanding bonds in minerals it’s sufficient to know that electron orbits exist in distinctive groups known as ***shells***.

Chemical elements like helium, neon, or argon whose atoms have just enough electrons to fill all of the shells are chemically almost entirely inert. Elements whose atoms have their outermost electron shell almost filled have a strong tendency to take on one or a few electrons to fill the shell, and in so doing they acquire a negative electric charge. Similarly, elements whose atoms have only one or a few electrons in their outermost shell lose those electrons to revert to a filled-shell configuration, and in so doing they acquire a positive electric charge. Such atoms with electric charges are called ***ions***.

According to what is known in physics as Coulomb’s law, unlike electric charges attract one another and like charges repel one another. Certain combinations of positively charged and negatively charged ions can become packed together in a regular three-dimensional array in such a way that the sum of all the forces, attractive and repulsive, among all of the ions in the structure is attractive, meaning that the structure is a stable one and can exist as a mineral. Almost all minerals are of this nature, and are called ***ionic crystals***; the bonds in such a crystal are called ***ionic bonds***. Ions can also be formed from two or more atoms that share their electrons in such a way that each atoms effectively has a filled-shell configuration. The bonds among such atoms are called ***covalent bonds***. Covalent bonds are much stronger than ionic bonds, and the covalently bonded atoms form a single ion that is available for ionic bonding with other ions.

It turns out that in terms of electric charge the silica tetrahedron has a net negative charge. The silica tetrahedra are bonded to various positively charged atoms, mainly of the elements aluminum, iron, magnesium, calcium, potassium, and sodium. These ionic bonds are weaker than the strong covalent bonds within the silica tetrahedra. When a mineral is broken, it breaks along these ionic bonds rather than through the bonds within the silica tetrahedra.

There is an important further complication to silicate minerals, which makes them highly diverse (and almost unique in the universe!): individual tetrahedra can be joined together by sharing one or more (two, three, or even all four) of their oxygens. The phenomenon is called ***polymerization***. By polymerization, zillions of tetrahedra can be joined into chains or sheets that stretch all

the way through a macroscopic mineral grain. In some minerals, all of the oxygens of the tetrahedra are shared, resulting in a complex three-dimensional network of shared tetrahedra.

There are other important minerals besides silicate minerals. Oxide minerals are perhaps the most important kind, in the context of this course. Oxide minerals are those in which negatively charged oxygen atoms are bonded with various positively charged ions. The most important oxides are *iron oxides* and *aluminum oxides*. These oxides of iron and aluminum are generally unspectacular in their appearance, but they are voluminous and important in Earth-surface environments. Carbonate minerals are also important near-surface non-silicate minerals. In carbonates, a negatively charged carbonate ion, consisting of one carbon atom covalently bonded with three oxygen atoms, is ionically bonded with calcium, magnesium, and/or iron ions; calcite and dolomite are common carbonate minerals.

Conclusion

To conclude this section, I'll list just the very most important kinds of minerals in the Earth's upper crust. It's important to know something about these minerals, because the mineralogy of solid Earth materials is of great importance in determining how those Earth materials are affected by their physical and chemical environment.

- *ferromagnesian silicates ("femags"):* This collection of minerals (olivines, pyroxenes, amphiboles, biotite, and certain metamorphic minerals) contain iron and/or magnesium, in various proportions, as their main positive ions. They tend to be dark in color. They are abundant in igneous and metamorphic rocks—but not in sedimentary rocks, because they tend to weather rapidly.
- *micas:* Micas consist of sheets of polymerized silica tetrahedrons, each with three of their four oxygens shared, with various positive ions between the sheets. The main micas are muscovite (a potassium mica), which is common in igneous, metamorphic, and sedimentary rocks, and biotite (a ferromagnesian mineral, with iron and magnesium rather than potassium). Biotite is common in both metamorphic and igneous rocks, but it weathers rapidly, so it's uncommon in sediments and sedimentary rocks.
- *feldspars:* Feldspars are the most abundant kind of mineral in the Earth's crust. They are network silicates with either potassium (these are called potassium feldspar, or just Kspar for short) or varying proportions of sodium and calcium (these are called plagioclase) in their structure. Feldspars are the major minerals of igneous rocks and are common in many metamorphic rocks. Potassium feldspar is common in sediments and sedimentary rocks, but plagioclase is not, because it weathers readily.
- *quartz:* Quartz, with the simple formula SiO_2, is a network silicate mineral with a three-dimensional network of silica tetrahedra, with all of the oxygens shared with adjacent tetrahedra. It is common but not abundant in igneous rocks, is variably abundant in metamorphic rocks, and is very abundant in sediments and sedimentary rocks.
- *iron oxides:* Iron oxides, of which there are several forms, are common only in sediments and sedimentary rocks. In soils, there are several poorly crystallized hydrous iron oxide minerals.
- *aluminum oxides:* In poorly crystallized form, these are characteristic of soils formed by deep weathering of aluminum-bearing rocks under intense weathering conditions.
- *carbonates:* The calcium carbonate minerals calcite and aragonite, along with the calcium–magnesium mineral dolomite, are the main minerals in chemically precipitated carbonate sedimentary rocks, and they are common as well in sediments in the warm and shallow oceans.
- *clay minerals:* Several minerals with sheet-silicate structures are common as very fine-grained plate-shaped particles, some even smaller than a micrometer (one-thousandth of a millimeter; too small even to be seen with a powerful optical microscope). Clay minerals are the main constituent of most fine-grained muds.

This page titled 4.2: Minerals is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.2: Minerals** by John Southard is licensed CC BY-NC-SA 4.0. Original source: https://ocw.mit.edu/courses/12-090-the-environment-of-the-earths-surface-spring-2007/.

4.3: Rocks

Rocks, in the form of what is called ***bedrock***, are exposed at the Earth's surface over wide areas, and they everywhere underlie the unconsolidated surficial materials, called ***regolith***, you will learn about later in this chapter. Moreover, owing to their physical and chemical breakdown under the conditions of the Earth's surface environment they are the source of that surficial material. There's no room in this course for a systematic and detailed treatment of rocks—that's the province of an introductory course in geology—but you should know at least some general things about them.

Rocks are naturally occurring aggregates of minerals. Again, that succinct definition is not very revealing of the nature and variety of rocks. I suppose that it verges on common knowledge that there are three kinds of rocks: *igneous*, *sedimentary*, and *metamorphic*. These three categories are for the most part distinct, although there is some fuzziness to the boundaries.

Igneous rocks are those that form by cooling and solidification of ***magma***. Magma is the term for melted rock in the Earth's interior. At certain times and in certain places in the Earth's shallow interior, down to a hundred or so kilometers, the rocks of the upper mantle or lower continental crust melt, to form magma. The melting is only partial, commonly up to fifteen to twenty percent of the rock, but the magma collects and then moves upward owing to its buoyancy (it's slightly less dense than the surrounding rock). It either becomes parked in subsurface spaces, called magma chambers, there to cool to form coarse-grained igneous rocks like granite or gabbro, or it rises all the way to surface to be ejected from volcanoes, either as flowing liquids or as already solidified particles.

Most of the igneous rocks you are likely to encounter as bedrock outcrops on the continents are of three general kinds (although this is a great oversimplification of the wide range of igneous rocks):

- coarse-grained and light-colored rocks like ***granite***, which consist mainly of feldspar with some quartz;
- fine-grained and dark-colored rocks like ***basalt***, which consist mainly of ferromagnesian minerals and feldspars; and
- rocks, called ***tuff***, that form by lithification of volcanic ash, the solid particles that are ejected explosively from certain kinds of volcanoes.

Sedimentary rocks are those that form from sediments that are deposited at the Earth's surface and become lithified (turned into solid rock, by a variety of processes) as they are buried more and more deeply in the earth's crust. Sedimentary rocks vary widely in composition and origin. There are two major realms of sediments and sedimentary rocks: *siliciclastic* and *chemical*. Siliciclastic sediments and rocks are those that consist of particles of minerals or rock that are derived from weathering of bedrock on the continents. Chemical sediments and rocks are those that are precipitated from natural surface waters, mainly in the ocean. By far the most important chemical sediments are carbonate sediments. Most carbonate sediment is produced by carbonate-secreting organisms, like clams or corals, although some is precipitated inorganically, in somewhat the same way that rock candy grows in a beaker on your windowsill.

Most of the sedimentary rocks you are likely to encounter are of the following kinds. (Again, this is a highly simplified list, in large part because these major sedimentary rock types grade imperceptibly into one another.)

- *sandstones:* Sandstones consist of sand-size particles of various minerals, mainly quartz but also potassium feldspar, as well as sand-size fragments of fine-grained rocks.
- *shales:* Shales consist mostly of very fine-grained mineral particles of quartz and clay minerals.
- *limestones*: Limestones consist mainly of the calcium carbonate mineral calcite. Also abundant are dolostones, which consist mainly of the carbonate mineral dolomite.

Metamorphism refers to changes in mineral composition or rock geometry that occur in solid rocks with increasing temperature and pressure. These changes produce a rock called metamorphic rock. It is important to remember that *changes producing metamorphic rocks take place while the rock is solid*. Any part of the rock that is melted eventually cools to form an igneous rock.

The basic idea behind changes in mineral composition during metamorphism is that each mineral has associated with it a range of temperature and pressure in which it is stable, in the sense that it remains in existence indefinitely. When a mineral is out of its stability range, it is liable to be converted into, or replaced by, one or more other minerals that are stable at those temperatures and pressures. A simple way of phrasing this is that *new minerals grow at the expense of the original minerals*. These changes in mineral composition act very slowly, because the rock remains solid and atoms can be exchanged between adjacent minerals grains

only by atomic diffusion—although very small percentages of pore solutions along grain boundaries can expedite the transport of atoms.

Deformation (change in shape) is an important factor in the development of many metamorphic rocks. For most kinds of metamorphic rocks, change in mineral composition is accompanied by strong shearing deformation. Strong shearing tends to obscure or even obliterate original geometrical features like stratification in a sedimentary or volcanic rock. Also, most metamorphic rocks show some degree of what is called foliation: development of planar features in the rock, like layering or a tendency for the rock to split along parallel planes.

Below are brief descriptions of the most common metamorphic rocks. As with the sedimentary rocks, there are all gradations among these types.

- *slate:* Slate is produced by low-intensity metamorphism of shale. Individual crystals are too small to see, but the rock has a strong tendency to split along parallel planes, as a result of growth of new mica crystals of sheet-silicate minerals all parallel to one another.
- *phyllite:* Phyllite is produced when slate undergoes further metamorphism. Phyllite typically has a silky sheen.
- *schist:* Schist is formed by metamorphism of many igneous and sedimentary rocks. It has clearly visibly flakes of sheet-silicate minerals (muscovite, biotite, and/or chlorite), and it tends to split along parallel planes in the same way as slate and phyllite.
- *gneiss:* Gneiss consists mainly of granular minerals like feldspar, quartz, and ferromagnesian minerals. Gneiss is often prominently layered.
- quartzite: As the name implies, quartzite consists mainly of quartz. It forms mostly by metamorphism of quartz-rich sedimentary rocks.
- *marble:* Marble consists mainly of calcite. The calcite crystals tend to be large enough to see with the unaided eye, having grown larger during metamorphism without change in overall mineral composition.

This page titled 4.3: Rocks is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.3: Rocks** by John Southard is licensed CC BY-NC-SA 4.0. Original source: https://ocw.mit.edu/courses/12-090-the-environment-of-the-earths-surface-spring-2007/.

4.4: Bedrock

By ***bedrock*** I mean *solid rock, exposed at the Earth's surface or buried at shallow depths, that is connected continuously with rock at greater depths in the* crust (Figure 2-4). A great many geologists spend most of their professional lives studying the bedrock of the Earth's continents.

In some places on the continents, bedrock is exposed at the surface, in the form of what are called ***outcrops*** or ***exposures***. There are natural exposures (to which the term *outcrop* is typically applied) and artificial exposures. Natural outcrops range in area from small, human-scale patches peeping out from the surrounding blanket of regolith, to whole mountainsides. Favorite places to find good outcrops are stream beds, hill slopes, mountain tops, and sea cliffs. These days, highway road cuts make the best artificial exposures; in times past, railroad cuts were prime exposures. Building foundations are good but temporary.

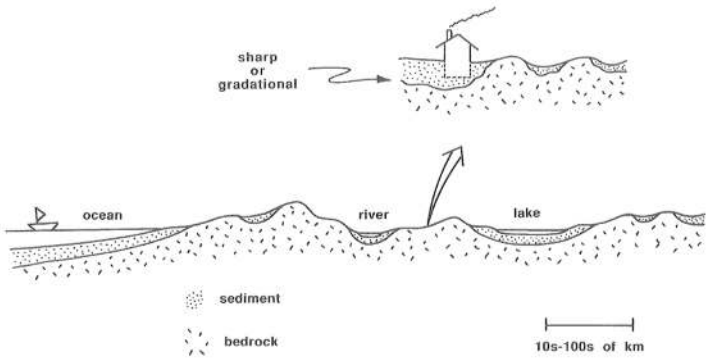


Figure 2-4. The bedrock of the continents

The percentage of continental surface where bedrock is exposed varies greatly, from nearly one hundred percent, in regions with rigorous climate and steep slopes, to zero, in regions with warm humid climates and little topographic relief. (The term ***relief*** is used for *the magnitude of differences in elevation of the land surface in a given region.*) The extent of exposure of bedrock depends mainly on the competition between two opposing processes: the rate of production of new regolith by weathering of exposed bedrock, and the rate of removal of existing regolith by the erosive action of running water or moving glacier ice or by downslope movement of regolith by the pull of gravity. Climate is important because the rate of production of regolith by weathering is much greater in warm and humid climates than in cold and arid climates. In some regions with a warm, humid climate and low relief, one needs to dig or drill tens or even hundreds of meters down to find fresh bedrock!

This page titled 4.4: Bedrock is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.4: Bedrock** by John Southard is licensed CC BY-NC-SA 4.0. Original source: https://ocw.mit.edu/courses/12-090-the-environment-of-the-earths-surface-spring-2007/.

4.5: Weathering

2.5.1 **Weathering Introduction**

Weathering is the term used for the *chemical decomposition and physical disintegration of bedrock at and just below the earth's surface*. Weathering acts upon all bedrock near the surface, although with greatly varying nature and rate depending upon a number of factors.

Rock at depth tends to be brought upward toward the surface, by gradual unroofing by weathering at the surface and removal of the weathering products. It's a matter of relative motion: the bedrock at depth may be either motionless (relative to, say, the center of the earth) or moving slowly upward by uplift or downward by subsidence, but in any case *it's moving upward relative to the earth's surface if material at the surface is being stripped away*. Astherock nears the surface, it tends to be subjected to what's called *near-surface alteration*: it finds itself at much lower pressures and temperatures than at greater depth, and surface-derived waters, which tend to be oxygenated and carry dissolved CO2, have the opportunity to cause certain chemical reactions in many constituent minerals, like ferromagnesian silicates and feldspars. In a sense, you can think of such near-surface alteration as a precursor or forerunner to weathering in the stricter sense. But despite such near-surface alteration, the rock is still solidly bedrock, and nowhere near a condition that could be described as regolith.

Weathering, in the usually accepted sense, acts upon rock that is close enough to the surface to be affected not only by active circulation of surface- derived waters but also by such things as temperature changes and plant growth. But in thinking about weathering you don't have to restrict yourself to the actual rock surface, upon which the sun shines!

Weathering is traditionally subdivided into two sets of processes:

- *chemical weathering*, the chemical decomposition of some or all of the constituent minerals or the bedrock.
- *physical weathering*, the mechanical breakdown or disintegration of the bedrock itself into particles large and small.

This is a useful distinction, but here's a very important fact to understand: *chemical weathering and physical weathering tend to act simultaneously and reinforce one another*.

2.5.2 Chemical **Weathering**

Chemically, *the earth's surface is a very reactive place*. To begin with, in most places and at least at certain times *the surface is an aqueous environment*, and water is not only a very effective solvent in itself but also a necessary medium for a great many chemical reactions.

Owing to the high content of oxygen in the atmosphere, in most places *the earth's surface is a well oxygenated environment*, as in the shallow subsurface wherever the concentration of organic matter is not so great and the permeability so low (as is common in bogs and swamps) that the oxygen is consumed in oxidation of organic matter before it passes downward to deeper depths. (If you're unsure about the nature of oxidation, there will be some background material on it later.) Because so many of the common rock-forming minerals contain ferrous iron (that is, iron in the reduced state, and thus susceptible to oxidation to the insoluble ferric state), *oxidation reactions are an important aspect of chemical weathering*.

In the same way, because the atmosphere contains a substantial percentage of carbon dioxide (CO_2), *surface waters invariably contain appreciable dissolved CO_2*. You know already that CO_2 readily dissolves in water. That dissolved CO_2 reacts with the water itself to form a weak acid, *carbonic acid*, according to the reaction

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \tag{4.5.1}$$

At equilibrium, about one percent of the dissolved CO_2 is in the form of carbonic acid. Most minerals are less stable in an acidic medium than in a neutral medium, so the dissolved CO_2 enhances the weatherability of many minerals.

Specifically, what's going on with Reaction 1? A molecule of carbon dioxide, which consists of a carbon atom tightly bound up with two oxygen atoms, combines with a molecule of water, which consists of an oxygen atom tightly bound up with two hydrogen atoms, in such a way that a single, new, larger molecule, with the chemical formula H_2CO_3, is produced. In the aqueous solution, this process is going on all the time, and at the same time the H_2CO_3 molecules are being changed back into CO_2 and H_2O molecules. Where the balance is struck depends on the imposed concentration of CO_2, which in this case comes either from the

atmosphere or from the chemical breakdown of organic matter. (You probably know that, by photosynthesis, plants take up water and carbon dioxide and convert it into complex carbohydrate molecules, and when the plant dies and the plant tissues are exposed to an oxygenated aqueous environment the reverse happens and the organic matter is broken back down into carbon dioxide and water.)

Carbonic acid is not the only acid common in natural near-surface waters: in most land areas, organisms (both plants and animals) living on or in the soil (that is, the uppermost part of the layer of regolith) use carbon dioxide and various nutrients to produce organic matter, which in its decay back to the constituent compounds tends to produce *a variety of weak organic acids*.

In the rest of this section, several kinds of chemical reaction will show the range of weathering reactions that are important in chemical decomposition of bedrock minerals. They involve *solution*, *oxidation*, *and carbonation*, in various ways. Keep in mind that although these particular reactions are important, *they are just representative reactions*: details vary depending upon the assemblage of minerals that are being weathered, and upon their particular compositions (which, remember, vary widely because of the wide range of ionic substitutions in silicate minerals).

Solution

Some minerals simply *dissolve in water*. Halite (sodium chloride) and gypsum (hydrated calcium sulfate) are good examples:

$$\mathrm{NaCl\,(s) \rightleftharpoons Na^{+} + Cl^{-}} \tag{4.5.2}$$

$$\mathrm{CaSO_4 \cdot H_2O \rightleftharpoons Ca^{2+} + SO_4{}^{2-} + H_2O} \tag{4.5.3}$$

This is the simplest kind of weathering reaction to understand, but not the most important. A more important, and more complicated, kind of solution involves calcite, the constituent of limestone:

$$\mathrm{CaCO_3\,(s) + CO_2 + H_2O \rightleftharpoons Ca^{2+} + 2\,H^{+} + CO_3{}^{2-}} \tag{4.5.4}$$

The reason this reaction is more complicated is that it involves the atmospheric carbon dioxide dissolved in the water. The CO_2 has to be involved because of the CO_3^{2-} ion both in the structure of the mineral and as a product of the solution and then dissociation of CO_2 in water. (Incidental note: this reaction can go in either direction, depending upon the concentration of CO_2: if CO_2 is added to the system, as in weathering environments, calcite is dissolved; if it's extracted from the system, as in certain shallow warm ocean environments, calcium carbonate sediment is precipitated. Think about that the next time you're lying on a white carbonate-sand beach in the Caribbean.)

carbonation of silicate minerals that don't contain aluminum or iron:

Most silicate minerals that don't contain aluminum contain both Fe^{2+} and Mg^{2+} ions in various proportions. Rather than writing a single weathering reaction for such minerals, it's neater to look at separate reactions for those with only magnesium and those with only iron. Such minerals, although they exist, are not common, but their weathering reactions are instructive. First look at the weathering of olivine, which is a common mineral in igneous rocks, and one that's very susceptible to chemical weathering. Most olivines in igneous rocks contain some iron in addition to magnesium, but let's assume, for simplicity, that we are dealing with a pure magnesium olivine:

$$\mathrm{Mg_2SiO_4 + 4\,CO_2 + 4\,H_2O \longrightarrow 2\,Mg_2{}^{+} + 4\,HCO_3{}^{-} + H_4SiO_4} \tag{4.5.5}$$

(I've used the single arrow to the right, rather than the double arrow, here and below because the reverse of these reactions is not actually how the minerals get built under deep Earth conditions.)

Note that "silica" (that is, SiO_2 in some form or other) is extracted from the original silicate mineral. In most natural waters, the dominant species of silica in solution is H_4SiO_4, a weak acid called *silicic acid*. It's in the form of a molecule that consists of four hydrogen atoms, four oxygen atoms, and one silicon atom, all tightly bonded together.

carbonation of silicate minerals that don't contain aluminum but do contain iron:

Now look at a ferromagnesian mineral with all Fe and no Mg. (We'll assume that the mineral is a pyroxene mineral; no details here.) In our oxygen- rich near-surface environment, it's a two-step process. The first step is like that expressed in Reaction 5 above, but the numerical values of the coefficients in front of the reactants and products are somewhat different, because we started

out with a mineral with a different chemical formula. Qualitatively, the result is the same: *the iron-bearing silicate mineral is broken down in the presence of water and carbon dioxide to put iron and silica in solution*. The dissolved iron is in the ferrous ("plus-two") form, which is freely soluble in pure water. It is converted in the presence of oxygen to ferric iron oxide and silica in solution:

$$FeSiO_3 + 2\,CO_2 + 3\,H_2O \longrightarrow Fe_2{}^+ + 2\,HCO_3{}^- + H_4SiO_4 \tag{4.5.6}$$

But that's not the end of the story, because the ferrous iron in solution reacts with

oxygen dissolved in the water, to precipitate an iron oxide mineral, Fe_2O_3:

$$2\,Fe_2{}^+ + 3\,O_2 \longrightarrow 2\,Fe_2O_3 \tag{4.5.7}$$

The iron in the Fe_2O_3 mineral is in the ferric ("plus-three") form, which is almost totally insoluble in pure water.

The representation on Reaction 7 is still a bit of a fake, however, because ferric oxides produced at the earth's surface are commonly hydrous or hydrated: one could describe them as hydrous ferric oxides. There are several closely related minerals of that kind; the most common are goethite, FeO(OH), and ferrihydrite, 5Fe2O3.9H2O). They all have colors that range from yellow through orange to brown; that's why weathered surfaces on bedrock are so commonly of those colors. In a sense, ferrous-iron-bearing rocks *rust* when exposed at the earth's surface.

If we had started with a real pyroxene, with a combination of Fe^{2+} ions, Mg^{2+} ions, and Ca^{2+} ions, the latter two would have been freed to solution while the first would have found its way into insoluble oxides; that's because magnesium and calcium are readily soluble in water but ferric iron is almost completely insoluble.

weathering of aluminosilicates:

Now to the biggest show of all in chemical weathering. The example here will be what is certainly *the single most important weathering reaction on earth: hydrolysis and carbonation of feldspar*. We'll use potassium feldspar as the example; weathering of plagioclase feldspar is similar.

$$2\,KAlSi_3O_8 + 2\,H_2CO_3 + 9\,H_2O \longrightarrow Al_2Si_2O_{25}(OH)_4 + 4\,H_4SiO_4 + 2\,K^+ + 2\,HCO_3{}^- \tag{4.5.8}$$

Comments:

1. The first product on the right is the clay mineral *kaolinite*. (There will be more material on clay minerals in a later section.) Other kinds of clay minerals are possible as solid products of weathering of feldspar as well, depending upon the details of the chemical environment.
2. The products in solution are silicic acid, as before, and potassium ions.
3. If the feldspar had been plagioclase, Na^+ and Ca^{2+} ions would have gone into solution rather than K+ ions.
4. The reaction could have been rewritten slightly to start with CO_2 rather than with H_2CO_3. The differences are immaterial. Those of you who are well versed in chemistry can try to rewrite the reaction by adding to it the following reaction: $2CO_2 + 2H_2O \Leftrightarrow 2H_2CO_3$.

In conclusion, if you don't remember anything else about chemical weathering, remember this:

when feldspar is weathered, by reaction with water and carbon dioxide, the result is a clay mineral, positive ions (K^+ and/or Na^+ and/or Ca^{2+}) in solution, and silica, mostly in the form of H_4SiO_4, in solution.

In intense weathering situations, in warm and humid environments, weathering of aluminosilicate minerals like feldspar can go a step further, to the ultimate weathering product:

$$2\,Al_2Si_2O_{25}(OH)\ 4 \longrightarrow 4\,Al(OH)3 + 4\,H_4SiO_4 + 5\,H_2O \tag{4.5.9}$$

The kaolinite, which itself was the weathering product of potassium feldspar, is stripped further of all of its silica, to produce a totally insoluble aluminum hydroxide mineral and more silica in solution. This is the ultimate weathering residue! There are several hydrous aluminum oxide minerals that come out of a reaction like this; the mineral *gibbsite* is the most common. Surficial materials that are especially rich in such insoluble hydrous aluminum oxides are called ***bauxite***. As you probably know, such weathering-residue mixtures of hydrous aluminum oxides are the world's supply of aluminum ore. Depending upon the composition of the original rocks, such ultimate weathering residues are commonly mixed with insoluble hydrous iron oxides as well.

One final point, which will be important for our later consideration of sediments and soils: all of the solid weathering products in the reactions above— clay minerals and various oxides of aluminum and ferric iron—are commonly *of finer particle size than the minerals from which they are produced during weathering*.

What are the controls on chemical weathering? Viewed on the large scale, there are four major controls on chemical weathering:

- *climate:* The intensity of chemical weathering increases with precipitation and temperature. The reason it increases with precipitation is that weathering requires an aqueous medium for the reactions. The greater the percentage of time a given mass of weatherable rock is saturated with surface water, the more chemical weathering it is likely to experience. The reason that the intensity of chemical weathering increases with temperature is that, as with all chemical reactions, the rate of the reaction increases with temperature. That's basically because the vigor of the thermal agitation of the atoms and molecules increases with temperature.
- *rock type:* Some kinds of rocks are more susceptible to chemical weathering than others. Rocks that contain minerals that were formed under conditions of the high temperatures and pressures of the deep crust or upper mantle are much more weatherable than rocks that contain abundant minerals like quartz, calcite, or muscovite, which are stable, or at least only slowly weatherable, under most Earth-surface conditions. For that reason, most igneous and metamorphic rocks are generally more susdceptible to weathering than most sedimentary rocks.
- *relief:* Regions with high relief usually also have steep slopes. Gravity-driven mass movements (more on that in a later chapter) then are effective at stripping away newly weathered material, thereby exposing fresher bedrock to continued chemical weathering. In regions with low relief, however, the layer of regolith thickens to the point where the underlying bedrock is largely sealed off from further contact with the surface environment of chemical weathering.
- *vegetation:* As mentioned earlier, when plant materials decay they release carbon dioxide and weak acids into the regolith environment. These substances acidify the surficial material, thereby tending to drive the various chemical weathering reactions discussed above to the right.

There's another way of looking at the controls on chemical weathering: from the standpoint of *what a small mass of weatherable material actually senses*. This is the small-scale view, in contrast to the foregoing largescale view. It boils down, largely, to four more specific physicochemical factors:

- mineral composition of the rock
- chemical composition of the water
- rate of passage of water through the rock • temperature

2.5.3 Physical Weathering

Most rocks are strong: you need to pound on them with a hammer to break them. The strongest rocks, like tough quartzites, yield only reluctantly to even the most vigorous hammer blows. Other rocks, like soft sandstones, you can break with your own hands. (In fact, there is a continuous gradation between loose and unconsolidated sediment and strongly lithified sedimentary rock, depending in large part upon depth of burial.) Rock strength is a complicated topic, which is important for the construction industry, as well as for geoscientists who try to understand rock fracture that causes large earthquakes.

Let me a bit more specific about the nature of the strength of solids like rocks. There are three ways of breaking solids: by *compression*, by *tension*, and by *shear* (Figure 2-5). In rocks, the compressive strength and the shear strength are both large, but the tensile strength is much less: generally, it's much easier to break a rock by pulling it apart than by squeezing it. (That's true of concrete, an artificial rock, as well.)

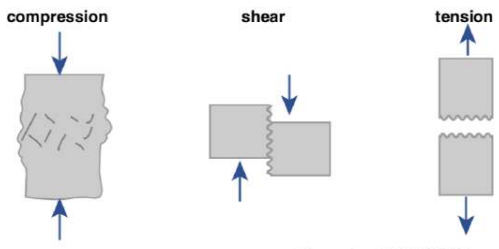


Figure by MIT OCW.

Figure 2-5. Three ways of breaking solids.

The easiest way to break a rock is to *wedge open a preexisting crack*. That's partly because of the low tensile strength, but also because of an effect known as *stress concentration* at the tip of the propagating crack (for reasons that are beyond the scope of this course). A simple analogy can serve well here: it's much easier to "break" a sheet of paper in tension by starting a cut and then pulling the edges of the cut apart than by grasping the sheet at opposite edges and pulling on it uniformly. When a rock fractures in tension, either rapidly as a result of a blow or an explosion, or slowly, by a variety of natural weathering processes, the crack enlarges as the tip of the crack propagates into the non-fractured rock.

Breakage of polycrystalline aggregates like rocks (or metals) is more complicated than breakage of homogeneous materials like glass. That's because *the crystal grains of the rock are joined together at the grain boundaries*, so the strength of the grain contacts needs to be taken into consideration as well as the strength of the mineral grains themselves. When a rock is stressed, extremely fine cracks tend to develop first along the grain boundaries.

Rocks at the surface tend to be broken down (disaggregated, disintegrated) by a variety of processes, whose importance varies greatly from place to place and from time to time. These physical processes (you might also call them mechanical processes) have traditionally been cited as important in physical weathering, but definitive quantitative studies of the various effects are discouragingly scarce. I'll set them out for you here without making much of an attempt to rank them in importance, except for some concluding comments at the end of this section.

sheeting: Sheeting is *a style of rock fracturing that develops sets of fractures just below the bedrock surface and concordant with it*. The reason such fractures tend to form is that as rock is brought up to the surface it is *unloaded*: the weight of overlying rock is reduced. The rock tries to expand slightly (rock has small but definite compressibility), and *it can expand only upward, not sideways*. This leads to development of fractures that are roughly parallel with the rock surface. Sheeting is most common in otherwise massive and non-fractured bedrock. The sheeting joints decreases rapidly in spacing downward from the surface, usually disappearing within several meters. The next time you see a photo of the mountain slopes in Yosemite, think in terms of sheeting fractures.

exfoliation: Exfoliation (Figure 2-6) is the *separation, during weathering, of successive shells, a few millimeters to some tens of centimeters thick, from massive bedrock*. The mechanical effect is somewhat like that of sheeting: as the thin near-surface layer of bedrock undergoes chemical weathering, the weathering-product minerals tend to be of lower density that the original minerals, which means larger volume, so the thin uppermost zone of the rock tries to expand. It can expand only upward, so fractures parallel to the rock surface are formed. In this situation, chemical weathering and mechanical weathering act together to disintegrate the bedrock. A related process leads to what's called ***spheroidal weathering***: the exfoliation and spalling is slightly more effective at edges and corners than on rock faces, so in an outcrop cut by sets of planar fractures in various orientations (such fractures are called ***joints***; they are ubiquitous in bedrock outcrops) *the edges and corners are rounded off* to produce a very spheroidal or "blobby" look to the outcrop. In advanced cases the spheroids are completely freed from the outcrop, and lie around on the surface or roll downslope.

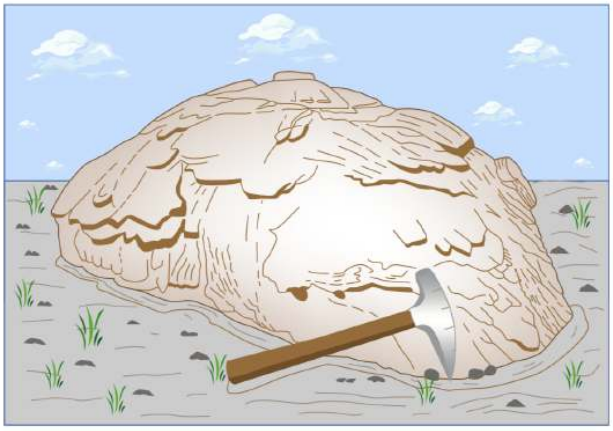


Figure by MIT OCW.

Figure 2-6. Disintegration of a surface boulder by exfoliation. (From Hunt, 1972.)

granular disintegration: Granular disintegration is another process that is in part a mechanical effect of chemical weathering. Many if not most rocks consist of a mixture of minerals, some susceptible to chemical weathering and others not. As the weatherable minerals expand upon incipient chemical weathering, while the non-weatherable minerals retain their original size, intergranular forces are set up, which tend to break the bonds between mineral grains. The mineral grains fall from the rock surface, one by one, to leave an overall smoothly rounded but granular-rough surface. The mass of loose grains at the base of the outcrop is called ***grus***. This kind of weathering is most common in crystalline rocks like granite.

thermal expansion and contraction: Expansion and contraction in response to changes in temperature can by itself aid in propagation of fractures in bedrock ranging in scale from intergranular fractures to widespread joints. Such a process is most effective in arid regions where the day-to-night temperature differences are great. It's unclear whether the process is effective in thoroughly dry rocks, or whether the presence of water is an important element in the process. Experiments on heating and cooling of rock samples, over a great many cycles to simulate many decades or even centuries of the natural environment, have not been successful in producing non-negligible fracturing, but maybe the conditions of the simulation have not been representative. It's well known, however, that fire causes breakdown of rock by sudden expansion of the surface layer exposed to the extreme heating.

frost wedging: Frost wedging (Figure 2-7) is important in climates where surface water is abundant and the day-to-night temperature range spans the freezing point many times through the year, as in New England in the late fall and especially the early spring. Water expands upon freezing, to the tune of about 9% in volume, an extremely large change. When water confined deeply in a thin crack in bedrock freezes, it can exert enormous side forces on the walls of the crack. The maximum possible force, which has been computed to be over 2000 kilograms per square centimeter (!), is not anywhere close to being achieved, however, because the ice does not seal the crack as strongly as the rock walls themselves, but the process is nonetheless very effective. Mountaintops with expanses of bedrock subjected to freeze–thaw cycles tend to be a jumble of frost-shattered rock; hikers in the Presidential Range in New Hampshire know that well.

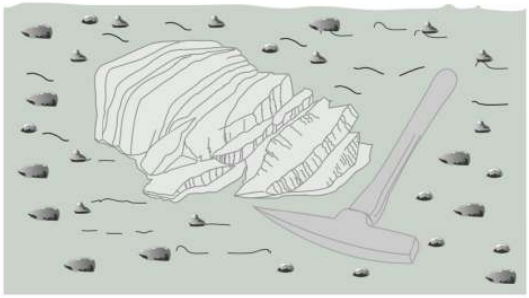


Figure by MIT OCW.

Figure 2-7. Disintegration of a surface boulder, probably by a combination of freeze-thaw, thermal expansion and contraction, and chemical weathering. (From Hunt, 1972.)

hydrofracturing: It's thought by some investigators that the frost-wedging effect is more complicated than just simple expansion upon freezing. Water as thin films in very small cracks may stay liquid even at very low temperatures, for subtle thermodynamic reasons. As the freezing front progresses inward from the rock surface, water can be forced under high pressures into microscopic cracks, causing propagation of fractures. The phenomenon is known as hydrofracturing.

"boulder lifting": It's often said that boulders "grow" in New England fields. Frost action provides a good explanation for the effect (Figure 2-8). As the water-saturated ground freezes from the top down, freezing reaches the base of a large boulder sooner than it reaches the surrounding soil, because the rock of the boulder has a much lower heat capacity than the water-rich soil around it and is therefore cooled faster. Freezing of the water beneath the boulder lifts or pushes the boulder upward a short distance. Upon thawing of the ice beneath the boulder, the boulder does not fall as far as it was raised, because of the tendency for collapse and flowage of soil and finer rock fragments into the space beneath the boulder. Given enough cycles, the boulder makes its appearance at the surface.

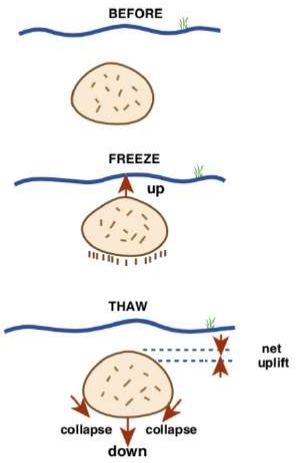


Figure by MIT OCW.

Figure 2-8. Uplift of a boulder by freezing and thawing of surrounding watersaturated regolith.

root growth: Root growth can be very effective at widening preexisting cracks in bedrock near the surface. All of you have seen plants growing heroically in cracks in rock. In many cases the plant just occupies an already large crack, but in other cases it's clear that the growing root has opened the crack.

Which of these various processes are the more important, and which are the less important? I can do no more than make an educated guess. (It depends, of course, on the particular setting of bedrock and climate, in any case.) In the right circumstances, *frost wedging* is clearly of great importance, but *granular disintegration*, and perhaps *exfoliation*, are probably of more general importance. Thermal expansion and contraction alone is generally considered to be less important than the other processes, except perhaps in particularly favored environments.

2.5.4 Weathering Profiles and Weathering Fronts

You've probably had the experience of picking up a broken stone and noticing that the outer part is colored reddish or brownish while the inner part looks a fresh gray. What's going on is that the stone, having been derived from some outcrop as a fresh piece and then perhaps rounded at least somewhat by transport, rested somewhere at the Earth's surface for a time long enough for the outer zone to have experienced some chemical weathering. The boundary between the outer, weathered zone and the inner, non-weathered zone is called the ***weathering front***. There's commonly a very abrupt change in the degree of weathering at the weathering front. In many cases the characteristic difference in color arises from the conversion of ferrous iron in minerals of the fresh rock to ferric iron oxides in the weathered zone.

The same effect tends to be present on a much larger scale when a bedrock landscape in an area of low relief is subjected to protracted weathering and the weathering products remain where they were formed. To see the effect, you need to have a vertical or steeply dipping cut down through the weathered rock, all the way down into fresh bedrock, which may be many meters or even tens of meters below the surface. Above the weathering front is a gradation from the most weathered material, at and near the land surface, down to the least weathered rock, at the weathering front. The succession of zones of degrees of weathering, from the surface down to the weathering front, is called a ***weathering profile***.

Weathering profiles vary greatly in their nature, depending upon the factors listed in an earlier section (largely a matter of rock composition and climate). Figure 2-9 shows a common weathering profile, of the kind that tends to develop on deeply in-place-weathered coarse-grained igneous rocks like granite. The material just above the weathering front, called ***saprock***, consists of partly chemically weathered minerals together with yet-unweathered minerals. The saprock preserves all of the original textures and structures of the parent rock, but its composition has been somewhat changed and its strength has been reduced. The saprock grades upward into material called ***saprolite***, which is more altered chemically than the saprock below but still retains much of the textures and structures of the parent rock. You might describe such rock as "rotten rock" (and that is in fact the etymology of the term saprolite). Above the saprolite is even more strongly altered material, which has lost even more of the coherence of the parent and

shows little of the original textures and structures. Above this, where the material has lost its coherence entirely and could be described as “loosed”, it is called ***regolith*** (see later section). The uppermost layer is the ***soil*** (see later section).

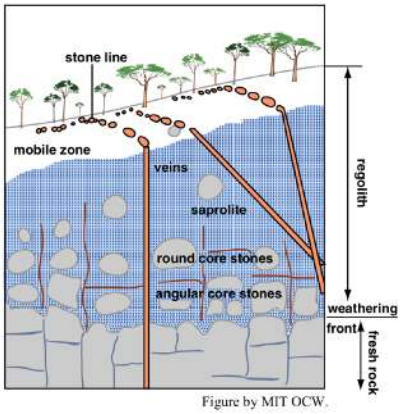


Figure 2-9. A typical weathering profile developed by deep, in-place weathering of coarse-grained granitic igneous rock. (From Taylor and Eggleton, 2001.)

Note in Figure 2-9 that *the geometry of the weathering front is strongly controlled by the presence of joints in the fresh rock*. The weathering can make better progress along the joints than in the non-jointed intervening rock. In the process, there’s a tendency for non-weathered masses to be left behind as the weathering front progresses downward. These rounded, irregular masses are called ***corestones***. They are similar to the weathering spheroids mentioned earlier, but they differ in being formed well below the land surface, whereas weathering spheroids develop at the rock surface as weathering proceeds at the same time that the weathering products are removed from the outcrop.

This page titled 4.5: Weathering is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.5: Weathering** by John Southard is licensed CC BY-NC-SA 4.0. Original source: https://ocw.mit.edu/courses/12-090-the-environment-of-the-earths-surface-spring-2007/.

4.6: The Mineral Particles of the Earth's Surface Materials

2.6.1

It seems a good idea, at this point, to describe the most common kinds of mineral particles in the ground under your feet—the materials that in the sections to follow are called regolith, sediment, and soil (regolith being the most inclusive term, as you will see). I think it's accurate to say that *most regolith consists mostly of mineral particles*, coarse and fine. A wide variety of kinds of minerals can be found in typical regolith, but only a very few are of any abundance.

2.6.2

In terms of relatively coarse mineral particles, at the top of the list is **quartz**. Quartz is common in most of the source rocks from which surface materials are derived, and it's largely immune to chemical weathering. This is a first-order fact: *most of the mineral grains in most of the regolith you're ever going to see (with the notable exceptions of fine muds) consist of quartz*. Quartz grains are mostly equant (meaning approximately equidimensional) in "overall" shape, and usually in between subangular (that is, a bit more rounded than a classically jaggedy angular particle) and well rounded in "local" shape. The degree of roundness depends partly on the shape of the particle upon being freed from the parent rock by weathering and partly on the extent of mechanical wear the quartz particle has experienced, and perhaps also on some dissolution if the quartz particle is exposed to acidic environments (because the solubility of quartz in water, although extremely low in pure water, increases markedly with increasing acidity of the water).

2.6.3

Potassium feldspar is also important as coarse particles in regolith, depending upon the composition of the parent rock. Potassium feldspar is more susceptible to weathering than is quartz, but commonly some or even much of it escapes weathering before it is deposited as sediment and passes out of the reach of near-surface processes. Survival of potassium feldspar to be seen in regolith is especially common in situations where weathering is curtailed, either by a cold and rigorous climate or by rapid stripping of the weathered material from steep slopes. Commonly, potassium feldspar in regolith is partly or mostly weathered to kaolinite, and under magnification has a "punky" or "chalky" appearance.

2.6.4

Rock fragments are also common in regolith. This is especially the case for sediments. Imagine a raging river eating its way into fractured or partly weathered bedrock: it can entrain fragments of even fresh rock, large and small, and transport them to sites of eventual deposition. Glaciers can do the same thing even more effectively. Identifying large rock fragments is easy; also important, however, are sand-size fragments of various common fine-grained rocks. It takes careful microscopic work to identify such fragments.

2.6.5

Muscovite (the most common form of mica, but not the only one) as coarse particles is common in regolith as well. Micas are sheet-silicate minerals. Sheet silicates are those in which the silica tetrahedra are polymerized by sharing of three of the four oxygens to form a sheet structure, with hexagonal symmetry, and with all of the unshared oxygens pointing in the same direction. Figure 2-10 is a drawing of the sheet structure, looking down on the side with the unshared oxygens. There's more to the mica structure than that, however: the basic sheet unit consists of two such silica sheets, each with their unshared oxygens pointing inward, with aluminum ions in between the two silica sheets. (If you want more detail on mica structure, see the clay mineral "advanced topic" below.)

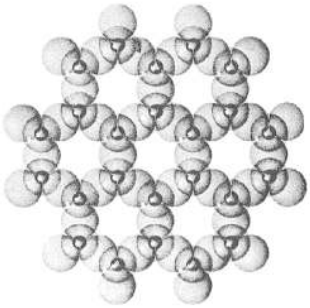


Figure 2-10 Polymerization of silica tetrahedra to form the sheet structure, a basic building block of sheet silicates.

2.6.6

The mica crystal is very weak in the direction parallel to the sheets of the structure, so in regolith it's usually in the form of plates or flakes. When it's coarse, so it can be seen with the unaided eye or with a hand lens, it is easy to identify. In sediments it's usually noticeably coarser than the equant particles like quartz, because its platy shape makes it easier for flowing fluid to transport it

2.6.7

In terms of relatively fine mineral particles in regolith, at the top of the list is the group of **clay minerals**. These are a group of minerals rather than a single mineral. Clay minerals are mainly sheet-silicate minerals. Some, called illite or smectite, have the same structure as muscovite, and nearly the same composition. Owing to various substitutions of some ions for others, though, the range of compositions of illite and smectite is wide. In fact, sedimentologists who have to deal with clay minerals often just refer to *illite—smectite* (or I–S for short). Kaolinite and chlorite, with yet different sheet structures, are other common clay minerals. For more detail on the nature of clay-mineral structures, see the following optional "advanced topic". The proportions of the various clay-mineral types in a given sample of regolith vary depending upon two major factors: *source-rock composition*, and *weathering environment*

2.6.8

An important thing for you to know about clay minerals is that *clay- mineral particles are all very small*, mostly less than a few micrometers. They range in size down to very small fractions of a micrometer, into what is called the colloidal size range. There will be more on colloids in the later section on soils, where they are especially significant.

Advanced Topic: Clay Minerals

I. Introduction

The word "clay" is used, somewhat ambiguously and confusingly, in three different but largely overlapping senses:

(1) As a *mineral* term, for a group of minerals, mostly but not entirely sheet silicates, that are produced from weathering of aluminosilicate minerals (mainly feldspars). They are almost always very fine, from the sub-micrometer range up to several micrometers, and they are mostly platy in shape, although some are curvy and curled. The principal kinds are *kaolinite*, *smectites*, *illites*, and *chlorites*. (I've used the plural for the last three because there's a very wide compositional range in each of them.)

(2) As a *size* term. According to the official grade scale for sediment sizes (see the later section on sediment), all particles finer than 1/256 of a millimeter, or about 4 micrometers, are termed clay.

(3) As a *material* term ("clay material", or just "clay"), which has a sticky and tenacious consistency. It's the stuff that potters use.

Most clay-mineral particles are of clay size; most clay-size particles are clay-mineral particles; and most clay material consists of clay-mineral particles of clay size!

As you can imagine, *clay minerals are not easy to study*: even in the coarser size range it's difficult to see them microscopically. Standard techniques for study include electron microscopy, which allows you to see the particle shapes beautifully, and x-ray diffraction, which allows you to identify their mineralogy, after a fashion. More sophisticated instrumental techniques have been

developed more recently. Clay mineralogy is worth a whole course in itself. What I'll do here is just give you a cursory account of the most important clay minerals and their occurrence.

II. Structure

Two structural units are involved in clay-mineral structures:

octahedral sheets: two planes of closely packed oxygens and hydroxyls (OH^-) with a plane of Al, Fe, and/or Mg between. The Al, Fe, and/or Mg are in octahedral coordination with the O and OH; that is, each of the former is equidistant from six of the latter, which are arranged as an octahedron around the former. Figure 2-11 shows a rough sketch of the arrangement. Figure 2-12 is a more detailed view of the geometry, perpendicular to the sheet. The oxygen-to-oxygen distance is 2.94 Å, and the thickness of the sheet is 5.05 Å. If Al is in the octahedral positions, only two-thirds of the positions are filled; if Mg and/or Fe are in the octahedral positions, all of the positions are filled.

tetrahedral sheets, in which silica (SiO_4) tetrahedra are polymerized by sharing of three of the four oxygens to form a sheet structure with hexagonal symmetry. Figure 2-13 shows two diagrammatic representations. The oxygen-to-oxygen distance is 2.55 Å, and the thickness of the sheet is 4.65 Å.

These two layers, or sheetlike arrangements, fit together, one on top of the other, in such a way that the vertices of the tetrahedra in the tetrahedral sheet point toward the octahedral sheet. The unshared oxygens of the tetrahedral sheet and the oxygens of the octahedral sheet are the same; they are shared for double duty. The hydroxyls that form part of the octahedral arrangement fit into the hexagonal holes of the tetrahedral sheet.

The dimensions of the two sheets are almost but not quite right for this shared arrangement. There has to be some strain in the lattices to make them go together to share oxygens. This seems to be accommodated by alternate cocking of the tetrahedra, to contract the tetrahedral sheet a little, and stretching and thinning of the octahedral sheet to make it a little bigger. I'll schematize this two-layer structure as shown in Figure 2-14.

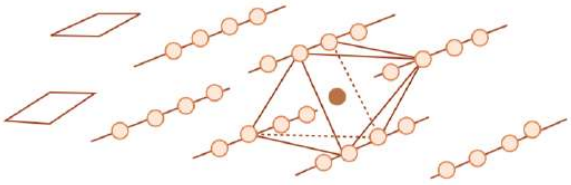


Figure 2-11. Sketch of the arrangement of atoms in the octahedral sheet of clay- mineral structures.

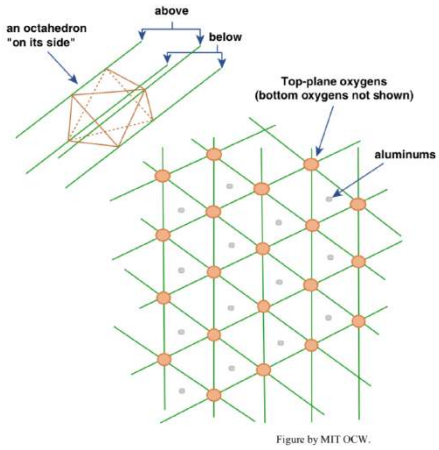


Figure 2-12. A more detailed view of the geometry of the octahedral sheet of clay-mineral structures.

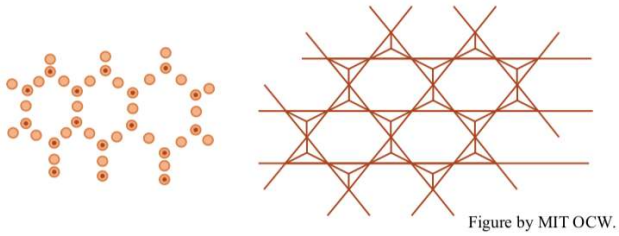


Figure by MIT OCW.

Figure 2-13. Diagrammatic representations of the arrangement of silica tetrahedra in the tetrahedral sheet of clay-mineral structures.

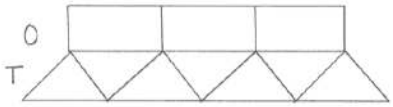


Figure 2-14. Highly diagrammatic sketch of the combination of octahedral and tetrahedral sheets in two-layer clay-mineral structures.

It's also possible to construct a three-layer arrangement in which a tetrahedral layer is sandwiched between two tetrahedral layers. Figure 2-15 shows the highly schematic way I'll represent this three-layer structure.

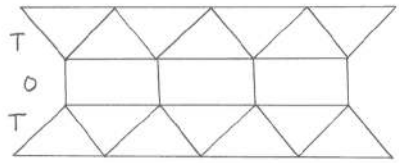


Figure 2-15. Highly diagrammatic sketch of the combination of octahedral and tetrahedral sheets in three-layer clay-mineral structures.

III. The Major Clay Minerals

Kaolinite:

Kaolinite (Figure 2-16) is the main two-layer mineral. It's fairly simple chemically: $Al_4(Si_4O_{10})(OH)_8$, with no isomorphous substitutions in either the tetrahedral sheet or the octahedral sheet; it's the purest of the clay minerals. Only two-thirds of the octahedral positions are filled, by aluminum; minerals like that are called *dioctahedral*. There's no net charge on the tetrahedral–octahedral layers; the layers are held together only by a kind of weak bonds called van der Waals bonds. The repeat distance normal to the layers is 7.2 Å, so this group of minerals is sometimes called *seven-Ångstrom clays*. The crystals can get relatively large, because there's very little stress in the structure; kaolinite is the coarsest of the clay minerals. Kaolinite is formed mainly by weathering of feldspars; production of kaolinite is favored by acidic and warm environments.

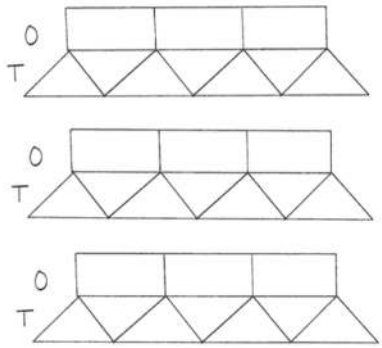


Figure 2-16. Highly simplified sketch of the structural organization of kaolinite.

Illite:

Illite (Figure 2-17) is a hydrous muscovite-like clay mineral. It has the same three-layer structure as muscovite, but less than the one-in-four replacement of Si by Al that's characteristic of muscovite (the figure is between one-in-five and one-in-six). So there isn't the same charge deficiency as in muscovite, and not as many K+ ions in the interlayer positions. H3O+ ions are also present in the interlayer positions. Also, substitutions in the octahedral layer are greater and more random; there's lots of variability. Illite is never found well crystallized, only as clay-size particles. The particle size is typically less than one micrometer. This is because there's considerable stress in the layers. Illite is closely similar in structure to the common muscovite you see in sedimentary, metamorphic, and igneous rocks; it's just much finer-grained. Illite is formed mainly by weathering of potassium feldspar in temperate weathering environments. With increasing intensity of weathering, kaolinite is formed instead.

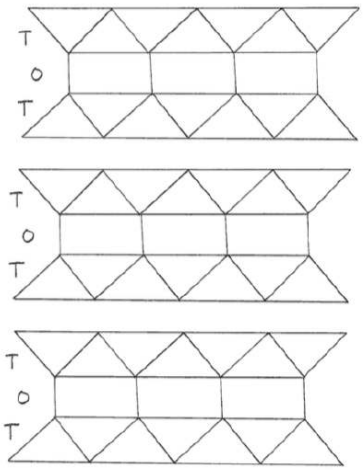


Figure 2-17. Highly simplified sketch of the structural organization of illite and smectite.

Chlorite:

Chlorite (Figure 2-18) is a kind of mixed-layer clay mineral. Its structure is an alternation of trioctahedral TOT layers, with Mg^{2+} and Fe^{2+} in the octahedral positions, and layers with the composition $(Mg^{2+}, Fe^{2+})_3(OH)6$, giving a repeat spacing of about 14 Å. There's a great range and diversity of ionic substitutions in each of the three ion sites, as well as differences in the geometry of stacking of the sheets Chlorite is produced in abundance only in relatively mild weathering environments, where it survives from the source rocks, although in the form of finer particles. So it's characteristically a high-latitude product.

2.6.9

Various oxides of iron and aluminum are especially important in the highly weathered soils of warm and humid regions, where (as you saw in the section on weathering) silicate minerals are broken down further to oxides. The main aluminum oxide is **gibbsite**. The formula for gibbsite, $Al(OH)_3$, shows that it is a hydroxide mineral. Gibbsite is formed by further weathering of silicate clay minerals, in which all of the silica is stripped out of the clay minerals; see Reaction 9 in the section on chemical weathering, above. The main iron oxide mineral in soils is **goethite**, FeOOH. It's goethite that imparts the characteristic yellow-orange color to weathered iron-bearing rocks and regolith. Recall from the section of chemical weathering that ferrous iron is freed from ferrous-iron-bearing silicate minerals and then oxidized to minerals like goethite. These aluminum and iron oxides (there are others, as well) in regolith are in the form of extremely small particles, in the colloidal size range. (See the later section on soils for more on colloids.)

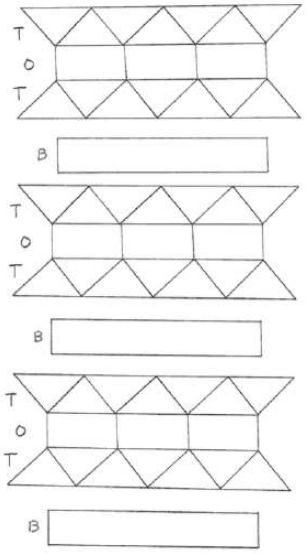


Figure 2-18. Highly simplified sketch of the structural organization of chlorite.

This page titled 4.6: The Mineral Particles of the Earth's Surface Materials is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.6: The Mineral Particles of the Earth's Surface Materials** by John Southard is licensed CC BY-NC-SA 4.0. Original source: https://ocw.mit.edu/courses/12-090-the-environment-of-the-earths-surface-spring-2007/.

4.7: Regolith

The Nature of Regolith

What is regolith? The term ***regolith*** is used for *the layer or mantle of fragmental and unconsolidated rock and mineral material, whether residual or transported, that rests on bedrock*. One might quibble that the definition excludes fragmental material that happens to be locally cemented by surficial processes but is otherwise closely related to nearby unconsolidated material. (You'll be hearing more, much later in the course, about one such kind of material: caliche.) The definition is not specific about age; most regolith is young, geologically and even by human standards, but some is much older, not having been subjected to consolidation, for one reason or another, for geologically long times.

For most of us geologists, continental bedrock is our bread and butter: we inspect it minutely, hammer on it, sample it, photograph it, and think about it day and night. But in most parts of the world, finding bedrock to inspect is not easy, and in many places almost impossible: it is covered by a layer of regolith. When was the last time you walked across an outcrop of bedrock? When we are out and about in the natural world, away from the ubiquitous pavement that's so characteristic of our civilization, we are mostly walking on regolith. If only for that reason alone, study of the Earth's regolith should be high on the list of desiderata for Earth scientists. A good case can be made that *the Earth's mantle of regolith is for humankind by far the most important component of the solid Earth beneath our feet*: it feeds us, in the form of soil. In fact, however, there is a tendency for regolith studies to be low on the priority list. (I think that there lurks in the minds of many geologists the notion that bedrock studies hold far more interest, and glory, than regolith studies.)

Regolith originates by the complex of processes collectively termed weathering, which was the subject of the preceding section. What happens to regolith once it is produced? You can detect from the foregoing definition that regolith is either *residual* or *transported*. The terms are largely self-explanatory, are they not? In areas of low relief and with a climate conducive to deep weathering, a thick mantle of regolith, in some places many tens of meters thick, lies where it has been produced. In most places, however, the regolith we see has been moved from its place of origin, nearby or far away, by various agents of transportation.

Keep in mind, however, that it's usually not a matter of an episode of production and then a later episode of transportation: generally, both production and removal are going on at the same time. The thickness of the mantle of regolith depends in great part on the relative magnitude of the rate of production and the rate of removal. In some situations, as in regions with steep slopes, high relief, and rigorous climate, regolith is stripped away as fast as it is produced; in areas with gentle slopes and climates conducive to deep weathering, the layer of regolith is very thick.

Agents That Mobilize Regolith

Regolith resting on a sloping surface is pulled inexorably downward by the force of gravity (or, more accurately, by *the downslope component of the force of gravity*) (Figure 2-19). Such movement of the regolith may be imperceptibly slow or at speeds in excess of a hundred meters per second. An entire later chapter will be devoted to the subject of downslope movements of near-surface material by gravity, collectively termed ***mass movements*** (absolutely no relation to sociopolitical mass movements).

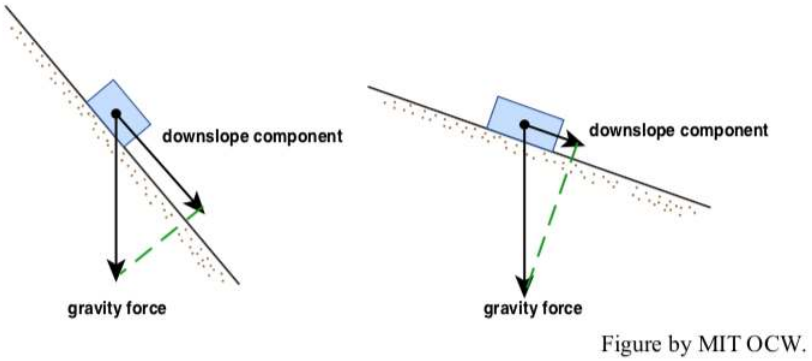


Figure 2-19. The downslope component of gravity acting on the layer of regolith. A) On a steep slope. B) On a less steep slope.

The principal agent for the transport of regolith, aside from the direct pull of gravity, is *the flow of water in streams and rivers*. Of course, river flow is itself a consequence of the downslope pull of gravity—but regolith transport by flowing water involves physical processes and effects that are fundamentally different from the direct pull of gravity. The mobilization and transportation of regolith by flowing water is the subject of part of the later chapter on streams and rivers. Movement of regolith by the action of land-based glaciers is of great importance in certain Earth-surface environments as well. Wind can also be a significant transporting agent, as described in the final chapter.

Kinds of Regolith

The main distinction in kinds of regolith is between ***residual regolith*** (also called ***sedentary regolith***), on the one hand, and ***transported regolith***, on the other hand (Table 2-1). With respect to transported regolith, several kinds are recognized, depending upon the agent of transportation. Transported regolith is collectively termed ***sediment***. There will be much more material, later in this chapter as well as in later chapters, on sediment, and how it is eroded, transported, and deposited. Colluvial deposits (called ***colluvium***) are those transported down steep slopes by the pull of gravity. Mainly, these are talus deposits (also called scree deposits) on steep slopes, and avalanche and landslide deposits of various kinds. Alluvial deposits (called ***alluvium***) are those transported and deposited by rivers and streams. In addition, there are glacial deposits and eolian (wind-blown) deposits.

regolith	residual	soils of various kinds; residual subsoil materials
	transported	colluvium: talus, scree, etc.
		alluvium
		glacial sediment
		eolian sediment

Table 2-1. Classification of regolith. (After Merrill, 1897.)

This page titled 4.7: Regolith is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.7: Regolith** by John Southard is licensed CC BY-NC-SA 4.0. Original source: https://ocw.mit.edu/courses/12-090-the-environment-of-the-earths-surface-spring-2007/.

4.8: Sediment

When regolith is mobilized or entrained, by flowing water, or by the wind, or by moving glacier ice, it is called ***sediment***, because, as the term implies (the Latin verb *sedere* means to sit), it eventually comes to rest again as a deposit. Here I will touch upon only a few aspects of sediments. You will hear much more about sediment movement, and sediment deposits, in later chapters, especially the chapters on rivers, on glaciers, and on coasts.

One of the characteristic things about sediment is that *it tends to become fractionated, by size, during transport*. The basic reason is that a given transporting agent, like water or the wind, moving at a given speed, can move sediment particles up to a certain size but no larger—although another important effect is that the water or the wind moves finer particles faster than coarser particles. The end result of such fractionation is that the sorting of the particles in the deposit is usually much better than in the original material that was the source of the transported material. (The ***sorting*** of a sedimentary deposit describes the spread of particle sizes around the mean particle size: in a well sorted sediment the particle sizes cluster closely around the mean, whereas in a poorly sorted sediment the particle sizes range widely around the mean.)

For convenience and clarity in communication, sedimentary geologists have adopted a series of terms for sediment particles over the entire range of particle sizes (Figure 2-21). Such a subdivision is called a ***grade scale***. It's based on powers of two, so it's geometric (or, what comes out mathematically to be the same, logarithmic), rather than arithmetic, for reasons that are obvious but not easy to justify: for example, the difference in behavior between a 1 mm particle and a 2 mm particle is bound to be far greater than the difference in behavior between a 101 mm particle and a 102 mm particle, right? You can see from Figure 1-20 that in geoscience such things as mud, silt, sand, and gravel are not just qualitative concepts but precisely defined entities!

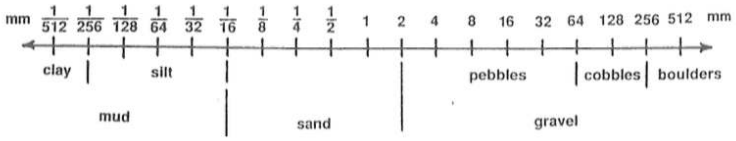


Figure 2-20. The Udden–Wentworth grade scale for sediment particle size.

The logic behind the grade scale in Figure 2-20 is fine for well-sorted materials, but how about poorly sorted materials? Various classifications have been proposed, but the one that's in most common use (with some variations) among regolith specialists is shown in Figure 2-21. Here you see what "loam" (so commonly pronounced "loom" locally!) really is, in the technical sense.

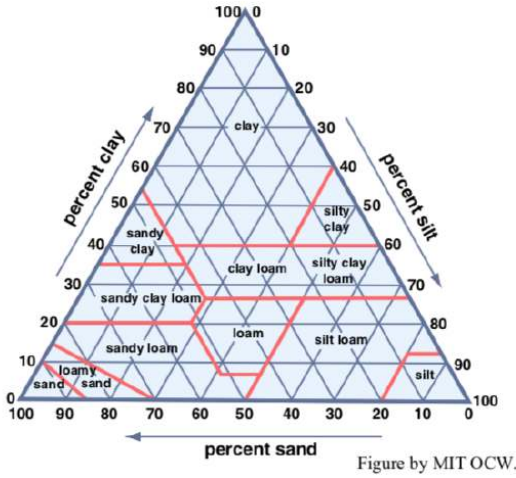


Figure 2-21. Conventional system for classification and terminology of sediment mixtures. (From FitzPatrick, 1986)

Advanced Topic: Measuring Sediment Size

1. Scientists or engineers who are responsible for studying surface deposits usually have to measure or characterize, in some way, the distribution of particle sizes. For large particles, in the gravel size range, it's possible to measure the size of each piece

separately, with calipers, although that's a challenging endeavor unless the particles is very well rounded, and even then the matter is not entirely straightforward (as it would be if the particles were perfect triaxial ellipsoids). For smaller particle sizes, in the silt to fine pebble range, the classic way of measuring particle size is by ***sieving***.

2. Here is how particle size is measured by sieving:

- make up a stack of special sieves, from coarsest on the top to finest on the bottom, with a lid on the top and a pan on the bottom;
- put a small sample of deposit in the top sieve;
- shake the stack until the particles have found their size bin;
- weigh each of the fractions on a balance;
- compute, for each sieve size, the total weight percent finer than that size;
- plot the results on graph paper with size along the horizontal axis, logarithmically, and, along the vertical axis, weight percent finer than, from 0% to 100%, arithmetically.

An example of the result, called a ***cumulative distribution curve*** (common in many other applications of statistics as well), is shown in Figure 2-22.

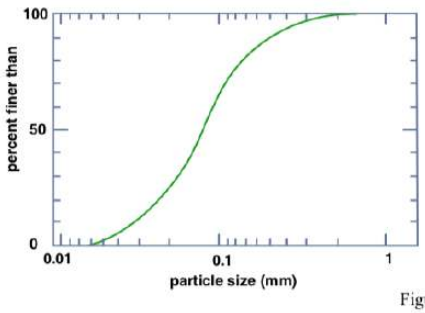


Figure by MIT OCW.

Figure 2-22. A typical cumulative distribution curve of sediment size.

3. The problem with sieving is that it works well only for sizes down to about 0.05 mm. For sediment finer than that, there are various techniques that involve suspension of the sample in water, settling of that suspension, and sampling during settling. The basic principle is that the coarser sizes settle faster than the finer sizes. Devices that are used to measure fine particle sizes in this way are called ***settling tubes***. The governing principle is that each particle has a terminal fall speed that is a function of its size. In recent years, a variety of more sophisticated instruments have been developed to measure particle size, either optically or sonically, by sensing each particle as it passes through a thin tube in suspension. Such instruments are now obtainable "off the shelf". There is still the tricky matter of intercalibrating between results from such instruments and traditional settling tubes.

This page titled 4.8: Sediment is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.8: Sediment** by John Southard is licensed CC BY-NC-SA 4.0. Original source: https://ocw.mit.edu/courses/12-090-the-environment-of-the-earths-surface-spring-2007/.

4.9: Soils

2.9.1 Introduction

Along with oxygen and water, soils are essential to human existence. Almost all of our food comes, directly or indirectly, from crops grown in the soil. The only significant exception, of course, is food from the oceans. Hydroponically grown crops are an insignificant exception—and the infrastructure and cost involved in hydroponic agriculture are likely to doom it forever to a minor role in human food supply. All of our lumber, most of our paper products, and much of our clothing ultimately come from the soil. Our dependence on soils will continue, but the world stock of soils cannot increase. In fact, it is *decreasing*, because of soil erosion, as well as urbanization (and suburbanization). As our reserves of fossil fuels continue to shrink, energy resources from biomass (living plant material) are likely to increase.

The meaning of the term "soil" varies greatly depending upon the field of endeavor in which it is used. *Agriculturalists* use the term narrowly to mean the uppermost layer of surface deposits, in which plants grow and agriculture is practiced. The *soil scientist* has a somewhat wider view: soil is the uppermost layer of surface deposits, almost always extending well below the agriculturalist's soil, which has developed upon underlying material by the influence of various soil-forming processes acting over long times. The *geologist* has an even broader view of soils as surface materials developed by weathering processes that have broken down bedrock into regolith, and the *civil engineer* (who practices the discipline of soil mechanics, and who is responsible for the siting and design of structures with foundations) takes a similarly wide view of soil as all materials at the surface that can be excavated without blasting. In the following sections, we'll look at the nature of soils mainly from the standpoint of the soil scientist (what is the nature of soils, how do they develop, and how are they classified?).

The study of soils as natural bodies is called ***pedology*** (and its practitioners call themselves soil scientists, or pedologists). You can also think in terms of the ***pedosphere***. (There is a tendency among science educators nowadays to think in terms of "spheres": geosphere, hydrosphere, biosphere, atmosphere, cryosphere. These terms are useful, in a broad and loose sense, to describe zones or domains of related materials and processes. I think that the suffix "*-sphere*" began to be used because in a very general sense these domains are like spherical shells around the Earth. I myself think that "sphere-ology" tends to be overdone on the part of the "spherophiles", but it remains a useful pedagogical device.)

Soils can be thought about and studied on a wide range of scales: from large regions of continents, in the context of major soil types, to the distribution of soils in a local area of varied topography and substrate composition (Figure 2-23A), to the vertical profile of soil at a given point on the land surface (Figure 2-23B), to the texture and structure of the soil at a point in such a profile (Figure 2-23C), to the details of physics, chemistry, and biology of the soil on the colloidal and atomic scale (Figure 2-23D).

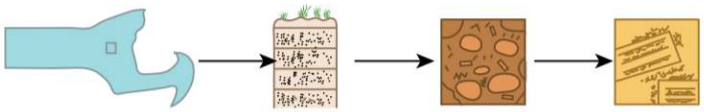


Figure by MIT OCW.

Figure 2-23: Thinking about soils on a wide variety of scales. (Inspired by Brady and Weil, 2002.)

Soils have a great variety of constituents. Here is a list of the most important kinds of soil constituents:

- mineral particles
- colloids
- organic matter
- microorganisms
- soil solutions
- soil gases

Most *soils consist mainly of mineral particles*. Such soils are often called ***mineral soils***. Some soils, however, consist mostly of organic matter. In general, soils contain the same kinds of mineral particles as do sediments; go back and review the section on sediments above. Most of the soils you are likely to encounter consist mainly of quartz, plus variable percentages of potassium

feldspar and rock fragments, in the coarse fraction, and mainly particles of clay minerals, iron oxides, and aluminum oxides, in the fine fraction. The remaining constituents in the list above are described in the following sections.

2.9.2 Colloids

Colloids are *mixtures of materials in which one substance exists as extremely small particles, called colloidal particles, which are dispersed in a second substance, which you can think of as the medium*. In the case of colloidal particles in a uniform liquid or gaseous medium, the term colloidal solution is commonly used. In the context of soils, we're dealing with small particles consisting of minerals or organic matter that are dispersed in the aqueous pore solutions in the soil or are aggregated together in great numbers, or cling to the larger soil particles. Colloids are important in a great many materials, both natural and manufactured.

Most clay-mineral particles have a platy shape, and they have a tendency to stick together in aggregates of many particles, held together by certain short-range electrostatic forces. Such forces are small, but so are the particles— and it's important to realize that because *the ratio of surface area to volume increases with decreasing particle size*, clay-mineral particles are in a sense "almost all surface and very little volume". The state of aggregation of the clay- mineral fraction is widely variable, and the macroscopic strength and consistency properties of the deposit vary greatly corresponding to that. Also, as you will soon see, the small size of the particles, together with the electrical charges characteristic of their surfaces, make the tiny particles chemically extremely reactive.

Colloidal particles are in a kind of "no person's land" between true solutions and undoubtedly solid particles. At one extreme are ions and molecules in an aqueous solution. The smallest ions are of the order of atomic dimensions, about 0.1 nanometers (that is, about 10-10 meters). Many molecules that consist of a large number of atoms are much larger, but they still behave like smaller ions in true solutions. At the other extreme are particles larger than a few micrometers, which is large enough to be seen with light microscopes. Such particles are large enough to settle in the surrounding fluid under the influence of gravity.

Typical soils contain several important kinds of colloids. Here's a briefly annotated list of the most important:

- ***clay-mineral particles:*** There are several types; see the earlier background section on clay mineralogy. These types differ in their physical and chemical properties, partly because of their differing size but mainly because of their differing surface charges and their differing susceptibility to take on or release positive ions and water molecules to and from their internal surfaces between the TOT layers.
- ***iron and aluminum oxides:*** These are especially important in the highly weathered soils of warm and humid regions, where silicate minerals are broken down further to oxides. The main aluminum oxide is *gibbsite*, and the main iron oxides are *goethite* and *ferrihydrite* (no details of mineralogy here).
- ***humus:*** These are especially abundant in the upper parts of soils, nearest the surface. Humus colloids are not crystalline structures: they are complex molecular chains and rings of carbon atoms bonded to hydrogen, oxygen, and nitrogen atoms. They are among the smallest of colloidal particles in soils. They have an enormous capacity to adsorb water onto their surfaces.

The great importance of soil colloids arises from their chemical reactivity. There are two reasons for that: *surface area*, and *surface charge*. The following elaborates on those points.

- ***surface area:*** Chemical reactions between a solid and the fluid with which it is in contact take place at the surface of the solid, at the interface between the solid phase and the liquid phase. The rate of such a chemical reaction depends upon a number of important factors, aside from the nature of the reaction itself. Temperature is one of those important factors: in general, the higher the temperature, the faster the reaction. That's because the energy of thermal vibration of the atoms and molecules in both the solid phase and the liquid phase increases with temperature. Also important, however, is *the area of the surface at which the reaction takes place, per unit volume of the reacting system*. The extremely small size of colloidal particles means that their surface area, per unit bulk volume of material, is spectacularly large.
- ***surface charge:*** At any given tiny region inside a mineral particle, the electric charges associated with the ions of the structure balance out. At the surface of the particle, however, the picture is different—because the ions right at the surface are bonded only to the ions below them. That means that there is a kind of "unsatisfied" electric charge at the surface of the particle. With regard to clay-mineral particles, the surface charge is generally negative on the broad faces of the exposed TOT sheets and positive along the narrow edges of the plates. For macroscopic mineral particles, like sand grains, these surface charges are

almost entirely negligible, but when the particles are extremely small their enormous area-to-volume ratio makes these surface charges very important.

Background: Surface Area

Take a nostalgia trip back to secondary-school math and think about the formulas for the volume and the surface area of some regular geometric solid. A sphere is a good example. The volume of a sphere is $(1/6)\pi D^3$, where D is the diameter of the sphere. In scientific parlance, *the volume goes as the cube of the diameter*. The surface area of the sphere is πD^2, so the surface area goes as the *square* of the diameter.

What's important here is not the particular formulas, or the values for volume and area of a particular sphere they give, but *how the ratio of surface area to volume varies with the diameter of the sphere*. You can easily see from the two formulas given above that the ratio of surface area to volume of any sphere is $[\pi D^2]/[(1/6)\pi D^3]$, or, simplifying, $6/D$. Ignore the factor 6 there; what's important is the $1/D$ part. The ratio of surface area to volume goes as $1/D$. That means that *as the diameter decreases, the ratio of surface area to volume gets bigger*.

Now for a revealing example of this $1/D$ effect. Think about a sphere 10 cm in diameter—about the size of a softball, a grapefruit, or a ball of pizza dough. It has a surface area of about 0.0314 square meters (or, in scientific notation, 3.14×10^{-2} m^2). Now compare that with a sphere 1 mm in diameter. That sphere has a surface area of only 3.14×10^{-6} m^2, far smaller, by a factor of ten thousand. But suppose you took the entire volume of the 10 cm sphere and converted it into an equivalent volume of 1 mm spheres. With some further simple math, and using the formula for the volume of a sphere, you would determine that the 10 cm sphere is equivalent to 106 (that's one million) of the little ones. Now the total surface area of the little ones is (106) (3.14 x 10-6) m^2, or 3.14×100 m^2. That doesn't sound like a lot, but compare it with the surface area of the original 10 cm sphere, which was 3.14×10^{-2} m^2. We see, then, that the aggregate area of the little spheres is a hundred times the area of the volume-equivalent big sphere! This is a concrete example of the $1/D$ effect: we reduced the diameter of the sphere by a hundred (ten centimeters to one millimeter), and the surface area of the equivalent volume increased by a factor of a hundred!

The combination of surface area and surface charge is a potent one. We need to deal with two different but related effects: coagulation, and ion exchange. Figure 2-24 is a cartoon that shows the various phenomena discussed in the following paragraphs.

coagulation: Clay-mineral particles, with their negative charges on faces and positive charges on edges, tend to clump together by electrostatic attraction, in an edge-to-face arrangement. (Keep in mind that like charges repel and unlike charges attract.) This tendency for coagulation is one of the important factors that make clay-rich soils sticky and tenacious.

ion exchange: As you will see in a later chapter, on water, the water molecule has a positively charged "side" and a negatively charged "side". Because of that, water molecules tend to be attracted to the charged surfaces of the colloidal particles. Picture a swarm of water molecules, perhaps up to several molecules thick, at the clay-particle surface, jiggling around and jostling one another, now and then trading places with other water molecules from the solution beyond. That layer of water molecules acts as a kind of "cushion" for the particle. Included in the layer of loosely bound water molecules are positive ions, which are attracted to the negatively charged faces of the clay particles. Also, smectite clays take up and give off water and positive ions to and from their internal surfaces. These loosely bound water molecules are the last to be extracted by plant roots as a soil dries between rains. You know that potassium is one of the most important plant nutrients. Its continued abundance in the soil is largely the result of its attraction to colloids, together with recycling in organic matter. Because of this combination of attraction to colloids and recycling by plants, under natural conditions the storehouse of available potassium, released by chemical weathering of primary minerals, is maintained at a workable level for plant growth in the face of continuing loss in solution to downward-percolating water from the surface.

2.9.3 Organic Matter

The uppermost parts of surface deposits—the soil of the soil scientist and especially of the agriculturalist—usually has an abundance of organic matter as well as inorganic mineral matter. It's not easy to specify what's meant by organic matter. It comprises a wide and complex variety of chemical compounds that a chemist would call organic compounds. Organic compounds are those that consist of carbon atoms, bonded to one another and to atoms of other elements, mainly oxygen and hydrogen but

including such elements as nitrogen and sulfur as well, in a great variety of ways. Organic compounds are both natural and laboratory-synthesized.

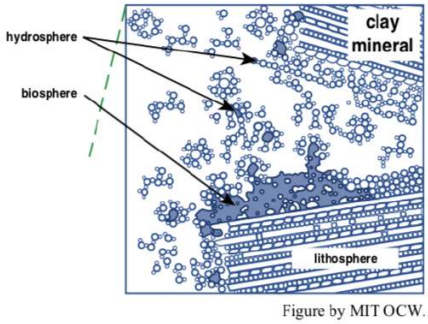


Figure by MIT OCW.

Figure 2-24. Submicroscopic view of a clay-mineral colloid particle in its aqueous environment.

The organic matter of soils is *ultimately derived from plant growth*. (Living plants and animals are not conventionally considered to be organic matter: you have to wait until the organism dies!) Nonliving plant tissues themselves are organic matter, as are the great variety of organic compounds that result from partial decay of that plant material. The tissues of dead animals and their decay products also represent soil organic matter, although not nearly in the same abundance as plant-derived organic matter.

A deep understanding of soil organic matter would necessitate a much greater familiarity with organic chemistry (the branch of chemistry that deals with organic compounds) than is appropriate for this course. Here we can only touch upon some of the most important considerations and kinds of materials.

2.9.3.2 The Role of Soil Organic Matter in the Carbon Cycle

The element carbon is a key substance in the workings of the Earth's near-surface environment. That comes about because of its important role in living organisms, soil organic matter, the atmosphere and the oceans in the form of gaseous and dissolved carbon dioxide, and carbonate sediments. There will be a more detailed treatment on the Earth's carbon cycle in a much later chapter on geochemical cycles. It must suffice here to point out that the soil plays a major role in that carbon cycle. It's been estimated that the Earth's soil contains almost twice as much carbon as the stock of living vegetation and the atmosphere combined.

Plants take up carbon dioxide gas from the atmosphere and convert it, by sunlight-powered photosynthesis, into the organic compounds that constitute plant material. When the plants die, some of the dead plant material is decomposed back to carbon dioxide and water by oxidation. In part that decomposition involves non-biological chemical reactions, but in great part it is mediated by the metabolic activity of soil microorganisms. The end result, however, is the same: production of CO_2. Some of that CO_2 remains dissolved in soil water and some is released into the atmosphere.

2.9.3.3 Kinds of Soil Organic Matter

There are many kinds of organic matter. These compounds range from easily decomposable to very difficultly decomposable. A typical plant consists of easily transformed compounds like polysaccharides and lipids, stiffened with the much more resistant materials cellulose and lignin, the principal components of woody plant tissue. Lignins, of which there are various kinds, are the slowest plant tissues to decompose. Only a few kinds of organisms, mainly fungi, can break down lignins.

The ultimate residue left behind by breakdown of plant materials is called ***humus***. What's called humus comprises a number of very stable organic compounds. Humus is always dark in color, the reason being that the chemical structures in the humus absorb nearly all wavelengths of visible light. Humus is so fine grained that its particles (which in reality are large molecules) are at the fine end of the colloidal size range. These compounds, whose basic building blocks are carbon hexagonal ring structures (called aromatic compounds, in organic chemistry) with large molecular weights (that is, containing very large numbers of polymerized atoms in their structures). They are the organic materials that are the most resistant to microbial decomposition. The water-holding capacity of humus, mentioned in an earlier section, is several times greater than that of clay-mineral particles.

Decomposition of humus is a slow process. The typical half-life of humus (that is, the time it takes to decompose half of an initial mass of the material) is measured in many decades, or even a few centuries! An important effect that slows the decomposition is

complexing of the molecules of the humus with inorganic clay-mineral particles. By poorly understood processes, this association protects the humus from decay.

2.9.3.4 The Role of Organic Matter in Soils

With respect to plant growth, humus is an essential soil component. It's common knowledge among gardeners that plants grow better in soils rich in humus. There are various reasons for that, both direct and indirect. It's not just a matter of supply nutrients: humus is not particularly rich in nutrients (despite what many organic gardeners have been led to believe). Most of the nutrients needed by plants are inorganic ions dissolved in aqueous soil solutions, although humus does play some role in the nutrient supply for plants.

It's the indirect effects of humus on soil properties that are the most beneficial aspects of humus. Humus can be viewed as a soil conditioner. These effects involve physical, chemical and biological properties of the soil In terms of *physical effects*, humus reduces the stickiness and cohesion of clay-rich soils, and it promotes the formation of granular texture, so prized by agriculturalists because it enhances the tillability of the soil. In terms of *chemical effects*, humus is even more effective at adsorbing and holding onto nutrient ions than are clay-mineral particles. In terms of *biological effects*, humus is the most important substrate for the life activities of soil microorganisms; it provides most of the food for detritus-feeding soil organisms.

If you are willing to be overwhelmed by the multitude of ways that organic matter affects soil properties, plant productivity, and environmental quality, look at Figure 2-25.

2.9.4 Soil Organisms

The last really important component of the deposit comprises soil organisms: the plants and animals that spend part or all of their lives in the soil. The soil is one of the Earth's most important ecosystems. (For a brief brush-up on ecology and ecosystems, see the background section below.) Every handful of soil contains literally billions of organisms—almost all of them microorganisms, too small to be seen with the unaided eye. The microbiology of soils is a fully developed and very important discipline in itself, and I can only touch upon it here.

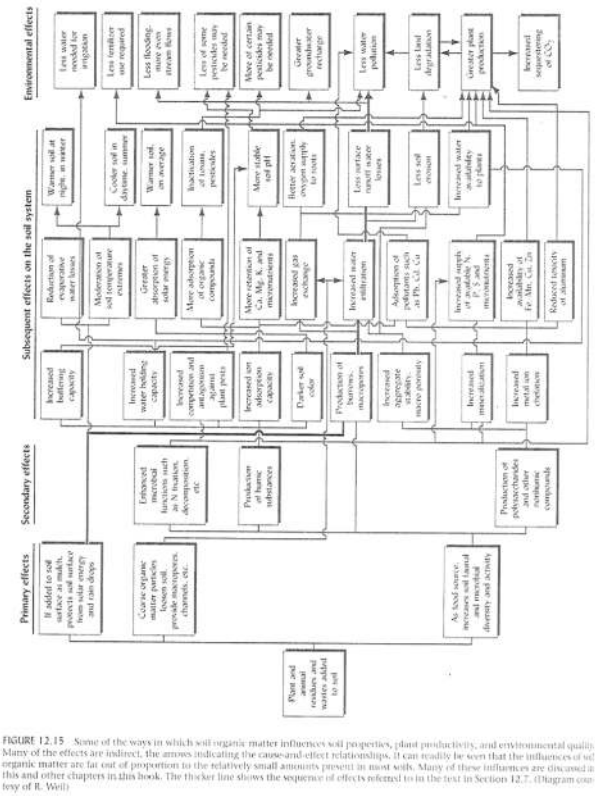


FIGURE 12.15 Some of the ways in which soil organic matter influences soil properties, plant productivity, and environmental quality. Many of the effects are indirect, the arrows indicating the cause-and-effect relationships. It can readily be seen that the influences of soil organic matter are far out of proportion to the relatively small amounts present in most soils. Many of these influences are discussed in this and other chapters in this book. The thicker line shows the sequence of effects referred to in the text in Sections 12.7. (Diagram courtesy of R. Weil.)

Figure 2-25. How organic matter affects soil properties, plant productivity, and environmental quality. (From Brady and Weil, 2002.)

https://geo.libretexts.org/@go/page/22343

Background: Ecology and Ecosystems

1. The term ***ecology***, which tends to be used too loosely in the popular media these days, is the study of *the interrelationships between organisms and their environment*. The largest entity studied in ecology is the ***ecosystem***, which consists of *some chosen part of the physical and chemical environment and all of the organisms within it*. An ecosystem involves all of the physical, chemical, and biological processes and interactions that operate within the given part of the environment. An ecosystem can be as large as the entire biosphere or as small as a puddle of water at our feet.

2. What are the elements of an ecosystem? A ***habitat*** is *the local environment in which a given organism lives*. Ecosystems usually involve a number of different and distinctive habitats. A similar but different concept is that of an ***ecological niche***, which is *the position of the organism in its habitat*, including its way of life and its role in the ecosystem. Most habitats are occupied by several species, each with its own ecological niche. Each species in an ecosystem is represented by a number (usually, but not always, large) of individual organisms, called a ***population***. A population of two or more species that occupy a given habitat is called a ***community***. There may be more than one community in a given ecosystem.

3. Within any ecosystem, there typically are many kinds of interactions among constituents, both living and nonliving. These interactions can be viewed in terms of flows of matter and energy through the ecosystem. Organic compounds are synthesized from the environment by ***producers***, which in all but the most specialized ecosystems are photosynthesizing plants. The producers are in turn consumed by plant-eating animals, called ***herbivores***. Some of the herbivores are in turn consumed by ***carnivores*** (or by ***omnivores***, which each both plants and animals). Other elements of the ecosystem are ***parasites***, which feed on living organisms without killing them, and ***scavengers***, which feed on dead organisms.

4. The sequence of species ranging from the producers, at one end, to carnivores that no other carnivores eat, at the other end, is called a ***food chain***. Often the term ***food web*** is more appropriate, because the actual situation is more complicated than a simple linear arrangement of species. The tissues of producers, herbivores, and carnivores that are not consumed by species higher in the food chain are broken down by organisms called ***decomposers***, which are usually bacteria.

5. When we think of ecosystems, we tend to think of tropical rainforests, or broad expanses of grasslands on which large mammals roam around and get photographed by ecotourists. In a real and important sense, however, *a shovelful of natural soil is an ecosystem*, and a very complex one at that.

One way of classifying soil organisms is into plants and animals. There's more to that than it might seem, however, because although we all learned, when we were small, that life is divided into an animal kingdom and a plant kingdom, most biologists these days recognize *five* kingdoms. But that's another story we need not deal with in this course. There are two more useful ways, for us, of classifying soil organisms:

- *heterotrophs vs. autotrophs:* ***autotrophs*** are organisms that generate their own tissues from inorganic raw material, by use of (mainly) solar energy. Plants are autotrophs. ***Heterotrophs*** are organisms that use organic matter produced by other organisms as food.
- soil organisms can be classified as ***macroorganisms*** (organisms with size ranging down to that of, say, earthworms), ***mesooorganisms*** (organisms ranging down to a small fraction of a millimeter), and ***microorganisms*** (organisms of microscopic size, smaller than about a tenth of a millimeter).

The numbers of kinds of organisms per unit volume of a soil increases strongly with decreasing organism size. A typical soil might contain several species of vertebrates and several species of earthworms, along with as many as several tens of species of insects, dozens of species of mesoorganisms, but thousands of species of microorganisms. Moreover, the numbers of organisms per unit volume increases spectacularly with decreasing size: although, obviously, numbers vary widely depending upon the type and condition of the soil, numbers of macroorganisms and mesoorganisms are large but not overwhelming, whereas the numbers of microorganisms, per cubic meter, say, might be in the trillions!

A whole chapter should be devoted here to the nature and significance of the life activities of soil organisms. Interesting as they are, the activities of what have been called ***ecosystem engineers*** (vertebrate burrowers, earthworms, ants, termites, and the like) in modifying the physical nature of the soil pale in comparison with the life activities of soil microorganisms. Microorganisms play

the essential role of decomposing soil organic matter (mainly plant residues) and converting it to carbon dioxide, water, dissolved ions, and the ultimately resistant residues called humus earlier in this section.

Several major kinds of microorganisms are well represented in soils. The most important are *algae*, *bacteria*, and *fungi*, and various kinds of *protozoans*. Algae, which are photosynthetic plants, are abundant only in the uppermost parts of deposits, very near the sunlit surface. Fungi and protozoans are important but restricted to the shallow zones. Bacteria, on the other hand, range widely downward to great depths.

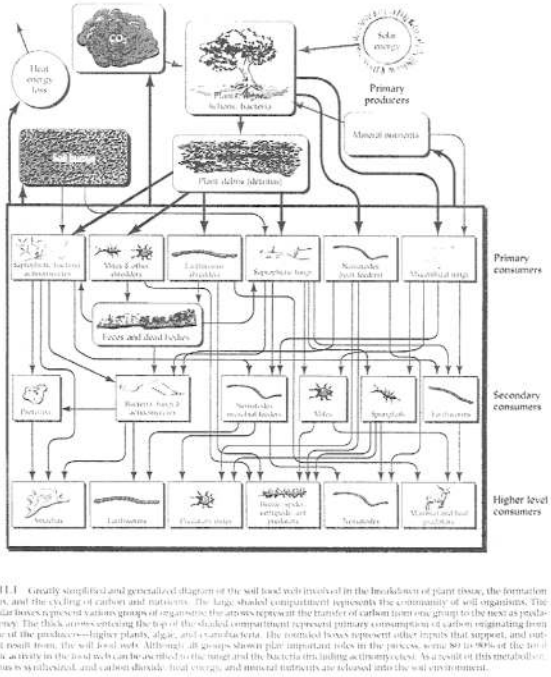


Figure 2-26. The soil food web. (From Brady and Weil, 2002)

Figure 2-26 shows a diagram of the soil food web. I've preserved the original caption for you. Note that it's described as "greatly simplified and generalized"!

2.9.5 Soil Solutions and Soil Gases

All soils are porous to some extent (that is, they contain empty spaces, called ***pore spaces*** or ***void spaces***, among the various solid constituents). It stands to reason that the pore spaces must contain either a gas or a liquid, or some of both. Whether the pore spaces are occupied by solely by liquid or whether they are occupied mainly by a gas depends upon where the region of soil lies relative to the groundwater table: below the groundwater table the soil is fully saturated with an aqueous pore solution, whereas above the groundwater table the pore spaces are occupied mainly by gas.

The reason I have used the term "aqueous pore solution" rather than just "pore water" is that the water in the pore spaces contains a great variety of dissolved substances, in varying concentrations. Just after a rainstorm, as the rainwater percolates downward to replenish the water table, the concentrations of dissolved materials is very low (except for dissolved atmospheric gases), but as time goes on some of the constituents of the soil particles are leached out to go into solution. The chemistry of such soil solutions is a important part of soil science.

Above the groundwater table, the pores spaces are filled mainly with gas. The composition of soil gases is not greatly different from that of the atmosphere, with one major exception: owing to continuing decay of organic matter, the concentration of carbon dioxide is typically substantially higher than in the overlying atmosphere.

Even in the soil zone above the groundwater table, the soil is seldom entirely dry: some residual water clings tenaciously to the soil particles. I will defer discussion of that until the chapter on groundwater.

2.9.6 Processes of Soil Development

The development of a soil (in the agriculturalist's and soil scientist's sense) depends upon five major (and interacting) factors, as shown in Figure 2-27. Of these, you might think that the parent material is the most important. It's generally believed, however, that climate is even more important than parent material. The relative importance of climate and parent material depends to a great extent upon time of development: in the initial stages of soil development, parent material is certainly the more important, but, as time goes on, soils developed upon different parent materials in a given climate tend to become more and more similar. Obviously, important differences as a function of parent material must remain: if for one soil the parent material is rich in quartz, and for another soil the parent material is a limestone or a mafic igneous rock like basalt, the soils can never converge.

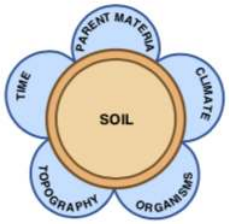


Figure by MIT OCW.

Figure 2-27. The five factors of soil development. (From FitzPatrick, 1986.)

Topography is generally of secondary importance; its role lies largely in governing the downslope creep of already developed soil, which tends to "freshen" the uppermost layer of the surface deposit by exposing less weathered material beneath.

The importance of organisms lies in their ability to add the all-important organic matter to the soil and to process that organic matter back into inorganic compounds, including nutrients for yet more plant growth.

I might also point out that *the five factors are not entirely independent of one another*; for example, organisms depend strongly on climate, and topography depends quite strongly upon climate as well, and also upon parent material and time.

This is a good place to point out that *some soils develop directly upon bedrock, but in many if not most cases, soil develops upon surface deposits, called sediments earlier in this chapter, that were derived from bedrock somewhere else and transported to the site of soil formation*. In some places, the surface deposit is grossly "out of place". That's true of almost all glaciated areas, New England being no exception. It's also true of areas of deposition from streams and rivers. In such areas, soil formation is usually nowhere near the "ultimate" or "equilibrium" soil, of the kind you would find in a low-latitude, non-glaciated area of low relief, where bedrock is weathered very deeply over a very long period of geologic time and the soil is highly developed.

(But just because a soil has had a long time to develop doesn't mean that it will be rich and deep: in the rain forests of the humid tropics, intense leaching removes nutrients released by chemical weathering so rapidly that the soil is thin, and has only a thin layer rich in organic matter at the surface; the lush stand of plants is in tenuous existence, recycling nutrients very efficiently. When such vegetative cover is removed, the soil rapidly becomes irreversibly barren.)

A variety of physical and chemical processes act to develop soil from fresh regolith. One of the most important of these we have already dealt with: weathering of preexisting bedrock and/or further weathering of relatively fresh and unweathered surface deposits. Another important process is vertical transport of matter in either dissolved or colloidal form by upward or downward flow of soil water. Such processes come under the soil scientists' term ***translocation***. Translocation also includes vertical mixing or churning by macroorganisms, most importantly earthworms and certain kinds of insects. Biological processes, most importantly ***humification*** of plant material, and chemical and organism-aided processes of ***cycling*** of nutrients and other compounds (most importantly nitrogen and carbon) are further soil-forming processes.

The nature of *water flow in the vertical direction in the soil*, which is largely a function of climate (intensity and frequency of rainfall; the effect of temperature, humidity, and vegetative cover upon evaporation from the soil), determines whether matter in solution undergoes net downward transport or net upward transport in the soil. This is the aspect of climate that's most important in soil development, along with the nature and rate of growth of vegetation.

2.9.7 Soil Profiles

Soil scientists classify soil on the basis of the sequence and nature of vertical zones or layers, called ***horizons***, in the soil. Each of the enormous variety of soil types has *a characteristic succession of such horizons*, each with its thickness, composition, physical properties, and transitions to overlying and underlying horizons. The particular succession of horizons is called a ***soil profile***.

To have a good picture of a soil profile, you need to go into your backyard and dig a trench, carefully, about a meter deep and wide enough for a good side view, with one side planar and almost vertical and shaved off carefully for good viewing with minimal smearing of soil material on the viewing face. Excavations for foundations, or even big holes for planting of trees or shrubs, can serve the purpose well. For the best and most representative results, you need to be sure to dig at a spot that hasn't been filled or excavated in recent times, by humans, either before or after the time of European settlement, and hasn't been cultivated, either. In New England, woodlands that served as pasture land rather than as tilled fields are the best places to see undisturbed soil profiles.

It's dangerous for me to describe one or a few soil profiles for you, because there is such a multiplicity of important soil types (see the next section), but for concreteness, and a good start in thinking about soil horizons and classification, here's *a description of a typical New England* soil (Figure 2-28). The two conditions such a soil experiences that are the most important factors in its development are (1) *strong net downward movement of water through the profile*, and (2) *acidic waters*. In the traditional classification of soils, such a soil would be called a ***podzol*** soil.

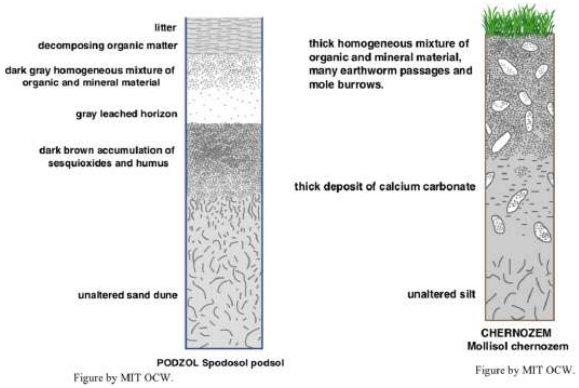


Figure 2-28 [left]. A vertical profile through a typical New England soil. (From. Figure 2-29 [right] . A vertical profile through a typical North American mid-continent soil. (From FitzPatrick, 1986.)

Right at the surface is a layer of forest litter: leaves and twigs that have fallen recently, within the last few years. That grades down into a layer of organic matter in a progressively more advanced state of humification with depth. Together these uppermost layers, called the *H horizon* if usually water-saturated or the *O horizon* if not usually water-saturated, are the part of the soil that's by far the richest in organic matter and living organisms—and most conducive to cultivation, in which case the original organic-rich layer is tilled and mixed downward to form the best medium for plant growth. Its thickness commonly ranges from just a few centimeters upward to ten to twenty centimeters. It passes rather abruptly into a layer consisting mainly of mineral grains, usually mostly quartz (if the soil is well developed and the parent material had a fairly high percentage of quartz), with a characteristic light gray color. Much, or most, or in extreme cases even all, of the other silicate minerals have been leached out of this layer by decomposition by net downward passage, from the surface, of waters rendered acidic by (1) dissolved CO_2 from the atmosphere and (2) organic acids produced in the uppermost organic-rich layer. This horizon is an example of what's called the *A horizon*. The A horizon passes downward, again fairly abruptly, into a layer, called the *B horizon*, with a characteristic color, usually combinations of yellow or orange or red or brown. Here is where iron leached out of the material of the A horizon is reprecipitated as hydrous iron oxides within the pore spaces among the existing mineral grains, and fine clay-mineral particles produced by weathering reactions in the A horizon have been deposited. (The general term for mobilization and removal of such things as iron and aluminum from a soil horizon is ***eluviation***; their deposition in a given horizon is called ***illuviation***.) The B horizon passes gradually into underlying material that's relatively little affected by soil processes; it's called the *C horizon* (but not universally considered to really be a soil horizon!).

The precise nature of the physicochemical processes that mobilize, transport, and deposit iron and, especially, aluminum are not yet entirely well understood. Iron, when weathered from parent minerals, is probably complexed with organic compounds and then precipitated as insoluble oxides lower down in the soil profile where the carrier organic compounds are degraded. The aluminum

presents more of a problem. The aluminum is not actually put into true solution, but neither can it make its way down the profile as large clay-mineral particles. It probably travels either as a colloid (i.e., particles so fine as to be in the range between true solution and fine mineral particles) or physically as very fine, submicron-size clay-mineral particles.

Where the average water table lies far below, the B horizon tends to be diffuse; where the average water table is not far below, the B horizon tends to be more concentrated, and its colors stronger. (In areas where the average water table lies close to the surface, the B horizon tends instead to be gray from deposition of compounds containing iron in the reduced, ferrous state, as well as clay minerals.)

Now for something entirely different: a capsule description of a typical soil from the mid-continent of North America (Figure 2-29) Such soils are common in a wide area including western Kansas, western Nebraska, eastern Colorado, most of the Dakotas, and on into south-central Canada. In the traditional terminology, such soil would be called ***chernozem soils*** (from the Russian for "black soils"), ***chestnut soils***, or ***brown soils***. In such areas, with a semiarid (rainfall 15–25 inches) and strongly seasonal climate and dominantly grassland vegetation (now largely converted to agricultural use, especially in the eastern parts of the area), the balance between precipitation and evaporation is such that there is little net downward movement of water in the soil, with two important consequences: (1) the soil is largely a *closed system*, in the sense that chemical constituents produced by weathering remain in the soil rather than being carried away in solution in groundwater, and (2) the pH of soil waters tends to be *alkaline* rather than acidic. In such soils, upwards of a meter thick, the upper layer (A horizon) of the soil, beneath the surface organic layer, are a deep, homogeneous, and friable mixture of organic matter and mineral particles, not strongly leached, which is kept thoroughly mixed by the activity of burrowing organisms. This layer grades downward into a lighter-colored material with less organic matter, and then passes downward into a layer (the B horizon) in which calcium carbonate has been precipitated by mobilization of calcium ions in the upper layers and precipitation at deeper levels. The carbonate is precipitated as soil moisture is drawn back up toward the surface. In areas where the parent material contains carbonate minerals, the carbonated layer is better developed, but carbonate minerals tend to be deposited in the lower part of the soil profile even in the absence of carbonate minerals in the parent material. Why? Because of the common presence of calcium ions in parent silicate minerals, and the presence of carbon dioxide in the surface waters, to supply the carbonate ion.

2.9.8 Classification of Soils

Now, on to the troublesome and confusing matter of soil classification. Early classifications, before the middle of the 1900s, relied heavily upon terms and concepts developed in the late 1800s by Russian soil scientists, who were true pioneers in the scientific study of soils. These terms are common still, but in recent decades a systematic and very detailed classification system, developed and continually refined by soil scientists at the U.S. Department of Agriculture, has come into widespread use, despite its complexity and almost absurdly non- intuitive terminology. Also, other newer classifications are in use in other parts of the world.

(The "almost absurdly" in the preceding paragraph a value judgment, I know, but it's a widely shared view. Just for laughs, I am appending, as Figure 2-31, from the thick tome on soil taxonomy published by the Soil Survey Staff of the Department of Agriculture, the official names of the "great groups", which are the third stage in the taxonomic hierarchy. Each of the 185 great groups is in turn subdivided into subgroups. There are nearly a thousand subgroups! This classification is a marvel of rational and synthetic classification and terminology. Also in Figure 2-30 is a typical page from the same source, giving a description of some of the subgroups of one of the great groups. Another other interesting point about this classification is that, in contrast to any other classification I know of, in any field of science, they insist upon capitalizing every soil name: Durorthidic Xerorthents, rather than durorthidic xerorthents.)

I think that what would be most valuable for you at this point is to have soil classification stripped to its barest essentials. I will do that for you in two ways. First, Figure 2-31 gives the soil "orders" of the USDA classification system. Even this most boiled-down presentation needs a few words of explanation. A ***pedon*** (rhymes with "head on") is the basic soil unit: the column of soil, with its various horizons, in a small representative area of the order of a square meter. The ***epipedon*** is just the surface horizon of the pedon; it's usually rich in organic matter. Second, Figure 2-32, in three parts, gives (A) a simple classification of the earth's climatic zones, and corresponding to that, (B) the major vegetation types and (C) the major soil groups. The typical New England soil, described in the preceding section, is a podzol developed in a climate with severe winters (!) and with spruce–fir forest. (Minor note: the climate of southern New England is gradational into the colder end of the "humid" box in Figure 2-32 and the box

for gray brown podzolic soil in Figure 2-32. The native forests in southern New England are a mixture of conifers and deciduous hardwoods.)

Names of orders, suborders, and great groups

Order	Suborder	Great group
Alfisols	Aqualfs	Albaqualfs.
		Duraqualfs.
		Fragiaqualfs.
		Glossaqualfs.
		Natraqualfs.
		Ochraqualfs.
		Plinthaqualfs.
		Tropaqualfs.
		Umbraqualfs.
	Boralfs	Cryoboralfs.
		Eutroboralfs.
		Fragiboralfs.
		Glossoboralfs.
		Natriboralfs.
		Paleboralfs.
	Udalfs	Agrudalfs.
		Ferrudalfs.
		Fragiudalfs.
		Fragiosudalfs.
		Glossudalfs.
		Hapludalfs.
		Natrudalfs.
		Paleudalfs.
		Rhodudalfs.
		Tropudalfs.
	Ustalfs	Durustalfs.
		Haplustalfs.
		Natrustalfs.
		Paleustalfs.
		Plinthustalfs.
		Rhodustalfs.
	Xeralfs	Durixeralfs.
		Haploxeralfs.
		Natrixeralfs.
		Palexeralfs.
		Plinthoxeralfs.
		Rhodoxeralfs.
Aridisols	Argids	Durargids.
		Haplargids.
		Nadurargids.
		Natargids.
		Paleargids.
	Orthids	Calciorthids.
		Camborthids.
		Durorthids.
		Gypsiorthids.
		Paleorthids.
		Salorthids.
Entisols	Aquents	Cryaquents.
		Fluvaquents.
		Haplaquents.
		Hydraquents.
		Psammaquents.
		Sulfaquents.
		Tropaquents.
	Arents	Arents.
	Fluvents	Cryofluvents.
		Torrifluvents.
		Tropofluvents.
		Udifluvents.
		Ustifluvents.
		Xerofluvents.
	Orthents	Cryorthents.
		Torriorthents.
		Troporthents.
		Udorthents.
		Ustorthents.
		Xerorthents.
Histosols	Fibrists	Borofibrists.
		Cryofibrists.
		Luvifibrists.
		Medifibrists.
		Sphagnofibrists.
		Tropofibrists.
	Folists	Borofolists.
		Cryofolists.
		Tropofolists.
	Hemists	Borohemists.
		Cryohemists.
		Luvihemists.
		Medihemists.
		Sulfihemists.
		Sulfohemists.
		Tropohemists.
	Saprists	Borosaprists.
		Cryosaprists.
		Medisaprists.
		Troposaprists.
Inceptisols	Andepts	Cryandepts.
		Durandepts.
		Dystrandepts.
		Eutrandepts.
		Hydrandepts.
		Placandepts.
		Vitrandepts.
	Aquepts	Andaquepts.
		Cryaquepts.
		Fragiaquepts.
		Halaquepts.
		Haplaquepts.
		Humaquepts.
		Pinaquepts.
		Plinthaquepts.
		Sulfaquepts.
		Tropaquepts.
	Ochrepts	Cryochrepts.
		Durochrepts.
		Dystrochrepts.
		Eutrochrepts.
		Fragiochrepts.
		Ustochrepts.
		Xerochrepts.
	Plaggepts	Plaggepts.
	Tropepts	Dystropepts.
		Eutropepts.
		Humitropepts.
		Sombritropepts.
		Ustropepts.
	Umbrepts	Cryumbrepts.
		Fraglumbrepts.
		Haplumbrepts.
		Xerumbrepts.
Mollisols	Albolls	Argialbolls.
		Natralbolls.
	Aquolls	Argiaquolls.
		Calciaquolls.
		Cryaquolls.
		Duraquolls.
		Haplaquolls.
		Natraquolls.
	Borolls	Argiborolls.
		Calciborolls.
		Cryoborolls.
		Haploborolls.
		Natriborolls.
		Paleborolls.

Figure_2-30.1.png" />

Figure 1-30. A) Orders, suborders, and great groups of the soil classification system developed by the Soil Survey Staff at the US Department of Agriculture. B). A random page from the system of soil taxonomy of the US Department of Agriculture. (From Soil Survey Staff, 1994)

https://geo.libretexts.org/@go/page/22343

States, but the subgroup is provided for use where needed.

Epiaquic Tropohumults.—Soils in this subgroup have some yellowish brown or olive brown colors in the upper horizons and have redder mottles inside the peds. The hue becomes redder with depth. The rainfall is seasonally high, and the upper horizons become saturated. The reducing conditions are commonly but not everywhere too mild to produce mottles of low chroma, but the hue in the upper horizons approaches that of Aquults. These soils are in the areas of high rainfall in Puerto Rico but are not extensive. Their slopes are mostly strong, and the soils are mostly cultivated.

Epiaquic Orthoxic Tropohumults.—These soils have some yellowish brown or olive brown colors in the upper horizons and have redder mottles inside the peds. The hue becomes redder with depth. The reducing conditions are commonly too mild to produce mottles of low chroma. The clay in these soils has low activity. The soils have strong slopes and are in the areas of high rainfall in Puerto Rico, but they are not extensive.

Humoxic Tropohumults.—These soils have a clay fraction that has low activity. They are common in the United States in Hawaii, where they formed on steep slopes or fans from basic rocks. They are largely used for pasture, but where slopes are favorable some of them are cultivated.

Orthoxic Tropohumults.—These soils have a clay fraction that has low activity, and they tend to be warmer than most of the Typic Tropohumults. They are not extensive in either Puerto Rico or Hawaii. Their slopes are mostly steep, and the soils are used mainly for pasture or tree crops.

Ustic Tropohumults.—These soils have an ustic moisture regime but are otherwise like Typic Tropohumults in defined properties. They are not known to occur in the United States, but the subgroup is provided for use if needed elsewhere.

Ustoxic Tropohumults.—Soils in this subgroup have an ustic moisture regime and a clay fraction that has low activity. They are not extensive in the United States.

Udults

These are the more or less freely drained, humus-poor Ultisols (plate 9D) in humid climates in mid or low latitudes that have well-distributed rainfall. Most of them have light-colored upper horizons, commonly a grayish horizon that rests on a yellowish brown to reddish argillic horizon. A few that developed from basic rocks have a dark brown or reddish brown surface horizon that rests on a dark red or dusky red argillic horizon. Some have a fragipan or plinthite, or both, in or below the argillic horizon.

The Udults developed in sediments and on surfaces that range from late Pleistocene to Pliocene or possibly older. Many are cultivated, either with the use of soil amendments or in a system in which they are cropped for a very few years and then are returned to forest to allow the trees to regather in their tissues the small supply of nutrients. Most of these soils have or had a forest vegetation, but some have a savanna that probably is anthropic. The moisture regime is udic. The soil temperature regime is mesic, isomesic, or warmer.

Definition

Udults are the Ultisols that
1. Are never saturated with water or have a redder hue or higher chroma than Aquults;
2. Have <0.9 percent organic carbon in the upper 15 cm of the argillic horizon and have <12 kg organic carbon in a unit volume of 1 square meter to a depth of 1 m exclusive of any O horizon; and
3. Have a udic moisture regime.

Key to great groups

FCA. Udults that have a fragipan in or below the argillic horizon. **Fragiudults, p. 360**

FCB. Other Udults that have plinthite that forms a continuous phase or constitutes more than half the volume in some horizon within the upper 1.25 m of the soil. **Plinthudults, p. 366**

FCC. Other Udults that have an argillic horizon that has <10 percent weatherable minerals in the 20- to 200-micron fraction in the upper 50 cm and have a clay distribution such that the percentage of clay does not decrease from its maximum amount by more than 20 percent of that maximum within 1.5 m of the soil surface, or the layer in which the percentage of clay is less than the maximum has skeletans on ped faces or has 5 percent or more plinthite by volume. **Paleudults, p. 364**

FCD. Other Udults that have
1. An epipedon that has a color value, moist, of less than 4 in all parts; and
2. An argillic horizon that has a color value, dry, of less than 5 and not more than 1 unit higher than the value, moist. **Rhodudults, p. 367**

FCE. Other Udults that have an isomesic or warmer soil temperature regime. **Tropudults, p. 367**

FCF. Other Udults. **Hapludults, p. 362**

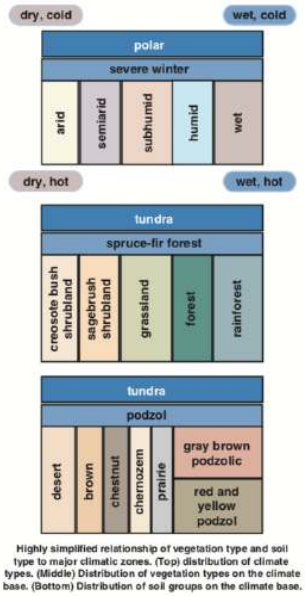
Fragiudults

These are the Udults that have a fragipan in or below the argillic horizon. They were formed mainly in loamy sediments or in residuum. The fragipan normally has an upper boundary about 50 to 75 cm below the soil surface. Perched ground water is present above the pan at some period during the year, and there are commonly thick gray skeletans at the top of the pan.

The Fragiudults in the United States are principally on gentle slopes throughout the southeastern states. Their temperature regime is mesic or thermic. Their vegetation in the United States has been forest, either of conifers or of broadleaf deciduous trees. Some were considered Planosols and others Red-Yellow Podzolic soils in the 1938 classification or the 1949 revision.

Figure_2-31.png" />

Figure 2-31. Orders of the US Department of Agriculture soil classification system. (Modified from Winegardner, 1996.)



Highly simplified relationship of vegetation type and soil type to major climatic zones. (Top) distribution of climate types. (Middle) Distribution of vegetation types on the climate base. (Bottom) Distribution of soil groups on the climate base.

Figure by MIT OCW.

Figure 2-32. Highly simplified relationship of vegetation type and soil type to major climatic zones. (Top) distribution of climate types. (Middle) Distribution of vegetation types on the climate base. (Bottom) Distribution of soil groups on the climate base. (From Hunt, 1972.)

References

Brady, N.C. and Weil, R.R. 2002, The Nature and Property of Soils: Prentice Hall, 960 p.

FitzPatrick, E.A., 1986, An Introduction to Soil Science: Longman, 255 p.

Merrill, G.P., 1897, A Treatise on Rocks, Rock Weathering and Soils: Macmillan, 411 p.

Hunt, C.B., 1972, Geology of Soils; Their Evolution, Classification, and Uses: Freeman, 344 p.

Soil Survey Staff, 1994, Keys to Soil Taxonomy: US Department of Agriculture, Soil Conservation Service, 306 p.

Taylor, G., and Eggleton, R.A., 2001, Regolith Geology and Geomorphology: Wiley, 375 p.

Winegardner, D.L., 1996, An Introduction to Soils for Environmental Professionals: Lewis Publishers, 270 p.

This page titled 4.9: Soils is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.9: Soils** by John Southard is licensed CC BY-NC-SA 4.0. Original source: https://ocw.mit.edu/courses/12-090-the-environment-of-the-earths-surface-spring-2007/.

4.10: Soil Chemistry

Darshani Kumaragamage, Jim Warren, and Graeme Spiers

Learning Objectives

On completion of this chapter, students will be able to:

1. Define and describe the importance of the major soil chemical properties in sustaining life on Earth
2. Describe and compare the structures, charge development, and properties of organic and inorganic soil colloids
3. Explain the influence of soil minerals on soil properties
4. Define cation exchange capacity, explain the importance of cation exchange and anion adsorption, and identify the factors influencing ion exchange in soils
5. Discuss the causes of soil acidity and alkalinity, characteristics of strongly acidic and alkaline soils, and their effect on plant growth
6. Describe the causes of soil salinity, major characteristics of saline, saline-sodic and sodic soils and reclamation of salt-affected soils
7. Describe the importance of redox reactions in nutrient dynamics and greenhouse gas emissions
8. Explain the basic principles in measuring cation exchange capacity, pH, electrical conductivity and redox potential in soils, and the interpretation of their values

INTRODUCTION

Soil chemistry is a branch of soil science that deals with the chemical composition, chemical reactions and chemical properties in soils. Abiotic phases of the soil include solids (organic matter and inorganic minerals), liquids (soil water), and gases (soil air), while the biotic phase consists of living organisms. Soil, therefore, is a dynamic living system in which ions and molecules constantly may move from one phase to the other, while interacting with each other.

Soil chemical properties are based on the concentrations, and/or proportions of dissolved species in soil water and/or on the ion exchange complex. Soil chemical properties such as cation exchange capacity (CEC), pH, redox potential (Eh or pe), and electrical conductivity (EC) are important as they influence nutrient availability, plant growth, fate of pollutants, biological activity, etc. For example, CEC is a measure of the amount of negative charge capable of adsorbing cations on mineral or organic surfaces, usually expressed in centimoles of charge [cmol(+)] per kg of soil. Soil pH is a measurement of the concentration of H^+ in the soil solution, while the Eh (or pe) is related to concentrations of oxidized/reduced species of redox-sensitive elements such as iron (Fe^{3+}/Fe^{2+}) . Electrical conductivity (EC) is another important soil property that provides an indication of the abundance of soluble salts in a soil.

Most of the important chemical properties of soils are controlled by reactions that occur between the soil solution and surfaces of colloidal (particles <0.002 mm diameter; Brady and Weil, 2010) soil mineral particles and soil organic matter. In the first part of this chapter we will focus on the chemistry of the soil solution and the colloidal fraction. The remaining sections of the chapter will focus on important soil chemical properties; namely, CEC, anion exchange, soil pH, EC, and Eh, with emphasis on their importance in sustaining life on Earth.

SOIL Solution

Soil water, including dissolved solutes, is referred to as the soil solution and is the lifeblood of soil. All chemical reactions, mineral precipitation/dissolution reactions, ion exchange reactions, redox reactions, and nutrient uptake by plants occur in, or are mediated by the soil solution. Without soil water, very few chemical or biological reactions would occur in soil; in fact, there would be no life on this planet.

The soil solution is very complex and contains a large variety of cations and anions (as both as free ions and complex ions) as well as dissolved organic molecules, usually in low concentrations. For soil nutrients to be taken up by plant roots, they must exist in a dissolved form in the soil solution. At least 17 elements are required to support the growth of most plants (see Chapter 7), with an additional four elements essential for the growth of some plants (Havlin et al., 2013). For example, some important cations and anions that plants absorb from soil solution are: nitrogen as ammonium (NH_4^+) and nitrate (NO_3^-), phosphorous as several

phosphate species (e.g., HPO_4^{2-} and $H_2PO_4^-$), potassium (K^+), calcium (Ca^{2+}), magnesium (Mg^{2+}), and sulphur as sulphate (SO_4^{2-}). All nutrients are absorbed from the soil solution by plant roots.

The soil solution also contains other non-nutritive ions and molecules, some of which are innocuous, while some are toxic, depending on concentration. These include some trace elements (e.g., lead, arsenic), pesticides (e.g., chlorpyrifos, glyphosate), antimicrobial ions and compounds (e.g., disinfectants, antibiotics), and complex organic molecules. These may also be taken up by plant roots. Nutrients and other chemicals in the soil solution may also be lost from the soil through surface runoff and/or downward leaching, particularly under excessive rainfall and irrigation events.

SOIL COLLOIDS

The soil colloidal fraction is defined as that fraction of the soil made up collectively of small (<0.002 mm) inorganic and organic particles (Brady and Weil, 2010). The high specific surface area and the presence of electric charges on most particles are responsible for important soil characteristics such as swelling capacity, the ability of the soil to retain water, cation and anion adsorption and ion exchange.

The importance of colloids in soil relates to the relationship of their size with the amount of surface area available to react with the soil solution. Table 5.1 shows the relationship between particle size and the magnitude of surface area associated with particle size range for a clay loam soil. Specific surface area is defined as the surface area per unit mass of the solid particle and expressed as $m^2 g^{-1}$ of soil. As can be seen in Table 5.1, even though this soil contains equal proportions of sand, silt, and clay by mass, by far the majority of the surface area is associated with the smallest size fraction.

Table 5.1. Relationship between particle size range and specific surface area of soil particles.

Particle Size Fraction	**Size Range**	**Median Diameter**	**Percent of Clay Loam Soil [a]**	**Specific Surface Area [b]**	**Percentage Surface Area**
	(µm)	**(µm)**	**(%)**	**($cm^2 g^{-1}$)**	**(%)**
Sand	2000-50	1000	33.3	22.6	0.2
Silt	50-2	25	33.3	90.9	0.8
Clay	<2	1	33.3	11,320	99

[a] Assumed a clay loam soil with equal proportions of sand, silt and clay. [b] Calculated based on uniform spherical particles with median particle diameter.

Mineral phases in the colloidal or clay-size fraction are primarily phyllosilicate and oxy-hydroxide minerals (Figure 5.1). Organic colloids are also in the clay-size range and are composed of small particles of humus that are relatively stable and resistant to further biological decay. These diverse colloidal particles may either exist independently in the soil environment or they may occur in association in organo-mineral complexes (Newman and Hayes, 1990).

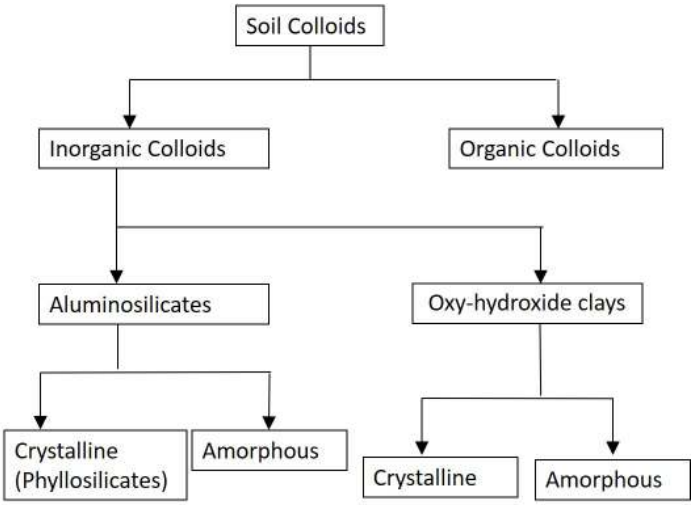


Figure 5.1. A classification of soil colloids. © Darshani Kumaragamage, Univ. of Winnipeg is licensed under a CC BY (Attribution) license.

Phyllosilicates

It is important to understand that there are differences among aluminosilicate clay minerals (also called phyllosilicates or layer silicates) that make up the vast majority of mineral colloids and that these impart substantial differences in important soil properties, such as ion exchange.

Phyllosilicates (*phyllon* meaning leaf or sheet in Greek) are crystalline minerals, with a structure consisting of layers of repeating atomic units. There are two fundamental units composed of layers of oxygen and hydroxyl groups bonded with silicon, aluminum, magnesium and other cations. The fundamental layers are composed of tetrahedral and octahedral sheets that occur in different combinations to form the different clay minerals.

Tetrahedral sheets are composed of repeating tetrahedron units, each typically containing a silicon atom surrounded by four oxygen atoms: one apical oxygen atom and three basal oxygen atoms forming a four-sided (tetrahedral) configuration (Figure 5.2A). Tetrahedral sheets extending in two dimensions are formed by sharing the basal oxygen atoms between silicon atoms (Figure 5.2B).

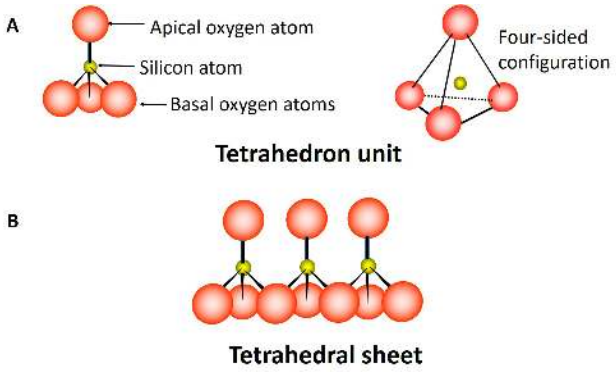


Figure 5.2. Arrangement of atoms in a silica tetrahedron (A) and linked in a tetrahedral sheet (B). © Darshani Kumaragamage, Univ. of Winnipeg is licensed under a CC BY (Attribution) license.

The octahedral sheet is composed of multiple octahedron units (Figure 5.3A), each unit containing aluminum (or magnesium) atoms surrounded by six oxygen atoms and/or hydroxyl groups. Three oxygen atoms (or hydroxyl groups) occur in the lower plane, the remaining three oxygen atoms (or hydroxyl groups) occur in the upper plane with the aluminum or magnesium atoms sandwiched in between, forming an eight-sided (octahedral) configuration. These units combine by sharing oxygen atoms to form an octahedral sheet extending in two dimensions (Figure 5.3B).

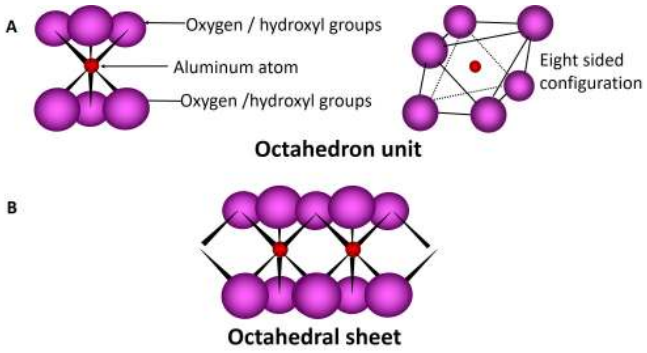


Figure 5.3. Arrangement of atoms in an octahedron (A) and in an octahedral sheet (B). © Darshani Kumaragamage, Univ. of Winnipeg is licensed under a CC BY (Attribution) license.

Tetrahedral and octahedral sheets combine to form unit layers of different phyllosilicates through the sharing of the apical oxygen atom of the tetrahedral sheet with oxygen atoms in the octahedral sheet. Phyllosilicates composed of one tetrahedral sheet and one octahedral sheet are called 1:1 type phyllosilicates; those composed of one octahedral sheet sandwiched between two tetrahedral sheets are called 2:1 type phyllosilicates (Figure 5.4). The ideal unit cell formula for a 1:1 clay mineral is $Si_4Al_4O_{10}(OH)_8$, while that of a 2:1 clay mineral is $Si_8Al_4O_{20}(OH)_4$.

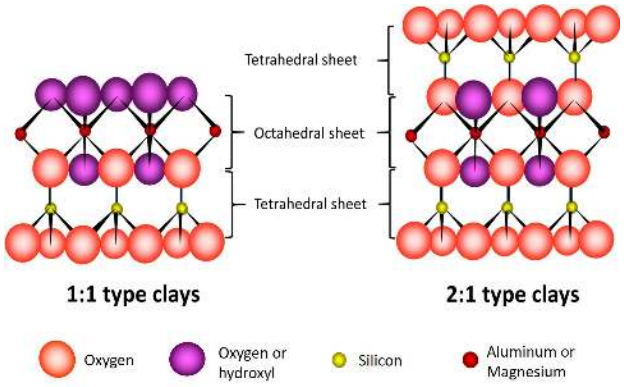


Figure 5.4. Layer arrangement of 1:1 and 2:1 silicate clay minerals. © Darshani Kumaragamage, Univ. of Winnipeg is licensed under a CC BY (Attribution) license.

Phyllosilicates consist of many layers (1:1 type or 2:1 type) stacked one on top of the other with different forms of bonding between the layers. The region located between two adjacent layers is called the interlayer region. The basic chemical behaviour and physical properties of the different phyllosilicates are dictated in large part by the characteristics associated with the interlayer region of the clays.

Isomorphous Substitution

In addition to their large surface area (see Table 5.1), the phyllosilicates are also electrically charged due to substitution of the central cations within the tetrahedral or octahedral sheets by other ions of similar size (see Table 5.2) but different valence. This imparts a net negative charge to the structure of the mineral.

Isomorphous substitution is defined as “replacement of one atom by another of similar size in a crystal structure without disrupting or changing the crystal structure of the mineral” (Canada Department of Agriculture, 1976). When a substituting cation is of a lesser valence than the cation it is replacing, there is an excess of negative charges associated with the oxygen atoms surrounding it and therefore a net negative charge is imparted to the structure. For example, trivalent aluminum (Al^{3+}) may substitute for some of the tetravalent silicon (Si^{4+}) in the tetrahedral sheet. The lesser charge of the aluminum compared to silicon creates a charge deficit and a negative charge within the tetrahedral sheet of the structure. Similarly, in the octahedral sheet, Mg^{2+}, Fe^{2+}, or other divalent cations, may substitute for some trivalent Al (Al^{3+}), resulting in an extra negative charge associated with the octahedral sheet of the structure. These negative charges attract positively charged cations to the surface of the minerals to neutralize the negative charge. These charge neutralizing cations are located mostly within the interlayer region of the phyllosilicates. The abundance of negative charge sites (i.e., the **charge density**), the location of the negative charges (tetrahedral

vs. octahedral sheet), and the type of cation in the interlayer region neutralizing the charge (e.g., Na^{+}, Ca^{2+}, Mg^{2+}, K^{+} or other) determines the degree to which the phyllosilicates expand or swell, and their capacity to retain cations.

Table 5.2. Radius of common cations and location in phyllosilicate structures

Cation	Radius (nm)	Location
Si^{4+}	0.041	tetrahedra
Al^{3+}	0.05	tetrahedra, octahedra, interlayer
Fe^{3+}	0.064	octahedra
Fe^{2+}	0.076	octahedra, interlayer
Mg^{2+}	0.065	octahedra, interlayer
Ca^{2+}	0.099	Interlayer
Na^{+}	0.095	Interlayer
K^{+}	0.133	Interlayer

Phyllosilicate Minerals

There are about 50 species of phyllosilicate minerals, but only five are common in Canadian soils (see Chapter 14). These are sketched in Figure 5.5 and described as follows:

Kaolinite (Figure 5.5A) is a 1:1 type clay mineral held together by hydrogen bonds in the interlayer region. Hydrogen bonding occurs between the hydroxyl in the octahedral sheet of one unit layer, and the (basal) oxygens in the tetrahedral sheet of the adjacent layer. The hydrogen bonds hold the layers closely together to form a very stable structure. This combined with very little isomorphous substitution inhibits the intrusion of cations and water into the interlayer region. This mineral is non-expandable since water molecules cannot enter the interlayer space. Kaolinite exhibits very little swelling when wet, or shrinkage when dry. The interlayer is not exposed and therefore the specific surface area is low. With almost no isomorphous substitution, kaolinite minerals have a low surface charge and low cation exchange capacity.

Can You Dig It!

Rocks that are rich in kaolinite are known as kaolin which is derived from the Chinese word "Gaoling" (literally meaning 'High Ridge'), which is the name of a village in southeastern China where kaolinite was mined historically. Kaolinite in its pure form is white and its minimal shrink/swell characteristic and ability to not crack when dried made it highly prized for making porcelain and "fine China" dishware.

Smectite (Figure 5.5B) is a group of expandable 2:1 type phyllosilicates. Montmorillonite and beidellite are the two most common species of the smectite group of minerals in Canadian soils. The amount of isomorphous substitution in both minerals is moderate compared to other 2:1 type phyllosilicates (e.g. hydrous mica), and the two species are differentiated based on the location of the layer charge in the structure. Most of the layer charge in montmorillonite is a result of isomorphous substitution in the octahedral sheet, whereas in beidellite the layer charge derives mostly from isomorphous substitution in the tetrahedral sheet. Hydrogen bonding does not occur in 2:1 clay minerals since there are no hydroxyl groups associated with the basal oxygens of the tetrahedral sheets of adjacent layers facing each other. Thus, the structural layers are only held together by electrostatic forces of whatever neutralizing cation happens to be in the interlayer region to neutralize the negative charges associated with isomorphous substation within the layer structures. Exchangeable cations in the interlayer are usually surrounded by six water molecules in octahedral coordination. Smectite minerals have a high cation exchange capacity. Since the layers are loosely bound, both water molecules and cations can freely enter the interlayer space. Therefore, these clays have a high specific area, and are expandable; swelling when wet and shrinking when dry.

Vermicullite (Figure 5.5B) is a 2:1 type phyllosilicate similar to smectite but with a layer charge intermediate between hydrous mica/illite (see below) and the smectites. The mineral has limited expandability, with moderate shrink/swell characteristics. Most isomorphous substitution occurs in the tetrahedral sheet imparting a high surface charge and a high cation exchange capacity. The

structure is capable of trapping K^+ ions in the interlayer region similar to hydrous mica/illite, but as the layer charge is slightly lower, the trapped K^+ can be released back into the soil solution unlike hydrous mica/illite.

Hydrous Mica also called **illite** (Figure 5.5C) is another 2:1 type clay mineral, Hydrous mica is derived from the partial weathering of mica, a primary mineral common in the granitic rocks of the Canadian Shield. The structure of hydrous mica/illite is similar to the smectites and vermicullite, except that the layer charge is higher with isomorphous substitution occurring almost exclusively in the tetrahedral sheet where Al^{3+} is substituted for Si^{4+}. Potassium almost exclusively occupies the interlayer region, neutralizing the layer charge and holding the structure together. The potassium ion (K^+) is almost an exact fit for the hexagonal shaped holes that occur in the tetrahedral sheets of the 2:1 layers, and thus gets even closer to the charged sites and strongly binds adjacent layers together. This prevents entry of water into the interlayer region and renders minerals of the mica group non-expandable. So even though the layer charge is higher than the smectites and vermiculite, the cation exchange capacity of hydrous mica/illite is low, since the charge is neutralized by K^+.

Chlorite (Figure 5.5D) is a 2:1:1 (or 2:2) type phyllosilicate. The mineral has a 2:1 type structure, but with an additional octahedral sheet—composed of an aluminum, magnesium or iron hydroxide layer in six-fold (octahedral) coordination—in the interlayer region between the 2:1 type layers. This interlayer hydroxyl layer does not share oxygens with the main 2:1 layer, instead it is hydrogen bonded to the 2:1 layer; consequently, the mineral is non-expandable. The amount of isomorphous substitution in the structure is variable, but the charges are neutralized by the positively charged octahedral sheet in the interlayer, thus the ion exchange capacity is low.

Amorphous aluminosilicates are yet other aluminosilicates found in soil but these are not phyllosilicates. These are described as amorphous or as minerals of short-range order, meaning they lack the defined structural units of the phyllosilicates. Allophane and imogolite are examples of amorphous aluminosilicate minerals. These minerals are formed in abundance in volcanic ash soils found in countries such as Japan and New Zealand, but they are also being found in very small amounts in some Podzolic and Solonetzic soils in Canada. Amorphous aluminosilicates are non-expandable. They are extremely small (<4 nm diameter) with a very high specific surface area and a very high ion exchange capacity.

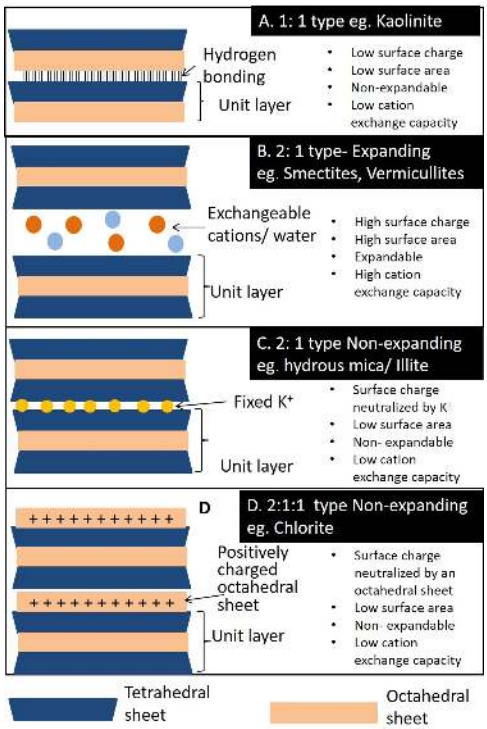


Figure 5.5. Arrangement of tetrahedral sheets and octahedral sheets in common phyllosilicates and some key properties. © Darshani Kumaragamage, Univ. of Winnipeg is licensed under a CC BY (Attribution) license.

Oxy-hydroxide Minerals

Oxy-hydroxide minerals (sesquioxides) are mainly oxides, hydroxides and hydrous oxides of iron, aluminum, and manganese, and are among the youngest mineral phases found in soil. Some are well crystallized (e.g., gibbsite, hematite, goethite), while others are amorphous (e.g., ferrihydrite). Crystalline oxide clays are composed of oxygen and/or hydroxyl groups, usually in an octahedral arrangement with iron, aluminum, or manganese. The octahedral structures of these minerals are held together by H-bonding. Non-crystalline (amorphous) oxy-hydroxides are complex iron, aluminum, and manganese minerals that are typically precipitated very quickly from the soil solution in response to wetting and drying cycles. Precipitation of these minerals occurs when the solubility product for a combination of cation and anion concentrations is exceeded, and usually other solutes including carbonate, phosphate and sulphate may be trapped in the structure during the rapid precipitation. Amorphous oxy-hydroxides are metastable and convert slowly to their more crystalline forms.

Both crystalline or amorphous oxy-hydroxide minerals do not carry permanent negative charges but do have a high surface charge associated with surface hydroxyl groups. As such, the exchange capacity of oxy-hydroxides is very pH dependent with the variation due to protonation of hydroxyl groups at low pH producing positive charges, and dissociation of hydroxyl groups at high pH producing negative charges. Oxy-hydroxide minerals are **amphoteric** due to their variable cation exchange capacity; meaning they typically have a net cation exchange capacity at high pH and a net anion exchange capacity at low pH.

Oxy-hydroxides are usually found in small quantities in young Canadian soils, but are more common in highly weathered soils found in warm, humid regions in other parts of the world. They are highly reactive and have a very high specific surface area. Presence of these minerals significantly influences soil properties, such as ion adsorption and exchange.

Organic Colloids (Humus)

Organic colloids, generally known as humus, consist of highly decomposed organic matter. These colloids have a complex structure, and are chemically composed mainly of carbon, hydrogen and oxygen. On the molecular scale, humus may be described as a complex polymeric mixture of random organic components and is probably one of the least understood aspects of soil chemistry. Organic colloids may be highly reactive chemically, with high water holding capacities allowing absorption of water. Compared to inorganic colloids, organic colloids have greater specific surface area and higher surface charge. Surface charges are predominantly negative and pH dependent arising from partial dissociation of hydroxyl (-OH), carboxyl (-COOH), and phenolic groups, which are associated with central carbon structures of varying size and complexity. On a per unit mass basis, organic colloids impart a greater influence on soil properties than inorganic colloids even though their abundance in soil is generally far less than that of the inorganic colloid fraction.

Charges on Soil Colloids

Charges on soil colloids arise from the following two mechanisms:

1. **Isomorphous substitution** within the crystalline structure of phyllosilicates – leading to formation of mostly negative, permanent charges. As described earlier in this chapter, isomorphous substitution is the replacement of one atom by another of similar size in a crystal lattice without disrupting or changing the crystal structure of the mineral. The replacement of a higher valent cation with a lower valent cation (e.g., substitution of Mg^{2+} for Al^{3+} in the octahedral sheet) results in an unbalanced negative charge on the mineral surfaces. These negative charges are permanent and thus are not affected by pH.
2. **Protonation/Deprotonation** of functional groups on the surfaces of oxy-hydroxide clays, the edges of phyllosilicate clay minerals and organic compounds, give rise to positive and negative charges, referred to as pH dependent (variable) charge. Protonation of functional groups occurs under acidic (low pH) conditions whereas dissociation of hydroxyl protons predominates under basic (high pH) conditions (Figure 5.6). Thus, the relative amounts of positive and negative surface charge vary depending on soil pH, with negative charges predominating under high pH conditions and positive charges predominating at low pH conditions. The pH at which the number of positive charges is equal to the number of negative charges is referred to as the point of net zero proton charge or zero point of charge. Humus particle surfaces are similar to inorganic colloids, containing functional groups (e.g., carboxyl [-COOH], phenolic [R-OH]) that can dissociate (deprotonate) depending on the pH, giving rise to a negative charge on organic colloids. As the pH increases, dissociation is greater, and the negative charges increase. In acidic soils, conditions favour the non-dissociated functional group, thus negative charges are low.

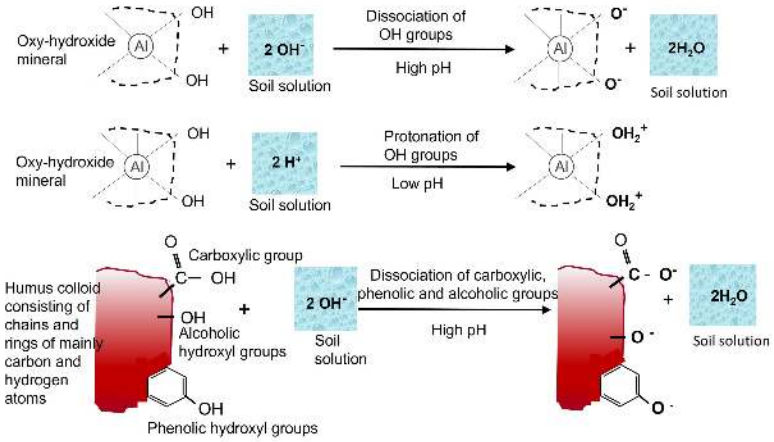


Figure 5.6. Variable pH dependent charges in oxy-hydroxide clays and soil humus particles. © Darshani Kumaragamage, Univ. of Winnipeg is licensed under a CC BY (Attribution) license.

Influence of Soil Colloids on Soil Properties

The amount and type of inorganic and organic colloids present in the soil has a profound influence on soil properties. For example, soils containing a high proportion of humus have high water holding capacities. Soils having a high proportion of freely expanding clay minerals, such as smectite, shrink when dry and swell when wet. Vertisols, fine textured soils containing >60% clay of which at least half is smectite, are typically very sticky when wet and form wide cracks when dry.

Phyllosilicates, oxy-hydroxides, and humus particles have electrical charges on their surfaces, which contribute to the total charge in the soils, and thus to the cation exchange capacity. The ranges of CEC associated with the different soil colloids are presented in Table 5.3. For example, a soil containing 40% smectite clay with CEC of 100 cmol(+) kg^{-1} may contribute up to 40 cmol(+) kg^{-1} to the total CEC of the soil. Given its high charge density, humus can be a significant contributor to the total CEC even in mineral soils—particularly under neutral to high pH conditions where pH dependent negative charges dominate. Depending on the characteristics of the humus, the pH, and interactions of humus with other soil particles, a 3% humus content (assuming a CEC of 400 cmol(+) kg^{-1}), may contribute up to 12 cmol(+) kg^{-1} to the CEC of the soil. The contribution of iron oxy-hydroxide clays to CEC is also pH dependent. However, compared to humus and phyllosilicates with high CEC, the contribution of iron oxy-hydroxides towards soil CEC is negligible. Conversely, oxy-hydroxides together with 1:1 type layer silicates may exhibit positive charges under acidic conditions and are thus a main source of **anion exchange capacity**.

Table 5.3. Ranges of CEC in different soil colloids. Adapted from Brady and Weil (2010) and Soil Survey Staff (2014)

Colloid	CEC Range [cmol(+) kg^{-1}]
Kaolinite	2 to 16
Chlorite	10 to 40
Hydrous Mica/Illite	20 to 40
Montmorillonite	60 to 100
Vermiculite	100 to 150
Organic matter (humus)	150 to 400
Amorphous aluminosilicates	5 to 350
Fe/Al oxy-hydroxides	~0 to 3

CATION EXCHANGE IN SOILS

Negatively charged surfaces on soil colloids attract cations by electrostatic forces. Cations retained by soil colloids may be eventually released to the soil solution through exchange with other cations in the soil solution. The cations most commonly held

by soil colloidal particles are Ca^{2+}, Mg^{2+}, K^+, Na^+, H^+ and Al^{3+}. Of these, Ca^{2+}, Mg^{2+}, K^+, and Na^+ are so-called base-forming cations, while H^+ and Al^{3+} are acid-forming cations. Exchangeable H^+ and Al^{3+} are found in increasing abundance in acidic soils.

Cation exchange is a process by which cations in the soil solution exchange with cations attached to the surfaces of clay minerals and humus. Cation exchange occurs on a charge-for-charge basis as illustrated in Figure 5.7.

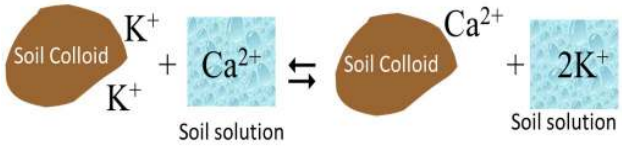


Figure 5.7. A schematic representation of a cation exchange reaction in soil. © Darshani Kumaragamage, Univ. of Winnipeg is licensed under a CC BY (Attribution) license.

Cation exchange reactions are **stoichiometric**. Thus, for divalent-monovalent exchanges (e.g., exchange between Ca^{2+} and K^+) one divalent cation exchanges with two monovalent cations. The reaction is rapid (almost instantaneous) and reaches equilibrium quickly. The cation exchange reaction is usually reversible, unless the exchanged cation is fixed in the interlayer space of a phyllosilicate (e.g., K^+ in hydrous mica/illite).

Cation exchange capacity (CEC) is defined as the sum of total exchangeable cations that a soil can adsorb (Brady & Weil, 2010). The CEC is usually expressed in centimoles of positive charge per kilogram of exchanger (cmol(+) kg^{-1}) or millimoles of positive charge per kilogram of exchanger (mmol(+) kg^{-1}). Thus a soil with a CEC of 10 cmol(+) kg^{-1} can adsorb 10 cmol of a monovalent cation (e.g., Na^+) or 5 cmol of a divalent cation (e.g., Ca^{2+}). In this example, if Na^+ exchanges with another monovalent cation (e.g., K^+), 10 cmol of K^+ would exchange with 10 cmol of Na^+. However, if the exchange is with a divalent cation (e.g., Ca^{2+}) only 5 cmol of Ca^{2+} will exchange with 10 cmol of Na^+.

Since it is the negative charges (including pH dependent negative charges) in the soil colloids that are responsible for the CEC of a soil, the amounts of inorganic and organic colloids, the type of clay minerals present, and the soil pH are the primary determining factors controlling the CEC of a soil.

Can You Dig It!

Ion exchange in soils is important to sustain life on Earth. Because of ion exchange, plant nutrients are retained in soil and released when needed by plants. Thus ion exchange occupies a central position unique to soils in their ability to supply plant nutrients and act as a filter to retain contaminants from reaching groundwater. Ion exchange retains toxic compounds, reducing their mobility, thus, protecting water resources from pollution.

Determination of Cation Exchange Capacity

The CEC of soil is generally estimated by saturating all cation exchange sites with an index cation and then replacing and measuring the amount of the index cation released to solution. The first step is to saturate a known mass of oven-dried soil with an index cation (e.g., Na^+ or NH_4^+). This is accomplished by equilibrating the sample with a solution containing the index cation buffered at pH 7.0 (usually sodium acetate or ammonium acetate). After ensuring that all cations are displaced by the index cation, the excess index cation is removed from the soil sample by leaching with a non-polar solution such as 95% ethanol. The amount of index cation held by the exchange sites in the sample must then be determined by leaching the soil with a solution containing a different cation (e.g., K^+) to replace the index cation on the exchange sites. Then the concentration of index cation in the leachate is measured and used to calculate the CEC of the soil (Hendershot et al., 1993).

Base Saturation

The proportion of CEC saturated by base-forming cations (typically Ca^{2+}, Mg^{2+}, K^+, and Na^+) is known as base saturation. **Base saturation percentage** is calculated using the equation:

(1) $$Base\ saturation\ \% = \frac{\sum\ Exchangeable\ base-forming\ cations}{CEC} \times 100$$

Both the CEC and the sum of exchangeable base-forming cations should be expressed in the same units (e.g., cmol(+) kg^{-1}). Since the dominant exchangeable base-forming cations in soils are Ca^{2+}, Mg^{2+}, K^+, and Na^+, the equation is usually written as:

(2) $$Base\ saturation\ \% = \frac{exch.\ Na^{+} + exch.\ K^{+} + exch.\ Ca^{2+} + exch.\ Mg^{2+}}{CEC} \times 100$$

Note that for divalent cations such as Ca^{2+} and Mg^{2+}, exchangeable cations in cmol (+) should be calculated considering the charge; i.e, 1 cmol Ca^{2+} = 2 cmol of charge (+). Base saturation is positively related to soil pH. If the pH is low (<7) then the base saturation is also low. And, as soil pH increases, base saturation also increases approximately linearly between pH 4 and 7.

ANION ADSORPTION AND EXCHANGE

Soils are able to retain anions in a similar fashion to retaining cations, particularly in the lower pH range when the surfaces of some colloids (e.g., oxy-hydroxides) have positive charges. Retention of anions through adsorption varies from one anion to another. Some anions are retained through non-specific adsorption mechanisms, while others are retained through specific adsorption mechanisms.

Non-specific adsorption refers to an anion that is retained by positively charged sites and the anions are weakly held through electrostatic forces. This type of adsorption is common for anions like chloride (Cl^{-}) and NO_3^{-}. These anions are weakly held on positively charged surfaces and they can freely exchange with other anions in soil solution; consequently, they may be easily lost through leaching. For non-specific adsorption to occur, the colloidal surfaces must be positively charged; thus, soil containing minerals with pH dependent charges have a high anion exchange capacity, particularly at low pH. Soils of the tropics contain high proportions of oxy-hydroxides and kaolinite, which exhibit positive charges under low pH conditions, thus in these soils, anion adsorption through non-specific mechanisms can be high.

Specific adsorption occurs when the ion becomes chemically bonded to the surface of the colloid. Mechanisms are specific to the anion being adsorbed. For example, phosphate anions (e.g., $H_2PO_4^{-}$) may be specifically adsorbed to certain colloidal surfaces through exchange of a surface hydroxyl on the colloid for an oxygen atom that is part of the anion (Figure 5.8). The strongly bonded anion is not mobile in soils and is less available to plants. Specific adsorption is only minimally reversible, resulting in permanent loss of the ion from the soil solution and from uptake by plants. In extreme cases, the original mineral surface may become completely coated with the adsorbed ion. The common anions that are retained in soil through specific adsorption are phosphate, sulphate, silicate, and carbonate.

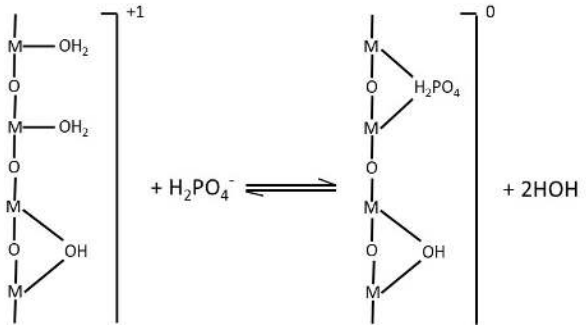


Figure 5.8. Mechanism of specific adsorption of phosphate to colloidal surfaces. Adapted from D. Shreeja (n.d.) by Darshani Kumaragamage. © Darshani Kumaragamage, Univ. of Winnipeg is licensed under a CC BY (Attribution) license.

SOIL pH

Soil pH is a measure of the **activity** (cf. concentration) of hydrogen ions in the soil solution and describes the acidity or alkalinity in the soil. For our discussion herein we define activity as “effective” concentration, whereby activity is roughly equal to concentration in dilute solutions but generally deviating from concentration (typically becoming less than the concentration) as the total concentration of salts dissolved in the soil solution increases. For current purposes, we assume that activity and concentration are equal.

The standard pH scale spans from 0 to 14, with a pH value of 7 being neutral. Pure water has a pH value of 7.0. Values of pH less than 7 are acidic and become increasingly acidic as values approach zero. Values of pH greater than 7 are alkaline, or basic, and become increasingly alkaline as values approach 14. The pH scale is logarithmic, which means that for every one-unit change in pH there is a 10-fold change in the activity of hydrogen ions in the soil solution. For example, a soil with pH 6 has ten times more

hydrogen ions in solution than a soil at pH 7, and a soil with pH 8 has ten times less hydrogen ions than a soil at pH 7. A comparison of soil pH values for some soils with some common acids and bases are shown in the Figure 5.9.

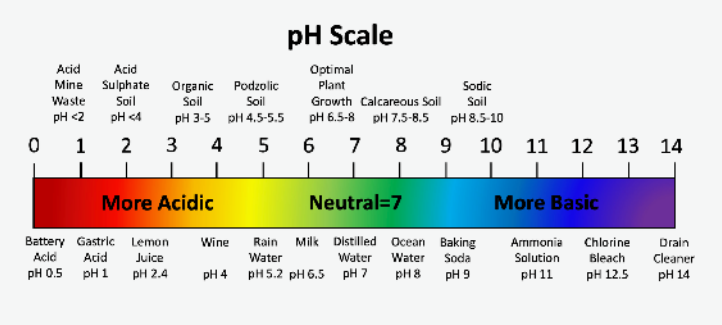


Figure 5.9. The pH scale with some common materials and soils. Adapted from Brady and Weil (2010). Adapted by Lewis Fausak and licensed under a CC BY (Attribution) license.

The primary factors influencing the natural soil pH are the geological parent materials of the soil and reaction with carbon dioxide (CO_2) dissolved in meteoric water. Organic acids derived from the decay of vegetation also influence soil pH. Generally, mineral soils in Canada are formed on one of two very broadly defined parent materials: those developed on granite-based parent materials, commonly referred to as acidic parent materials, and those developed on carbonate-based parent materials, commonly referred to as alkaline parent materials. Granitic-based parent materials are derived mostly from rocks of the **Precambrian Shield**, while carbonate-based parent materials, typically limestone and dolostone (dolomite), are derived from Paleozoic aged rocks or younger. Soils formed from granitic-based parent materials are generally in the acidic pH range. Conversely, soils formed on carbonate-based parent materials typically have soil pH values in the neutral to alkaline range.

Rainwater equilibrates with atmospheric gases, particularly CO_2, and has a pH of approximately 5.6. Other factors such as atmospheric pollutants (e.g. sulphur dioxide which hydrolyzes to sulphuric acid) typically drive the pH lower. The general trend for all soils is to be driven to the acidic end of the pH scale with increasing length of exposure to precipitation, which may cause leaching of basic cations. Granitic parent materials generally contain very little basic mineral material and therefore the pH of granitic-based soil is easily driven to the acid end of the pH scale. Podzolic soils, typically developed on granitic parent materials, have pH values in the range of 4.5 to 6.0. Although it may appear somewhat counter intuitive, the pH of soils containing carbonate minerals are buffered against the acidifying effects of dissolved CO_2 in the soil solution. Calcareous soil materials typically have pH values in the range of 7.5 to 8.5.

Soil pH may be buffered in some soils. That is, although pH reflects the concentration of H+ in solution at the time of measurement, there are additional sources of H^+ on exchange sites referred to as “exchangeable acidity” that buffer and maintain the pH in solution. Through exchange of H^+ between the soil solution and exchange sites, high CEC soils can buffer soil pH, which means they will resist a change in pH when an acidic or basic compound (e.g., a fertilizer) is added. Acidity, measured immediately in the soil solution, is referred to as “**active acidity**”, while acidity on the exchange sites is referred to as “**reserve acidity**”. Some soils are highly pH buffered with a large amount of reserve acidity (e.g., soils with high clay and/or organic matter content), while others are not (e.g., sandy soils with low organic matter content).

Soil pH affects plant nutrient availability. However, not all nutrients are available at a specific pH level. Soil pH values in the range of 6.5-8.0 are considered optimum for the availability of most nutrients to plants. Generally, extreme pH values are detrimental to nutrient uptake and plant growth. For example, acidic pH values increase the solubility of aluminum, manganese and iron, which are potentially toxic to plant roots when in high concentrations in the soil solution. These elements also may react with plant nutrients to render them unavailable to plants. Similarly, alkaline pH values result in excess amounts of calcium in the soil that also react with nutrients (e.g., phosphate) so they are unavailable for plant use. General ranges of nutrient availably with pH in mineral and organic soils are shown in Figure 5.10.

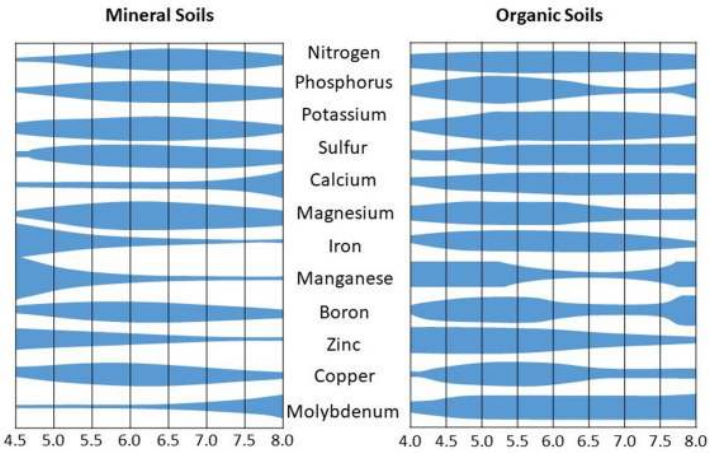


Figure 5.10. Nutrient availability in mineral and organic soils influenced by soil pH. Adapted from Munroe (2018) by Lewis Fausak and licensed under a CC BY (Attribution) license.

Acids and Bases

Soil pH values reflect the various reactions involving acid-base pairs in water, which itself dissociates:

$$H_2O \leftrightarrow H^+ + OH^-$$

An aqueous solution is said to be neutral (pH = 7) when the concentration of protons [H^+] exactly equals the concentration of hydroxyl ions ([OH^-]), with both equal to 10^{-7} moles per litre.

An acid is any chemical substance that can donate a proton when dissolved in water, while a base is any substance that can accept a proton. When an acid loses a proton it forms the conjugate base:

$$HL \leftrightarrow H^+ + L^-$$

In the above expression HL is an acid capable of donating a proton (H^+) to solution and L^- is the conjugate base, capable of accepting a proton. In this expression, the base (L^-) or conjugate base is also referred to as a ligand. The strength of an acid is the measure of its ability to dissociate in water and release protons to solution. A strong acid (e.g., sulphuric acid) is a substance that completely dissociates into protons and ligands.

A base is any chemical substance that can generate hydroxide ions (OH^-) or takes up H^+ ions in solution forming its conjugate acid.

$$B + H_2O \leftrightarrow OH^- + BH^+$$

Humified soil organic matter, although perhaps still poorly understood in terms of specific structural/chemical makeup, is the source of surface functional groups that influence the chemistry of the soil. Some of the most important organic functional groups include carboxylic (–COOH), phenolic-OH, alcoholic-OH, carbonyl (C=O) and quinone groups, as well as amine (-NH_2) and thiol (-SH) groups. The carboxylic acid, phenolic-OH and alcoholic–OH functional groups contribute to acidity and ion-exchange. These groups behave as weak acids in soil acting as a source of reserve acidity and serve as cation exchange sites. Other functional groups, such as aromatic acids, dissociate and function in a similar fashion.

Soil pH controls many of the chemical processes that take place in the soil, especially plant nutrient availability. Most soils have pH in the range of 5 to 8 with the optimal pH range for most agronomically important plants between 5.5 and 7.5, though many plant species have adapted to pH values outside of this range. For example, blueberries prefer soil with pH in the range of 4.3 to 5.5, growing very poorly in soils with pH in the alkaline range (pH>7).

If soil contains significantly more Na^+ in the soil solution compared to Ca^{2+} or Mg^{2+}, there may be a risk of sodium buildup. This ion buildup is of greater concern in arid or semi-arid regions (e.g., Canadian prairies) than in more humid regions (e.g., eastern Canada). Soils with naturally high sodium levels, or those that have received large quantities of sodium bicarbonate (Na_2CO_3) through irrigation, may have pH levels of 8.5 or higher. This high pH value is due to the presence of Na_2CO_3 precipitates in the soil which produce the elevated pH. This process is a characteristic of Solonetzic soils discussed later in this chapter.

Measurement of Soil pH

The pH of a soil is likely the most frequently measured soil chemical property. Soil pH measurements are typically made using a commercial pH meter equipped with a combination glass pH electrode. Combination electrodes are single electrodes that include both a glass electrode and a reference electrode. There are numerous methods for soil pH measurement, which depend on the application. Methods vary for the soil to solution ratios used; addition of an electrolyte and sample resting time prior to measurement. The methods include, but are not limited to measurements in: saturated paste; 1:1, 1:2, 1:2.5, 1:5, and 1:10 soil:water suspensions; 1:2 and 1:5 soil:0.01M $CaCl_2$ suspensions; 1M KCl; and 1.0M NaF (Soil Survey Staff, 2014; Hendershot et al., 2008). The soil pH measurement is that of the solution in contact with the soil solids and not of the soil solids themselves (Hendershot et al., 2008). Soil pH values measured in electrolyte (e.g., 1:2 soil to 0.01 M $CaCl_2$) suspensions are usually more stable and reproducible than those taken in water because the concentration of the electrolyte is constant. However, the addition of an electrolyte generally produces a decrease in soil pH compared to water; e.g., values measured in soil:water suspensions are typically 0.5–0.6 units higher than those in 0.01M $CaCl_2$, and 0.7 units higher than those in 1M KCl. This is because of the inclusion of exchangeable (reserve) acidity introduced to solution through ion exchange reactions involving Ca^{2+} or other introduced ion with Al^{3+} and/or H^+ ions on the exchange sites.

Although the pH of a soil containing a carbonate mineral should ideally be 8.2, soil pH measurements for these soils are commonly in the 7.6 to 8.5 range. This measured pH is a result of several factors. First, in addition to exchangeable ions such as Ca^{2+}, Mg^{2+}, Na^+ and K^+, the ion exchange sites on soil also contain small amounts of exchangeable aluminum and iron, which on entering the soil solution hydrolyze and react to reduce the measured soil pH. Second, although soils may contain carbonates, the mineral dolomite [$Ca,Mg(CO_3)_2$] reacts (dissolves) much slower compared to calcite ($CaCO_3$), and the pH may not stabilize within the timespan (e.g. within an hour) of the measurement. Third, typical procedures for pH measurements (e.g., Hendershot et al., 2008) recommend allowing prepared sample suspension to stand for one hour prior to pH measurement. This procedure provides a standard reaction time allowing easy adaptation for routine high sample throughput operations in commercial soil laboratories that follow standard ISO17025 accredited analytical methods. Although this time frame is sufficient to allow for ion exchange equilibration, it may or may not be sufficient to allow full equilibration for dissolution of carbonate and equilibration with CO_2 in the atmosphere. Other chemical reactions may also have an influence on the measured pH.

Soil pH Adjustment

The natural tendency is for soil to become more acidic through the natural acidity of rainwater, decay of organic matter and through application of fertilizers such as urea and anhydrous ammonia. However, soil pH may also be increased through additions of agricultural limestone. The timing and amount of agricultural "**lime**" applied depends on the crop to be grown, with lime application being commonly advised when soil pH values decrease below 6.0 (Munro, 2018). Finely ground limestone (primarily comprised of the mineral calcite; $CaCO_3$) or dolostone (primarily comprised of the mineral dolomite; $Ca,Mg(CO_3)_2$) are the most frequently used liming products, though many other products with acid neutralizing capacity may be used.



Images of the Martindale area of Sudbury in 1981 **(top)** and 2018 **(center)** after application of the "Regreening Recipe", which includes manual application of lime **(bottom)**. Photo credits: K. Winterhalder and P. Beckett.

The Sudbury Regreening Program

A century of sulphur dioxide fumigation and nickel particulate deposition by base-metal smelters, coupled with fire, soil erosion, and enhanced frost action left thousands of hectares of barren land with eroded soils and stunted, open birch-poplar woodlands in the Sudbury area.

The relict soils of the area are coarse textured, eroded Podzols that are acidic, metal-rich, nutrient poor and drought-prone—all of which have contributed to limit plant recolonization. The drought vulnerability of seedlings germinating in the soils was further enhanced by aluminium toxicity limiting root growth, and further exacerbated by enhanced frost action in the medium textured soil that lacked the protection of surface humus forms previously lost as a result of wind and water erosion.

A revegetation program that detoxified the soils enough to allow for successful colonization was developed based on field research in the 1970's. This research demonstrated that the surface application of ground dolomitic limestone [primarily $CaMg(CO_3)_2$] at 10 tonnes per hectare adjusted the soil pH from less than 3.0 to between 5.2 and 5.6. Application of the liming treatment, with or without an accompanying fertilizer and/or a grass-legume seed application, led to the natural colonization by woody species including birch, aspen, and willows. This colonization has further been aided by planting a mixture of native coniferous and deciduous species together in groups to form a "seed bank" for future colonization – with nearly 10 million trees being planted in the Sudbury region between 1976 and 2020.

Adjustment of soil pH towards the acid end of the scale may be required when growing certain crops such as potatoes, blueberries, azaleas or some other horticultural crops. Downward pH adjustment is typically accomplished with the addition of elemental sulphur, sulphur containing compounds or ammonia containing fertilizers; however, application of amendments to intentionally decrease soil pH is very uncommon.

SOIL SALINITY AND SODICITY

Soil **salinity** levels are controlled by the concentration of total dissolved mineral salts present in the soil solution. Salts occur naturally in both soil and water and may, under certain conditions, accumulate in soil. The accumulation of soluble salts in the soil is known as **salinization**. Dissolved salts that contribute to soil salinity include the cations Na^+, Ca^{2+}, Mg^{2+}, and K^+, and the anions Cl^-, SO_4^{2-}, HCO_3^-, CO_3^{2-}, and NO_3^-.

Sodicity is related to salinity, but refers specifically to the concentration of Na^+ in the soil solution relative to concentrations of Ca^{2+} and Mg^{2+}. The degree of sodicity, estimated by calculation of the **sodium adsorption ratio** (SAR), is a measure of the concentration of Na^+ in saturated soil extract relative to Ca^{2+} and Mg^{2+} concentrations. The SAR is calculated as the ratio of the Na^+ concentration divided by the square root of average [Ca^{2+} + Mg^{2+}] concentrations. The square root of [Ca^{2+} + Mg^{2+}] concentrations is used since these are divalent ions while Na is monovalent. Sodicity can be quantified as **exchangeable sodium percentage** (ESP); i.e., the exchangeable Na^+ expressed as a percentage of cation exchange capacity. If the SAR of the soil equals or exceeds 13, or if the ESP of the soil equals or exceeds 15%, the soil is termed sodic.

$$SAR = \frac{[Na^+]\ (mmol\ L^{-1})}{(0.5 \times [Ca^{2+} + Mg^{2+}])^{\frac{1}{2}}\ (mmol L^{-1})} \quad (3)$$

$$ESP = \frac{[Na^+]\ (cmol(+)\ kg^{-1})}{CEC\ (cmol(+) kg^{-1})} \times 100 \quad (4)$$

Development of Salinity in Soils

Salts in soils originate from natural and anthropogenic sources. Natural sources include weathering of primary minerals, atmospheric deposition, sea water intrusions, discharge of saline groundwater, and brines from natural salt deposits. Major anthropogenic sources of salts include irrigation with waters containing high salt concentrations (see Figure 5.11), fertilization with synthetic and organic fertilizers with high salt content including manure, biosolids, and compost. Road salts applied during winter in Canada to melt ice and improve road conditions, are also a localized source of salinity to soils.

Soluble salts accumulate in soils for various reasons:

- Dry climatic conditions – Soluble salts accumulate in arid and semiarid areas where rainfall is not sufficient to leach the salts through the soil profile. High evaporation losses from these soils result in the upward movement of water, which may bring salts from the deeper soil layers that then remain at the soil surface as visible crusts when water evaporates. Evaporative losses can range from 50–90% in arid regions resulting in 2- to 20-fold increase in levels of soluble salts.
- Drainage – Poor drainage (due to high water table or low subsoil permeability) can increase the salt content.
- Irrigation water quality – Salinity problems are common in irrigated lands as most irrigation waters contain varying concentrations of soluble salts. The water added to the soil is utilized by the crop or evaporates directly from the moist soil, leaving the dissolved ions behind to form saline crusts.
- Coastal regions – Soils in coastal areas may have high salt accumulation through sea water intrusions and saline groundwater.
- Other management practices – Salt accumulation may take place due to other management practices such as overuse of synthetic fertilizers, continuous application of manures, or yearly application of road salts that are washed into road-side ditches and/or blown onto nearby fields and vegetation by wind.

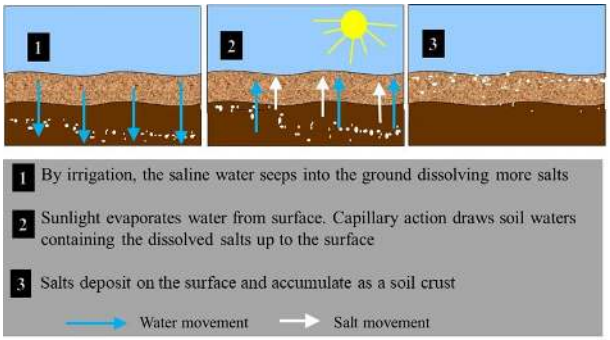


Figure 5.11. Salinization of soils by irrigation with saline water. © Graeme Spiers, Laurentian University is licensed under a CC BY (Attribution) license.

Measurement of Soil Salinity

The total dissolved salt concentration of a soil can be measured by extracting a soil sample with water and measuring the concentrations of individual salts in the extract. In practice however, measuring individual salts is time consuming, cumbersome and expensive; thus the electrical conductivity (EC) of a solution, which increases in direct proportion to the overall abundance of dissolved salts, is employed as an alternative index to assess and describe soil salinity. Soil extract EC is the preferred measurement to assess the overall salinity status of the soil. The measurement of EC—accomplished with a conductivity meter in soil testing

laboratories—is reliable, rapid and cost effective. The EC of a soil is determined in a saturated extract (i.e., in the soil solution extracted from a saturated soil paste) and expressed in units of decisiemens per meter (dS m^{-1}) or millisiemens per centimeter (mS cm^{-1}). Soils with an EC value of saturated soil extract >4 dS m^{-1} are classified as saline.

Properties of Salt-Affected Soils

Salt-affected soils can be classified as **saline**, **saline-sodic** or **sodic**, depending on the total concentration of Na^+ salts in the soil. Saline soils have high salt content but low concentrations of Na^+. Sodic soils have relatively low salt content but high concentrations of Na^+. Saline-sodic soils have high concentration of both salts and Na^+. Figure 5.12 illustrates the classification of salt-affected soils based on the EC, SAR and ESP.

Saline soils have high concentrations of total dissolved salts, with EC values greater than 4 dS m^{-1} (Figure 5.12). These soils have relatively low amounts of Na^+ in the soil solution or on the exchangeable complex, and thus SAR values are less than 13 and ESP values are less than 15%. The soil pH of saline soils is usually less than 8.5. The high salt concentrations in these soils promote flocculation of soil colloids (i.e., individual colloids are attracted together forming flocs or clusters).

Sodic soils have low total concentrations of dissolved salts, with EC values less than 4 dS m^{-1}. These soils, however, have relatively high concentrations of Na^+ in the soil solution and on the exchange complex relative to Ca^{2+} and Mg^{2+}. Therefore, SAR values are greater than 13 and ESP values are greater than 15% (Figure 5.12). Sodic soils typically have pH values greater than 8.5 due to the hydrolysis of Na_2CO_3. The colloids in these soils are dispersed due to the swelling in response to the high proportion of exchangeable sodium. Swelling and dispersion is greater when the exchange complex has a greater proportion of hydrated, monovalent ions particularly Na^+. When hydrated Na^+ ions come between clay particles, the electrostatic forces that bind the clay particles together are neutralized and additional layers of water molecules force the particles apart, thus dispersing the soil matrix materials. The high pH also disperses soil organic matter. Alkaline sodic soils classified as Black Solonetz have dispersed, dark coloured humic material coating the soil peds.

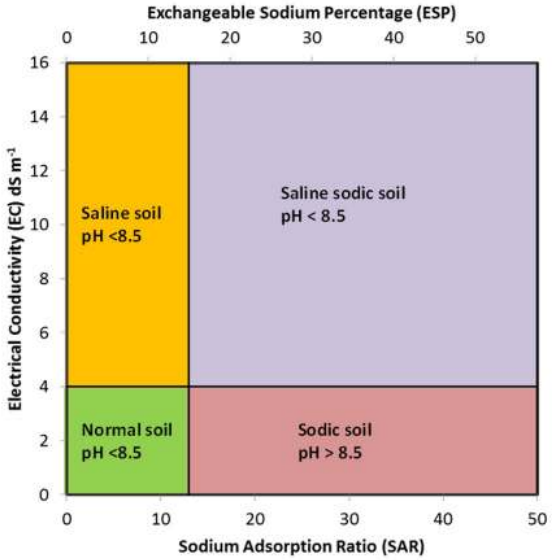


Figure 5.12. Classification of salt-affected soils. Adapted from McKenzie and Woods (2010) by Jim Warren and redrawn by Darshani Kumaragamage, Univ. of Winnipeg. © Jim Warren and Darshani Kumaragamage is licensed under a CC BY (Attribution) license.

Soils having high concentrations of dissolved salts as well as high concentrations of Na^+ are termed saline-sodic soils. Such soils have EC values greater than 4 dS m^{-1}, SAR values greater than 13 and ESP greater than 15%. Saline-sodic soils usually have pH values less than 8.5. These soils are flocculated because of the high concentration of dissolved salts (Sparks, 2003).

Effects of Soil Salinity and Sodicity on Ecosystem Eealth

Soil salinization, one of the most devastating global environmental concerns, is a major cause of land degradation and contributes to desertification of dry ecosystems. Worldwide, salinization of soils leading to salinity and sodicity have caused major reductions in cultivated areas and crop productivity.

Both salinity and sodicity affect soil properties, plant growth, and microbial activity. High salt concentrations will adversely affect plant growth as the dissolved salts lower the osmotic potential of the soil solution. Since plant roots take up water primarily by osmotic mechanisms, they are unable to absorb water from highly saline soils, with some plants being more sensitive to saline conditions than others. This resultant condition is known as “physiological drought” since plant roots are unable to take up water even when the water content in the soil may be adequate for normal growth. Similarly, microbes in the soil may be affected because of an inability to obtain water due to osmotic effect.

Table 5.4. Sensitivity of some plants to saline conditions

Sensitive	Moderately Sensitive	Moderately Tolerant	Tolerant
(EC < 4 dS m^{-1})	**(EC 4-7 dS m^{-1})**	**(EC 7-9 dS m^{-1})**	**(EC 9-12 dS m^{-1})**
Vegetables	Alfalfa	Fescue	Barley
Maple	Corn	Aspen	Wheat
Tomatoes	Grasses	Oats	Willow
Potatoes	Clover	Soybean	
Note: plants capable of surviving EC in the range of 12–15 dS m^{-1} are considered halophytes; however, very few, if any, plants can survive EC values >15 dS m^{-1}			

The high concentration of sodium in sodic soils favours swelling of expandable clay minerals and dispersion of clay particles, resulting in poor soil structure, reduced permeability and infiltration due to blockage of the soil pores, and formation of surface crusts. Low infiltration and permeability may also increase soil erosion and may lead to saturation of the soil surface during irrigation and rain/snowmelt events, in turn restricting air entry into the soils. This leads to anaerobic conditions in soils which also negatively affects plant growth.

Salt-Affected Soils: Distribution Globally and in Canada

Soil salinization is a global problem. Salt-affected soils occur on all continents, mostly in arid and semi-arid areas. Salinization is of lesser concern in humid areas as rainfall is sufficient to leach excess salts from the soil profile. Salinized areas in the world increase at a rate of about 10% annually because of natural and anthropogenic causes. Based on the 1970–1980 world soil map, the total area of salt affected soils worldwide is estimated as 397 million ha of saline soils and 434 million ha of sodic soils (FAO, 2020). Figure 5.13 illustrates the distribution of saline and sodic soils worldwide (Zheng, 2014). Globally land degradation due to salinization has been reported in the Aral Sea Basin (Amu-Darya and Syr-Darya River Basins), the Indo-Gangetic Basin in India, the Indus Basin in Pakistan, the Yellow River Basin in China, the Euphrates Basin in Syria and Iraq, the Murray-Darling Basin in Australia, and the San Joaquin Valley in the United States (Qadir et al., 2014).

There are an estimated 6 to 8 million ha of salt-affected soils in Canada, located primarily in the Canadian prairies (Miller and Brierley, 2011). In 1995, the risk of salinization was rated as moderate to very high affecting an estimated 34% of farmland area in the Prairie Provinces (Acton and Gregorich, 1995), with the estimate decreasing to 12% of farmland area by 2001 (Eilers et al., 2010). This decrease in the risk of salinization potential in agricultural lands in the Prairie Provinces is largely due to the reduction in summer fallow, and conservation practices such as conversion to permanent cover.

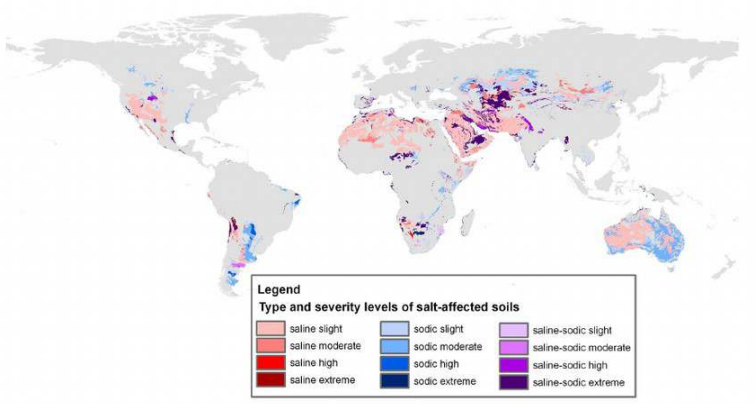


Figure 5.13. Global distribution of saline and sodic soils. Source: Zheng (2014), used with permission.

Can You Dig It!

Surface and subsurface salinity in the Canadian Prairies – A threat to soil quality and sustainable ecosystem productivity

Soil salinization is a typical process for the Canadian prairies and adjacent areas of the Boreal Plains. It is estimated that at least 1,431,00 ha in the region are affected by salinity. The two images below show white saline crust, which often occurs adjacent to shallow wetlands (sloughs) on the prairies. The salts are dominantly calcium and magnesium sulphates (gypsum and epsom salts) derived from evaporation of saline groundwater discharge from the bedrock below the regional glacial till veneer. The surface expression is the result of a gently rolling topography combined with a high water table in the wetland.



Surface expression of a severely salt affected soil from southwest Manitoba **(left**; Photo credit: Robert G. Eilers). Chernozemic to Solonetz catena near Drumheller, Alberta **(right**; Photo credit: G.A. Spiers). Note the saline crust in the lowland area.



Roots on the surface of a Brown Solodized Solonetz column impacted by salt efflorescence at base (Photo by G.A. Spiers)

The image to the right shows the effects of subsurface salinity. The most common expression of salt-affected soils is irregular growth of crop plants across a field. The soil surface may look normal, but salinity is often greater at depth in the root zone. Plants typically germinate and grow after seeding due to available moisture at the surface. As the soil dries and plant water consumption increases, the roots penetrate deeper searching for water. At some depth they can encounter higher levels of salt in the subsoil, which acts as a barrier to accessing moisture. The result is stunted or irregular plant growth, and possibly death due to lack of moisture.

Reclamation of Salt Affected Soils

Methods of removing salts from surface soils include scraping, surface flushing and leaching. Since the salts are highly soluble, leaching with good quality water is the most effective and commonly used method of reclaiming saline soils. While saline soils are aggregated and have high infiltration rates, sodic and saline-sodic soils are usually dispersed and have low infiltration rates. Thus, it is difficult to leach the salts and sodium with water from sodic and saline-sodic soils since they are typically dispersed. Thus, to reclaim sodic and saline-sodic soils, application of soil amendments such as gypsum ($CaSO_4 \cdot 2H_2O$) is usually needed. The added Ca^{2+} promotes soil aggregation and improves water infiltration when monovalent Na^+ is replaced by divalent Ca^{2+}. The displaced Na^+ and excess salts can then be leached by irrigation with good quality water. However, because leached salts may accumulate in the groundwater, saline drainage water may be diverted through subsurface drains. Other ways of removing salts from salt-affected soils include growing salt tolerant crops (e.g., barley, sunflower) and planting trees that will take up the dissolved cations and anions from soil solution. Preventing salinization by using best management practices (e.g., use of good quality irrigation water in agricultural lands, managing water below root zone through subsurface drainage) is a better approach than allowing salinity to increase in soils and then attempting to reclaim the salt-affected soils.

OXIDATION-REDUCTION (REDOX) REACTIONS

Oxidation-reduction reactions, or redox reactions, involve the transfer of electrons and the conversion of an element from one valence state to another. Understanding the behaviour of redox sensitive elements in soil is central to understanding the behaviour of Gleysolic and Organic soils, and gleyed soil subgroups that occur within all mineral soil orders in Canada (SCWG, 1998). Internationally, about 130 million ha of land, accounting for about 10% or all arable land on the planet, is temporarily flooded for rice production generating reduced soil conditions and emphasizing the importance of understanding redox reactions (IRRI, 1997).

Redox reactions are typically written as half reactions (i.e., the electron and change in oxidation state are explicitly stated in the equation). The reversible reaction is always to be written as a reduction reaction with the electron (e^-) on the left side of the equation:

$$\text{oxidants} + e^- \leftrightarrow \text{reductants}$$

where oxidants are chemical species capable of accepting electrons and reductants are chemical species capable of donating electrons. For example:

$$Fe^{3+} + e^- \leftrightarrow Fe^{2+}$$

Redox measurements are useful to provide an understanding of chemical behaviour in wet and poorly drained soils (e.g., Gleysolic or Organic soils and gleyed subgroups; see Figure 5.14) as a high redox potential (>300 mV) indicates good availability of O_2 (aerobic conditions), whereas a low redox potential (<300 mv) indicates limited availability of O_2 (anoxic conditions).



Figure 5.14. Poorly drained Gleysol from Delta, BC during wet fall/winter. © Maja Krzic, UBC is licensed under a CC BY (Attribution) license.

Key redox sensitive elements in soil include oxygen (O_2), nitrogen (NO_3^-, NO_2^-, NO_x, N_2O, N_2), manganese (Mn^{4+}, Mn^{2+}), iron (Fe^{3+}, Fe^{2+}), sulphur (SO_4^-, S^{2-}), and carbon (CO_2, CH_4). These are the main species commonly involved in redox reactions in soils and constitute a reaction sequence indicating the tendency of each element to become reduced:

Oxygen > Nitrogen > Manganese > Iron > Sulphur > Carbon

Typically, elements on the left of this sequence dominate the prevailing oxidation state of the other elements to the right of it. For example, the presence of oxygen dominates electron acceptance over all other elements in the sequence and thus dictates that each of the other species must donate their electrons and be converted to their oxidized form. Thus, in the presence of oxygen (i.e., oxidizing environments) the stable valences of the other elements are NO_3^-, Mn^{4+}, Fe^{3+}, SO_4^{2-}, and CO_2.

Table 5.5. Half reactions, standard electrode potentials (E°), typical Eh range in soils and mediating micro-organism group for dominant redox species in soil

Half Reaction	E° (millivolts)	Eh range (millivolts)	Micro-organism
$O_2(aq) + 4\ H^+ + 4\ e^- \leftrightarrow 2\ H_2O$	1230	600 to 350	Aerobes
$2\ NO_3^- + 12\ H^+ + 10\ e^- \leftrightarrow N_2(g) + 6\ H_2O$	1240	400 to 200	Denitrifiers
$Mn(VI)O_2(s) + 4\ H^+ + 10\ e^- \leftrightarrow Mn2+ + 2\ H_2O$	1220	300 to -50	Mn-reducers
$Fe(III)(OH)_3(s) + 3\ H^+ + e^- \leftrightarrow Fe^{2+} + 3\ H_2O$	930	50 to -50	Fe-reducers
$SO_4 + 9\ H^+ + 8\ e^- \leftrightarrow S2- + 4\ H_2O$	250	-150 to -200	Sulphate reducers
$CO_2(aq) + 8\ H^+ + 8\ e^- \leftrightarrow CH_4(aq) + 2\ H_2O$	170	-150 to -250	Methane fermenters

The notation used to identify redox species in in the aqueous phase (i.e., soil solution) is different than that use to identify elements occurring in the solid phase. For example, iron species in the aqueous phase are typically noted as Fe^{2+} and Fe^{3+}, whereas iron species present in soil minerals and other solids are expressed as Fe(II) and Fe(III).

Generally, oxidative decomposition (decay) of organic matter is mediated by various microorganisms to derive energy (see Table 5.5). Consequently, organic matter is the primary source of electrons in soil. Conversion/decay of organic matter varies considerably from complete conversion to CO_2 (gas) to partial and incomplete oxidation and conversion to intermediate products (i.e., conversion to humus), but all involve the reduction of redox species to their reduced form. Because the composition of soil organic matter is highly complex (see Chapter 3) there is no general chemical formula attributed to it (van der Park, 2006).

Figure 5.15 is an example of an Eh-pH diagram, or Pourbaix diagram, a graphical representation showing the approximate ranges for Eh and pH in soil. Soil pH values generally range from about 4 to 10, with Eh bounded on the top and bottom by the redox stability of water. The Eh varies with pH according to Nernst equation by -59.16 mV per pH unit at 25 ℃ (298.15K). Figure 5.15 also indicates the approximate ranges for oxic conditions (O_2 reduction reaction is dominant and mediated by aerobic organisms); sub-oxic conditions (nitrogen, iron and manganese reduction reactions are dominant and mediated primarily by facultative organisms); and anoxic conditions (sulphur and carbon reduction reactions are dominant and mediated by anaerobic organisms) in soil.

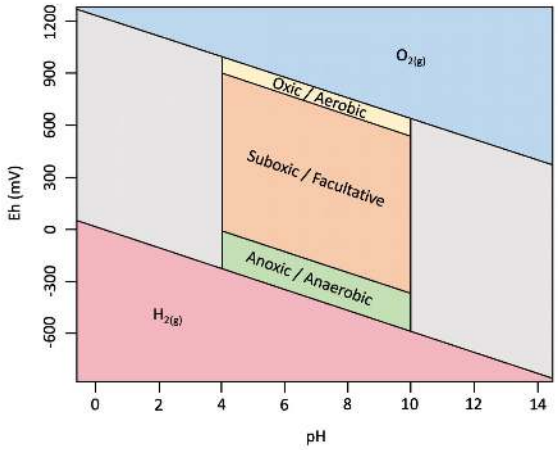


Figure 5.15. Common Eh and pH ranges for soils including approximate ranges for oxic, sub-oxic and anoxic environments. Figure created by James Warren, adapted by Lewis Fausak and licensed under a CC BY (Attribution) license.

Oxygen

In non-saturated (aerobic) soils, oxygen occurs in both the soil gas phase (O_2 g) and in the soil solution (aqueous) phase (O_2 aq). Although similar in composition to that of the atmosphere (about 20% O_2 by volume), the concentration of oxygen in soil air is typically somewhat lower as oxygen in the soil gas phase is constantly consumed by plant roots and microbial respiration. Under oxygenated (non-saturated) conditions, oxygen is effectively in infinite supply capable of accepting electrons and completely oxidizing soil organic carbon to CO_2 (g). Continued introduction of oxygen to topsoil due to cultivation practices is the primary reason for the reduction in soil organic matter levels as a result of oxidation reactions.

Although the concentration of oxygen in the soil gas phase is usually high, the concentration of dissolved oxygen in water is far less, being generally on the order of 10 mg L^{-1} (at 15 ℃) but varying with temperature. As dissolved oxygen becomes consumed, it is replenished by the normally abundant pool of oxygen in the soil gas phase. Aside from the large difference in the abundance of oxygen, the mobility of oxygen is roughly 10,000 times slower in water compared to oxygen in the gas phase. Consequently, as soils become increasingly saturated with water the oxygen supply from air becomes increasingly limiting until all, or almost all, of the soil pore space becomes filled with water, and oxygen transport is possible only via the aqueous phase. The redox potential decreases as the soil effectively becomes sub-oxic or even anoxic/anaerobic (see Figure 5.15).

Can You Dig It!

Seasonally anoxic soils during spring snowmelt flooding



Flooded farmland in southern Manitoba during the spring snowmelt period. Photo credit: Don Flaten.

Most farmlands in the Canadian prairies get flooded during the spring snowmelt period. Seasonal flooding can make the soil sub-oxic or anoxic as the redox potential decreases. This will have many environmental impacts. For example, phosphorus concentration in floodwater may increase when Fe(III) and Mn(IV) are reduced to divalent forms releasing the phosphate associated with them. Released phosphorus may end up in lakes causing algal blooms. Anoxic conditions may also enhance emissions of greenhouse gases such as nitrous oxide (N_2O) and methane (CH_4).

Nitrogen

The primary reason for the depletion of available oxygen from soil is saturation with water. As discussed above the amount of dissolved oxygen in soil water is very limiting and diffusion of oxygen through water is very slow. As oxygen becomes limiting/depleted, typically under temporary saturated conditions, nitrate-reducing microbes become active initiating the process of denitrification. The process of denitrification is a major pathway for the loss of fertilizer nitrogen applied as a nutrient source for crops. Nitrate (NO_3^-) is reduced to nitrogen gas (N_2) with nitrite (NO_2^-), nitric oxide (NO) and nitrous oxide (N_2O) as intermediate species. Soluble nitrite is a short-lived aqueous intermediate while nitric oxide (NO) and nitrous oxide (N_2O) intermediates are involved in the destruction of stratospheric ozone (O_3) layer. Nitrous oxide is also an important greenhouse gas with a global warming potential (GWP) 265–298 times that of carbon dioxide.

Manganese

As nitrate becomes depleted, manganese reducing bacteria take over. Although the total content of manganese in soil is typically only about 750 to 1000 mg kg^{-1} (Ure and Berrow, 1982; Bowen, 1979) it is still a major redox species in soil. Manganese exists in soil mainly as Mn^{2+}, or as Mn(VI). The soluble form prevalent in anoxic soil is Mn^{2+} (aq), which when oxygen is introduced oxidizes and precipitates manganese as Mn(IV), usually as non-crystalline (amorphous) oxy-hydroxides. Manganese oxides appear as black coatings on soil particles, or as black nodules, some as large as 2 cm in diameter. When present, manganese solids act as scavengers for nutrient anions (e.g., phosphate).

Iron

Iron is the fourth most abundant element in soil and is responsible for diverse mottles characterized by the drastic differences in colouration between reduced (Fe^{2+}) and oxidized (Fe^{3+}) valences in wet and water-logged soils. Strongly reduced soils with Eh values in the range of -200 mV typically have gleyed (dull grey or practically white) colour due to the presence of reduced iron. This reduced soil condition means that dissolved oxygen as well as nitrate and manganese are essentially absent and that most dissolved iron is present as the Fe^{2+} species, with most iron compounds (responsible for the grey colour) occurring as Fe(II) oxides and hydroxides (e.g., ferrous hydroxide [$Fe(OH)_2$]. When the water table lowers or oxygenated water infiltrates into a reduced soil (perhaps through a root channel, open pore or fracture), ferrous iron transfers electrons to the oxygen resulting in the formation of oxidized [Fe(III)] compounds. These are observed as mottles occurring as reddish blotches on an otherwise dull grey background. Mottles are a very common morphological feature in Gleyed sub-groups or Gleysolic soils that are periodically wet or have a fluctuating water table. Mottles are diagnostic features in the identification of moderately well, imperfectly, and poorly drained soils.

Sulphur

Although there are several oxidation states for sulphur, only two oxidation states are prevalent in soil. Under aerobic conditions sulphate (SO_4^{2-}) is the stable form with an oxidation state of +6, while sulphide (S^{2-}) with an oxidation state of -2 is the stable form under strongly reducing conditions (McBride, 1994). Sulphate is the primary form of sulphur taken up by plants, precipitating primarily as the mineral gypsum ($CaSO_4 \cdot 2H_2O$) when there is sufficient dissolved calcium (Ca^{2+}) present in the soil solution. Sulphate may be reduced to sulphide which has very low solubility in water under anoxic conditions forming either odourous hydrogen sulphide gas (H_2S with the aroma of rotten eggs), or insoluble metal sulphides such as FeS_2 if soluble metals are present.

The introduction of oxygen into sulphide containing soils results in the re-oxidation of metal sulphides:

$$FeS(s) + {}^{15}/_{4}\, O_2 + {}^{7}/_{2}\, H_2O \leftrightarrow Fe(OH)_2 + 4H^+ + 2\, SO_4^{2-}$$

This reaction is common in reclaimed coastal soils high in sulphides, in Northern Prairie soils derived from sulphide containing shales (referred to as acid sulfate soils) and in mine wastes (tailings and waste rock) containing pyrite and other metal sulphides.

Carbon

Ultimately if sulphate is limited, then methane producing bacteria continue the process of organic matter decomposition under anaerobic conditions as follows:

$$2CH_2O \leftrightarrow CO_2 + CH_4$$

Methane production typically occurs under stagnant, water saturated conditions in wetlands. Under these conditions the organic decay process is slow and incomplete, resulting in the accumulation of soil organic matter and the formation of peatlands and organic soils.

Measurement of Redox

The direct measurement of redox (oxidation-reduction) potential in soil is a quandary. Fundamentally, redox measurement is extremely simple, being similar to pH measurement in mechanics, but very different from pH with respect to fundamental chemistry and interpretation. Redox measurements or measuring the "concentration of electrons" is not analogous to pH because free electrons do not exist in soil solutions in the same way as hydrogen (H^+) ions. Redox is typically measured as a voltage (usually reported as volts or millivolts), which represents a measure of the potential for electrons to be transferred from one element to another. This potential measure provides an indication of which redox species are the most receptive to gaining electrons at the time of the measurement. Actual times for electron transfer are typically much slower than the time required to complete a redox measurement, with reactions being mediated by specific biological organisms as discussed above. The redox measurement thus indicates the tendency or likelihood that a redox reaction will occur, while the reaction itself may not truly occur to any meaningful extent as the measured redox condition may be only temporary.

Regardless of difficulties with measurement, redox measurements provide critical data for the description and analysis of wet or seasonally saturated soils. Measurements are actually recordings of voltage (Eh) between a reference electrode and a sensor electrode inserted into a soil. The most common sensor electrodes are usually constructed of platinum wire (Pt electrode). The voltage results from an exchange of electrons between a redox couple such as ferrous (Fe^{2+}) and ferric (Fe^{3+}) iron during the process of reduction and oxidation. In soils that have fluctuating wet and dry conditions, wide seasonal fluctuations in Eh values occur (Fiedler et al., 2007).

Although free electrons do not exist in soil, the concept of free electrons—analogous to describing free-proton activity and pH values—is useful as a first step in describing and understanding soil redox status. Similar to pH, "electron activity" may be express as pe (analogous to pH) which is the negative common logarithm of electron activity:

(5) $pe = -log(e^-)$

Values for pe are not measured directly, but derived from Eh through the Nernst equation:

(6) $Eh\ volts = E^o - \frac{RT}{nF}\ ln\frac{reductants}{oxidants} = 0.05916\ pe$

where E^o is the standard electron potential for the half-reaction, R is the molar gas constant (8.31 joules/mole/K), T is absolute temperature (298.15 K); F is the Faraday constant (96,490 coulombs/mole) and n is the number of electrons transferred in the half reaction.

Although the measurement of Eh using a platinum electrode is a simple procedure, it is also subject to measurement uncertainty. Principle issues (Sposito, 1989) include:

- voltage measurement in soil is usually due to more than one-half reaction;
- concentrations of redox species in solution may be too small to provide electron transfer detectable by the Pt electrode;
- several redox species are not readily reactive;
- Pt electrodes are easily contaminated or 'poisoned' by oxides or other coatings; and,
- the salt content of the soil solution (and its ability to conduct electrons) may differ greatly from the calibration buffer.

These uncertainties mean that Eh measurements are considered qualitative measures that provide a general indication of the soil redox status sufficient to differential oxic, suboxic, and anoxic environments (Figure 5.14). The Eh data should not be used to numerically quantify the concentrations of redox species but can be used to determine which redox couples are most active at the time of measurement. Quantitative measurements for redox species should involve direct measurement of the species of interest; for example, dissolved oxygen or ion specific electrode measurement, or multi-species techniques such as ion chromatography.

SUMMARY

- Soil chemistry involves the study of the chemical composition, chemical properties and chemical reactions involving soil solids (organic matter and inorganic minerals), liquids (soil water), gases (soil air), including interactions with plants and soil

organisms. Soil chemical properties have huge implications for global human and environmental health. Through the influence on fate of nutrients and contaminants, soil chemical properties play a role in crop production and food security, biodiversity, water and air pollution, global climate change through greenhouse gas emissions etc. Knowledge of soil chemistry is fundamental in understanding the fate of nutrients and contaminants in the soil system, predicting their environmental impact and finding solutions to mitigate impacts.

- Soil chemical properties are largely influenced by the type and quantity of inorganic minerals present in the soil and the organic matter (humus) content. Inorganic colloids mainly include phyllosilicates (layer silicate minerals), amorphous aluminosilicates and oxyhydroxide minerals.
- The presence of surface charges in both inorganic minerals and humus particles is responsible for ion adsorption and exchange in soils. Most charges are permanent (not dependent on soil pH) and negative, while others vary with pH and may be negative or positive.
- Negative charges on soil particles can attract and adsorb cations, which can be released back to the soil solution through ion exchange. Cation exchange capacity, defined as the sum of total exchangeable cations that a soil can adsorb, is an important soil chemical property that influences the retention of cationic plant nutrients supporting plant growth and retention of cationic pollutants minimizing groundwater pollution.
- Soil pH is a measure of acidity and alkalinity, which controls many of the chemical processes that take place in the soil including availability of plant nutrients. The primary factor influencing the natural soil pH are the geological parent material from which soils are derived. Liming is a practice commonly used to raise the pH of acidic soils.
- Soil salinity gives an indication of the salt content of the soil and is determined by measuring the electrical conductivity of a saturated soil paste extract. Depending on the electrical conductivity and the relative concentration of sodium in soil solution or exchangeable complex, salt-affected soils are classified as saline, saline-sodic and sodic soils. Irrigation with saline waters, particularly in arid climates, is one of the major anthropogenic causes of salinity. Excess salinity in soil can decrease plant available water and affect plant growth.
- Oxidation-reduction (redox) reactions in soil involve the conversion of an element from one valence state to another. Key redox sensitive elements in soil include oxygen (O_2), nitrogen (NO_3^-, NO_2^-, NO_x, N_2O, N_2), manganese (Mn^{4+}, Mn^{2+}), iron (Fe^{3+}, Fe^{2+}), sulphur (SO_4^{2-}, S^{2-}), and carbon (CO_2, CH_4). Redox reactions influence dynamics for certain plant nutrients (eg., nitrogen, sulfur, iron, manganese) and gaseous losses from soils, including greenhouse gas emissions (nitrous oxide and methane).

STUDY QUESTIONS

1. Explain why montmorillonite has a higher shrink-swell capacity than kaolinite.
2. Define isomorphous substitution. Briefly explain how isomorphous substitution causes negative charge development on clay mineral surfaces.
3. Explain why Vertisols in Canada are typically very sticky when wet and form wide cracks when dry.
4. Define cation exchange capacity. Why is soil cation exchange capacity important in protecting and enhancing biodiversity and ecosystem health?
5. Consider the structure and properties of phyllosilicates summarized in Figure 5.5 and answer the following True/False questions.
    - Presence of interlayer hydrogen bonds makes kaolinite a relatively non-expandable clay mineral.True or False
    - 2:1 type clay minerals are composed of one tetrahedral sheet sandwiched between two octahedral sheets.True or False
    - There are no interlayer hydrogen bonds in 2:1 type clay minerals since hydroxyl groups are not exposed.True or False
    - Montmorillonite and illite (hydrous mica) are both 2:1 type phyllosilicates. Montmorillonite freely expands while illite has a limited expandability.True or False
    - Illite (hydrous mica) has a high layer charge due to isomorphous substitution but its cation exchange capacity is relatively low.True or False
6. A soil with a cation exchange capacity of 42 cmol(+) kg^{-1} has the following cations on the exchange complex: Na^+ = 7 cmol(+) kg^{-1} soil, Ca^{2+}= 12 cmol(+) kg^{-1} soil, K^+= 6 cmol(+) kg^{-1} soil, and Mg^{2+}= 8 cmol (+) kg^{-1} soil). Calculate the base saturation of this soil.
7. Explain how a change in soil pH influences the cation and anion exchange capacities of soils.

8. Differentiate between active and reserve acidity.
9. “Soil pH is greatly influenced by the parent material from which soil is formed”. Explain the statement giving examples of parent materials in Canada to support your answer.
10. Briefly explain how soil colloids buffer soil pH changes caused by acidic or basic inputs, such as nitrogen containing fertilizers and liming amendments.
11. Briefly explain how soil pH influences nutrient availability to plants.
12. Define saline, sodic and saline-sodic soils. Outline their major characteristics and effects on ecosystem health.
13. Briefly explain how irrigation under arid climates can influence soil salinity development.
14. A soil with a cation exchange capacity of 20 cmol(+) kg^{-1} has exchangeable Na content of 8 cmol(+) kg^{-1}.
    - Calculate the exchangeable sodium percentage (ESP) of the soil.
    - If the electrical conductivity of the saturated extract of this soil is 5.6 dS m^{-1} (decisiemens per meter), how would you classify the soil? (Hint: See Figure 5.11)
    - How can such a soil be reclaimed?
15. Identify the key redox sensitive elements in the soil. Give one oxidized form and one reduced form for each of these elements.
16. Describe how the redox state of the soil influences nutrient availability and greenhouse gas (nitrous oxide and methane) emissions.

WORKED EXAMPLES

Describe the different mechanisms by which negative charges are developed on surfaces of soil colloids.

Two main mechanisms are responsible for the development of negative charges to surfaces of soil colloids. One is isomorphous substitution in phyllosilicate minerals where central cation in the octahedral or tetrahedral sheet in the mineral is substituted by other cations. The substitution occurs with ions of similar sizes and a negative charge is developed when a higher valent cation is replaced by a lower valent cation. For example, if the Si^{4+} in the tetrahedral sheet is replaced by Al^{3+} (as in illite or hydrous mica), it will give rise to a surface negative charge. Similarly, in the octahedral sheet, Al^{3+} may be substituted by Mg^{2+} which creates a negative charge on the mineral. The negative charges due to isomorphous substitution is permanent and not dependent on pH.

The second mechanism of development of negative charges on soil colloids is the deprotonation (dissociation) reactions depending on pH. Oxyhydroxide clays and the edges of phyllosilicate and amorphous silicate clays have hydroxyl groups which can dissociate at high pH developing a negative charge on surface. Organic colloids (humus) also can carry pH dependent negative charges due to dissociation of functional groups such as carboxyl and phenolic groups. These charges increase with increasing pH.

Explain why the clay mineral montmorillonite has a higher CEC than kaolinite (hint: see Figure 5.2)

Montmorillonite is a 2:1 phyllosilicate clay mineral with 2 silicon tetrahedral sheets enclosing an aluminum octahedral sheet. Isomorphic substitution, e.g. of magnesium (Mg^{2+}) for some of the aluminum (Al^{3+}) in the octahedral sheet results in a permanent negative charge. In addition, montmorillonite clays expand when wetted, as water enters the interlayer space between adjacent tetrahedral sheets. This increases the available surface area for cation adsorption, since in montmorillonite has both external & internal surfaces available for cation adsorption. The combination of net negative charge due to isomorphic substitution and high specific surface area gives montmorillonite a high CEC.

Kaolinite, on the other hand, is a 1:1 phyllosilicate clay mineral, which consists of single silicon tetrahedral sheets alternating with single aluminum octahedral sheets. There is very limited isomorphic substitution, and therefore the kaolinite has a limited number of negative charges. Additionally, being a 1:1 phyllosilicate clay mineral, which does not expand, kaolinite has lower effective surface area on which cations can be adsorbed. Thus kaolinite has a lower CEC.

Fill in the characteristics of the phyllosilicates in the table below.

Mineral	Type	Isomorphous substitution	CEC	Surface area	Expandability
	(1:1, 2:1, 2:1:1)	(Low/ High)	(Low/ High)	(Low/ High)	(Low/ High)
Kaolinite	1:1	Low	Low	Low	Low
Montmorillonitee	2:1	High	High	High	High
Illite (hydours mica)	2:1	High	Low	Low	Low
Chlorite	2:1:1	High	Low	Low	Low

Calculate the CEC of the soil

A soil has 3% organic matter, 20% montmorillonite, and 17% kaolinite with the remaining 60% being sand and silt sized primary mineral grains. Calculate the cation exchange capacity of the soil in cmol (+) kg^{-1} (Note: assume cation exchange capacity of organic matter = 200 cmol(+) $kg^{-1,}$ montmorillonite = 100 cmol(+) kg^{-1} and kaolinitie = 10 cmol(+) kg^{-1}).

CEC of the soil = [0.03 x 200 cmol(+) kg^{-1}] + [0.20 x 100 cmol(+) kg^{-1}] + [0.17 x 10 cmol (+) kg^{-1}] = 27.7 cmol(+) kg^{-1}

Identify four major factors that contribute to soil acidity and briefly explain their contribution.

Major factors contributing to soil acidity are acidic parent material, high rainfall, application of ammonium fertilizers, and decay of organic matter. Soils derived from granite-based parent materials, commonly referred to as acidic parent materials are generally acidic in reaction. The natural acidity of rainfall, because of the dissolved carbon dioxide, is a cause of soil acidity. In areas of high rainfall basic cations in soils are leached, increasing the acidity. Application of ammonium fertilizers may also contribute to acidity because the nitrification of ammonium release protons. The decay of organic matter release organic acids contributing to soil acidity.

Name two natural and two anthropogenic causes of soil salinity and briefly describe how they contribute to salinity development.

Two natural causes of soil salinity are the dry climatic conditions and poor drainage. Soluble salts accumulate under dry climatic conditions where rainfall is not sufficient to leach the salts through the soil profile. High evaporation losses from the soils may bring the salts from deeper soil layers with the upward movement of water, which remain at the soil surface.

Under poor drainage conditions (due to high water table or low subsoil permeability), salts are not removed through drainage and leaching, which will result in accumulation of salts on the surface when water evaporates from the soil.

Two anthropogenic causes of salinity are irrigation with saline waters, particularly in dry regions, and overuse of fertilizers. Irrigation water will move downward and dissolve more salts. When the evaporation is high under a dry climate the water moving up will bring dissolved salts. Water will evaporate leaving the salts on the surface. Fertilizers (which are salts containing different nutrients) can increase the salt content of soils if used inappropriately.

Consider the following data for four mineral soils that all have a clay loam texture:

Soil Property	Soil #1	Soil #2	Soil #3	Soil #4
pH	8.8	8.1	7.5	7.6
electrical conductivity (dS/m)	2.7	12.2	1	4.3
base saturation (%)	90	90	60	86
cation exchange capacity	12	11	19	9
(cmolc/kg)				
exchangeable sodium (%) or ESP	17.3	16.8	2.1	2.5
Salinity classification	Sodic	Saline-sodic	Non saline	Saline

A) Determine the salinity classification (i.e., saline, sodic, saline-sodic, or non saline) for each of these 4 soils. [Hint: see Figure 5.12]

B) Which of the above soil(s) would most likely be in a dispersed state? Briefly explain your answer based on the information provided.

Soil #1 is most likely be in dispersed state since it has a high ESP of 17.3% while the salt concentration is low (with EC of 2.7 dS m^{-1}).

C) What are the implications of a dispersed soil for plant growth?

Since individual soil particles are in dispersed state, there is no flocculation and consequently no aggregate formation. The consequence of this is that soil will have poorly developed soil aggregate structural units and be dominated by micropores, which prevent air & water movement and also impede root growth. Thus, plant growth will be impaired.

Identify the common oxidized and reduced forms of the following elements in soils: manganese, iron, sulfur, and nitrogen.

- manganese: Mn^{4+}, Mn^{2+} (reduced)
- iron: $Fe^{3}+$, Fe^{2+}
- sulphur: SO_4^{2-}, S^{2-}
- nitrogen: NO_3^-, NO_2^-, NO_x, N_2O, N_2

REFERENCES

Acton D.F., and Gregorich L.J. 1995. The health of our soils: toward sustainable agriculture in Canada. Agriculture Agri-Food Canada, CDR Unit, Ottawa.

Bowen H.J.M. 1979. Environmental chemistry of the elements. Academic Press, London, New York.

Brady N.C. and Weil, R.R. 2010. The Nature and Properties of Soils 13th Ed. Pearson Education.

Canada Department of Agriculture; Research Branch 1976. Glossary of Terms in Soil Science, Publication 1459.

Eilers W., MacKay R., Graham L. and Lefebvre A. (eds.). 2010. Environmental Sustainability of Canadian Agriculture: Agri-Environmental Indicator Report Series — Report #3. Agriculture and Agri-Food Canada, Ottawa, Ontario.

FAO 2020. Food and Agriculture Organization of the United Nations. FAO Soil Portal. http://www.fao.org/soils-portal/en/

Fiedler S., Vepraskas M.J., and Richardson J.L. 2007. Soil Redox Potential: Importance, Field Measurements and Observations. Advances in Agronomy 94:1-54.

Havlin J.L., Tisdale S.L, Nelson W.L., and Beaton J.D. 2013. Soil Fertility and Fertilizers: An Introduction to Nutrient Management, 8th ed. Upper Saddle River, NJ: Prentice Hall.

Hendershot W.H., Lalande H., and Duquette M. 2008. Soil Reaction and Exchangeable Acidity. Pages 173-178 IN Carter, M.R. and Gregorich, E.G. (eds.) Soil Sampling and Methods of Analysis. 2nd Edition. Canadian Society of Soil Science. CRC Press, Baton Rouge.

Hendershot W.H., Lalande H., and Duquette M. 1993. Soil Reaction and Exchangeable Acidity. Pp. 167-176 in Carter M.R. (ed.) Soil Sampling and Methods of Analysis. 1st Edition. Canadian Society of Soil Science. CRC Press, Baton Rouge.

International Rice Research Institute (IRRI). 1997. Rice Almanac, 2nd Ed. International Rice Research Institute, Manila.

McKenzie R.H. and Woods S.A. 2010. Management of sodic soils in Alberta, Agdex 518-20, Alberta Agriculture. https://www1.agric.gov.ab.ca/$depart...18-20.pdf?Open Element

McBride M.B. 1994. Environmental Chemistry of Soils. Oxford University Press. New York. 406pp.

Miller J.J. and Brierley J.A. 2011. Solonetzic soils of Canada: Genesis, distribution and classification. Can. J. Soil Sci. 91(5):889-902.

Munroe J. (ed.) 2018. Soil Fertility Handbook. Publication 611 3rd edition, Ontario Ministry of Agriculture, Food and Rural Affairs. 236pp.

Newman A.C. and Hayes M.H. 1990. Some properties of clays and of other soil colloids and their influences on soils. In: De Boodt M.F, M.H. Hayes, A. Herbillon (Eds). Soil colloids and their associations in aggregates. Springer Science & Business Media; Boston. MA, USA. pp. 39-55.

Qadir M., Quillérou E., Nangia V., Murtaza G., Singh M., Thomas R.J., Drechsel P., and Noble A.D. 2014. Economics of salt-induced land degradation and restoration. In: Natural Resources Forum 38 (4):282-295.

Shreeja D. (n.d.) Phosphate Fixation in Soil: 3. Reactions of Anion Fixation. Available at https://www.soilmanagementindia.com/...-fixation/2252

Sparks D.L. 2003. Environmental Soil Chemistry. Academic Press, London, UK.

Sparks D.L. 2019. Fundamentals of Soil Chemistry. Encyclopedia of Water: Science, Technology, and Society, pp 1-11.

Sposito G. 1989. The Chemistry of Soils. Oxford University Press. New York. 277pp.

Soil Classification Working Group (SCWG), 1998. The Canadian System of Soil Classification. 3rd Edition, Research Branch Agriculture and Agri-Food Canada, Publication 1646.

Soil Survey Staff. 2014. Kellogg Soil Survey Laboratory Methods Manual. Soil Survey Investigations Report No. 42, Version 5.0. R. Burt and Soil Survey Staff (ed.). U.S. Department of Agriculture, Natural Resources Conservation Service. 1001pp.

Ure A.M., and Berrow M.L. 1982. The Elemental Constituents in Soils. Pages 94-204 IN Bowen, H.J.M. (ed.). Environmental Chemistry. Volume 2. Royal Soc. Chem. London.

Van der Perk M. 2006. Soil and Water Contamination from Molecular to Catchment Scale. Taylor and Francis, London 389pp.

Zheng T. 2014. Numerical Analysis of Modeling Concepts for Salt Precipitation and Porosity-Permeability Evolution during Brine Evaporation. Doctoral dissertation, Universität Stuttgart, Germany.

About the Authors

Darshani Kumaragamage, Professor, Department of Environmental Studies and Sciences, University of Winnipeg, Manitoba, Canada.



Darshani Kumaragamage (licensed under a CC-BY-NC-ND license)

I obtained a BSc and a MPhil in Agriculture from the University of Peradeniya, Sri Lanka, and a PhD in Soil Science from the University of Manitoba, Canada, specializing in soil physical chemistry. I have over 30 years of teaching experience at the University level and have taught a variety of courses in Soil Science including Introduction to Soil Science, Soil Chemistry, and Soil Fertility as well as courses in Environmental Science. My research interests include assessing, monitoring, and mitigating environmental impacts of fertilizers and animal manure, with special focus on phosphorus transformations and mobility in soils under different climatic and management scenarios. I have authored/coauthored 65 peer-reviewed journal publications.

My favorite memory of working with soil was when I conducted a field research in Manitoba on a very sticky, chernozemic, clay soil on one rainy day. I had to take soil samples from many plots, and as I stepped into the field, about one kilo of soil got stuck onto my boots. With the next step, it almost doubled, and with each step I took, my boots got heavier and heavier until I literally had to drag my feet to go from one plot to another.

C. James (Jim) Warren, Ph.D. P.Geo., Land Resource Specialist, Ontario Ministry of Agriculture Food and Rural Affairs, Guelph, Ontario, Canada.



Jim Warren (licensed under a CC-BY-NC-ND license)

I am a P.Geo. and an environmental soil scientist with a diverse background in research, consulting and teaching. I am a graduate of both the University of Guelph (B.Sc. (Agr), Soil Science), and the University of Alberta (M.Sc. Soil Science; Ph.D. Soil Chemistry). I conducted research in agronomy, geochemistry, and mineralogy at the University of Guelph and University of Waterloo before becoming a private consultant in acid mine drainage. I am currently employed as a Land Resource Specialist conducting soil survey and pedology work with the Ontario Ministry of Agriculture Food and Rural Affairs, and am also an Adjunct Professor with the University of Guelph. I have delivered courses in soil science, soil mineralogy, soil chemistry, geochemistry, and pedology as a sessional lecturer at four universities (Alberta, Guelph, Waterloo, and Toronto). I have authored or co-authored over 100 publications in refereed journals; book chapters, conference proceedings and reports.

How did I get interested in soils? Well, I grew up on a farm where, as a teenager, I spent a lot of time on a tractor doing tillage operations and had a lot of time to think. One of the questions I asked myself during long days in the field was: “What makes plants grow?” With that I spent the next several decades trying to understand soils to answer that question. The quest continues . . .

Graeme Spiers, Chair of Environmental Monitoring and Professor, School of the Environment, Department of Earth Sciences / Department of Biology, Laurentian University, Sudbury, Ontario, Canada.



Graeme Spiers (licensed under a CC-BY-NC-ND license)

I teach in the School of the Environment, the Department of Earth Sciences and the Department of Biology, Laurentian University. I obtained a B.Sc. in Earth Science and Botany at the University of Waikato in New Zealand, and M.Sc. and Ph.D. degrees from the University of Alberta, specializing in Pedology, Clay Mineralogy and Chemistry. With commercial laboratory and consulting experience across Canada, I have been associated with research programs at the Universities of Alberta, Guelph, Waterloo, Laurentian and New Brunswick. My research has resulted in numerous technical and scientific publications and presentations in analytical chemistry, aquatic chemistry, environmental geology, ecology, plant biology and soil science. Prior to attending university on a full-time basis, I was a New Zealand dairy farmer for 10 years, earning my undergraduate degree as a long-term hobby between milkings. In Sudbury, I am involved in research initiatives examining the effects of both historical and current anthropogenic metal emissions on soils, rivers and lakes, as well as vegetation within the Sudbury Smelter Footprint. I am actively involved in a variety of research programs across Canada, and actively collaborates with researchers in Russia (Moscow State University and the Russian Academy of Sciences), Australia, New Zealand, China, Peru and South Africa.

This page titled 4.10: Soil Chemistry is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Canadian Society of Soil Science via source content that was edited to the style and standards of the LibreTexts platform.

4.11: Soil Biodiversity and Ecology

Nathan Basiliko, Kari Dunfield, and Mario Tenuta

Learning Objectives

On completion of this chapter, students will be able to:

1. List and describe the major types and groups of biota that live in soils, including their size, distribution, and taxonomy
2. Define the metabolic strategies of soil biota in terms of how they harness energy and nutrients from the environment
3. Describe some of the major challenges and technological breakthroughs in how we study and learn about soil biota
4. Describe how biological communities are integral drivers of organic matter turnover, nutrient cycling, and soil fertility that are covered in detail in Chapters 3 and 7
5. Describe some of the important trophic interactions (i.e., who eats whom) among biotic groups in soils

INTRODUCTION

The mysterious community beneath your feet!

Since Schatz and Waksman's 1943 discovery of the lifesaving antibiotic streptomycin produced by soil bacteria, soil biology has grown to be one of the most exciting and cutting-edge disciplines. Microbiologists have gone on to discover numerous other vital classes of antibiotics produced by soil microbes. Have you ever had a bacterial or fungal infection and been prescribed an antibiotic?- thank soil microbes. But those discoveries only begin to scratch the surface of the value of the biotic communities that make soils their home. Soils are now recognized as arguably the most biodiverse habitat on Earth. The exceedingly complex physical and chemical makeup of soils you encountered in Chapters 3-5 has given rise to vast niche diversity and, thus, the evolution of millions, if not billions, of unique taxa of soil biota. These communities dynamically shape the soil environment, and in the process influence such vitally important ecosystem services ranging from plant and crop growth and health, to water quality, even the regulation of the global climate system through the storage of atmospheric carbon and exchange of atmospheric greenhouse gasses.

Can You Dig It!



Figure $PageIndex1$: Professor Waksman working on soil enrichment cultures with students.

As unlikely as it may sound, a soil scientist won the Nobel Prize in Physiology or Medicine for curing tuberculosis, once one of the world's most devastating respiratory diseases. Ukrainian born Selman Waksman spent most of his career as a professor at Rutgers University in New Jersey (U.S.A.), and was a pioneer in the early 20th century studying soil microbiology and organic matter.

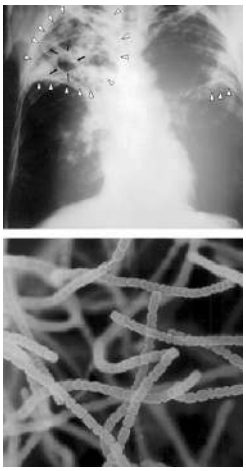


Figure $PageIndex2$: X-ray image of a human lung with severe *Mycobacterium tuberculosis* (top). Micrograph of filamentous *Streptomyces griseus* (bottom; the width of each filament is ~0.5 micrometers) the soil bacterium that produces streptomycin- the first antibiotic to effectively defeat *M. tuberculosis*.

His group was one of the first to tackle the complex chemical nature of soil humus (see Chapter 3), including how soil microorganisms play roles in its formation and decay. Waksman's research group also focused on the unique physiological properties of the fungi and bacteria they were culturing from soils, including that some produced powerful antibiotics.

Ph.D. student Albert Schatz was supervised by Waksman and characterized Waksman's large collection of Streptomyces cultures obtained from local soils and composts for the ability to kill the pathogen Mycobacterium tuberculosis, which led to eventual isolation of streptomycin, the first antibiotic that effectively killed the pathogen without also killing the patient! The story goes that, according to Schatz, Waksman was too afraid to work with the Mycobacterium cultures for which there was no cure at the time, and never ventured to the lab where Schatz was handling the TB-causing microbe. The pair published on the seminal discovery of streptomycin together- and streptomycin was later produced commercially saving countless lives. This discovery was not without controversy, however, as Schatz was initially shut out of any patent royalty payments, and the 1952 Nobel Prize in Physiology or Medicine was awarded solely to Waksman. The apparent rationale was that the difficult task had been obtaining the extensive soil bacteria cultures, while the selection procedures to test for antibiotics were more routine.

Key Terminology

Biological communities are made up of multiple populations interacting with each other. A population refers to all of the same species in a defined area (e.g., the population of humans in Canada; the population of redback salamanders living in the soil of a 1-hectare forest stand; etc.). An ecosystem refers to a biological community dynamically interacting with its abiotic environment; in soils this abiotic environment is the organic and mineral solid components and water and gasses in pore spaces. The concept of species is complicated for many groups of soil microorganisms and small invertebrate fauna. For these, DNA sequence similarity or uniqueness of taxonomic marker genes in relation to sequences in other organisms often defines unique or same taxa (or "species", if only loosely defined).

In this chapter we introduce the range of soil organisms, from sub-micrometer diameter (that is, less than 1/1,000,000 m) unicellular bacteria and archaea to the much-loved (but also occasionally loathed) night-crawler, to vertebrate fauna. What are their strategies for survival and growth, and how do they interact with each other? How do soil microbes and fauna dynamically interact with their physical environment (building on Chapter 4) and with plants? We will explore soil biota as key players driving the cycling of carbon and organic matter you read about in Chapter 3, and the cycling of essential nutrients like nitrogen that you will read more about next in Chapter 7.

An important theme will emerge in this chapter: we presently know very little about the breadth and depth of soil biodiversity, nor do we understand the exact functions of many populations and communities. This is especially true for the soil microorganisms and microarthropods, however, in the past decade, molecular biology tools for detecting and characterizing biodiversity have begun to give us a picture of the millions to billions of "known unknown" organisms dwelling in soils. This makes soil biology an exciting,

breakthrough field; however, we can almost certainly guarantee that if you are reading this chapter 10 years or so from now, at least some of what we now “know” will have changed!

SOIL BIODIVERSITY

Who lives in the soil?

One small handful of healthy surface soil contains more organisms than the whole population of humans on Earth, up to 10 billion individual cells in a single gram! Soil biota range from microscopic unicellular bacteria and archaea that lack nuclei and other organelles to warm-blooded borrowing small mammals like voles and woodchucks, as well as the hidden half- give or take- of plants. The range of sizes of organisms in soil ecosystems spans across nearly ten orders of magnitude! For example, the smallest soil bacterium is ~0.5 micrometers in diameter (0.000005 m), soil-dwelling amphibians and mammals might typically be ~10 cm long (0.1 m), plant roots are typically in the 0.1 to 10 m length range, while certain pathogenic soil forest soil fungi grow thread-like mycelia that extend up to 1 km (1000 m) and beyond! Traditionally soil scientists have lumped soil biota together roughly by some combinations of size and function (see Figure 6.1). Although it is useful to be aware of these groups because the terminology still is in use, here we introduce broad groups of soil biota a bit differently, beginning with the prokaryotes, that are the bacteria and archaea. The term prokaryote refers to single cellular organisms that lack internal membrane-bound organelles (e.g., nucleus, mitochondria, chloroplast), while eukaryote refers to organisms whose cell(s) contain(s) at least a nucleus, and often times other organelles.

Soil scientists and ecologists often use the following groups to describe soil biota:

- **Microflora:** bacteria and archaea (the two domains of life termed prokaryotes, that is, organisms lacking nuclei and other organelles), algae and fungi.
- **Macroflora:** plant roots.
- **Microfauna:** protozoa (single cellular eukaryotes that are motile), small nematodes, and small arthropods.
- **Mesofauna:** metazoans (referring to the animal phylum) with body sizes typically between 0.1 and 2 mm as well as the larger predatory and omnivorous nematodes.
- **Macrofauna:** all metazoan invertebrates and members of the chordate phylum (mammals, amphibians) larger than 2 mm.

However, this scheme can be confusing. For example, whereas the group name **microflora** implies “little plant”, only the algae and cyanobacteria are – like plants – photoautotrophs while many soil bacteria and all soil fungi are heterotrophs. In another example, although certain fungi (yeasts) are unicellular, others form large macroscopic (to massive) fruiting bodies and mycelia (see Figure $PageIndex4$).

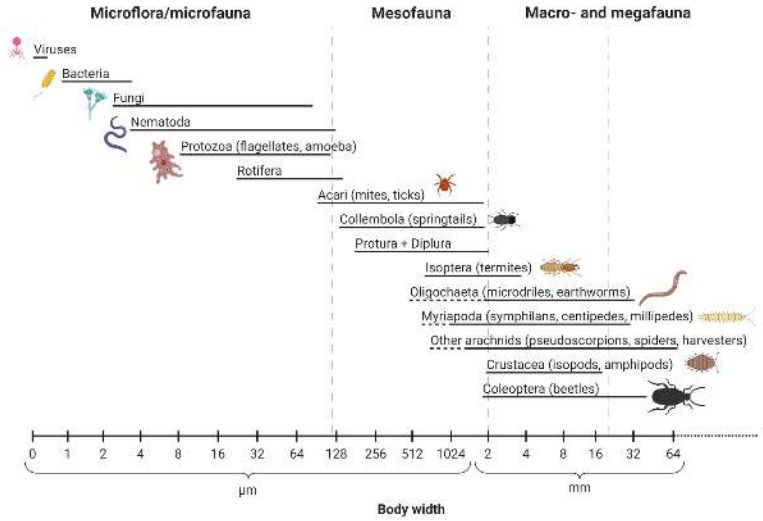


Figure $PageIndex3$: Representation of the main taxonomic groups of soil organisms based on body-size. © K. Dunfield; modified after Swift et al. (1979) and licensed under a CC BY (Attribution) license. Figure created using BioRender.com (www.biorender.com).



Figure $PageIndex4$: Despite fungi sometimes being referred to as “microflora”, branching threads of soil fungi called mycelia **(A)** and their fruiting bodies, including of the conifer tree pathogen *Armillaria ostoyae* **(B)**, are very macroscopic (i.e., visible to the human eye). Indeed a colony of *A. ostoyae* growing in the soil of a Douglas fir forest in Oregon (USA) covers multiple square kilometers and might be the largest organism (by area at least) on Earth! Photo credits: **(A)** Lex vB. (https://commons.wikimedia.org/wiki/F...C3%A9lium).jpg). **(B)** W.J. Pilsak. (https://commons.wikimedia.org/wiki/F...ia_ostoyae.jpg). Images licensed under a CC BY-SA (Attribution ShareAlike) license.

Bacteria and archaea

Being similar in size (typically just a one to a few micrometers in diameter), single cellular, and lacking organelles, it took until the early 1990s before bacteria and archaea were widely recognized as being clearly distinct groups from one another. Although they share some traits, bacteria and archaea diverged from each other evolutionarily a very long time ago. This divergence led to bacteria and archaea having fundamentally different cell envelopes (cell membranes, cell walls, and other outer layers), protein synthesis machinery, mechanisms for packaging their DNA, among many other major physiological differences. In fact, based on DNA sequence analyses, they are as dissimilar to each other as each of them is to higher plants or animals! Carl Woese from the University of Illinois (U.S.A) was the principal scientific force behind the first major redrawing of the tree of life (showing relatedness among all organisms) based on DNA sequences and proposed just three main branches representing the domains Bacteria, Archaea, and Eukarya. Despite their fundamental differences, both bacteria and archaea are small and can interact with micro-scale habitats in soil particles, and both have the ability to evolve quickly to new environments, conditions, and resources with common mechanisms of horizontal gene transfer. This means that they can have some niche overlap in soil environments, and a diverse range of metabolic strategies exists across both domains (Table 6.1), including:

- source of energy (ATP);

- source of molecules to provide electrons to the electron transport chain to produce metabolic energy as ATP; and
- source of carbon to build new organic molecules.

In surface soils, bacteria predominate over archaea in number and biomass. However, there are exceptions where lithotrophic archaea carry out very specific reactions of ammonia oxidation (see Chapter 7) in the nitrogen cycle, or producing methane in flooded, oxygen-free soils. The roles of many other soil archaea are not well known. The interesting strategy of obtaining energy (electrons) from the environment through lithotrophy literally means “eats mineral matter” (see Table 6.1).

Table $PageIndex1$. Common strategies (and related terminology, all ending with “-troph” roughly meaning food) for soil biota to acquire energy for ATP generation and carbon to build biomass. Note that only certain bacteria and archaea are lithotrophs, acquiring their electrons for redox and ATP generation from reduced mineral compounds

Source of energy/electrons	Source of carbon for building biomolecules	Termed:
Sunlight harnessed to split H_2O	CO_2 (inorganic carbon is chemically reduced or “fixed” to organic forms using the Calvin Cycle)	Photoautotroph (phototroph)
Reduced mineral (i.e., non-organic) compounds	CO_2 (inorganic carbon ‘fixed’ to organic forms using the Calvin Cycle or other assimilation pathways	Chemolithoautotroph
Reduced mineral compounds	Organic matter	Chemolithoheterotroph
Organic matter	Organic matter	Chemoorganoheterotroph (heterotrophs)

Many soil bacteria are heterotrophic (see Table 6.1). Most utilize organic carbon molecules that enter the soil ecosystem mainly as plant litters or from plant roots and are already in a chemically-reduced state (that is, carbon is in reduced form as organic molecules as opposed to carbon dioxide). Cyanobacteria that are commonly found in marine and aquatic habitats can also inhabit wet surface soil environments exposed to sunlight. Cyanobacteria are also known to live symbiotically and in mutualistic relationships with a number of plant species. Some lichens on the surface of soils also contain cyanobacteria as the photosynthetic symbiont. Unlike eukaryotic organisms, a good proportion of soil bacteria and some archaea have the ability to carry out cellular respiration in the absence of molecular oxygen (O_2), instead utilizing oxidized forms of other elements like nitrogen (N), sulfur (S), manganese (Mn) and iron (Fe) as the terminal electron acceptor in the production of ATP (the main energy currency in cells). These anaerobic respiration pathways often generate relatively more ATP than fermentation reactions, which certain eukaryotes like yeasts can carry out when O_2 is lacking, but also dramatically influence nutrient budgets in soils. Because of their small size, bacteria and archaea can reside in soil pores as small as 1 micrometer in diameter. Many are attached to the soil matrix, however, some remain **motile**, moving around and between particles via films of soil water.

Fungi

Soil fungi (generally divided into categories based on form as single cellular yeasts and multicellular, filamentous molds) are heterotrophs, more specifically, eukaryotic chemoorganotrophs (see Table 6.1). They play a very important role in the decomposition of woody and other cellulose and lignin-rich plant litters that are low in nutrients like nitrogen (see Chapter 3). Thus, they tend to dominate decomposition processes over bacteria in many forest soils, whereas bacteria may become dominant in more near-neutral pH and alkaline soils. Together, fungi and bacteria are often regarded as the main heterotrophic microbial decomposers in soils. Although fungal cell numbers are generally lower than bacteria in soils, their cell sizes are typically about ten times as large, and in many cases, the contribution to decomposition rates can be about equal between bacteria and fungi in slightly acidic soils. Fungi typically can withstand lower water availability (A_w; Chapter 4), while certain bacteria can outperform fungi in low- and no-oxygen (anoxic, also called anaerobic) conditions.

Certain fungi, called mycorrhizal fungi live symbiotically with most plant species (collectively the fungi and the plant hosts are called mycorrhizae). Specific mycorrhizal fungi are able to colonize the roots of specific plant species and extend their filamentous tufts of hyphae (the larger, often quite-macroscopic structures of hyphae are called mycelia), and at first glance might look like

plant roots themselves (Figure $PageIndex5$). Usually, the association is mutualistic in mycorrhizae, where the plant supplies sugars as sources of energy to the fungal partner, and in turn, the fungus aids the plant in acquiring nutrients from soil, warding off would-be pests or pathogens with antibiotic defense compounds, and helping improve soil structure (see Chapter 4) in the rhizosphere (i.e., the zone near plant roots) through the secretion of sticky protein and polysaccharide compounds. There are different types of mycorrhizal associations, and these are plant-specific, including ectomycorrhizal **fungi** that surround the outer layers of cells of plant roots, and arbuscular mycorrhizae that are generally obligate symbionts of plants (i.e., they can't grow without their plant host) and that extend through cell walls of the root cortex (Figure $PageIndex5$). Mycorrhizal mycelial networks in soils can be pervasive and connect roots from multiple plants- and occasionally between multiple plant species, and facilitate energy (sugars) and nutrient (nitrogen and phosphorus) flow between them! A soil internet of sorts.



Figure $PageIndex5$: Close up photo (**top left**) of ectomycorrhizal mycelium (white) associated with black spruce tree roots (brown). © A.D. Picard (https://commons.wikimedia.org/wiki/F...orhizes-01.jpg). Roots of Douglas Fir (**bottom left**) colonized by ectomycorrhizal fungi giving the appearance of 'stubby' roots © T. Trofymow. The roots feed the mycorrhizal fungi sugars and in turn, the fungi scavenge soil nutrients, provide protection from pests and pathogens, and improve the soil structure in the rhizosphere. The sexual fruiting structures of the fungus is evident on the soil surface as *Lactarius rubrilacteus* mushrooms (**top right**) © T. Trofymow. A root segment of flax is stained to show colonization of arbuscular mycorrhizal fungi (**bottom right**), with hyphae of the fungi evident as thread-like structures running between root cells. Arbuscules are the large deeply coloured blue appendages of the fungi. Arbuscules are in individual root cells, they don't kill the cell but provide phosphorus in exchange for energy from the plant © C. Welsh. Images by André Picard; T. Trofymow, Canadian Forest Service; and Cathy Welsh, University of Manitoba are licensed under a CC BY-SA (Attribution ShareAlike) license.

The life cycles of different soil fungi vary. Yeasts remain unicellular but can exist in either **(1n)** haploid or **(2n)** diploid forms under different conditions and divide through mitosis in either 1n or 2n state. The 1n body form is the main state of fungi, with the 2n state being very small and present only during sexual reproduction (genetic combination) when conditions are poor (e.g., moisture stress) or resources, like readily decomposable plant materials, are scarce. Introducing genetic variations might allow them to cope better with stress. The 2n state undergoes meiosis to return the fungus to the 1n haploid state. Stages of multicellular filamentous mould lifecycles can include fruiting body phases (mushrooms, puffballs) and single cellular spore phases. The majority of fungal growth, however, occurs by mitotic division of the 1n haploid state. Specific fungal species can be saprophytic (degrade soil organic matter and dead plant materials) or pathogenic on plants (cause plant diseases). Some of the most devastating plant diseases are caused by fungi such as rusts of cereals and grasses, and the honey fungus causing white rots of many deciduous and coniferous trees. The width of fungal hyphae means they can occupy soil pores 5 micrometers in diameter and greater.

Water moulds

Water moulds are also called oomycotes and are filamentous (hyphal forming) multicellular eukaryotes more related to the giant kelp of the oceans than fungi! They are called water moulds because they prefer living in water and in water-saturated soil. As they have hyphae, they are often confused with fungi. The normal body state is diploid (2n). They can reproduce asexually (without genetic combination) by mitosis to produce 2n motile zoospores that germinate into hyphae. Zoospores need water-filled pores to

use their flagella for motility. Sexual reproduction occurs by genetic combination of a male and female 2n hyphal strand followed by meiosis to produce a 2n non-motile oospore. Oospores are very thick-walled and can rest in the soil for years and withstand the stress of drying and freezing. Under wet and resource-abundant conditions (e.g., high density of root tips of crop seedlings), oospores germinate to produce many motile 2n zoospores. Species of oomycetes can be saprophytic or plant pathogenic. Plant pathogens can be nasty, such as the late blight pathogen of potato that helped cause the Irish Potato Famine of 1845-1849, Aphanomyces root rot of pulses, and Sudden Oak Death that is devastating oaks in California. Similar to fungi, the width of the hyphae of water moulds means they can occupy soil pores 5 micrometers in diameter and greater.

Green algae

Algae are non-plant, photosynthetic organisms (or "photoautotrophic"; see Table 6.1) whose cells contain organelles (i.e., they are eukaryotes). They are very important primary producers in aquatic ecosystems (e.g., lakes), and in aquatic and marine habitats span from single cellular to large seaweeds. Unicellular algae are found in wet areas on the surface of some soils where they receive sunlight, but typically are in very low abundance, and generally, algae don't represent an important input of newly fixed carbon through photosynthesis to terrestrial ecosystems relative to plants. Green algae reside on wet soil surfaces and surface water-filled pores greater than 50 micrometers in diameter.

An important exception to this is in lichens that can be found in soil biological crusts and colonizing rocks, bare soil surfaces, and standing and fallen woody plant parts. Lichens are interesting cases where two species join together to form an entirely new "species". Lichens represent an important mutualistic symbiosis (i.e., different organisms living together) and have both a "photobiont" and a "mycobiont". In about 90% of lichen "species", the photobiont is an alga, while the mycobiont is always a fungus. In this mutualistic relationship, the photobiont uses sunlight to fix CO_2 into organic matter (i.e., supplying food), while the fungal partner provides most of the shape and physical strength. The fungi in lichens (chemo-organo-heterotrophs; Table 6.1) also obtains water and essential nutrients from the environment (e.g., nitrogen, phosphorus) in the partnership, and provides the physical anchor the keep the lichen in place. In ~10% of lichens, the photobionts are photosynthetic bacteria called Cyanobacteria. Recently, some lichens have been identified as being a tripartite symbiosis between an alga, cyanobacterium, and fungus. Lichens have the ability to live in very harsh environments, and are important early colonizers of barren soils, and even rocks (Figure 6.4)! Lichens that contain a cyanobacterium also have the neat ability to biologicallly fix atmospheric nitrogen gas to ammonia, which is usable by the fungal partner and input of plant-available nitrogen to the soil. Thus, lichens serve many important ecological roles as colonizers providing cover in exposed and new soils such as Regosols (Chapter 8) preventing erosion, building soil through the input of organic matter and plant-available nitrogen to the soil, increasing chemical weathering of mineral surfaces by the production of organic acids, and initiating succession of plants on soils.



Figure $PageIndex6$: The lichen *Xanthoria elegans* growing in cracks of graywacke rock along the shore of Hudson Bay at Churchill, Manitoba. Lichens are often first colonizers of exposed land surfaces and important primary producers and biological nitrogen fixers in those environments. This lichen can survive extremely harsh conditions. In fact it survived being 18 months in space outside the International Space Station. That's one tough organism. © Mario Tenuta, University of Manitoba is licensed under a CC BY (Attribution) license.

Protozoa

Protozoa are single cellular eukaryotic (they have a nucleus and often other organelles) heterotrophic (they eat organic matter; see Table 6.1) microbes that can move about in water (they are motile). Although ranging in size, a typical protozoan in soil might be ten times larger than a soil bacterium, but still microscopic. Broadly, protozoa are grouped by how they move; amoeba shift their cytoskeleton and pull themselves with pseudopods, ciliates like paramecium are covered in short hair-like cilia that bend in unison to propel the cell, and flagellates use one or multiple long whip-like flagella for motility. Although much remains to be learned about the diversity and roles of protozoa in soils, many feed on bacteria, and in some cases fungi, and thus can play important roles in structuring the microbial community through this predation and contribute to the mineralization of carbon (Chapter 3) and nutrients such as nitrogen and phosphorus (Chapter 7) from the bodies of grazed microorganisms or small particles of soil organic matter. Protozoa are typically found in larger numbers in the rhizosphere where both organic matter and bacteria are abundant compared to root-free areas of soils. The populations of protozoa can oscillate with those of bacteria in soil, where populations increase in response to the growth in bacteria such as after rewetting of soil or addition of animal manures to the soil. As swimmers by definition, protozoa typically reside in the films of waters on soil particles and in water-filled pore spaces, and numbers are low in drier soils. Protozoa reside in water-filled pores greater than 50 micrometers in diameter.

Nematodes

Nematodes are multicellular eukaryotic, nonsegmented roundworms. They constitute as much as 90% of the number of multicellular animals in soil, and their numbers often exceed several million per m^2 of A horizon (Chapter 2) soil! They are not worms such as earthworms; those are segmented worms and a very different group of soil organisms. Nematodes vary with species from 0.5 mm to 5 mm in length. Nematodes are long and narrow with either tapered or rounded ends (head and tail). Compared to the organisms we have learned about so far, nematodes are very advanced in the degree of specialization of their body components, having a mouth and esophageal system for feeding, digestive tract, anus for the elimination of wastes, glands and openings for osmoregulation with soil water, male or female reproductive organs, openings and various chemo and tactile sensory structures, and nervous system and rudimentary brain. Nematodes move through water-filled or near water-filled soil pores greater than 50 micrometers in diameter.

Can You Dig It!

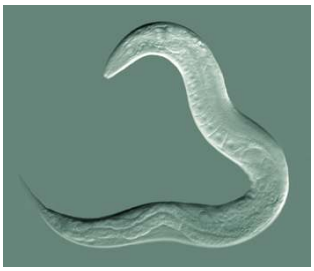


Figure $PageIndex7$: *Caenorhabditis elegans*, adult hermaphrodite © B. Goldstein; licensed under a CC BY-SA (Attribution ShareAlike) license.

***Nematodes** are very hardy and fascinating animals. The most famous nematode is Caenorhabditis elegans, which is thought to have been first collected from a stockpile of animal manure in England. This nematode is a bacterial feeder and present in humid temperature soils with a plentiful supply of plant residues. Most C. elegans individuals are hermaphroditic worms, self-fertile females. It can be dried or frozen for decades and still be revived. It has been to space on many missions by NASA, was the first multicellular organism to have its genome sequenced, its cell development from egg to adult fully characterized, and first and only organism to have its nerve cells and nervous system completely mapped out. C. elegans, as scientists call the nematode, has been the subject of many breakthroughs in recent science history, including research leading to three Nobel Prizes, 2002 and 2006 in Physiology or Medicine, and 2008 in Chemistry.*

A nematode species can be assigned to a specific trophic group being plant, bacterial, fungal, algal, omnivorous, or predatory feeding (Figure $PageIndex8$). Plant feeders are herbivores, and some can cause very damaging diseases such as the soybean cyst nematode, root-knot of tomato and potato early dying, but most plant-feeding nematodes do limited damage to plants. Omnivorous nematodes can feed on fungal hyphae, fine roots and hairs, other nematodes and mesofauna. Predatory nematodes feed on nematodes and mesofauna.

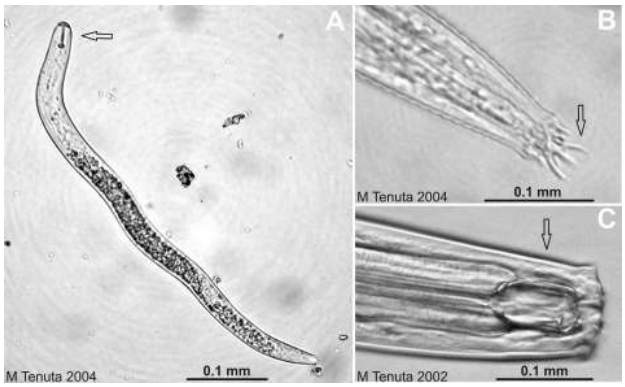


Figure $PageIndex8$: Micrographs of nematodes. It is fairly easy to determine what soil mesofauna feed on based on their feeding structures or 'mouths'. **(A)** a root lesion nematode (*Pratylenchus penetrans*) that is a plant-parasitic nematode with a 'pin-like' spear structure called a stylet (arrow) is protruded into root cells for feeding. **(B)** The head area of a bacterial-feeding nematode of the genus Acrobeles. Bacterial-feeders have small mouth cavities called a stoma where bacteria are ingested. The arrow points to cool appendages called probole of uncertain purpose. They may aid in scraping bacteria biofilms for feeding. **(C)** The head area of a predatory nematode that preys on other nematodes and mesofauna. The arrow points to a large mouth cavity called a buccal cavity. Within the cavity is a spur called a 'stationary tooth' that helps rip the surface of prey allowing their contents to be ingested.
© Mario Tenuta, University of Manitoba is licensed under a CC BY (Attribution) license.

The ecological roles of nematodes involve introducing plant fixed carbon to the soil food web, increasing mineralization, particularly that for nitrogen from their food sources, regulating population and community diversity of food webs, and reducing levels of plant pathogenic and plant damaging nematodes through predation.

Nematodes are excellent bioindicators of the effect of management practices on soil health as the presence and population of individual species respond to populations of food sources in the soil food web with some species of nematodes being sensitive to soil management practices (e.g., tillage, fertilizers, pesticides, moisture, rotations, etc.) and stresses (e.g., contaminants such as metals). Many nematodes are parasites of plants. Some nematodes are r-strategists that respond quickly in population to levels of favourable food resources (e.g., bacteria) and soil conditions (e.g., moisture and temperature). Others are K-strategists that thrive in soils with little stress and fluctuating conditions; these are slow-growing, have long life spans for a nematode (one to many years). Then there are nematodes intermediate to r- and K-strategists that are usually most abundant in soil.

Microarthropods

Arthropods have exoskeletons, paired legs, and jointed appendages (e.g., think of insects and spiders). In soils, *microarthropods* are often dominated by two groups: springtails (of the sub-class Collembola, which is often used as a scientific name) and oribatid mites, or Oribadita (Figure $PageIndex9$). Collembola and Oribatida fall in the body-size category of soil "mesofauna" (see Figure $PageIndex3$), ranging from ~0.1 to 2mm in size, and similar to all of the groups of biota described above, are found in soils globally. Typically between 10 and 100 unique species of microarthropods exist throughout the profile of one square meter of soil and on decomposing vegetation aboveground, and microarthropod communities can contain up to 10,000 individuals throughout the profile of a one-m^2 of soil. Soil microarthropods perform important functions through their feeding activities as fungivores (grazing on fungi), herbivores (e.g., grazing on small plant roots), decomposers (grazing on soil organic matter and plant litters), and predators (eating other microarthropod). Similar to protozoa and nematodes, some species of soil microarthropods have highly specialized feeding behaviours (e.g., specialists grazing on very specific types of fungal hyphae), while many of the herbivores and decomposers are generalists. Soil microarthropods can selectively influence the structure of the biological community (e.g., by selectively feeding on certain fungi or other arthropods), distribute microbial cells through the soil profile, and their fecal pellets can promote the formation of aggregates, which together with burrowing activities, improves soil structure (see Chapter 4). Microarthropods can reproduce sexually or asexually, and Collembola are typically thought of as early colonizers of recently disturbed soils (r-strategist) while Oribatida tend to colonize and reproduce slower and maximize numbers in older-growth ecosystems (*K- strategist*).

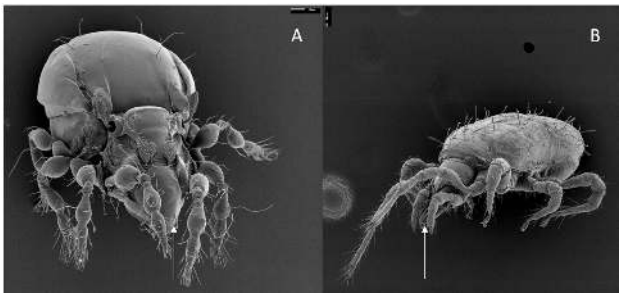


Figure $PageIndex9$: Scanning electron micrographs of a fungal-feeding mite of the genus Quadroppia **(A)** and a predatory mite of the genus Ololaelaps **(B)**. The piercing/chewing structure called a chelicera is shown with arrows for each. For the fungal-feeding mite, the chelicera is very small because fungal hyphae are small themselves. Thus the chelicera is not large enough to be seen here. In contrast, the chelicera of the predatory mite is needs to be larger to pierce the hard bodies of mites it preys on. © Heather Proctor, University of Alberta is licensed under a CC BY (Attribution) license.

Other invertebrates

Soils and associated plant litters are also home to a very large diversity of other larger arthropods (ants, beetles, and others) with many generalist and specialist detritivores, fungivores, herbivores, and even carnivores (e.g., centipedes; scientific name Myriapoda). Annelids (worms; scientific name Oligochaeta) are also important macroscopic invertebrates found in soils, including a variety of earthworms. Earthworms are often revered by farmers and gardeners as being both indicators of healthy soils and contributors to soil health, but these sorts of earthworm paradigms don't tell the full story (Figure $PageIndex10$). On the one hand, earthworms, notably the nightcrawler (*Lumbricus terrestris*) create deep borrows and casts that can improve soil structure, increase pore spaces in compacted soils, and speed up nutrient turnover from plant litters. In intensively managed ecosystems with repeated tillage, harvesting and organic residue management (e.g., crop stover, livestock manure), earthworms arguably do help. However, in unmanaged or less intensively managed ecosystems like long-harvest-rotation temperate and boreal forests, earthworms do more harm than good. There are very few native earthworm species in Canada; the exceptions are on Vancouver Island and in Yukon. The reason is that Canada was nearly entirely glaciated (97%) during the last glacial maximum of the Pleistocene Epoch (ending ~12k years ago), and earthworms have likely only become widespread following the proliferation of resource exploitation by European settlers that involved the movement of soil on tillage equipment and improper management of worms as fishing bait. Particularly in forest ecosystems, earthworm invasion- particularly by *L. terrestris* changes the structure of the soil profile, including largely eliminating organic horizons that are important habitat for plant seedlings, fungal networks, and fauna, including certain birds that nest in the forest floor. Additionally, nutrient retention in mature forest ecosystems is particularly high; long-lived trees producing mainly woody biomass need only small amounts of nutrients like nitrogen and phosphorus to grow each year, and correspondingly, the release of nutrients through decomposition in soils occurs slowly (see Chapter 7). Earthworm invasion leads to very rapid stimulation of microbial decomposer activity and the release of a finite pool of nutrients at a rate too fast for trees to use. It is no surprise that earthworms like *L. terrestris* have earned the title "ecosystem engineer", alongside the likes of beavers, given the severity of the changes they facilitate to soils. For example, Charles Darwin estimated earthworms in a meadowed grass soil in England transported 5 cm of soil from below the surface soil to the surface over ten years!



Figure $PageIndex10$:A sign at Canada's most visited zoo praising earthworms (the burrowing nightcrawler, *Lumbricus terrestris* middle photo) for turning "dirt" into something valuable… hey, should have consulted a soil scientist! Although earthworms can play beneficial roles in heavily-managed agricultural and garden soils, they are non-native in Canada, highly invasive, and have adverse effects in many forest ecosystems. On the right are photos from an experiment showing ingestion leaf and burial activity of a single worm over just a few weeks in an Ontario sugar maple forest floor. Canada's forests and soils are not adapted to such a rapid turnover of organic matter and nutrients. © Nathan Basiliko and Michael McTavish, Laurentian University is licensed under a CC BY (Attribution) license.

Vertebrates

The biomass of vertebrates in soils is quite low compared to smaller biota, however, they can have an important influence in ecosystem functioning by way of altering soil physical and chemical properties (via tunnelling, mixing soils, and their excrement) and by selectively or indiscriminately feeding on other soil biota. A range of mammals (gophers, prairie dogs, groundhogs, etc.), amphibians (e.g., salamanders), and reptiles (skinks, lizards, turtles, etc.) spend at least portions of their lifecycles in burrows or under litter in the surface of soils. Even bird species like the ovenbird spend a portion of their life nesting in the forest soils of Central Canada and the North Central USA. Some vertebrates like the redback salamander are useful indicators of physical (tillage, forestry-related activities) and chemical (e.g., pollution) disturbance because they are reliant on stable habitat and food supply in the form of a robust arthropod community, and because they breathe through their skin.

Viruses

Viruses are smaller than even bacteria (0.05 to 0.2 micrometers), are not organisms and not "alive" in the accepted meaning of the term. However, despite being non-cellular entities, they are composed of biomolecules and play immense roles in disease and genetics in most, if not all, organisms and ecosystems. The simplest viruses are made up solely of genetic material (DNA or RNA) and a protein shell called a capsid. Virus particles are entirely dependent on their hosts- they "hijack" the host's genome with their genetic code and instruct the host's metabolic machinery (i.e., ribosomes for protein synthesis and DNA or RNA synthases for nucleic acid synthesis) to build 10s to 100s of new virus particles. This inevitably leads to cell death, and in the case of single cellular organisms like bacteria or protozoa, kills the host. Viruses that infect microorganisms are also called phages and broadly grouped based on the type of organism they infect (e.g., bacteriophage for bacteria, mycophage for fungi). Some viruses are vectored (transmitted) by soil organisms. For example, some plant-feeding nematodes harbour plant viruses in their esophagus that infect plants with feeding by the nematode. Little is known about the ecological importance of viruses in soil. However, because viruses are often specific to a species of an organism, they may be more diverse than any group of organisms.

Soils can serve as a reservoir for plant and animal viruses that cause disease; however, the role of viruses in ecosystem functioning is much less clear. The number of virus particles per gram of surface horizon soil can be as high as one billion. This is like concentrations seen in surface waters of aquatic ecosystems- and in the past two decades, virus roles in aquatic ecosystem functioning have become well understood. By lysing phytoplankton (cyanobacteria and unicellular algae, the primary producers in aquatic ecosystems) and heterotrophic bacteria, viruses are responsible for shaping biotic communities, and for releasing as much as one-quarter of all carbon fixed into biomass from CO_2 during photosynthesis. Soil viruses may play similar roles, but with much higher biological *richness* (the number of unique species) and much more species *evenness* (i.e., without one or a few dominant taxa) in soils than waters, soil viruses will very likely have their own nuances in how they shape biological communities and ecosystem functioning.

Plant roots

Roots are soil biota, influencing soil development and properties as much as any group of organisms we have explored in this chapter. The biomass of roots is often higher than for all other soil organisms combined in soil that is vegetated. For example, a soil covered in perennial grasses can typically have 10 to 90 t ha^{-1}, compared to fungi (2-5) and bacteria (1 to 2), and up to 2.5, 0.5, 0.2 t ha^{-1} for earthworms, protozoa and nematodes, respectively. The size of roots varies, meaning they occupy a variety of soil pore sizes, handy for obtaining nutrients and water in the soil. Root hairs are about 10 micrometers in diameter, fine roots up to 1 mm, and coarse roots above 1 mm in diameter.

Plant roots are very important in the ecology of soils. They are a major source of organic carbon materials and hence energy for soil heterotrophs and saprotrophs, they lower soil moisture and nutrients levels, consume oxygen through respiration, and alter soil chemical (e.g., soil pH) and physical (e.g., promote aggregation and macropore formation) properties. For more information on soil chemical and physical properties, see Chapter 3, Chapter 4, and Chapter 5.

Distribution of soil biota

Where does the bulk of soil biota reside within soil profiles? The simplest strategy for answering is to “follow the carbon”. Because most of the non-plant soil community are heterotrophic (see Table $PageIndex1$), and reliant on organic matter for both carbon and a source of electrons for generating ATP/energy, numbers of most of the types of soil biota introduced above are highest near the surface of soils and near plant roots in the rhizosphere and Ah horizon of soil (Figure $PageIndex11$). Even lithoautotrophic organisms (Table \(PageIndex{11}\:) tend to cluster around the organic matter inputs and other community members, as they are still dependent on the soil for many of their non-carbon nutrients (e.g., reduced mineral nutrients such as ammonium [NH_4^+] or ammonia [NH_3] for nitrifying bacteria) for growth and metabolism. That said, even in lower abundance, microbes and other heterotrophs alike are still found throughout soil profiles if there are resources available. The importance of microbial biofilms in and around primary soil particles and peds is also worth mentioning. Biofilms form when even distantly related microbes signal to each other to produce sticky extracellular compounds that provide stable habitat (and trophic interactions within the biofilm), and thus allow microbial communities to build their own niches. Although motile soil microorganisms are important- including the protozoans that are defined as being motile, most soil bacteria are attached to soil particles or reside in biofilms.



Figure $PageIndex11$: Fence post showing rot where the wood was in contact with the Ah horizon of soil. The Ah horizon has the most presence of soil organisms and biological activity © Mario Tenuta, University of Manitoba is licensed under a CC BY (Attribution) license.

What about the larger-scale distribution of soil organisms? For example, are the same species and communities found in similar soils on different continents? This is an intriguing question that, in some regards, is not perfectly answered yet. For larger organisms, in the absence of humans facilitating biological invasions by intentionally or unintentionally introducing species, the answer is probably “no”. Think of the nightcrawler example above- there are still “invasion fronts” of these earthworms in forests of Canada and the USA today following their introduction to North America post-colonization by settlers. What about for the very smallest organisms that tend to dominate the non-plant biological activity in soils- the bacteria, archaea, protozoa, and unicellular phases of fungi? Pioneering Dutch environmental microbiologists Martinus Beijerinck in the early 1900s and later Lourens Baas-

Becking are credited with the hypothesis that “Everything is everywhere, but the environment selects”. In essence, there is a global ‘seeding’ of microorganisms given their small size (e.g., could be transported globally on small dust particles in the atmosphere), and they flourish when conditions and resources are right for them, regardless of where those conditions and resources are. Body-size-threshold paradigms have been proposed to explain (lack of) constraints on the natural distribution of biota (including soil biota). However, given that highly-mobile humans have touched nearly every soil environment on Earth, the role of introduced/exotic species is arguably equally as important.

How do we know who lives in the soil?

Studying what we cannot easily see- both because it is buried, but also because it is small- makes fields of soil biology/ecology both challenging and exciting. Analytical artifacts and biases are present in studies of any of the groups of soil biota we introduced above, but on most fronts, soil biologists are rapidly advancing the techniques used to study soil populations, communities, and dynamic interactions. It is surprisingly recently that we have discovered what we don’t know about the scope of soil biology. This perhaps began for soil bacteria with what is called “the great plate count anomaly”. Here, it was noted that the number of cells that could be extracted from soil and viewed/counted using a microscope was orders of magnitude greater than the number of colonies that could be grown on microbiological media in the lab from the same sample (Figure 6.9). The term anomaly was probably too harsh. Extracting and viewing cells from soil is relatively free from selective bias, however, trying to simulate soil conditions and resources in a synthetic medium in the lab is extremely challenging- it is no surprise that only a small fraction of the soil bacterial cells that could be viewed could be grown in the lab. Think back to what you have already learned in Chapter 3, Chapter 4, and Chapter 5 in this book. The chemical, physical, and ecological diversity and complexity in soils are immense at very small scales (i.e., humus molecules with molecular weights in the tens of kg per mole range; a single gram of clay with more surface area than a soccer pitch!). This small-scale complexity has given rise to virtually limitless biological niches that soil microbes and other organisms have evolved to exploit. Just within the bacterial realm a gram of surface soil in a productive ecosystem contains 10s of billions of individual cells from 10s to 100s of thousands of unique taxa. Likewise, the amount of biomass of soil biota is also staggering. A cultivated field can typically have 20 t ha^{-1} of roots, which is about the weight of 25 cows and 15 t ha^{-1} of other soil organisms being the weight of 20 cows! That alone is an overwhelming prospect, and this single section in a single chapter of an introduction to soil science textbook could comprise multiple textbooks on its own. Below we touch on just a few examples of both the challenges and excitement of methods for answering questions on soil biodiversity and ecology.

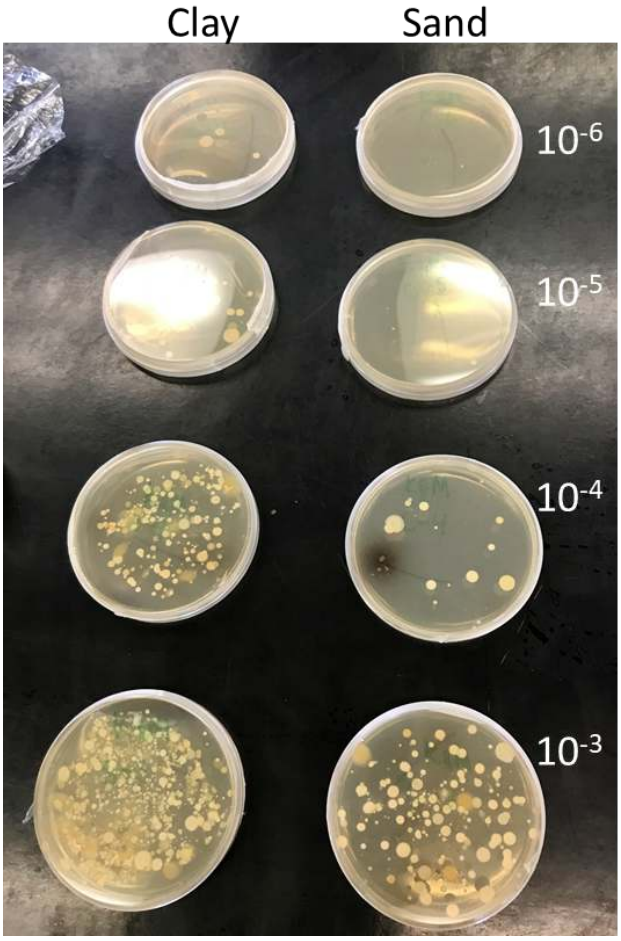


Figure $PageIndex12$: Dilution spread plating of soil suspensions onto agar medium to growth and count population levels of fast-growing aerobic heterotrophic bacteria. Agar plates after six days produced by senior undergraduates for a course are shown. 0.1 mL of 10^{-3} to 10^{-6} dilution suspensions were transferred to plates. Colonies produced are where one bacterial cell from the suspension came to rest and grew. Notice the number of colonies decreases as the dilution increases of the soil suspension. The students were able to estimate the number of bacteria in the soil from the 10^{-4} plate of sand and 10^{-5} plate of clay. They estimated the populations to be 22,000,000 for a gram of the clay soil and 1,800,000 for a gram of the sand soil. This is typical that clay soils have higher microbial populations and activity. Despite that these numbers might appear large, this approach for counting bacteria still drastically underestimates the numbers of bacteria present in soil, because it selects only for the ones that can grow on media in the laboratory. © Mario Tenuta, University of Manitoba is licensed under a CC BY (Attribution) license.

In-situ observations

Depending on the target population or community, different methods exist for studying soil biota in-situ. The soil “contact slide” approach involves inserting a glass slide into soil for a number of days to weeks and then observing attached microorganisms with microscopy. Invertebrates can be observed with pitfall traps, which are buried smooth-surface containers that are open on the top and allow soil fauna to fall in, but not climb out. Often times pitfall traps are partially filled with a preservative like ethanol to kill and prevent decomposition of the trapped fauna. (Find out how to make your own pitfall trap in the You Can Dig It! Activity #1). A particularly interesting and simple technique is commonly used to measure earthworms in soils: mustard powder (the same you would find in the spice section of your local grocery store) is mixed in a weak solution and poured over a defined area of soil. The mustard irritates the worms and they flee their borrows to the surface temporarily, allowing them to be counted, weighed, and identified to species, and there appears to be no long-term negative physiological effects from this contact (see You Can Dig It! Activity #2). Plant roots can be observed and tracked in-situ using clear tubes inserted in soils and a “rhizotron” camera that is inserted periodically in each tube. Image analysis software helps in identifying the same roots over time and quantifying their growth and death. Finally, for groups of soil biota that perform a specific task (often called “guilds”), rates of activities can be

measured by quantifying changes in either reactants or products over time. For example, in upland soils, aerobic methane-oxidizing bacteria consume the trace atmospheric gas methane, an important global greenhouse gas even though its concentration in air is only approximately 0.00002%. Placing a chamber over a defined area of soil and measuring the decrease in methane trapped in the chamber over time gives a reasonable estimate of how active the methane-oxidizing bacteria are in the soil under field conditions.

Isolation approaches

A common approach in soil biology has been to isolate species from their soil environment and community so they can be studied directly (Figure 6.10). For microorganisms, the goal is often to obtain a pure culture (i.e., a population growing on an artificial medium in the lab) where physiological traits can be measured, including what resources are used, what products produced, and how changing conditions like temperature influence growth rates and survival. Much has been learned by cultivating microbes, but the approach is highly selective, and selecting one species from a community of thousands to millions is challenging. In the past few decades, biomarker fingerprinting approaches described below for characterizing soil microbes- particularly those using DNA sequences to identify specific taxa- have helped guide isolation and cultivation approaches towards environmentally relevant organisms and ignore “weeds” that inherently grow well under simple laboratory conditions.



Figure $PageIndex13$: Isolation approaches for studying soil biota. **(Left)** A streak plate approach where microbial cells extracted from soil have been dragged across a 10 cm diameter petri dish with a solid medium in decreasing concentration. Note the isolated colonies that represent single populations. Source: US Centers for Disease Control and Prevention. © J. Gathany. (https://upload.wikimedia.org/wikiped...a_Plate_01.png). Like many mesofauna, nematodes are recovered from soil water suspensions **(Right)** using Baermann funnels filled with water and with soil submerged and wrapped in tissue. Nematodes move out of soil into water and fall by gravity to the end of the rubber tube. © M. Tenuta. For an excellent set of videos with Professor Terry Tollefson from the University of Saskatchewan carrying out soil faunal extractions and visualization, visit https://www.youtube.com/watch?v=TCenRcKbf7U and https://www.youtube.com/watch?v=VuHznslr8aI © US Centers for Disease Control and Prevention (CDC)/ James Gathany; Mario Tenuta, University of Manitoba is licensed under a CC BY (Attribution) license.

Isolation of soil invertebrates differs from the isolation of microorganisms in that individuals are large enough that they can be observed and characterized on their own with microscopy (i.e., growing populations on artificial media is not needed to study them). Nematodes can be extracted with a Baermann Funnel system, where the soil is placed on a cloth mesh in the funnel and saturated. Nematodes are washed downward and through the cloth, while soil particles remain. A common approach for isolating collembolans and mites (groups of soil “mesofauna” see Figure $PageIndex3$:) is the use of a Tulgren (sometimes also called Berlesse) Funnel extraction system. Here a light above a funnel filled with soil is a repellent for small soil fauna and they move downward until they drop through the stem of the funnel and into a container with a liquid preservative like ethanol below. Although individual specimens can be obtained with these approaches, taxonomic identification is still a major challenge and requires a great deal of time and expertise.

Biomarker approaches

Soil biota contains different unique biomolecules that can be extracted and characterized chemically to help study community composition, the relative size of a particular population or functional guild of organisms (e.g., nitrogen-fixing bacteria, see Chapter 7), or metabolic functioning of a soil community. These different biomolecules include cell membrane lipids that vary across broad groups of biota (for example, Archaea have a unique ether-lipid structure of their cytoplasmic membranes), sugar polymers that makeup cell walls (e.g., cellulose in plants and algae versus a compound called chitin in fungi and some invertebrates or a compound called peptidoglycan in bacteria), and very importantly nucleic acids (DNA and RNA). DNA can be extracted relatively easily from soil biota by pulverizing a soil matrix and binding the negatively charged nucleic acid mixture to a silica powder with

high cation exchange capacity where it can be washed and then re-dissolved in pure water. Often taxonomic “marker genes” are targeted for sequencing and quantification (Figure $PageIndex14$: and Figure $PageIndex15$). These are genes or other regions of an organism’s genome that are present in many or all organisms. For example, genes involved in producing a portion of the ribosome (protein synthesis factory present in every known organism) are commonly used to show similarity or dissimilarity in bacteria and archaea (16S rRNA gene), and fungi (18S rRNA gene or Internal Transcribed Spacer (ITS)), while a gene encoding for an oxidase enzyme involved in cellular respiration (Cytochrome Oxidase 1 (CO1) gene) is often used to show the relatedness of animal taxa. It is assumed that a more similar gene sequence between organisms means that they are evolutionarily more closely related. Techniques including the polymerase chain reaction (PCR) and high throughput DNA sequencing have become very accessible in research laboratories, and community fingerprints based on the DNA sequences of millions of individuals in a soil sample (i.e., how many of which individual taxa are in the soil) can be generated relatively cheaply and quickly (Figure $PageIndex14$: and Figure $PageIndex15$). There are other nucleic acid based approaches, for example using quantitative PCR (qPCR) for estimating the number of copies of a particular gene in an original soil DNA extraction- and here genes could be associated with very specific functional guilds, for example like encoding for a portion of the nitrogenase enzyme complex involved in nitrogen fixation. Gene transcripts (i.e., messenger RNA or mRNA) can also be quantified with a variation of qPCR to characterize gene expression related to particular biochemical function. “Shotgun” sequencing of DNA and RNA extracted from soils is also a very powerful tool for characterizing the community and biochemical potential and functioning of soil communities. Referred to as metagenomic or environmental genomic sequencing, random fragments of nucleic acids are sequenced and analyzed, giving a picture of the overall community with less selective bias than PCR-based sequencing approaches.

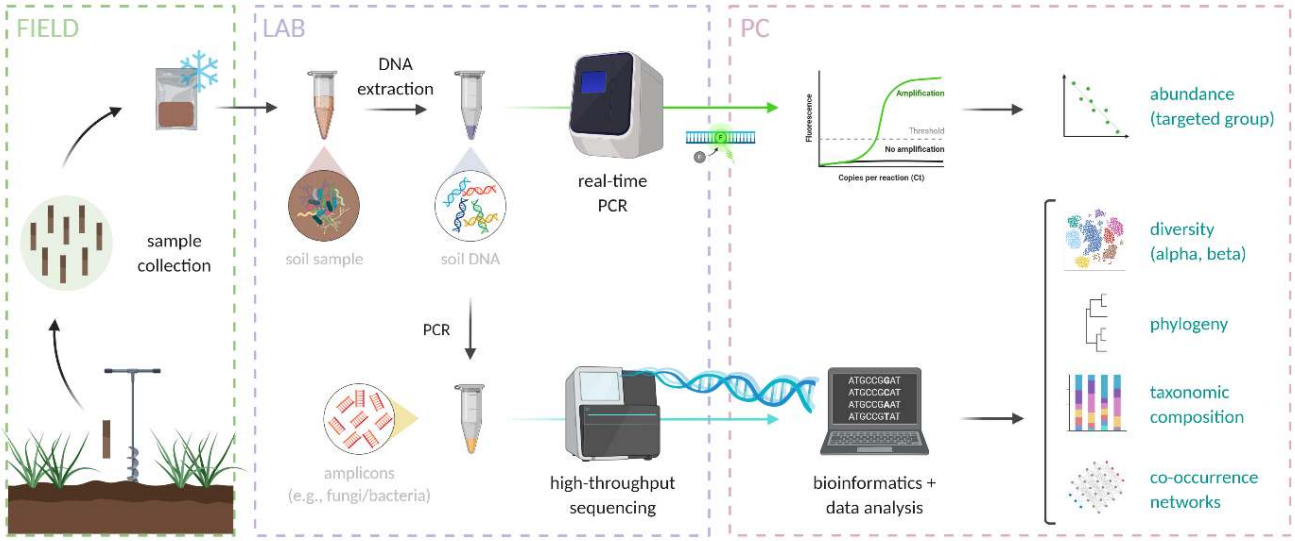


Figure $PageIndex14$: Schematic representation of steps involved in studying microbial communities in soil using DNA approaches. Representative soil samples are taken from the field and immediately stored on ice. Once taken back to the lab, DNA is extracted and stored at -80°C for further analysis. Quantitative PCR can determine the abundance of a target gene, providing an estimate of gene copies per gram of soil. High-thoughput sequencing of PCR amplicons of taxonomic biomarker genes (16S, ITS, CO1) can be combined with bioinformatics approaches to provide a picture of microbial diversity, phylogeny, taxonomic composition and even co-occurrence. Up to date protocols for soil DNA extraction, 16S and ITS sequencing on the Illumina platform, and initial bioinformatics can be found on the Earth microbiome project (EMP) https://earthmicrobiome.org/. © Kari Dunfield, Univ. of Guelph (created using Biorender) is licensed under a CC BY (Attribution) license.



(Figure $PageIndex{15}$: DNA and RNA sequence fingerprinting is common for characterizing soil community diversity and for identifying individual soil specimens or isolates. Gel electrophoresis micrographs **(Left)** showing polymerase chain reaction (PCR) to identify plant-parasitic nematodes. A. Amplification by PCR of nematode DNA shows a positive identification for the nematode, *Ditylenchus weischeri*, in lanes 1, 2, and 3. DNA of nematodes in lanes 4 through 8 do not have the gel band and are thus other species. B. DNA from nematodes in lane 5, 6, and 7 amplified in a different PCR reaction indicating those to be *Ditylenchus dipsaci*. © M. Madani. CC BY. Benchtop high-throughput DNA sequencing system **(Right)** capable of sequencing 1M bp of DNA at the cost of ca. $100 in the year 2020. For comparison, the sequencing cost of the first human genome less than 20 years earlier was over $100,000,000. © Konrad Förstner. (https://upload.wikimedia.org/wikiped..._sequencer.jpg) CCO 1.0 Universal. Images by Mehrdad Madani (University of Manitoba) and Konrad Förstner are licensed under a CC0 (Creative Commons Zero) license.

Although technologies for less selective nucleic acid extraction and quantification and inexpensive sequencing are ever-improving, a major challenge in the use of nucleic-acid sequence-based analyses of soil organisms and communities lies in the fact that very few taxa of soil biota globally have been described in detail (e.g., through isolation and cultivation approaches). For example, there are only approximately 10k identified unique strains of bacteria in culture collections globally, while up to potentially billions of unique bacteria across global soils. Fingerprinting approaches have taught us much about the diversity and ecology of soil biota over the past few decades, but at least at the time this chapter is being written, they have also shown us that there are many "known unknowns" waiting to be discovered!

SOIL ECOLOGY

What do soil biota do?

Ecology describes the interactions within a biological community and the dynamic interplay with the abiotic environment. Soil biota influence the soil environment in many ways that are covered throughout this book. For example, in Chapter 3 you encountered the still somewhat mysterious production and breakdown of soil organic matter that intimately involves soil communities, in Chapter 4 the role of biota in the formation of soil structure was highlighted, and next in Chapter 7 you will learn how key nutrient elements are transformed in soils, in large part a result of biological processes. Here in the present chapter, we take the approach of illustrating soil biological activities from the perspective of the organisms and communities. For example, the soil processes of nitrification and denitrification—where nitrogen is converted from one mineral form to another—are exceptionally important in terms of nutrient availability and loss from soils, and have regional and global importance in terms of how they affect water quality and the composition of the atmosphere. But if we step back and evaluate the directive of the microorganisms involved in these processes, they are- much more earnestly- just harnessing redox gradients in the environment to generate energy, grow, reproduce and successfully stay dormant in between periods of resource availability and proper environmental conditions. Next, we briefly present a number of important soil and related ecosystem functions carried out by soil communities in terms of the biochemical goals of the underlying guilds of organisms driving the processes that have broader ecosystem consequences. This complements the ecosystem-scale perspectives covered in other chapters.

Decomposition, cycling carbon, and methane

With the exception of plant roots, soil biota are predominantly heterotrophic and rely on fixed organic matter as the source of C for building biomass (anabolism) and as a source of energy (see Table 6.1). The initial input of food (i.e., organic matter) in soils is predominantly from plant litters and root turnover, however, organic amendments like manure can also be important large sources in crop and pasture systems. As you learned in Chapter 3, the quality and quantity of different plant litter influence rates of decomposition and formation of soil humus, and an important fraction of soil organic matter appears to be microbial by-products made from processing organic matter, rather than partially broken-down plant structures. In well-aerated soils, simple, and nutrient-rich compounds like sugars, starches, nucleic acids, and polypeptides can easily be mineralized to CO_2 for energy generation and

incorporated in biomass. More complex polymers like cellulose and chitin and aromatic polymers like lignin take longer to be converted back to mineral forms. Carbon dioxide is an important atmospheric greenhouse gas, and the role of soils in the global climate system and as mediators of global climate change is determined in large part by the rate of net carbon inputs to plant biomass minus the rate of release of CO_2 by heterotrophic respiration as plants subsequently die and decompose.

What happens to heterotrophs when soils become saturated and no longer contain O_2 rich air in pore spaces? In many soils, this happens periodically after a rain or snowmelt event; however, wetland soils are defined by prolonged saturation. Decomposition rates are much slower without O_2, and this is one of the factors that contribute to the formation of organic soils (see Organic Soil order in Chapter 8); essentially, rates of plant growth exceed rates of heterotrophic mineralization over long periods of time. In bogs and fens (i.e., peatlands, the predominant ecosystem types with organic soils), soil profiles consisting of partially decomposed wetland plants can extend many meters deep (Figure 6.13)! Without oxygen, heterotrophic soil biota must rely on other biochemical pathways besides aerobic respiration to generate energy. Obligate aerobes only carry out metabolism when O_2 is available, while obligate anaerobes cannot function in the presence of O_2. The term facultative anaerobe (or facultative aerobe) refers to an organism that can generate energy and grow under O_2 and O_2-free conditions. Fermentative organisms carry out glycolysis- generating 1 or 2 ATP molecules per glucose molecule consumed- and then fermentation reactions to regenerate the molecules needed for glycolysis to continue. Fermentations generate reduced compounds like alcohols, organic acids, and H_2 gas as waste products, and typically yield 18 to 36 times less ATP than glucose can provide an aerobe. Anaerobic respiration pathways are used by certain anaerobic soil bacteria and archaea to process plant litters and soil organic matter. In these reactions, compounds other than O_2 are used in cellular respiration as the terminal electron acceptors. Denitrification and sulfate reduction are important examples of anaerobic respiration; denitrification in soils is typically carried out by facultative anaerobes and can yield up to 1/2 of the ATP per unit substrate as aerobic respiration, while sulfate reduction is carried out by strict anaerobes and yields much less energy. Methanogenesis is an important terminal anaerobic decomposition step in wetland soils (See Figure $PageIndex16$:). A group of strictly anaerobic archaea utilize the waste products from fermentative soil bacteria (notably acetic acid and H_2) and produce methane (CH_4). Although energy yields from methanogenesis are relatively small, methane is a globally important greenhouse gas that is much more potent than CO_2 per molecule, and changing methane emissions from wetland soils represent an important soil role in biosphere feedbacks to global climate change. Groups of bacteria called methane-oxidizing bacteria or methanotrophs can consume methane and use it both as a source of carbon and energy. Indeed, methane is a good source of electrons for ATP generation (similarly, it is a good source of heat in natural gas!), but it has high activation energy, and most methane oxidation in soils occurs under aerobic conditions, with O_2 used as an electron acceptor. In wetland soils, methanotrophs reside in surface soil horizons where they can obtain methane produced under anaerobic conditions from below and O_2 from the atmosphere above; they can be an important “filter” that keeps a portion of methane from reaching the atmosphere and contributing to the global greenhouse effect. Interestingly in upland soils, particularly those not disturbed by agriculture, methanotrophs consume atmospheric methane, representing an important global sink for this greenhouse gas, consuming about 5-10% of all methane emitted annually.



Figure $PageIndex16$: A poor fen wetland ecosystem northwest of Sudbury, Ontario. The inset photo on the left shows bubbles of methane-rich gas (noted with a white arrow) escaping from the surface around a metal corer that was produced by anaerobic soil microorganisms. The inset photo on the right is core sample of the upper 50 cm of organic 'peat' soil with clearly intact plant parent material. The stainless steel "Russian" corer that was used to extract the core is also in the photo. The depth of the organic soil/peat deposit in this site is 6 m. © Nathan Basiliko, Laurentian University is licensed under a CC BY (Attribution) license.

Nutrient cycling

The same types of activities by heterotrophic soil biota that break down organic matter and mineralize carbon as CO_2 are also responsible for the release of important mineral forms of nutrients like nitrogen (as ammonium) and phosphorus (as phosphate). Uptake of nutrients for anabolism by non-plant soil biota is often termed immobilization, as the resources are at least temporarily unavailable to support plant growth, whereas plant root uptake of nutrients is termed assimilation. Certain groups of prokaryotes are also responsible for transforming mineral forms of nutrients as well. This includes the very specialized, energy-consuming process of nitrogen fixation where N_2 in the atmosphere is chemically reduced and converted into available N build into amino acids (see below and Chapter 7) and the energy-producing processes of oxidizing reduced mineral forms or reducing oxidized mineral forms of nutrients. In the case of nitrogen, this involves the chemolithotrophic (Table 6.1) process of nitrification where ammonium or ammonia is used as a source of electrons that are typically transferred to O_2 for ATP generation (producing nitrite and nitrate), and the anaerobic heterotrophic process of denitrification where nitrate is used as an electron acceptor during organic matter decomposition. These processes influence the availability of different forms of mineral N explored more in Chapter 7, and ammonium oxidation (the first step in nitrification) releases H^+ ions and lowers soil pH, a consideration for farmers when spreading ammonium-based fertilizer salts. Both processes of nitrification and denitrification are multi-step and mediated by multiple enzymes. The potent greenhouse gas nitrous oxide (N_2O) is an intermediate product and reactant in both. Although this means that soils can be either sinks or sources for atmospheric nitrous oxide, denitrification-related losses of nitrous oxide from fertilized agricultural soils represents one of the most important sources of this gas that contributes to both climate change and degradation of the stratospheric (upper atmosphere) ozone layer. Similar to nitrogen, sulphur and iron in reduced or oxidized forms can be both sources of electrons for chemolithotrophic microorganisms or electron acceptors for anaerobic heterotrophic respiration.

Phosphorus is essential for growth, because it is found in many critical biomolecules including nucleic acids, phospholipids, and ATP, and is often considered the second most limiting nutrient for productivity after N. Low soil P can decrease plant yields, while excess soil P that enters freshwater bodies due to surface runoff can cause a bloom in algal populations called eutrophication. The P cycle is unique because phosphorus does not exist in a gaseous form. Inorganic P present in the soil solution as orthophosphate ions ($H_2PO_4^-$ and HPO_4^{2-}) is the only available form for uptake, but is usually found in very low concentrations (around 1-10 µM), because it is very reactive, and binds to cations and clays. P is primarily recalcitrant in soils in inorganic (sorbed P, primary and secondary minerals) and organic (organic matter, microbial biomass, living biomass) forms. Microorganisms can aid in plant P uptake through several mechanisms:

1. iincreasing root interactions with orthophosphates by extending existing root systems through symbiotic associations with mycorrhizal fungi or stimulating root growth through the release of hormones (phytostimulation) by plant growth promoting

rhizobacteria (PGPR);
2. competing with plants for available P in the soil solution, immobilizing it into microbial biomass, and then regulating supply of P through microbial biomass turnover, and;
3. solubilizing inorganic P through a range of metabolic processes conducted by a wide range of bacteria and fungi. These mechanisms include proton pumps, or organic acids, which solubilize precipitated (solid) forms of P (e.g., Ca-phosphates), siderophores that chelate metal ions associated with complexed forms of soil P, and microbial enzymes such as phosphatases that can hydrolyse organic P or cellulolytic enzymes that mineralize organic matter (see Chapter 7 for an explanation of "P fixation" v available forms of soil P).

Microorganisms with these capabilities can be used to increase the availability of soil P to plants. A fungus, *Penicillum bilaiae* was first isolated by researchers at an Agriculture Canada Research Station in Lethbridge Alberta (Kucey, 1983) and was commercialized as the first phosphate inoculant in Canada, JumpStart™, registered under the Canadian Fertilizers Act, for use on wheat in 1990, and is now registered for most crops grown in Canada.

Pollutant transformation

Soil biota, particularly some bacteria and archaea, play important role in the cycling of metal and organic pollutants. Similarly to iron, certain metals like arsenic can serve as both electron donors and electron acceptors for certain prokaryotic microorganisms depending on the conditions and resources in soils (e.g., concentration of oxygen and available organic matter), and the bioavailability and toxicity of arsenic is altered as its oxidation state is changed through microbial redox processes. Unlike with iron, however, arsenic has no structural role in biomolecules and indeed causes oxidative stress to cells and is severely toxic to certain organ systems in complex organisms. Cadmium is an element that is often present in trace amounts in mined agricultural fertilizers and can accumulate over repeated fertilization to levels of concern for food and feed supply. Unlike with iron and arsenic, prokaryotic microorganisms in soils don't commonly harness reduced or oxidized forms for redox and ATP generation. The redox states of other metals like copper and nickel can be altered by certain soil prokaryotes for both energy generation and for anabolic use or detoxification. Copper, for example, is needed in small amounts for important microbial enzymes related to hydrolysis of organic matter and oxidation of methane by methanotrophic bacteria, while a trace amount of nickel is required for methanogenic archaea to produce methane. Both nickel and copper can be quite toxic in even moderate concentrations though-think that copper sulfate is the key additive in antimicrobial "pressure-treated" lumber, for example. For metals that are trace nutrients, much remains to be learned about how biota walk the line between utilization of and protection from the same elements.

In wetland soils, certain anaerobic bacteria and some archaea influence to the toxicity of mercury (Hg). In its elemental form, mercury is semi-volatile and can be transported globally in warm air masses. Primarily from coal burning, inputs of mercury deposition have more than doubled in many soils. Luckily elemental mercury (like you would find in an old thermometer) is not very bioavailable to organisms in soils, however in wetland soils where there is abundant organic matter and sulfate (e.g., like the peatland in Figure 6.13), anaerobic sulfate-reducing bacteria transform inorganic mercury into a bioavailable organic form called methylmercury. Methylmercury travels from soil water to surface waters where it bioaccumulates in organisms and biomagnifies in aquatic food chains.

Metal pollutant elements like arsenic, cadmium, and mercury cannot be fundamentally created or destroyed by soil biota, however, their toxicity and bioavailability can be changed by soil bacteria by altering redox states, or incorporation into a simple organic molecule like methylmercury. Organic pollutants that are typically synthesized for industrial processes or formed as inadvertent by-products during high-temperature industrial processes, can be fundamentally destroyed. Human-synthesized organic compounds like classes of triazine herbicides introduced in the 1960s and widely used in agriculture today initially were not biologically altered by soil bacteria. However, by the mid-1990s heterotrophic bacteria were isolated that had the ability to grow on the herbicides and break them down to a non-toxic form! Rapid evolution, including through horizontal gene transfer in bacteria, plays an important role in the ability to catalyze the breakdown of human-synthesized pollutant compounds. Soil bacteria certainly can't process all of our organic pollution. In the case of crude oil and other hydrocarbon spills in soil, bacterial enzymes exist to catalyze breakdown; however, they tend not to be very active because hydrocarbon fuel pollutants tend to have very little nitrogen and phosphorus in them that is also needed for bacterial growth.

Plant-microbe interactions

There are many important interactions between plants and microorganisms in soils. The term symbiosis refers to different organisms living together, and although it is sometimes used to imply a mutualistic relationship, symbioses can also involve commensal interactions where only one of the organisms benefits and the other is generally unharmed, and **parasitic** relationships where one organism is harmed and one benefits (e.g., see on pathogens and parasites below). The rhizosphere is an important zone were plants and microbes interact. Heterotrophic decomposers gain energy and carbon from plant root turnover and root exudates, and in turn, mineralize organic forms of nitrogen and phosphorus into inorganic plant-available forms. This is a broad case of generally mutualistic symbiosis, however, when decomposing low-nitrogen-content plant litters like woody tissues, the relationship can shift and microbes can actually immobilize soil nutrients for a period of time to grow on the high carbon, low nitrogen material, leading to plant nutrient deficiencies.

Other plant-microbe relationships are much more direct. Symbiotic nitrogen fixation involves plants "farming" nitrogen-fixing bacteria in root nodules- the bacteria are fed carbon compounds pulled from the plant's Krebs cycle, and in turn, the bacteria use a portion of the energy to convert atmospheric N_2 into a chemically-reduced bioavailable form for the plant that is pushed back into the Krebs cycle and subsequently amino acid synthesis, allowing the plant to produce high-N containing compounds like chlorophyll and rubisco (a key enzyme in photosynthesis). Only certain plants and microbes can form these mutualistic symbioses. Legumes pair with a group of bacteria called rhizobia, while members of some shrubs like alder and sweet gale pair with members of the bacterial phylum Actinobacteria. Nitrogen fixation is exceptionally costly- with up to 16 ATP consumed per single molecule of atmospheric N_2 reduced, and thus symbiotic N fixation is a good adaptation- as plants can readily produce energy-rich compounds through photosynthesis, whereas free-living heterotrophic soil N fixing bacteria cannot. The rates of N fixation between free-living and symbiotic root-associated bacteria are astoundingly different, with input rates of available N to soil jumping from 1 kg N ha^{-1} $year^{-1}$ up to 200 kg!

Another important plant root-microbe symbiosis that was introduced above is mycorrhizae (see Figure 6.3). Many plant species allow colonization of roots by mycorrhizal fungi and supply the fungus with sugars, while the fungus helps the plant acquire nutrients like phosphorus, wards off plant parasites, and improves the soil structure in the rhizosphere. The symbiotic relationship in mycorrhizae is usually mutualistic, and recent research has shown that both plants and fungi have important mechanisms to choose partners who are going to facilitate the other in the relationship. In this regard, a fungus cannot utilize carbon-rich food provided by the plant without also doing its part and helping the plant acquire important limiting nutrients from soil like phosphorus.

Trophic interactions

In soil, the transfer of organic energy and nutrients in the form of molecules in the body of organisms determines the kinds and population levels of organisms in those communities. The concept of the soil food web is used to describe the community of organisms in soil for all or part of their lives. Food webs can depict the transfer of energy or nutrients between organisms within a community, but usually, it focuses on the flow of organic energy. These flows can be extremely complex, governed by the feeding needs of organisms. The steps in energy transfer (feeding of one organism on another) are referred to as trophic interactions.

The basis of most soil food webs is the production of organic energy by autotrophic organisms (largely photosynthesizing plants and referred to as primary producers). This energy is introduced to the soil through the dropping of dead above-ground parts (leaves, stems, trunks, crop residues) and roots. Organic energy and nutrients in dead plant parts and roots are used or released by decay organisms that are heterotrophic and called decomposers. Many species of bacteria and fungi are decomposers. From here on, the trophic interactions are done by heterotrophic organisms collectively called consumers.

Have you ever wondered how large dead plant parts disappear in time in soil? The feeding action of microarthropods that chew and shred those parts into smaller pieces is key. They are called detritivores and their main feeding interest is consuming the decomposers on dead plant parts. Their activity produces smaller dead plant parts, which in turn contains a greater surface area for continued colonization and decay by decomposers.

Living roots are also fed upon directly by a wide range of organisms including many bacteria and fungi, root-feeding nematodes, and microarthropods collectively referred to as herbivores and sometimes primary consumers. Herbivores can be detrimental or beneficial to the health of plants. Some bacteria and fungi, called plant-pathogens, target roots to feed by either killing root parts

before feeding or feed on living cells, slowly draining roots of energy and leading to the eventual death of roots (Figure $PageIndex17$:).



Figure $PageIndex17$: Top: morphology of the soil fungal pathogen of potato, Verticillium dahliae, isolated from Manitoba fields. **(A)** Colony of the fungus after placement of a single spore onto potato dextrose agar 18 days earlier. **(B)** and **(C)** Conidiophore and whorl phialide bearing dispersal cells called conidia. **(D)** and **(E)** Clusters of fungal cells with darkened pigmentation for survival in harsh conditions of soil called Microsclerotia. **(F)** Close up of individual Conidia. © Oscar Molina; licensed under a CC BY (Attribution) license. **(Bottom)** A survival structure also called conidia, of the fungus that cause Early Blight of Tomato and Potato, *Alternaria solani* extracted from a potato field soil in Ontario. Conidia of this fungus has 6-10 cells. They are also pigmented to help survive in soil. © M. Tenuta; licensed under a CC BY (Attribution) license.

Soil fauna that kill parts of roots or compromise the health of plants are referred to as plant-pests (Figure $PageIndex18$). Many root-feeding nematodes and microarthropods are plant-pests. Beneficial herbivores obtain energy from roots and promote the health of those plants and thus called plant growth promoting rhizo-organisms (PGPR). They live in the rhizosphere (in the soil around or on and in roots) to obtain energy from plants. Examples of beneficial herbivores include biological nitrogen-fixing bacteria in the roots of legumes (e.g., bacteria of the order Rhizobiales in which the genera *Bradyrhizobium* forms nodules on soybean roots) and plants of the alder family (ex., bacteria in the genus *Frankia*: Figure 6.16). Examples of beneficial fungal herbivores are ecto- and arbuscular mycorrhizal fungi (Figure $PageIndex5$:).



Figure $PageIndex18$: Many nematodes are of great economic importance. Here roots of potato growing in the Andes Mountains in Peru are parasitized by females of the potato cyst nematode. The females are evident as small tiny white spots on roots. Because of the devastation this nematode causes in reduced potato yield and marketability (tuber surface is warted, blemished and spotted), soil and plants infested with it are legally prevented from entering many countries such as Canada. © Mario Tenuta (University of Manitoba) is licensed under a CC BY (Attribution) license.



Figure $PageIndex19$: Nodule of the nitrogen-fixing bacterium *Frankia alni* attached to the roots of alder (Alnus sp.). Each nodule can contain billions of symbiotic bacteria that benefit from a stable environment and organic matter to consume from the plant; in turn the bacteria fix large amounts of otherwise unavailable atmospheric N_2 for the host plant. © W. Cranshaw. (https://commons.wikimedia.org/wiki/File:Frankia_alni_at_Alnus_sp._(02).jpg); licensed under a CC BY (Attribution) license.

Numerous organisms feed on decomposers, plant-pathogens, and beneficial bacteria and fungi and are called microbial grazers or sometimes secondary consumers. They are considered carnivores as they do not feed directly on living or dead plants. Grazers often specialize to prey on bacteria or fungi, but rarely both. Feeding on bacterial grazers involves taking up the whole organism by the predator. Examples of grazers include amoeba, ciliate and flagellate protozoa, rotifers, bacterial feeding nematodes, and Collembola. Examples of fungal grazers include tardigrades (Figure $PageIndex20$:) and fungal-feeding nematodes, mites, and Collembola.

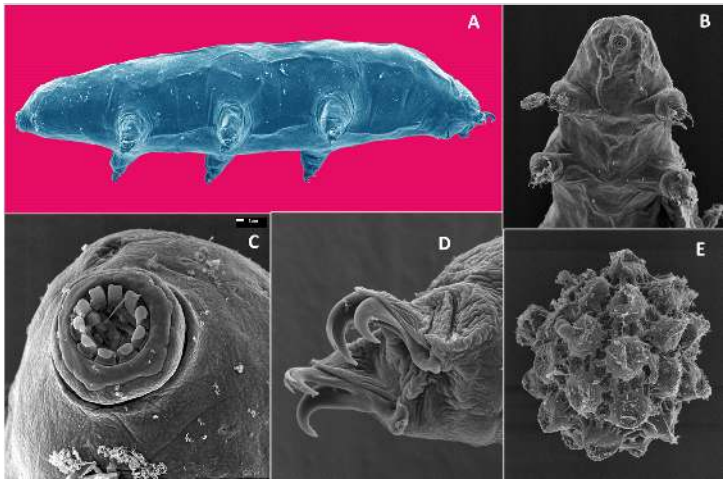


Figure $PageIndex20$: Scanning electron micrographs of tardigrades or water bears. A. Lateral (side) view. B. Ventral (tummy) view with head at top. C. Anterior (head) close-up showing the mouth where the stylet at centre would protrude from to feed. D. Tardigrades have claws at the end of their legs that are helpful to move in soil and perhaps latch onto prey. E. Tardigrade egg that has a heavy surface armour to survive many years in soil. Tardigrades are renown as 'extreme organisms', capable of surviving harsh conditions. Tardigrades that were dried and encased in amber crash landed on the moon in an ill-fated moon mission in 2019. Scientists believe the tardigrades will revive if placed in suitable conditions. © Matthew Boeckner, University of Alberta is licensed under a CC BY (Attribution) license.

Tertiary consumers feed upon microbial grazers and plant-pests. They can be fully carnivorous or omnivorous (feeding on combinations of fungi, soil animals and living roots and other plant parts. Examples of tertiary consumers include omnivorous nematodes, entomopathogenic nematodes (nematodes that kill and feed on microarthropods), nematode-trapping fungi (fungi that immobilize and feed on nematodes), and predatory nematodes and mites.

The trophic interactions of soil food webs can be described as diagrams showing trophic organisms or groups and the direction of energy flow. Soil food webs are analyzed by carefully sampling and processing soil and using methods for recovering and enumerating organisms present. The feeding preference of recovered organisms is either well documented as for those of economic importance such as pathogens and pests, or apparent—based on feeding structures of the organism. For example, predatory mites have a more massive piercing structure (chelicera) than fungal-feeding mites because of differences in prey size.

An example of a soil food web from the prairies of northeastern Colorado is shown in Figure $PageIndex21$: Organisms of similar trophic levels (preferred feed source) are given in boxes with the direction of food web energy flow in trophic interactions (food source/prey to consumer/predator) shown with arrows. The general trophic levels of primary producers, herbivores, decomposers, microbial grazers, and tertiary consumers are given. Note that the apex predator (top carnivore) in this food web are predaceous mites.

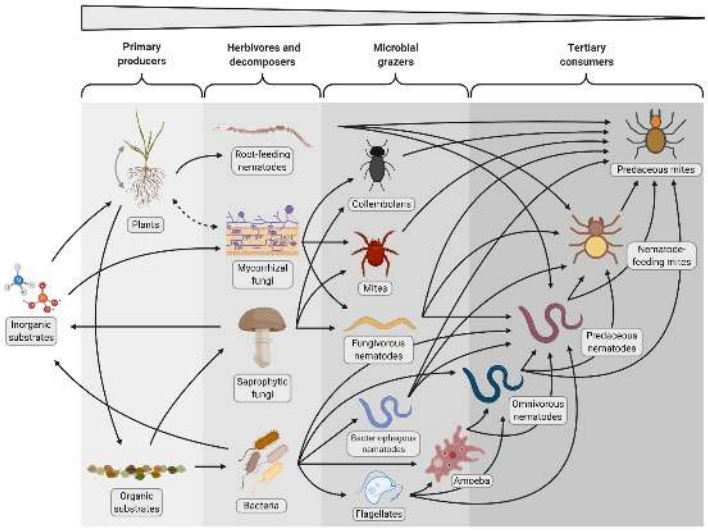


Figure $PageIndex21$: Representation of detrital food web in shortgrass prairie. Organic substrates consist of both labile and resistant substrates. Flows omitted from the figure for the sake of clarity include transfers from every organism to the substrate pools (death) and transfers from every animal to the substrate pools (defecation) and to inorganic N (ammonification). Triangle represents the ecological pyramid principle, demonstrating energy transfer up trophic levels. Note – this site was too dry for earthworms, therefore they are not represented. Adapted from Hunt et al. (1986). Adapted by Kari Dunfield, Univ. of Guelph, using Biorender (Biorender.com), and licensed under a CC BY (Attribution) license.

Primary producers and consumers in a food web have important roles in the soil, such as to balance populations of beneficial and detrimental organisms to plants. Food webs that are out of balance are prone not to promote the growth of plants. Balance in soil food web structure is controlled by bottom-up and top-down processes. Bottom-up processes involve the level of a resource determining the population of consumers in the soil food web. For example, where the levels of microbial and tertiary consumers are related to the density of roots in the soil, bottom-up control is evident (i.e., level of energy input via roots determines populations in higher trophic interactions). Top-down processes are where the levels of the predator determine that of the prey. In agriculture and forestry, top-down control of pest organisms is applied in biological control practices. Here the health of a plant affected by an herbivore is improved by the addition of predator of the later. Examples of biological control agents include entomopathogenic nematodes that kill lawn grubs or predatory mites to control populations of gnats and thrips.

Soil food webs conform to the ecological pyramid principle. Soil food webs usually have much higher amounts of energy in the biomass of primary producers (i.e. the plants), and the amount of energy/biomass decreases sequentially in decomposers, microbial grazers and finally tertiary consumers, whose overall biomass is smallest in a given area of soil. This observation is strong evidence that bottom-up control of soil food webs is important. Figure 6.18 shows that for a typical grassland plant, roots have the largest biomass of all soil organisms, followed by fungi and bacteria (presumably largely decomposers) then microbial grazers. When present in soils, earthworms, feeding on anything in ingested soil, likely derive most of their energy from detritus, bacteria and fungi.

The same reasons shaping the ecological pyramid above-ground is in effect in soil. Each trophic interaction (feeding activity) results in loss of organic energy as heat. This means there is less energy transferred up the food web. Typically, we don't think of organisms in soil as producing heat as mammals do. But if you have ever composted, you know temperatures can get as high as 65°C at the centre of large piles. Compost is an excellent insulator for heat, thus retaining heat produced by decomposers (Figure $PageIndex22$:). Other reasons for an ecological pyramid are predators expend energy seeking prey and predators can live longer than prey.



Figure $PageIndex22$: Decomposers during the active phase of composting (early on in composting when degradation is most rapid) generate great amounts of heat. Here the heat generated is evident in a compost pile of cattle manure where the air temperature on this day was -25°C but the internal temperature of the compost was 44°C (112°F). © Megan Westphal, University of Manitoba is licensed under a CC BY (Attribution) license.

SUMMARY

1. The very large diversity of chemical, physical, and ecological niches in soils have selected for diverse biota spanning from micrometer-sized bacteria to vertebrate fauna; soil biota dynamically responds to and shapes their soil environments.
2. Small size, high numbers, and lacking databases and expertise in taxonomy represent important challenges to our understanding of the scope of diversity in soils- there are many new discoveries to come!
3. Different metabolic strategies are used by soil organisms in terms of how they harness energy and nutrients from the environment; however, heterotrophic activity predominates.
4. These metabolic processes represent biotic directives of generating ATP, surviving, and reproducing, but from the soil ecosystem perspective, the associated resource exploitation essential drives processes such as organic matter and nutrient cycling (influencing soil fertility), among many others.
5. Soil food webs depict trophic interactions between organisms and the flow of organic energy through biota; trophic levels in soil are primary producers, herbivores, decomposers, microbial grazers, and tertiary consumers; and bottom-up control where the amount of organic energy input to soil from plants generally controls the structure (biomass) of soil trophic levels.

SUGGESTED READING

Coleman, D., Callaham, M. and Crossley, D. 2017. Fundamentals of Soil Ecology. 3rd Ed. Academic Press, London, UK.

De Wit, R. and Bouvier, T. 2006. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ. Microbiol. 8: 755-758.

Hunt, H.W., Coleman, D.C., Ingham, E.R., Ingham, R.E., Elliott, E.T., Moore, J.C., Rose, S.L., Reid, C.P.P., and Morley, C.R. 1987. The detrital food web in a shortgrass prairie. Biol. Fertil. Soils 3:57–68.

Kingston, W. 2004. Streptomycin, Schatz v. Waksman, and the Balance of Credit for Discovery. Journal of the History of Medicine and Allied Sciences 59:441-462.

Kucey, R. M. N. 1983. Phosphate-solubilizing bacteria and fungi in various cultivated and virgin Alberta soils. Can. J. Soil Sci. 63:671_678

Madsen, E.L. 2015. Environmental Microbiology: From Genomes to Biogeochemistry. 2nd Edition. Wiley-Blackwell, Hobokon, NJ.

Nair, P. 2012. Woese and Fox: Life, rearranged. P.N.A.S 109: 1019–1021.

Paul, E.A. 2015. Soil Microbiology, Ecology and Biochemistry, 4th Edition. Academic Press, Boston, MA.

Richardson, AE., and Simpson, R.,J. 2011. Soil microorganisms mediating phosphorus availability update on microbial phosphorus. Plant Physiol. 156: 989-996. DOI: 10.1104/pp.111.175448

Schmitt, C.L., and Tatum, M.L. 2008. The Malheur National Forest:Location of the World's Largest Living Organism [The Humongous Fungus]. United States Department of Agriculture- Forest Service. Available from http://www.fs.usda.gov/Internet/FSE_...ev3_033146.pdf [accessed 26 October 2020].

Simard, S., Perry, D., Jones, M., Myrold, D.D., Durall, D.M. and Molina, R. 1997. Net transfer of carbon between ectomycorrhizal tree species in the field. Nature 388:579–582

Swift, M.J., Health, O.W., Anderson, J.M. 1979. Decomposition in Terrestrial Ecosystems. University of California Press, Berkley, CA.

STUDY QUESTIONS

1) Why is the diversity of soil bacteria so high? What roles do soil properties you learned about earlier in the text and trophic interactions among soil biota play in this high level of diversity?

2) What are biomarker and isolation based approaches to studying soil biological communities? Why are biomarker approaches particularly useful for characterizing the diversity of small-sized biota in soils?

3) Give one example each and describe the nature of the relationship in a mutualistic and a parasitic symbiotic relationship between soil organisms.

4) Where is biomass typically largest in the soil profile and why? How does this relate to the dominant metabolic strategy (e.g., heterotrophy, lithotrophy, autotrophy) in belowground ecosystems?

5) Imagine a sterile soil entirely devoid of biota. Could this soil support plant group? Why or why not?

6) Describe differences in soil biota in a flooded (anoxic) wetland soil relative to a well-drained upland soil.

7) How does soil biota influence the cycling of carbon in both soils and the atmosphere?

8) Draw a hypothetical soil food web.

Can You Dig It!

Activities

Activity 1: Make a Pitfall trap to catch Soil Invertebrates and Insects

Adapted from: https://www.nhm.ac.uk/discover/how-to-make-pitfall-trap-to-catch-insects.html

You will need:

- a hand trowel
- a small plastic container or plastic cup
- a tray
- ID guides (optional)

1. Choose a location for your trap on flat ground near vegetation.
2. Use a trowel to dig a small hole.
3. Place a clean plastic container in the hole. Fill in any empty space around the container with soil. Make sure that the top of the pot is level with the ground, or you won't catch anything.
4. Leave your trap overnight.
5. Empty the trap into a tray to see what creatures wandered in. Use ID guides or apps such as iNaturalist supported by the Canadian Wildlife Federation and the Royal Ontario Museum Biodiversity Program https://inaturalist.ca/) to help you identify what kind of invertebrates they are.
6. Record your findings: take pictures, make a note of what you caught, the date and location.
7. Carefully release the creatures, returning them to a safe, sheltered place.
8. Return the area back to how you found it.

Adapted from: Christina Curell, Michigan State University Extension – June 29, 2016

https://www.canr.msu.edu/news/earthworms_can_be_an_indicator_of_soil_health

Materials that you will need to measure worms in the field:

- Tape measure
- 2 liters of tap water
- Hand trowel or shovel
- Container to collect worms
- Solution of 2 tablespoons of mustard powder dissolved in 2 liters of water

Once you have your materials gathered you are ready to count earthworms.

Step 1: Measure a square foot in the test area and dig down 12-inches.

Step 2: Collect and count the number of worms found. If possible, differentiate worms by type. For example, label as earthworms, red worms, etc.

Step 3: (Optional) Level out the bottom of the hole, and pour the mustard solution slowly. Deep burrowing worms should come to the surface within 5 minutes. Collect and count the worms that come to the surface.

Step 4: Count and record the total number of earthworms that are collected.

About the authors:

Nathan Basiliko, Professor and Canada Research Chair in Environmental Microbiology, Department of Biology, Laurentian University



Nathan Basiliko (licensed under a CC-BY-NC-ND license)

My research group focuses nearly entirely on soil biology and the roles in biogeochemical cycles. However, because my academic “home” for the past 7 years at Laurentian University has been in our Biology department, I don’t teach soil science courses anymore (I used to). I miss teaching introductory soil science in particular, but my current position does give me a new opportunity to squarely build soil science in to other courses- including introductory and advanced microbiology and ecosystem ecology courses. I hope that this has some broader impact by engaging students with soil science who otherwise would not take soil science courses.

Kari Dunfield, Professor in Applied Soil Ecology and Canada Research Chair in Environmental Microbiology of Agro-Ecosystems, School of Environmental Sciences, University of Guelph



Kari Dunfield (licensed under a CC-BY-NC-ND license)

Kari is a Professor in Applied Soil Ecology and Canada Research Chair in Environmental Microbiology of Agro-Ecosystems in the School of Environmental Sciences at the University of Guelph (Guelph, Ontario). Her research group works at the intersection of microbiology, ecology, and soil science, and takes an inter-disciplinary approach to find solutions to complex global environmental challenges, including, recycling nutrients and minimizing greenhouse gas emissions in agriculture systems; breaking down pollutants in brownfield sites; and understanding pathogen transport from agricultural systems into waterways. I love studying soil microbiology because it is a combination of field and lab based research. In my group you need to own a pair of steel toed boots and a lab coat, because in the morning you could be going out to the field, using soil probes and shovels to get soil and root samples, and in the afternoon you could be back in the lab doing delicate and sterile molecular biology techniques.

Mario Tenuta, Professor of Applied Soil Ecology and Senior Industrial Research Chair in 4R Nutrient Stewardship, Department of Soil Science, University of Manitoba



Mario Tenuta (licensed under a CC-BY-NC-ND license)

As a Soil Ecologist, the work is very diverse and full of new challenges and learning opportunities. My research involves tackling the issue of lowering the amount of greenhouse gases emitted from agricultural soils. For this we do field research to understand how microbes are involved in nutrient cycling, how soil conditions affect microbial activities and functions and how farmer practices of fertilizer and manure use affect microbes. There is lots of opportunity for laboratory analysis of microbes, nutrients in soil, nutrients in plants, and of crop productivity. We take that understanding to recommend practices farmers can use to lower emissions of important greenhouse gases such as nitrous oxide (N_2O). Soil is everywhere and the work also provides lots of opportunity to travel locally and the world doing research, teaching and development

The goal of inviting this team of authors was to ensure that a range of ecosystems (including forest, prairie, wetland, agricultural, permafrost-affected) and groups of soil biota was covered within their combined expertise. All three of the co-authors are active researchers in the field of soil biology and have each held or currently hold Canada Research Chair positions at their universities.

This page titled 4.11: Soil Biodiversity and Ecology is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Canadian Society of Soil Science via source content that was edited to the style and standards of the LibreTexts platform.

4.12: Soil Nutrient Cycling

Joann Whalen, Noura Ziadi, Jeff Schoenau, Maxime C. Paré, David Burton, and Tom Bruulsema

Learning Objectives

On completion of this chapter, students will be able to:

1. Know the forms of soil nutrients that can be acquired by plants
2. Know the factors controlling nutrient transformations into plant-available forms
3. Understand how to evaluate soil fertility with soil testing and plant analysis
4. Understand that inorganic and organic fertilizers can be applied to replenish the soil nutrient supply
5. Be familiar with the 4R Nutrient Stewardship approach to soil nutrient management
6. Be aware of the environmental implications of soil nutrient management
7. Be acquainted with some examples of soil fertility improvement in Canada

INTRODUCTION

Soils support plant growth. Seeds germinate in the warm, moist soil environment. The seedling emerges at the soil surface, which needs to be porous and uncompacted, with no surface crust. Elongation of the radical and development of the root system occurs easily in soil with a crumbly, loose structure. The root system has two vital functions. The first is anchorage. Many plants grow a thick, strong primary root that is approximately as tall as the stem. A plant that is firmly rooted in the soil is unlikely to be dislodged by normal rainfall events and windstorms. The second vital function is resource acquisition, since most of the water and many essential elements for plant growth are absorbed through the root hairs (Table 7.1).

Table 7.1. Essential plant elements, and the forms of nutritive cations, anions and molecules absorbed by plant roots (adapted from Havlin et al., 2014)

Essential element	Concentration	Nutritive form(s) absorbed		
	(μg g^{-1} dry mass)	Cation	Anion	Molecule
C	450000			CO_2 [1]
H	450000			H_2O [2]
O	60000			CO_2, H_2O
N	15000	NH_4^+	NO_3^-	N_2 [3]
K	10000	K^+		
Ca	5000	Ca^{2+}		
P	2000		HPO_4^{2-}, $H_2PO_4^-$	
S	2000		SO_4^{2-}	SO_2 [4]
Mg	1000	Mg^{2+}		
Fe	100	Fe^{2+}, Fe^{3+}		Soluble chelate [5]
Cl	100		Cl^-	
Mn	50	Mn^{2+}		Soluble chelate
B	20		$H_2BO_3^-$	H_3BO_3, soluble chelate
Zn	20	Zn^{2+}		Soluble chelate
Cu	6	Cu^{2+}		Soluble chelate
Mo	0.1		MoO_4^-	Soluble chelate
Ni	0.01	Ni^{2+}		Soluble chelate

[1] CO_2 gas is absorbed through leaves and converted to carbohydrates $(CH_2O)_n$ via photosynthesis.
[2] H_2O, as water vapour or liquid water, is absorbed through foliage and roots.
[3] Symbiotic and associative N_2-fixing prokaryotes that grow with the plant transform N_2 gas into NH_4^+ via biological nitrogen fixation, and then transported into plant cells.
[4] Less than 10% of the plant's sulphur requirement comes from SO_2 gas absorbed through leaves.
[5] Soluble chelates are readily absorbed by the root.

Soil fertility refers to the ability of soil to supply nutrient ions for plant growth. Soils have inherent fertility, which depends upon four factors:

1. parent material, which is weathered to release nutrients;
2. vegetation, which produces organic residues that are decomposed to release nutrients;
3. organisms that participate in weathering and are responsible for decomposition and nutrient cycling; and
4. climate, which determines the rates of biological, chemical and physical processes that contribute to or deplete the soil nutrient supply.

Inherent soil fertility is a good predictor of what plant species will grow in a given area under natural conditions, without human intervention. When we harvest agricultural crops to feed people and livestock, we remove some nutrients from the soil-plant system. Consequently, agroecosystems are usually fertilized to replenish or increase the soil nutrient supply. Selecting the right fertilizer and other soil amendments requires careful decision, since it is expensive to purchase and apply these materials to

agricultural land, and because overuse can have negative environmental consequences. Farms across Canada can follow nutrient management plans to guide the application of fertilizing materials to agricultural soils, for the best agronomic, economic and environmental outcomes.

NUTRIENTS IN THE SOIL-PLANT SYSTEM

Fine roots and root hairs grow in soil pores, which contain air and water. Soil pore water is a dilute electrolyte solution containing water-soluble nutrient ions. To satisfy their water requirements, plants absorb water from soil via osmosis and rely on aquaporins to selectively conduct water molecules in and out of cells, which regulates the cell volume and internal osmotic pressure. Water-soluble ions in the soil pore water enter the root hairs and diffuse through the epidermis. However, the root surface must maintain electrical neutrality. Any ion that enters the root causes the release of a counter-ion such as H^+, OH^- or HCO_3^- (Figure 7.1). For example, when a positively charged ion such as potassium (K^+, a cation) passes through the root epidermis, the root releases one H^+ ion into the surrounding soil pore water, and absorption of a calcium ion (Ca^{2+}) results in the release of 2 H^+ into the soil pore water. Anion absorption is similarly balanced, such that uptake of nitrate (NO_3^-) causes the root to release an OH^- ion or a HCO_3^- ion into the soil pore water.

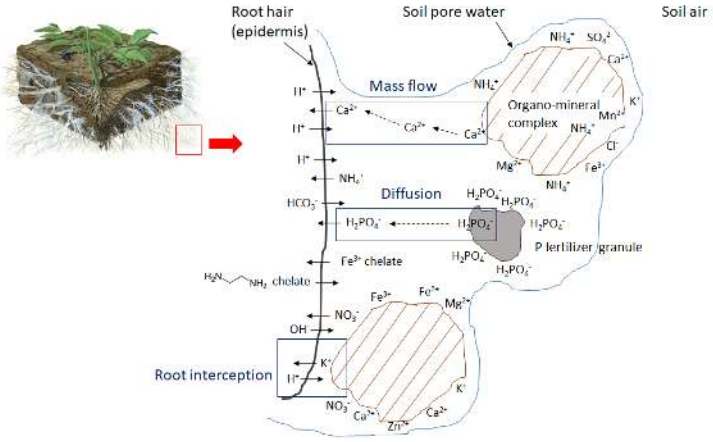


Figure 7.1. Nutrient ions move through the soil pore water and are absorbed by root hairs resulting in the release of counter-ions (H^+, OH^-, HCO_3^-) to maintain electrical neutrality at the root surface. Ions that travel along with water are absorbed by mass flow. Diffusion involves ion movement across a diffusion gradient, from high concentration to low concentration, near the root. Roots can intercept ions retained in organo-mineral complexes. Roots release chelates that bind to metals ions, increasing metal mobility and uptake by the root. © Joann Whalen is licensed under a CC BY (Attribution) license.

Once inside the root, water-soluble ions can move through the apoplast of the cortex, a transport pathway that transports water and dissolved substances along cell walls and extracellular space outside of plasma membranes. There are two other pathways for water-soluble ion movement: the symplastic and transmembrane pathways. In the symplastic pathway, water and solutes move along the cytosol and cross the plasma membranes of adjacent cells via the plasmodesmata. Alternatively, in the transmembrane route, the dissolved minerals and water move from cell to cell by crossing the cell wall, exiting one cell before they enter the next cell. The suberized layer surrounding the endodermis acts as a physical barrier, and ions must be transported through the plasma membrane via protein channels by binding with transporter proteins. Some of the transporter proteins responsible for nutrient ion flux in plants are listed in Table 7.2. Next, nutrient ions enters the xylem and are transported along with water to the stem, leaves, flowers, seeds and other components. Nutrient ions may be incorporated into structural compounds within these plant organs, or support metabolic processes in the cytoplasm. Furthermore, nutrient ions may be remobilized and translocated to other tissues or organs if they are needed in another part of the plant.

Table 7.2. Transporter proteins mediate the flux of nutrient ions across the plasma membranes of vascular plants. Adapted from Reid and Hayes (2003)

Nutrient ion	Transporter protein	Comments
NH_4^+	High affinity AMT transporter proteins	
NO_3^-	Nitrate transporter protein (NRT2) family	
$H_2PO_4^-$ and HPO_4^{2-}	PHT1 transporter protein	PHT1;1 through PHT1;9 were identified in *Arabidopsis thaliana*
K^+	K^+ transporter protein (Trk) family	
SO_4^{2-}	SO_4^{2-} transporter proteins	Plants also possess the cysteine (Cys) transporter, methionine (Met) transporter, and glucosinolate (GSL) transporter
Ca^{2+}	Ca^{2+} transporters, Ca^{2+}-ATPases and H^+/Ca^{2+} antiporters	Transporter proteins that move Ca^{2+} across the membrane, against the electrochemical gradient, are specifically referred to as an ion pump.

Plant uptake depletes the nutrient ion concentration in soil. Ions may move together with water through mass flow or be intercepted by growing roots that make contact with exchangeable ions on organo-mineral surfaces, a process referred to as root interception (Figure 7.1). Ions may also move across diffusion gradients from a zone of high concentration to a zone of low concentration in pore water close to the root (Figure 7.1). The concentration of nutrient ions in soil pore water is also affected by multiple, concurrent biological, chemical and physical processes, illustrated in Figure 7.2.

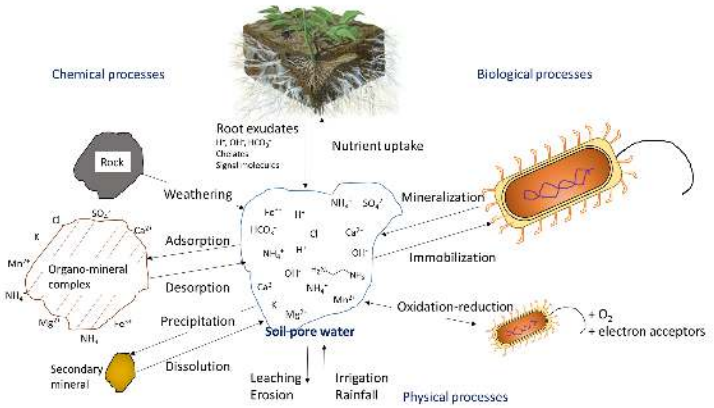


Figure 7.2. Biological, chemical and physical processes that affect the concentration of nutrient ions in the soil pore water. © Joann Whalen is licensed under a CC BY (Attribution) license.

Biological activity is responsible for the mineralization of organic compounds, which releases water-soluble nutrient ions. Mineralization requires a specific hydrolytic enzyme, produced by plant roots or soil microorganisms, to cleave the covalent bonds in an organic compound. Soil enzymes function in the presence of water, often outside of biological cells. Mineralization of soil organic matter and organic residues releases ammonium (NH_4^+), phosphate (HPO_4^{2-}, $H_2PO_4^-$) and sulfate (SO_4^{2-}) into the soil pore water. Immobilization refers to absorption of nutrient ions by organisms other than plants. Ions are immobilized in soil microbial biomass for a limited period of time (days to weeks) while the microorganism grows and reproduces, and are released at the end of their lifespan. The mineralization-immobilization-turnover concept applies to any nutrient that is covalently bound in organic matter, including carbon, nitrogen, phosphorus and sulphur. An illustration of the concept (Figure 7.3) shows how microorganisms produce enzymes to hydrolyze proteins, releasing NH_4^+ in the soil pore water, which is then immobilized inside growing microbial cells, and finally re-released into the soil pore water upon lysis of those cells.

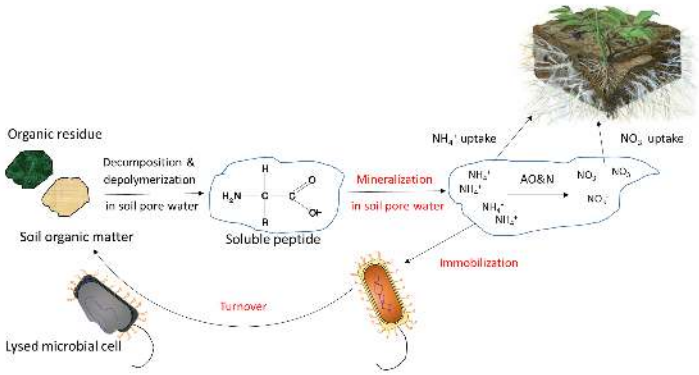


Figure 7.3. The mineralization-immobilization-turnover process. Organic residues, soil organic matter and lysed microbial cells undergo decomposition, which involves physical fragmentation and solubilisation before organic compounds are depolymerized by extracellular enzymes of microbial origin. Monomers such as the illustrated peptide then undergo further enzymatic degradation in the soil pore water, a process referred to as mineralization. The ammonium (NH_4^+) released into soil pore water may be taken up by plants, immobilized in microbial cells, or undergo microbial-mediated transformations called ammonia oxidation and nitrification (AO&N) that produce nitrate (NO_3^-). Similarly, NO_3^- is available for uptake by plants and microorganisms. The NH_4^+ and NO_3^- used by microorganisms for protein synthesis and other metabolic functions will be released when microbial cells are lysed, at the end of their lifespan. © Joann Whalen is licensed under a CC BY (Attribution) license.

The oxygen concentration in soil pore water affects the transformation of nutrient ions. Oxygen moves approximately 1000 times faster through air-filled pores than through soil water, so it is not uncommon for oxygen limitation to occur in soil pore water. Limited oxygen availability can stimulate denitrification, a biological reaction in which the nitrogen electron acceptors gain electrons and decrease their oxidation state, e.g.,

$$NO_3^-{}_{(aq)} \rightarrow NO_2^-{}_{(aq)} \rightarrow NO_{x(g)} \rightarrow N_2O_{(g)} \rightarrow N_{2(g)}$$

which also changes the chemical form of nitrogen from water-soluble NO_3^- to dinitrogen (N_2) gas. Chemical reduction of elements such as Mn, Fe, and sulfate in anoxic (low-oxygen) soil is another process that alters the chemical speciation and quantities of these nutrient ions in soil pore water.

Chemical reactions that add nutrient ions to soil pore water include weathering, which occurs when the broken, exposed surfaces of mineral surfaces are exposed to acidic compounds, such as carbonic acid. Dissolution refers to the solubilisation of nutrient ions from clays and colloidal materials in the presence of water. In dry soils, dissolution is less common because a low volume of soil pore water will increase the proximity of nutrient ions with reactive ions and molecules that can form precipitates. Surface complexation reactions on organo-mineral surfaces can either increase the concentration of nutrient ions, through desorption reactions, or remove ions from the soil solution through adsorption reactions.

The concentration of nutrient ions is highly dependent on the water content in soil pores. Irrigation and rainfall add water to the soil pores, thereby diluting the water-soluble nutrient ions. In humid regions that have more precipitation than evapotranspiration, soil pores may periodically reach 100% water-filled pore space. This will limit oxygen diffusion, which affects oxidation-reduction reactions, and impedes the growth of most plants, except those with morphological adaptations that allow them to tolerate inundation. Agricultural activities generally occur on land that has good natural drainage or where drainage systems are installed to encourage water movement below the cultivable topsoil. Drainage removes water, but it also leaches water-soluble nutrient ions through the profile, along with the water. Removal of excess water from topsoil via drainage or through evaporation will increase the relative concentration of water-soluble ions remaining in soil pore water.

SOIL FERTILITY AND PLANT NUTRITION

Plants acquire most of their essential nutrients ions from the soil pore water, a dynamic environment. In agriculture, much effort is made to understand the soil nutrient supply. Does a particular soil have enough nutrients to support the plant needs, or do we need to add fertilizer containing nutrients to avoid a plant nutrient deficiency?

Soil fertility testing is an accurate way to assess the nutrient supply in agricultural soil. The main purpose of soil fertility testing is to determine whether a soil contains sufficient nutrients to meet the needs of an annual crop during its lifespan (one growing season). In addition, soil fertility testing can determine whether there is an adequate nutrient supply for perennial crops during their

establishment and production phases, which span multiple growing seasons. Soil fertility testing relies on chemical and biological analyses by certified laboratories. The results of these tests are a good indicator of the plant-available nutrient supply. This concept relies on an understanding of two facts. First, the plant will acquire nutrient ions from the soil pore water throughout the growing season according to its nutrient demands. Second, the soil pore water will be replenished with water-soluble nutrient ions at a rate that reflects the combined biological, chemical and physical processes that contribute and remove ions from the soil pore water, during the growing season. For this reason, soil testing methods such as the Mehlich 3 extraction remove nutrient ions from soil pore water and other sources (Figure 7.4), a good approximation of the nutrient supply that will become available to plants during the growing season.

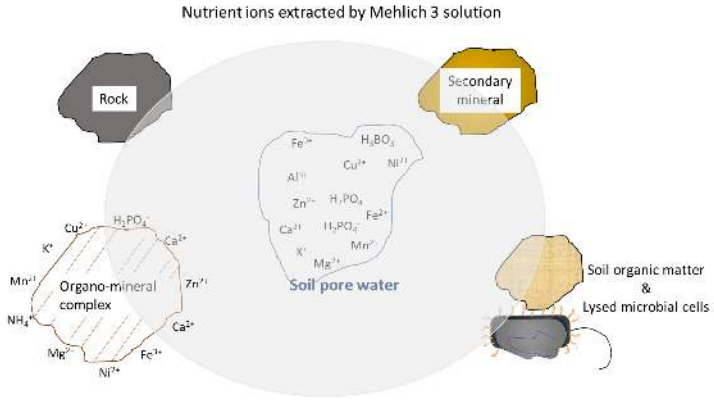


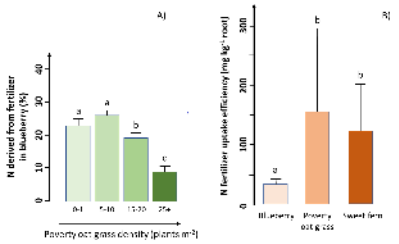
Figure 7.4. The Mehlich 3 extractant was developed by Dr. Adolph Mehlich to estimate the plant availability of many macronutrients and micronutrients in soil. In Eastern Canada, the fertilizer recommendations are based on the nutrient concentrations in Mehlich 3 extracts. The chemical solution extracts nutrients from the soil pore water, as well as some of those that are contained in rocks, in precipitated minerals such as iron phosphates and aluminum phosphates, adsorbed on cation and anion exchange sites, and contained in microbial biomass and organic matter. The chemical composition of the Mehlich 3 extractant is 0.2 N CH_3COOH + 0.25 N NH_4NO_3 + 0.013N HNO_3 + 0.015N NH_4F + 0.001 M EDTA (Mehlich, 1984). © Joann Whalen is licensed under a CC BY (Attribution) license.

Another purpose of soil fertility testing is to assess the probability of a profitable response to fertilizers and soil amendments. This evaluation is part of an integrated plan to improve the growing conditions for crops. Farms set yield targets for their crops, according to the production goals and economic returns they hope to achieve. The maximum yield that can be expected for a particular crop depends on its genetic and physiological characteristics, which govern the plant's photosynthetic capacity, its ability to acquire water and nutrients from soil, and its ability to convert these resources into grain, oilseed, starchy tubers, fruit or other marketable products. Furthermore, the maximum yield is influenced by the plant's response to environmental conditions during the growing season. Unfavorable weather can provide too much or too little rainfall for the crop. Heat stress resulting from too hot temperatures during summer months, or frost damage in the spring and fall months, are stressful for crops. Farms also follow good agronomic practices to control growth limiting factors. For instance, the number of plants occupying a given area is optimized to intercept all available solar radiation, but too-close spacing could cause competition among plants and should be avoided. Weeds, pests and disease-causing organisms must be controlled to prevent yield losses. Finally, soils need to provide adequate nutrients, at the right time during crop development, to achieve the expected yield and quality outcomes.

Can You Dig It!

When Blueberry Gets Wild with Nitrogen Fertilizer

The availability of nitrogen (N) in the soil solution limits plant growth and crop yields in most agroecosystems. That is why farmers often fertilize their fields with one or several forms of N. Nevertheless, the efficacy of N fertilization depends on several factors including the presence of weeds, which can compete with the crop for the acquisition of fertilizer-derived N. Inappropriate N fertilization can indeed be counterproductive as it can stimulate weed growth rather than that of the targeted crop.



N fertilizer uptake and uptake efficiency. © Marty et al. 2019 is licensed under a CC BY (Attribution) license.

Researchers conducted a study in a lowbush blueberry (Vaccinium angustifolium Aiton) commercial field of the Lac Saint-Jean region, Québec, Canada, to assess the efficacy of N fertilization in the presence of two common weed species (poverty oat grass and sweet fern). The researchers found a much higher aboveground biomass production and a larger amount of N in blueberry than in both weed species, indicating that blueberry captured most of the added N fertilizer. However, blueberry growth and fertilizer uptake decreased with increasing weed density (Figure A). When poverty oat grass density was more than 15 plants m^{-2}, this weed captured about the same amount of fertilizer as blueberry despite a much smaller belowground biomass, resulting in a significant negative impact on blueberry production. In addition, the researchers found that fertilizer uptake efficiency (the fertilizer uptake per gram of root) was about four times lower in blueberry than in both weed species (Figure B). ***In other words, compared to the weeds, blueberry requires four times more belowground biomass to acquire the same amount of N from fertilizer.***

The complete story is provided in Marty et al. (2019)

Since soil fertility testing quantifies the plant-available nutrient concentrations, it can be used to interpret the crop response to nutrient inputs. Soil fertility testing should be accompanied by measurements of the crop response to nutrient inputs in controlled environments such as growth chambers and greenhouses, and in the field. In soils with low to medium nutrient supply, crop growth will probably be improved by adding fertilizer. On the other hand, crops planted in soils with high to excessively high soil fertility rarely benefit from fertilizer application (Figure 7.5). Since fertilizers cost money to purchase, transport and apply to fields, it is not profitable to apply fertilizer when no yield improvement is expected, since this will not give an economic return on the investment.

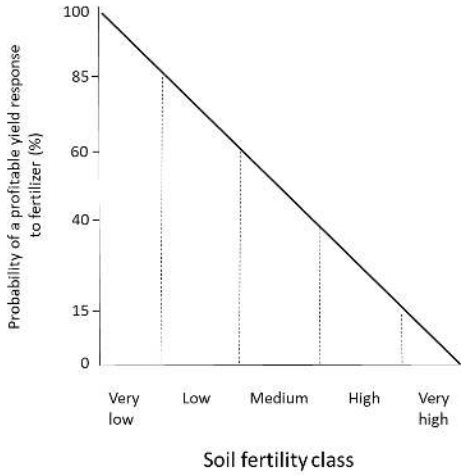


Figure 7.5. The probability of a profitable yield response from fertilizer declines as the soil fertility level goes from very low to very high. © Joann Whalen is licensed under a CC BY (Attribution) license.

Plant nutrition is determined by measuring the nutrient content of plants, either in a specific tissue or in the whole plant. When plants receive ample nutrients from the soil and fertilizer, if needed, their nutrient concentration will be in the sufficient range. When the soil nutrient supply is insufficient and no fertilizer is applied, the plant will be deficient in that nutrient. Plants that accumulate more nutrients than they need for their metabolic processes could have an excessively high or toxic level of nutrients (Figure 6). Nutrient toxicity is uncommon for macronutrients such as N, P, K, Ca, Mg and S because the active transport processes that transfers water-soluble nutrient ions through the endodermis and into the xylem are tightly regulated at a genetic level. Toxicity is sometimes observed when micronutrients such as B, Fe, Mn, Zn and Cu are applied at too-high rates to the foliage because these elements can damage the sensitive, photosynthetic cells in leaves. However, these micronutrients can be absorbed safely through

the root system because plants have defense mechanisms that control the uptake, accumulation and translocation of micronutrients and other metals. This involves retaining the ion in root cells, where they are detoxified through complexation with amino acids, organic acids or metal-binding peptides and/or sequestered into vacuoles. These mechanisms will prevent micronutrient translocation to the sensitive leaf tissues.

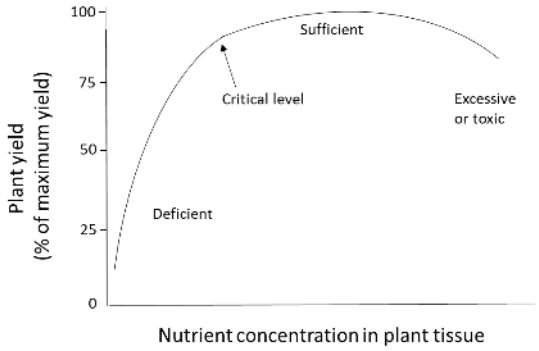


Figure 7.6. Plant yield is related to the concentration of essential macronutrients and micronutrients in plant tissue. As the nutrient concentration in plant tissue increases towards the critical level, plant yield increases. Above the critical level, the plant contains sufficient nutrients for normal growth and can continue to absorb nutrients without a yield increase. An excessive nutrient concentration in plant tissue can be toxic to the plant, resulting in yield loss or plant death. © Joann Whalen is licensed under a CC BY (Attribution) license.

Nutrient requirements are known for all agricultural crops. Crops that have less than the critical nutrient level for development will not grow normally, will be stunted, and they may display deficiency symptoms such as chlorosis (yellowing of the leaves). This may be a sign that the plant lacked nutrients needed to produce chlorophyll, since a well-nourished plant will produce dark green leaves for photosynthesis. Even if there are no visible symptoms of a nutrient deficiency, inadequate plant nutrition can produce other undesirable outcomes:

- weak stems that are susceptible to lodging during windstorms, hailstorms and intense rainstorms;
- few flowers or inflorescences;
- seed abortion;
- poor-quality seeds (low protein content, lacking essential amino acids, low lipid content); and
- loss of marketable yield due to little seed/grain production, small fruits and tubers.

The best way to determine if crop growth is limited by lack of nutrients is to sample and analyze plant tissue. This involves taking a sample of the plant at an appropriate growth stage, generally during vegetative growth or in the early reproductive growth stage. The plant sample is then sent to a certified laboratory for analysis, and the results can be interpreted by checking reference tables that present the sufficient level of the macro- and micro-nutrients in a particular crop (Table 7.3). Since every crop has unique nutrient requirements at each of its growth stages, the reference table has to be specific for the crop species of interest. Furthermore, plant tissue analysis is done early in the growing season because the results may indicate that fertilizers should be applied to correct or prevent a nutrient deficiency. Timely analysis of plant tissue can guide the farmer to make an in-season nutrient application that will prevent yield losses. However, the diagnosis from a plant tissue test might arrive too late to take corrective action. For example, the crop may be too large for fertilizer applicators to enter the field. Another possibility is that the deficient nutrient is not mobile in the plant, meaning it is too late in the plant's development to overcome the growth limitation associated with this nutrient deficiency. Therefore, soil fertility testing is the most important diagnostic tool to avoid a negative economic outcome that occurs when crops have inadequate nutrition during the growing season.

Table 7.3. Sufficiency ranges for macronutrients and micronutrients in corn plants, in relationship to corn growth stages (based on Jones, 2003)

Element	V4-V6 stage (30 cm tall)	V10 stage (prior to tasselling)	R1 stage (initial silk)
Nitrogen (%)	3.50-5.00	3.00-5.00	2.70-4.00
Phosphorus (%)	0.30-0.50	0.25-0.45	0.25-0.50
Potassium (%)	2.50-4.00	2.00-2.50	1.70-3.00
Calcium (%)	0.30-0.70	0.25-0.50	0.21-1.00
Magnesium (%)	0.15-0.45	0.13-0.30	0.20-1.00
Sulphur (%)	0.15-0.50	0.15-0.50	0.21-0.50
Boron (mg kg^{-1})	5-25	4-25	5-25
Copper (mg kg^{-1})	5-20	3-15	6-20
Iron (mg kg^{-1})	50-250	10-100	20-250
Manganese (mg kg^{-1})	20-300	15-300	20-200
Molybdenum (mg kg^{-1})	0.10-0.30	0.10-0.30	0.10-0.20
Zinc (mg kg^{-1})	20-60	15-60	25-100

FERTILIZERS AND SOIL AMENDMENTS

Fertilizer is any material that supplies water-soluble nutrient ions for crop production. Plants absorb nutrient ions from the soil pore water, without preference or discrimination among fertilizer sources. Therefore, any fertilizer that releases water-soluble nutrient ions could be beneficial to plants. Characteristics of some fertilizers that can increase the NH_4^+ and NO_3^- supply in soil are listed in Table 7.4.

Table 7.4. Characteristics of some nitrogen fertilizers that can supply NH_4^+ and NO_3^- for crops, based on IPNI (2012) and Munroe (2016)

Fertilizer	Guaranteed analysis[1]	Comments
Anhydrous ammonia	82-0-0	Applied as a compressed gas; requires special equipment and training
Calcium ammonium nitrate	27.5-0-0	Granular; water-soluble
Ammonium sulfate	21-0-0 + 24% sulphur	Granular; water-soluble
Urea	46-0-0	Granular; water-soluble Hydrolyzed by urease enzymes to release NH_4^+
Urea ammonium nitrate	28-0-0	Solution containing 50% urea and 50% ammonium nitrate
	Typical analysis[2]	
Liquid pig slurry	2.2-2.7-1.9	Solution (>90% water) with C:N ratio <5
Dairy slurry	1.2-1.8-2.9	Solution (>90% water) with C:N ratio <10
Chicken manure	2.4-1.1-1.0	Decomposed solid, C:N ratio <20
Beef cattle manure	0.8-1.8-2.6	Partially-decomposed solid, C:N ratio <25

[1]Guaranteed analysis is based on the %N-%P_2O_5-%K_2O in the fertilizer, on a dry weight basis.
[2]Typical analysis of manure is based on the %N-%P_2O_5-%K_2O, on a wet weight basis

Fertilizer sources can be classified based on how they are manufactured, but from the perspective of plant nutrition, it is more interesting to consider how quickly the water-soluble nutrient ions are released into the soil pore water. For example, granular monoammonium phosphate purchased from a fertilizer company dissolves in soil water to release water-soluble NH_4^+ and $H_2PO_4^-$ ions. Another solid fertilizer is chicken manure, which contains about 4% total nitrogen and 0.5% total phosphorus (equivalent to 1.1% P_2O_5). The total nitrogen is composed of uric acid, protein and NH_4^+, and the total phosphorus contains orthophosphate compounds (50-75% of the total P) and organic phosphorus compounds such as myo-inositol (25-50% of the total P). Because chickens do not produce phosphatase enzymes in their digestive tract, only 10 to 30 per cent of the organic phosphorus in their food is absorbed by the animal and the undigested phosphorus is excreted. When chicken manure is applied to soil, the organic phosphorus compounds dissolve in soil water and are degraded by phosphatase enzymes that were produced by plant roots and soil microorganisms. Phosphatase enzyme hydrolyze organic phosphorus to release phosphate (HPO_4^{2-}) ions into the soil pore water (Figure 7.7). The dissolution of granular monoammonium phosphate generally occurs faster than the hydrolysis of organic phosphorus compounds in chicken manure, which is why salt-based fertilizers are a ‘fast-release’ fertilizer and manure is considered a ‘slow-release’ fertilizer.

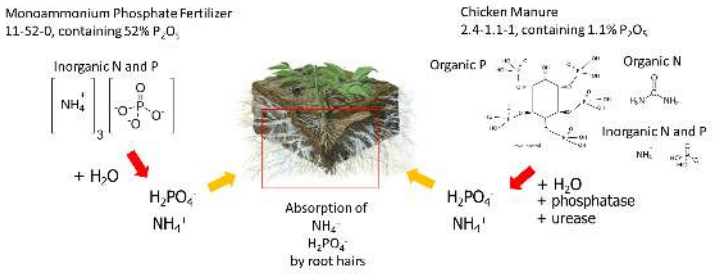


Figure 7.7. Commercial fertilizer such as monoammonium phosphate dissolves in water. Organic fertilizers such as chicken manure also contain NH_4^+ and $H_2PO_4^-$ ions, since inorganic forms can be >50% of the total nitrogen and total phosphorus in manure. Urease and phosphatase enzymes, produced by plant roots and soil microorganisms, are responsible for hydrolyzing organic nitrogen and organic phosphorus compounds in chicken manure. Soluble NH_4^+ and $H_2PO_4^-$ ions are absorbed through root hairs and used for plant metabolic processes. Water-soluble fertilizers are considered to be "fast-release" fertilizer and organic fertilizer is a relatively "slow-release" fertilizer. The numbers under each fertilizer are the guaranteed analysis based on the %N-%P_2O_5-%K_2O in the fertilizer. Monoammonium phosphate is a compound fertilizer that contains 11% N, 52% P_2O_5 and no potassium. Chicken manure is a compound fertilizer that, on average, contains 2.4% N, 1.1% P_2O_5 and 1% K_2O. © Joann Whalen is licensed under a CC BY (Attribution) license.

A soil amendment is a material that is added to improve the soil condition, but it does not necessarily provide nutrients for plant growth. Ground lime is often applied to acidic soils to increase the soil pH and reduce the concentration of Al^{3+} and Mn^{2+} in the soil pore water, since these elements can be damaging to crop roots. Neutralizing soil pH to around 6.0–7.0 provides more water-soluble NH_4^+, $H_2PO_4^-$ and K^+, and increases the activity of soil bacteria, which are beneficial for plant growth and maintain soil structure. Other soil amendments include mulch and biochar. These carbon-rich materials are not expected to be a substantial source of plant-available nutrients, but they contain organic matter that can retain soil moisture and exchangeable ions, build soil structure and provide substrate for diverse heterotropic microorganisms. Since carbon-rich amendments stimulate microbial growth, they must be used sparingly to avoid the possibility that fast-growing microorganisms will immobilize nutrient ions from the soil pore water, thereby depriving plants of plant-available nutrients.

SOIL NUTRIENT CYCLES

Agricultural crops absorb nutrient ions from the soil pore water to produce edible biomass (e.g., grain, oilseed, forage, fruit and tubers) and non-edible biomass, some of which is harvested for animal bedding and for other purposes (i.e., second-generation biofuels are made from non-edible stems and leaves). Fertilizer addition will replenish the plant-available nutrients, but the nutrient ions from fertilizer will not necessarily remain in the soil pore water because they are subject to the biological, chemical and physical reactions of soil nutrient cycles, described below.

Nitrogen Cycling

Reactions that contribute plant-available nitrogen

The soil nitrogen cycle (Figure 7.8) is well studied, due to the central importance of nitrogen for chlorophyll production, Calvin cycle reactions that take place in chloroplasts during photosynthesis, and protein synthesis in plants. One way that agricultural crops can obtain nitrogen is through a symbiotic or associative relationships with N_2-fixing prokaryotes. The best known of these are soil bacteria in the Rhizobium genera, but there are many others (Table 7.5). The N_2-fixing prokaryotes possess the nitrogenase enzyme, which can transform atmospheric N_2 into ammonia (NH_3), the precursor for protein synthesis, through the reaction:

$$N_2 + 8H^+ + 16ATP + 8e^- \rightarrow 2NH_3 + H_2 + 16ATP + 16P_i$$

The nitrogenase complex reduces N_2 to NH_3, and it also reduces H^+ to H_2. The energy (ATP) for this reaction comes from the breakdown, via glycolysis, of photosynthates that are transferred from the host plant to the symbiont. The NH_3 (g) product diffuses out of the bacteriod's protoplasm into the root cortex, where it is protonated with 2 H^+ ions, without additional energy, to yield:

$$NH_3 + 2H^+ \rightarrow 2NH_4^+$$

The NH_4^+ then binds to glutamate to produce glutamine, or is transformed into ureides such as allantoin, allantoic acid or citrulline, before it is translocated from the root to other parts of the plant.

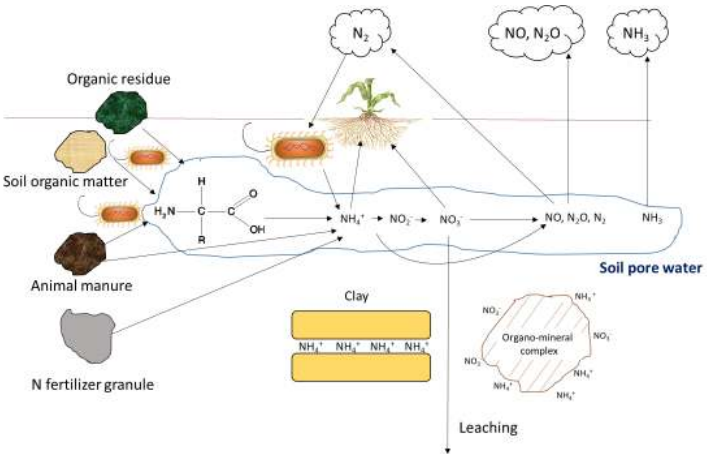


Figure 7.8. Simplified version of the soil nitrogen cycle. © Joann Whalen is licensed under a CC BY (Attribution) license.

Legumes in the Fabaceae family form a symbiosis with N_2-fixing bacteria, which inhabit nodules in their roots. The symbiosis is initiated when roots secrete flavonoids, which are chemo-attractants for the nodulating bacteria (rhizobia) to attach to the root hairs. The flavonoid signal activates the nodulation (nod) genes in the rhizobia, which synthesizes Nod factors that bind to surface protein receptors at the subapical root tip. This causes the root hair to curl and entrap the rhizobia, and the plasma membrane invaginates to form the infection thread, a tubular structure that extends from the root hair tip to lower cells of the root cortex. Rhizobia move through the infection thread and multiply, and root cortex cells proliferate into the nodule primordia, which will eventually become a fully-formed nodule enveloped in a membrane that prevents oxygen diffusion. This protective membrane prevents nitrogenase from being inactivated by oxygen. At this stage, the rhizobial bacteria become bacteroids capable of biological N_2 fixation.

Although N_2 is abundant (78% of the Earth's atmosphere), the biological N_2 fixation reaction is energetically demanding. In theory, 16 ATP are required to reduce one molecule of N_2 to $NH_{3\ (g)}$, but the actual energetic requirement is generally higher due to inefficiencies in the process. Therefore, biological N_2 fixation is a strictly controlled process. All other available forms of nitrogen, such as NH_4^+, NO_3^- and amino acids, will inactivate the nitrogenase complex and inhibit the expression of the nif genes that encode for the nitrogenase components. In addition, the reaction products $NH_{3\ (g)}$ and H_2 also inhibit nitrogenase activity. Consequently, most legumes acquire 40-75% of their nitrogen requirement from biological N_2 fixation, and the rest is absorbed as water-soluble NH_4^+ and NO_3^- from the soil pore water.

Associative N_2-fixing bacteria live in the rhizosphere (in the soil pores and on soil particles close to plant roots) or as endophytes (within the plant tissues) of grasses such as sugarcane, rice, wheat, sorghum, maize and others. Some common genera are Azospirillum and Herbaspirillum, and there are many others (Table 7.5). These bacteria do not form nodules. Instead, they live on the root surface or within the intercellular space and xylem vessels of stems and leaves. They obtain energy from plant photosynthates and transfer a portion of the NH_4^+ generated from N_2 fixation to the plant. There is considerable variability in their contribution to crop nitrogen requirements. In sugarcane, *Gluconacetobacter diazotrophicus* can supply 0-60% of the nitrogen requirement under field conditions. Cereals and oilseeds may obtain 1-25 kg N ha^{-1} y^{-1} from their associative N_2-fixing bacteria. The rest of the nitrogen needed by these crops must be absorbed from the soil pore water as NH_4^+ and NO_3^- ions.

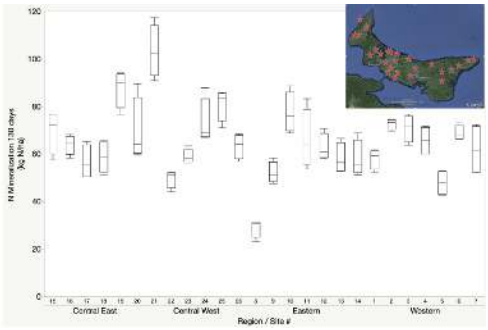
Table 7.5. Nitrogen-fixing prokatyotes that can supply nitrogen to plants. Adapted from Rascio and La Rocca (2013)

Class	N_2-fixing prokaryote	Host plant
Symbiotic	Allorhizobium spp. (fast-growing) Azorhizobium spp. (fast-growing) Bradyrhizobium spp. (slow-growing) Mesorhizobium spp. (intermediate-growing) Rhizobium spp. (fast-growing) Sinorhizobium spp. (fast-growing)	Legumes (Fabaceae family)
	Anabaena azollae (cyanobacteria)	Water fern (Azolla)
	Frankia spp. (actinobacteria)	Woody shrubs, e.g., in the families Betulaceae, Casuarinaceae and Myricaceae
Associative	Azospirillum spp. Bacillus spp. Beijerinckia fluminense Derxia spp. Enterobacter spp. Erwinia spp. Gluconacetobacter diazotrophicus Herbaspirillum spp.	Cereals (Poaceae family) Oilseeds (Brassicaceae family)

Can You Dig It!

The importance of measuring soil N supply: A Prince Edward Island example

Nitrogen is a very dynamic nutrient—both critical to crop production and with the potential for significant environmental impact. Thus it is essential that we use nitrogen as effectively and efficiently as possible. In determining the need for nitrogen additions, we cannot forget to account for the nitrogen that will be supplied by the soil. In semi-arid and subhumid agroecosystems, such as those in the Canadian Prairies, residual nitrate-nitrogen (NO_3-N) is an important source of plant N for the subsequent crop and should be measured in soil fertility tests and accounted for in fertilizer recommendations. In cool, humid agroecosystems such as those in Eastern Canada, overwinter loss of nitrogen is such that there is often limited carry over of nitrate (Zebarth et al., 2003). As a result, in-season nitrogen mineralization is an important crop input in both types of agroecosystems.

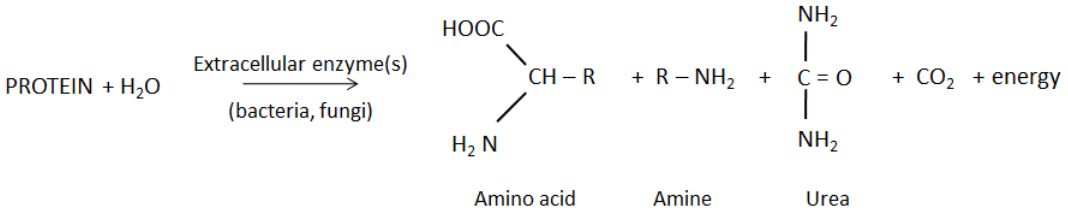


Nitrogen mineralization potential of PEI soils

There are a variety of approaches to estimate the nitrogen mineralization potential that range from measures of the quality of the organic matter to the activity of the microbial community (Dessureault -Rompré et al., 2011). The variation in nitrogen mineralization potential can vary substantially as a result of soil moisture, temperature, soil type, cropping system or soil management. A survey of nitrogen mineralization potential of fields in three-year potato rotations in Prince Edward Island demonstrated that potential N mineralization over a 130-day growing period could range from 30–110 kg N ha^{-1}. While there was on average greater N mineralization potential associated with central regions of the province, the greatest variation was from field to field, emphasizing the value of site-specific soil testing to quantify N mineralization potential.

https://geo.libretexts.org/@go/page/22346

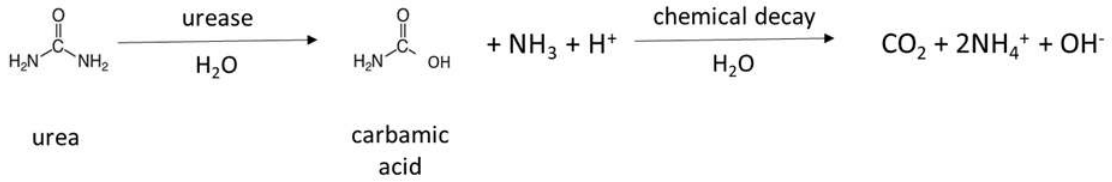
Biological mineralization of proteins, also referred to as nitrogen mineralization, is the main source of NH_4^+ in unfertilized soil. This naturally-occurring process releases water-soluble NH_4^+ from decaying plant, animal and microbial residues, which were recently added to soil or are part of the soil organic matter. Soil bacteria and fungi produce protease and peptidase enzymes, as well as amino acid hydrolase enzymes that cleave the amine groups from amino acids and urease, which degrades the urea molecule. The general mineralization reaction is:



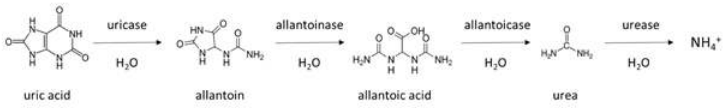
Nitrogen mineralization occurs more quickly with increasing soil temperature and is optimal when soil moisture reaches 60% water-filled pore space. Soil conditions that favour bacterial activity (pH from 6 to 7.5, low salinity) will stimulate mineralization, since bacteria have a faster generation time and greater capacity to produce protein-degrading enzymes than fungi.

Nitrogen mineralization is also affected by the C:N ratio in organic residues. The C:N ratio is the mass of the total organic carbon divided by the mass of total nitrogen in an organic substance. Organic residues with a larger proportion of organic carbon provide substrates for heterotrophic microorganisms, which use carbon for metabolic processes and biomass production, and respire CO_2. If their growth becomes limited by a lack of nitrogen in the organic residue, heterotrophic microorganisms will immobilize NH_4^+ and NO_3^- from the soil pore water. Organic residues with a C:N ratio > 20 such as woodchips, straw and grass-based residues tend to be rich in organic carbon and relatively impoverished in total nitrogen, which causes immobilization. Organic residues with a C:N ratio < 20 such as food-based compost and decomposed animal manure are relatively enriched in total nitrogen, so they will mineralize and increase the NH_4^+ concentration in soil pore water.

Many fertilizers are a source of NH_4^+ in soil. Granular fertilizer such as ammonium sulfate, $(NH_4)_2SO_4$, is water-soluble and dissolves to release NH_4^+ (Table 7.4). Urea fertilizer is hydrolyzed by the urease enzyme, which is produced by naturally-occurring soil microorganisms and plants, to release NH_4^+:



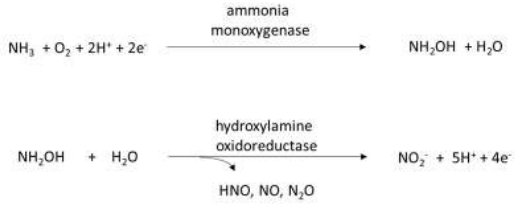
A similar reaction occurs in animal manure, which contains urea from animal urine. Chicken manure collected from laying facilities without bedding is enriched with urea-like molecules, due to the fact that 80% of the nitrogen excreted is in the form of uric acid. It takes several days to a week to transform uric acid to NH_4^+ in a warm, moist soil according to the general reaction:



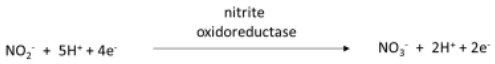
Pig manure that is collected by spray-washing the barns and stored in a liquid form also has a high proportion of NH_4^+, due to urea hydrolysis and protein degradation that occurs in the manure storage. For example, liquid pig slurry contains >90% water, has a C:N ratio of 3, and the NH_4^+ form represents 75% of the total N content. The high proportion of readily-degradable uric acid in chicken manure, and the large amount of water-soluble NH_4^+ in liquid pig slurry, makes these manures “fast-release” nitrogen fertilizer, compared to animal manure mixed with bedding, which has a lower water content.

https://geo.libretexts.org/@go/page/22346

Ammonia oxidizers and nitrifiers are responsible for transforming the NH_4^+ to NO_3^-. The ammonia oxidizers are a diverse group of prokaryotes, including bacteria and archaea, which possess the amo and hao genes that encode for the enzymes that catalyze the following 2-step reaction:



Nitrifiers are prokaryotes and eukaryotes that possess the nxr gene and are capable of converting nitrite (NO_2^-) into NO_3^-, as:



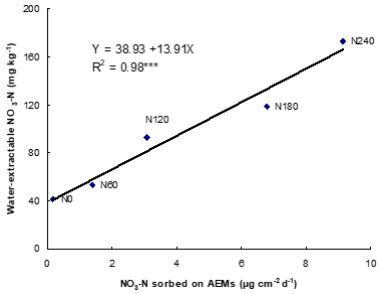
Ammonia oxidation and nitrification reactions occur quickly in field soils during the growing season. Consequently, NH_4^+ released from nitrogen mineralization or added in water-soluble fertilizer or animal manure will be transformed into NO_3^- within days to weeks. This understanding becomes important when deciding when to apply fertilizer, because we would like crops to absorb most of the water-soluble NH_4^- and NO_3^- from fertilizer as quickly as possible after the fertilizer application. Otherwise, reactions in the nitrogen cycle will remove these nutrient ions from the soil pore water, thereby reducing the amount of plant-available nitrogen for crops and increasing the risk of environmental pollution.

Can You Dig It!

Soil mineral nitrogen as predicted by ion exchange membranes

Ionic exchange membranes (IEMs) and plant root simulators were developed to determine N availability for agricultural and environmental studies. Indeed, IEMs contain positively or negatively charged surface functional groups that adsorb ions, such as NH_4^+ and NO_3^-, by electrostatic attraction. To maintain the equilibrium between the soil solid phase and soil pore water, IEMs continually remove NH_4^+ and NO_3^- that is mineralized from organic N or desorbed from organo-mineral surfaces.

One way to evaluate the NH_4^+ and NO_3^- concentration is to bury IEMs in the field for a period of time that depends on the objectives. Alternatively, the NH_4^+ and NO_3^- concentration can be assessed in a standardized laboratory environment. Once removed from the soil, IEMs are collected, washed with distilled water in the field to remove adhering soil particles and placed in individual tubes containing either NaCl (anionic exchange membranes) or HCl (cationic exchange membranes), and analyzed in laboratory for NO_3^- and NH_4^+, respectively. Soil nitrate and ammonium fluxes (supply rates) measured by IEMs are reported in µg cm^{-2} d^{-1} (Ziadi et al., 1999).



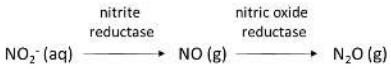
Relationship between soil NO_3-N following a water extraction and NO_3-N adsorbed on ionic exchange membranes in forages fertilized with various N rates in an experiment conducted at four site-years in eastern Canada (Ziadi et al., 1999). © Noura Ziadi is licensed under a CC BY (Attribution) license.

In general, there are significant correlations between IEMs and chemical extractions (Ziadi et al., 1999), and between crop N uptake and nitrates adsorbed on IEMs. Plant N uptake was more closely related to soil N as evaluated by IEMs than estimated by chemical extractions, suggesting that IEMs provide a better index of soil N availability than chemical extractions. Over all sites and treatments (mineral and organic N), NO_3^- and NH_4^+ adsorbed on IEMs increased with increasing N fertilizer rates. SoilNO_3^- and NH_4^+ adsorbed on IEMs varied largely with growing seasons and development stage of crop indicating the variation of soil N availability to plants. The in situ use of ionic exchange membranes as a soil testing procedure provide a more precise evaluation of N availability to different crops (forages, corn and potato) than standard methods. Based on its simplicity, rapidity, and low cost, IEMs have many practical advantages over chemical extractions for assessing soil N availability in eastern Canada.

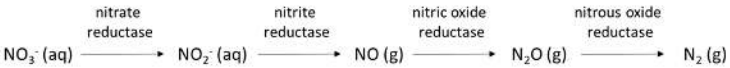
Processes that deplete plant-available nitrogen

The NH_4^+ and NO_3^- ions in the soil pore water are removed through biological, chemical and physical processes. Nitrogen uptake by plants and immobilization by microorganisms is driven by nitrogen requirements for protein synthesis and formation of genetic material. In plants, nitrogen is a major component of proteins of the Calvin cycle (e.g., about 20-30% of the leaf nitrogen is found in the RuBisCO enzyme that initiates carbohydrate production), the thylakoids and the chlorophyll molecule and thus essential for CO_2 fixation associated with photosynthesis.

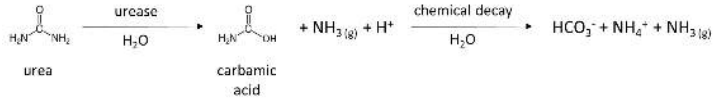
Soil biological reactions that use NO_2^- and NO_3^- as electron acceptors are common. These reactions permit soil bacteria and fungi to acquire oxygen when the soil environment is anoxic. High soil moisture content can create anoxic conditions, as can intense biological activity that consumes oxygen and produce carbon dioxide, such as in the rhizosphere or in the detritusphere (i.e., around decomposing organic matter). Nitrifier-denitrification is one such reaction, and is expected when the soil moisture content is from 50 to 80% of water-filled pore space.



Denitrification is another biological reduction reaction that consumes NO_3^-, and it is expected when the soil moisture content exceeds 80% water-filled pore space.



A chemical reaction that depletes the NH_4^+ concentration is $NH_{3\ (g)}$ volatilization, which occurs when NH_4^+ is deprotonated. During urea hydrolysis, this can be the result of carbonic acid formation:



Animal manure can lose up to 50% of the total nitrogen excreted by the animal as $NH_{3\ (g)}$ during manure storage and land application. Urea fertilizer is also susceptible to $NH_{3\ (g)}$ loss, especially in alkaline soils that have an excess of OH^- and HCO_3^- in the soil pore water, since H^+ associates with these counter-anions rather than with NH_3. Volatilization of $NH_{3\ (g)}$ is undesirable because the transfer of gaseous nitrogen from agricultural fields to the atmosphere reduces plant-available nitrogen for the crop. Eventually, the $NH_{3\ (g)}$ will return to the ground through wet and dry deposition, which results in the unintentional fertilization of non-agricultural land and aquatic systems.

Ammonium fixation is a chemical reaction that removes NH_4^+ from the soil pore water. This occurs when NH_4^+ adsorbs to exchange sites on organo-mineral surfaces. Ammonium fixation is greater in clayey soils, which have greater reactive surface area and more adsorption sites in the interlattices (2:1 clays only), than sandy soils. Due to its 1+ charge and small ionic radius, NH_4^+ is likely to bind to the interlattice, which reduces the amount of NH_4^+ in the soil pore water. This could account for a reserve of as much as 90–460 mg NH_4^+ kg^{-1} soil in fine-textured clays. In contrast, the NO_3^- molecule is loosely retained in the soil matrix due to its size and negative charge, and is more likely to leach through the soil profile and below the plant root zone. Nitrate leaching is affected by soil porosity and the water infiltration rate after snowmelt, rainfall and irrigation events. Finally, both NH_4^+ and NO_3^-

are susceptible to loss from soil through erosion and surface runoff processes. Wind and water erode soil particles that contain adsorbed NH_4^+, while runoff water transports NO_3^- from saturated soils through overland flow.

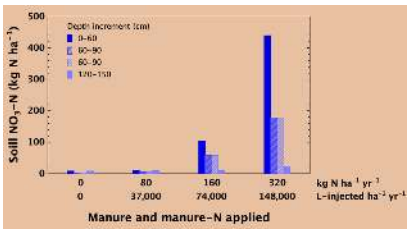
Can You Dig It!

Optimizing the nitrogen application rate of animal manure



Low disturbance injection of liquid swine manure at research plots in east-central Saskatchewan. Photo credit: Dr. Tom King. © Tom King is licensed under a CC BY (Attribution) license

Balancing the rate of nitrogen addition with removal in the harvested crop over time prevents excessive accumulation of nitrate in the soil. Unused nitrate can slowly migrate down below the rooting zone over time with deep percolating water. In a trial in east-central Saskatchewan, liquid swine manure was added every year for eight years at a recommended agronomic rate of 37,000 L ha^{-1} yr^{-1}, adding about 80 kg N ha^{-1} yr^{-1} which was about equal to the amount of N removed in the grain harvest each year. This rate resulted in little or no accumulation of nitrate in the soil profile, nor was there evidence of any deep leaching compared to an unfertilized control (lower right).



Fall nitrate concentrations in the soil profile (0-150 cm) after eight annual applications of liquid hog manure at Dixon, Saskatchewan (adapted from Stumborg et al. 2007). © Stumborg et al. adapted byTom King is licensed under a CC BY (Attribution) license.

In the unfertilized control, crop production over the years was greatly reduced due to lack of available nitrogen while the agronomic rate produced good yields. At annual application rates that were double (74,000 L ha^{-1} yr^{-1}) and especially quadruple (148,000 L ha^{-1} yr^{-1}) the recommended rate, soil nitrate contents in the top 60cm as well as the depths beneath were elevated compared to the agronomic rate and the unfertilized control. Large amounts of unused nitrate that accumulate in the soil profile over time can slowly migrate to greater depths. If leached deep enough, it can become unreachable by plant roots and be a potential source of ground water contamination. With appropriate application rate that matches crop need and removal, this can be prevented.

Phosphorus Cycling

Reactions that contribute plant-available phosphorus

Plant-available phosphorus is derived from biological and geochemical processes (Figure 7.9). The biological source is organic phosphorus, which is hydrolyzed by phosphatase enzymes from plant roots and soil microorganisms. Organic phosphorus compounds are abundant in plant residues and in undigested plant materials. Inositols are major phosphate storage compounds that contribute to cell wall biosynthesis, cell-to-cell communication, storage and transport of plant hormones and other stress-related molecules, and osmoregulation. Degradation of myo-inositol releases water-soluble $H_2PO_4^-$, as follows:

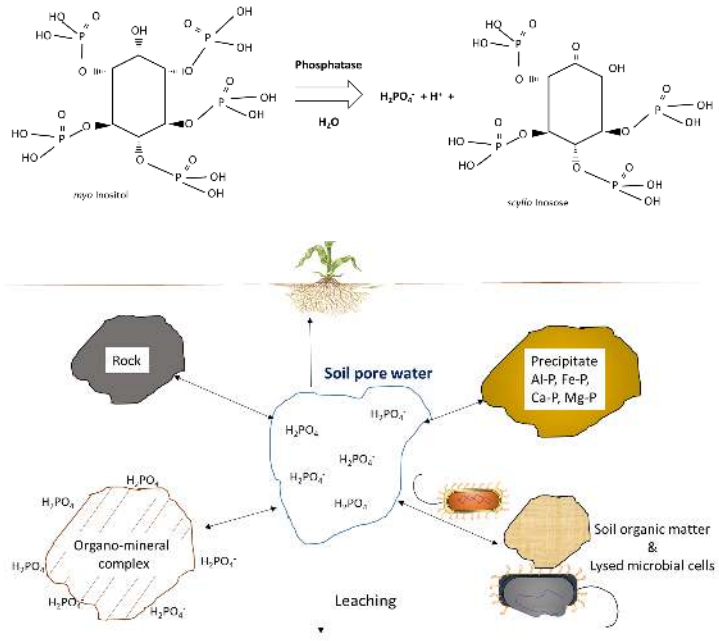


Figure 7.9. Simplified version of the soil phosphorus cycle. © Joann Whalen is licensed under a CC BY (Attribution) license.

Primary minerals such as apatite undergo weathering to release $H_2PO_4^-$ into soil pore water. Other geochemical sources include $H_2PO_4^-$ that is desorbed from clay surfaces or dissolved from secondary minerals, the iron, aluminum, calcium and magnesium phosphates. Weathering and dissolution processes require water. The dissolution of calcium phosphates and magnesium phosphates are accelerated by the presence of carbonic acid ($H_2CO_3 \leftrightarrow HCO_3^- + H^+$) and other acidic substances, however, iron phosphates and aluminum phosphates become less soluble in acidic soils. Inorganic acids that contribute to the solubilisation of $H_2PO_4^-$ from calcium phosphates and magnesium phosphates include H_2SO_4 and HCl. Organic acids produced by plant roots and soil microorganisms lower the pH in their surroundings, resulting in localized $H_2PO_4^-$ solubilization; in addition, organic acids can displace $H_2PO_4^-$ that is adsorbed onto clay and organo-mineral surfaces. Organic acids that are effective in this regard include acetic, lactic, malonic, malic, tartaric, oxalic, citric, ρ-hydrobenzoic and salicylic acids. Soil bacteria that produce organic acids, such as those from the genera Pseudomonas, Bacillus, Rhizobium and Burkholderia, can be isolated, cultivated and re-inoculated into soil to increase $H_2PO_4^-$ released via dissolution, but have not proven to be reliable at solubilizing $H_2PO_4^-$ under field conditions. Arbuscular mycorrhizal fungi also produce organic acids that increase water-soluble $H_2PO_4^-$, which they then absorb through their extra-radial hyphae. Since arbuscular mycorrhizal fungi are plant symbionts, some of the $H_2PO_4^-$ acquired by the fungus can be transferred to the plant host through low and high affinity phosphorus transporters.

Long-term impact of tillage practices and phosphorus fertilization on soil phosphorus forms as determined by ^{31}P nuclear magnetic resonance spectroscopy

Conservation tillage practices (minimum or no-till [NT]) are characterized by minimal soil disturbance and mixing and have been used more frequently in recent years to reduce off-site losses of nutrients associated with eroded particles, including phosphorus (P). However, by maintaining crop residues and fertilizers on the soil surface, the relatively immobile nutrients that do not readily move down the soil profile remain at or near the soil surface. Therefore, NT management systems often result in high concentrations of nutrients at the soil surface (0–5 cm) but sharply decreasing concentrations below this depth. The impact of tillage practices and P fertilization on soil P forms as determined using ^{31}P nuclear magnetic resonance (NMR) spectroscopy was investigated under a long-term corn-soybean rotation experiment established in 1992 in Québec that compared NT with conventional (mouldboard plow; MP) tillage. The results showed that:

- *total phosphorus varied with soil depth under the NT treatment— accumulating in the top unfertilized soil layer (0–5 cm) and in the 0–10 cm layer where P fertilizer was applied;*
- *pH, available P, and total carbon were significantly higher in the topsoil (0–5 cm) of the NT treatment than in the deeper layers (5–20 cm); total N also accumulated in the surface 10 cm;*

- *conversely, nutrient distributions were homogeneous throughout the soil profile under MP management—suggesting that stratification in NT results from the retention of crop residues at the soil surface;*
- *^{31}P-NMR spectra showed concomitant stratification of orthophosphate in NT plots that received 35 kg P ha^{-1};*
- *accumulation of scyllo-IP6 and nucleotides in the deeper layers was possibly due to their preferential movement through the soil column, whereas the pyrophosphate and DNA concentrations were greater at the surface (0–5 cm) soil layer than the deeper layers; and*
- *total C and N were similarly affected, suggesting that DNA was synthesized in greater amounts under NT owing to the higher organic matter accumulated at the soil surface in comparison to conventional tillage.*

Overall, it appears that the accumulation of labile inorganic P at the surface of the NT soil may increase the potential for soluble inorganic P loss in surface runoff and lead to the loss of organic monoesters draining through different hydrological pathways. ***More information available at Abdi et al. (2014).***

Reactions that deplete the plant-available phosphorus

The $H_2PO_4^-$ ions in the soil pore water are removed through biological, chemical and physical processes. Phosphorus uptake by plants and immobilization in microbial biomass are necessary for energy storage and transfer reactions, as well as the formation of nucleic acids and phospholipids.

Phosphorus retention reactions reduce the $H_2PO_4^-$ concentration in soil pore water because $H_2PO_4^-$ adsorbs to organo-mineral surfaces and to the surface of iron and aluminum sesquioxides, namely iron oxide (Fe_2O3), iron hydroxide ($Fe(OH)_3$) and aluminum hydroxide ($Al(OH)_3$). These reactions are common in acidic soils, but occur in all soils to some extent. In general, soils with a higher clay content will adsorb more $H_2PO_4^-$, particularly when they contain 2:1 expanding lattice clays such as smectite. The hydrated sesquioxides of iron and aluminum bind $H_2PO_4^-$ tightly and render it unavailable for plant uptake. In addition, $H_2PO_4^-$ precipitates with Fe, Al and Ca present in the soil pore water. This produces precipitates such as $Fe(OH)_2H_2PO_4$ and the variscite-like compound, $Al(OH)_2H_2PO_4$ in acidic soils. In neutral to alkaline soils, the $H_2PO_4^-$ precipitates with Ca to form slightly soluble dicalcium phosphate dihydrate, $CaHPO_4$. Further reaction with $Ca(OH)_2$ and $CaCO_3$ in alkaline soils leads to the formation of insoluble hydroxylapatite (e.g., $[Ca_3(PO_4)_2]_3$•$CaOH_2$) or insoluble carbonated calcium apatitie (e.g., carbonato-apatite; $[Ca_3(PO_4)_2]_3$•$CaCO_3$).

The phosphorus saturation concept is another way to describe the capacity of soil to retain chemically-bound $H_2PO_4^-$. Soil is shaken with an extracting solution such as Mehlich 3 solution, the amount of extractable P and Al is measured, and the ratio is calculated as:

(1) $$P\ saturation = \frac{P\ in\ Mehlich\ 3\ solution}{Al\ in\ Mehlich\ 3\ solution}$$

Soil with a high degree of phosphorus saturation has elevated $H_2PO_4^-$, relative to the amount of extractable Al that could bind with $H_2PO_4^-$. Therefore, more water-soluble $H_2PO_4^-$ remains in the soil pore water and could leach after snowmelt, rainfall and irrigation events. In addition, erosion of soil particles that contain adsorbed $H_2PO_4^-$ will transport particulate P, while surface runoff will transport the dissolved phosphorus compounds at the soil surface through overland flow. The magnitude of phosphorus loss through these pathways depends upon the concentration of water soluble $H_2PO_4^-$and particulate P in the surface soil, as well as the amount of leaching and severity of surface erosion processes.

Sulphur Cycling

Reactions that contribute plant-available sulphur

Plant-available sulphur can originate from weathering of parent material, atmospheric deposition, dissolution of gypsum ($CaSO_4$) and mineralization of organic matter (Figure 7.10). Plutonic rocks such as pyrite contain elemental sulphur (So) that is oxidized to SO_4^{2-} during the weathering process. Most oxidation reactions are mediated by autotrophic fungi and bacteria, such as Thiobacillus, which is abundant in the rhizosphere.

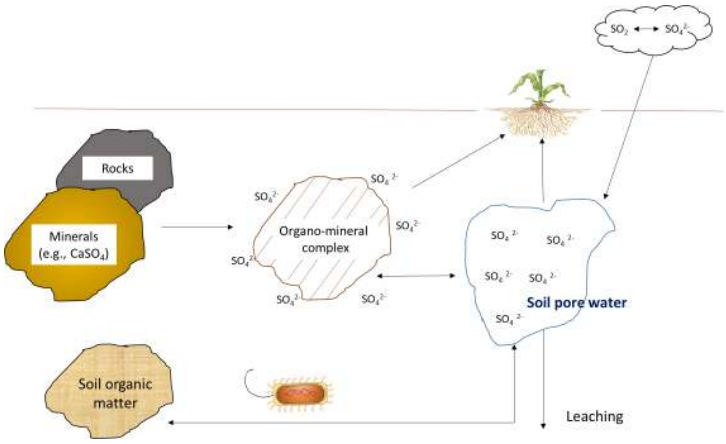
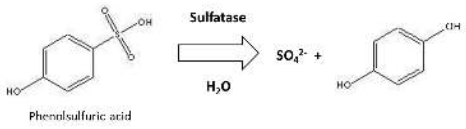


Figure 7.10. Simplified version of the soil sulphur cycle. © Joann Whalen is licensed under a CC BY (Attribution) license.

The conversion of So to SO_4^{2-} is extremely variable in soils, due to natural variation in the distribution of sulphur oxidizing microorganisms. In general, sulphur oxidation reactions occur rapidly in warmer soils with moisture content with around 60% water-filled pore space and soil pH from 6-7.5. In addition, SO_4^{2-} is released from the dissolution of hydrated salts in soil, such as $CaSO_4 \cdot 2H_2O$ and $MgSO_4 \cdot 7\ H_2O$.

Sulphur enters soil from atmospheric sources through dry and wet deposition. Several natural phenomena emit sulphur to the atmosphere, including particulate sulphur compounds that come from weathering of continental crusts (dust) and sea spray. Small particles of sulphur are transported by wind and deposited onto the surface of soil or vegetation. Gaseous emissions of sulphur dioxide, hydrogen sulfide, dimethyl sulfide and other gases from volcanic eruptions, swamps and coastal wetlands are another source. In the atmosphere, these gases tend to be oxidized to sulphuric acid and aerosol sulfate, and then return to the ground in rainfall and snow. These biogenic sources of sulphur account for <10 kg S ha^{-1} y^{-1} in terrestrial ecosystems. Since the Industrial Revolution, the anthropogenic sources of sulphur make a significant contribution to the soil sulphur cycle, and could exceed >50 kg S ha^{-1} y^{-1} in agricultural and forested lands in the proximity of industrial activities. Combustion of fossil fuels, biomass burning, waste incineration and the use of explosives all emit sulphur dioxide and therefore contribute through wet deposition to the sulphur content in soil. However, a reduction in air pollution has lowered the amount of sulphur released from these sources.

Organic sulphur is mineralized by soil enzymes – sulfatases – of plant and microbial origin to release SO_4^{2-}, as illustrated below:



Like other mineralization reactions, sulphur mineralization increases with higher soil temperatures and is optimal when soil moisture content is around 60% water-filled pore space and soil pH is near neutral. Sulphur mineralization is higher around plant roots, which produce sulfatase. Also, plants absorb some of the SO_4^{2-} in soil solution, which could limit the amount of SO_4^{2-} available for microorganisms and therefore induce them to synthesize sulfatase. When fresh plant or animal residues are added to soil, the organic carbon to organic sulphur (C:S) ratio of the material will determine its susceptibility to mineralization. Organic residues with a C:S ratio <200 are mineralized readily and release SO_4^{2-} into soil solution. However, organic residues with a C:S ratio >400 that have a low sulphur content could lead to SO_4^{2-} immobilization by microorganisms, to meet their metabolic requirements.

Reactions that deplete the plant-available sulphur

The SO_4^{2-} ions in soil pore water are removed through biological, chemical and physical processes. Biological uptake of SO_4^{2-} by plants and microorganisms is necessary for the synthesis of essential amino acids (methionine, cysteine, homocysteine, and taurine), for production of glucosinolates, vitamins, as an enzyme cofactor and for chlorophyll production. In anoxic soils, sulfur electron acceptors gain electrons and decrease their oxidation state, so SO_4^{2-} will be reduced to elemental sulphur So and then to S^{2-} by sulphur-reducing bacteria in the phylum Firmicutes, including the genera *Desulfotomaculum*, *Desulfosporomusa*, and *Desulfosporosinus*.

Several chemical reactions reduce the SO_4^{2-} concentration in soil pore water. In calcareous soils, the SO_4^{2-} can be co-precipitated with $CaCO_3$ to form a $CaCO_3 \cdot CaSO_4$ complex. This accounts for about 10% of the total sulphur in calcareous soils. Second, SO_4^{2-} binds to anion exchange sites on organo-mineral surfaces. However, SO_4^{2-} is not as tightly bound to anion exchange sites as HPO_4^{2-} and $H_2PO_4^-$. Consequently, SO_4^{2-} is susceptible to leaching and may be lost through erosion and runoff processes.

Potassium, Calcium, and Magnesium

The mineral elements—potassium, calcium, and magnesium—cycle through geochemical processes (Figure 7.11). Although plants and other organisms require substantial amounts of potassium, calcium and magnesium, these elements function in their ionic forms within biological tissues. For instance, potassium exists predominately as K^+ in the vascular solution of plant tissues and in the cytoplasm of plant cells. It is translocated along with water in living plants. It may exist between layers of organic compounds such as polysaccharides, proteins and lipids, but it does not bind covalently to those substances. Leaching of K^+ and other mineral elements such as Ca^{2+} and Mg^{2+} occurs when rainfall comes into contact with plant and animal residues, before biological decomposition begins.

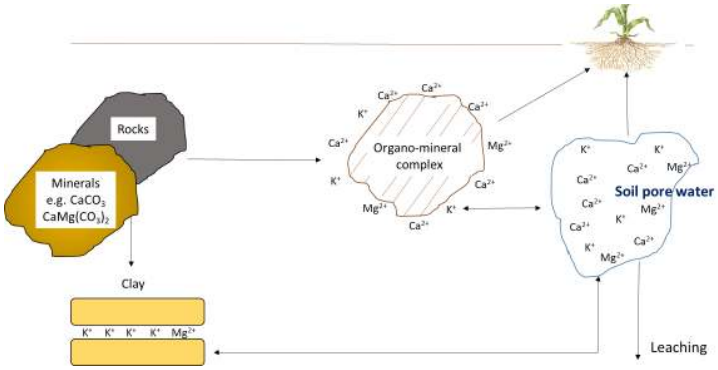


Figure 7.11. Simplified version of the soil potassium, calcium and magnesium cycles. © Joann Whalen is licensed under a CC BY (Attribution) license.

Reactions that contribute plant-available potassium, calcium and magnesium

The amount of K^+, Ca^{2+} and Mg^{2+} in the soil pore water and on the cation exchange sites depends on the nature and weathering of the parent material. Organic residues contribute minor amounts of water-soluble K^+, Ca^{2+} and Mg^{2+}, compared to parent material. Potassium originates from primary minerals, such as feldspar, microcline, mica and biotite, and can also be added as fertilizer.

Soil Fertility Response to Intensive Crop Nutrition



Grower's field in south-central Ontario monitored to assess impact of intensive nutrient management on soil fertility. © Tom Bruulsema is licensed under a CC BY (Attribution) license.

From 2002 to 2006, a corn-soybean rotation in a grower's field in south-central Ontario was monitored to assess changes in response to intensive management. The trial consisted of seven management combinations; three varying K rates at the grower's level of inputs, and four at a more intensive input level—varying both K rate and zone placement.

The soil is a London loam with good drainage. Initial soil fertility levels were 8 ppm Olsen P (low), and 107 ppm ammonium-acetate extractable K (medium). Five treatments used the conservation tillage practice of the grower: fall chisel plow with spring secondary tillage. The other two treatments used fall zone tillage followed by spring zone tillage. Each treatment plot consisted of a strip eight rows wide that ran the full length (300 to 450 m) of the field (see photograph at right).

Yields averaged over the five years increased modestly in response to both increased K rates and management intensity. No combination of input intensity and K rate was more economically viable than the grower's current level of management. There was no evidence of a greater requirement for K with higher input intensity.

Treatment	Fertilizer N-P_2O_5-K_2O (kg ha^{-1})	Seeding rate	Mean Corn Yield (Mg ha^{-1})	2006 Soil Test (ppm)	
				Olsen-P	AA-K
Control, zero K	147-71-3	normal	9.8	10	76
Control, grower K	147-71-33		10.2	9	77
Control, high K	147-71-200		10.6	7	140
Intensive, grower K	278-200-33	high	10.7	26	91
Intensive, high K	278-200-200		11.8	22	145
Intensive, zone-till, grower K	278 200 33		10.7	21	80
Intensive, zone-till, high K	278-200-200		11.1	17	119

Reference: Dean et al. 2007. Corn response to intensive crop nutrition. Better Crops 91:20-21.

Soil test levels changed little where P and K inputs were at the grower's rates but increased significantly where higher rates were applied. Grower rates resulted in net surpluses of P and net deficits of K, though the changes were not large enough to influence soil test levels. At the high rate of each nutrient, however, surpluses of P and K were large enough to substantially increase soil test levels for both nutrients. This work demonstrates that economically optimum crop yields require sufficient nutrient inputs to maintain soil fertility.

Calcium is released from weathering of primary minerals such as anorthite ($CaAl_2Si_2O_3$), followed by pyroxenes and amphiboles, with smaller amounts of Ca coming from the weathering of biotite and apatite. Substantial amounts of Ca^{2+} are released from the dissolution of secondary minerals such as limestone ($CaCO_3$), dolomitic limestone ($CaMg(CO_3)_2$), and gypsum ($CaSO_4 \cdot 2H_2O$), all of which are major sources of Ca^{2+} in Canadian soils.

Magnesium in soils originates from rocks such as biotite, dolomite, hornblende, olivene and serpentine. Magnesium is often a substituted cation in clay minerals, such as chlorite, illite, montomorillonite and vermiculite. Weathering of rocks and clays releases Mg^{2+} into soil pore water, as does dissolution of secondary Mg minerals such as dolomitic limestone ($CaMg(CO_3)_2$), epsomite ($MgSO_4 \cdot 7H_2O$) and bloedite ($Na_2Mg(SO_4)_3 \cdot 4H_2O$).

There is a chemical equilibrium between the amount of K^+, Ca^{2+} and Mg^{2+} in soil pore water and that on cation exchange sites. Smaller ions, particularly H^+, can replace these mineral elements on exchange sites, leading to desorption of K^+, Ca^{2+} and Mg^{2+} into soil pore water. The balance between exchangeable and soluble mineral ions is the key determinant of plant-available K^+, Ca^{2+} and Mg^{2+} concentrations. Since pH affects the charge on organo-mineral surfaces (i.e., neutral to alkaline soils have more negatively charged surfaces than acidic soils), the exchange capacity is related to soil pH, clay and organic matter content.

Reactions that deplete plant-available potassium, calcium and magnesium

The K^+, Ca^{2+} and Mg^{2+} ions in soil pore water are removed through biological, chemical and physical processes. Biological uptake of K^+ is essential for enzyme activation (more than 80 enzymes require K^+ as a cofactor), transport processes and osmotic regulation. The major functions of Ca^{2+} in plants are for maintenance of cell membranes, as well as cell division and elongation. An essential component in chlorophyll, Mg^{2+} is also a structural component in ribosomes and responsible for transfer reactions involving phosphate-reactive groups.

Adsorption reactions are the dominant chemical process that removes K^+, Ca^{2+} and Mg^{2+} from the soil pore water. Adsorption can occur on the outer surfaces ("o" position), inner surfaces ("i" position) and edge ("e" positions) of clay minerals. Furthermore, these ions can be adsorbed in the interlattice of 2:1 clays in mica, illite and vermiculite. Due to its 1+ charge and small ionic radius, K^+ tends to bind in the interlattice whereas Ca^{2+} and Mg^{2+} remain on outer surfaces of clay minerals. This reduces the amount of K^+ in the soil pore water due to potassium fixation, which can account for up to 20% of the total K contained in soils with 2:1 clay minerals. The K^+, Ca^{2+} and Mg^{2+} ions in soil pore water move out of the topsoil by leaching. This is associated with anion leaching, since anions that move through the soil pore water are accompanied by cations to maintain electrical neutrality. Erosion of clay-rich particles is another way that exchangeable K^+, Ca^{2+} and Mg^{2+} ions are depleted from the topsoil.

Micronutrient Elements

Weathering is the primary source of most micronutrients, including iron, copper, manganese, zinc, nickel, boron and molybdenum. Plant-available chlorine may be dissolved from salts such as NaCl and KCl, or it may enter the soil in rainfall and irrigation water. Micronutrients are generally concentrated in the vascular solution and cytoplasm of plants, so they are present in organic residues

and released during the initial physical fragmentation of decomposing organic residues. Consequently, plant and animal residues are a naturally-occurring source of essential micronutrients. Each micronutrient has specific functions in plants, listed in Table 7.6.

Table 7.6. Functions of micronutrients in plants, according to Jones (2003)

Element	Functions
Iron	Fe-containing proteins, including heme proteins and Fe-S proteins, are important for photosynthesis, chlorophyll production and ribonucleic acid synthesis. Cytochromes are Fe-containing heme proteins, and ferredoxin is the most well-known Fe-S protein.
Manganese	Important in redox processes, including electron transport in photosynthesis and detoxification of oxygen-free radicals. An enzyme cofactor for biosynthesis of secondary metabolites such as phenolics, lignins and flavonoids.
Copper	Cu-containing enzymes react with molecular O_2 and are important in photosynthesis, respiration, detoxification of superoxide radicals and lignification reactions.
Zinc	An enzyme activator that improves binding of enzymes and substrates involved in carbohydrate metabolism, protein synthesis and detoxification of superoxide radicals (O_2^-) to protect membrane lipids and proteins against oxidation.
Nickel	Required for the proper functioning of the urease enzyme that catalyses the reaction $CO(NH_2)_2 + H_2O \rightarrow 2\ NH_3 + CO_2$. Any crop that produces ureides requires Ni.
Boron	Forms stable complexes with organic compounds that contributes to cell wall formation and stabilization, lignification and xylem differentiation.
Molybdate	Involved in electron transfer reactions. Mo-containing enzymes catalyse biological N_2 fixation, and are responsible for pollen formation in plants.
Chloride	Mobile anion essential for osmoregulation (e.g., cell elongation, stomatal opening) and charge compensation (e.g., balance charges between cations and anions in the cytoplasm). Essential co-factor for photosynthesis.

Plant uptake of micronutrients is improved at higher temperatures and in plants with prolific root systems, since greater evapotranspiration will move the water-soluble forms of micronutrients to the root surface, where they can be absorbed. Soil moisture has to be sufficient to maintain micronutrient ions in the soil pore water, but not so high as to create anoxic conditions that affect ion absorption by roots. For instance, anoxic conditions prevent uptake of Fe^{2+} by roots, but may increase absorption of toxic Mn^{2+} ions. Soil pH is often critical. The amount of plant-available iron, copper, manganese, zinc and nickel declines as the pH increases to the neutral and alkaline range. For example, iron deficiency can occur in plants growing in calcareous soils with pH >7.5. Plants that are adapted to this environment will respond by altering root morphology to increase the growth of root hair, release protons to acidify the rhizosphere and make Fe^{3+} more soluble, reduce the Fe^{3+} to Fe^{2+} with a plasma membrane-associated Fe^{3+} chelate reductase enzyme, and transport Fe^{2+} across the plasma membrane with an iron regulated transporter. Adapted plants in the Poaceae family have a different strategy, which involves chelation of Fe^{3+} by the phytosiderophores before the Fe^{3+} soluble complex is absorbed by the root.

Plants need far less of the micronutrient elements than macronutrients (Table 7.1). Soil reserves are often sufficient to meet plant requirements, particularly when in soil with pH 6–7 and when adapted cultivars are planted. We can improve micronutrient nutrition by adding fertilizer containing the deficient micronutrient(s) to the soil or directly to the plant foliage. With foliar

application, a small amount of micronutrient fertilizer is sprayed onto the foliage in a solution with an adjuvant (to improve the adhesion of fertilizer solution to the leaf surface) during the plant's vegetative growth or early reproductive growth stages. Water-soluble micronutrients such as H_3BO_3, Cu^{2+}, Mn^{2+}, MoO_4^- and Zn^{2+} and soluble chelates are absorbed through pores in the leaf and move through the plant's vascular system to organs and tissues where they are required. Foliar fertilization is advantageous because it delivers a small amount of micronutrients and avoids adsorption and precipitation reactions that occur in soil. Due to the sensitivity of photosynthetic cells in leaves, the application rate and timing of micronutrients to foliage must be selected with care.

Non-essential Elements

Non-essential elements fall into two classes: (1) elements that are beneficial for some, but not all plants, and (2) elements that are not needed for plant growth. Elements that stimulate the growth of some plants include silicon, sodium and selenium. Silicon is beneficial for rice, sugarcane and other grass species. It accumulates preferentially in the stem tissues and appears to have a role in maintaining the structural integrity of the vascular system, which is required to transport water, nutrients, carbohydrates and other substances throughout the plant. Higher silicon concentrations are associated with greater drought tolerance. Sodium fertilization can improve the growth of some C4 plants, celery, spinach, sugar beet and turnips. Selenium appears to have general benefits related to oxidative stress tolerance and greater resistance to tissue damage caused by pathogens and herbivores.

Some elements can absorbed in quantities that are toxic to plants. In acidic soils with water-soluble Al^{3+} levels of 10 to 20 mg kg^{-1} or more, plants can be experience aluminum toxicity. The decline in plant root functions is related to oxidative damage of cell membranes, and less cell division and elongation at the growing tips due to Al^{3+} replacement of Ca^{2+}. In sensitive plants, the presence of Al^{3+} in leaf tissues may reduce stomatal conductance and chlorophyll content, which reduces the photosynthesis rate. Plants grown in acidic soils are also prone to manganese toxicity when they absorb more Mn^{2+} than required for metabolic functions, resulting in oxidative stress and destruction of indole-3-acetic acid (auxin), an important plant growth hormone. Manganese toxicity symptoms are expected to be more severe in anoxic soils because oxygen limitation will reduce several low-solubility compounds, such as MnO_2, Mn_2O_3 and Mn_3O_4, into Mn^{2+}. Liming acidic soils with $CaCO_3$ or another neutralizing substance that raises the soil pH will increase quantity of oxides and hydroxides, which form precipitates with Al^{3+} and Mn^{2+}. Since precipitated aluminum oxides/hydroxides and manganese oxides are not water-soluble, the risk of plant toxicity diminishes with increasing soil pH. Installing drainage systems that increase the soil oxygen level should reduce Mn^{2+} toxicity in plants.

In salt-affected soils, an abundance of sodium (Na^+), SO_4^{2-} or Cl^- creates salt stress for plants. Due to osmosis, root cells tend to lose water when the Na^+, SO_4^{2-} and Cl^- levels are greater outside the root than inside the root. This can lead to physiological disorders within roots and other plant tissues, such as ion toxicity, altered metabolic processes, oxidative stress, membrane disruption, and less cell division and expansion. Consequently, plants require more water to grow in salt-affected soils. Transpiration is responsible for the movement of water and the accumulation of water-soluble salts near the root. Evaporation also causes salt movement into topsoil, particularly in semi-arid and arid climates where annual water loss is greater than annual precipitation. Soil management can be effective in removing excess Na^+ and Cl^- from the root zone. Sprinkler irrigation combined with improved drainage could leach these salts through the soil profile and away from plant roots. Salt-tolerant crops should be planted in salt-affected soils.

STRATEGIES FOR MANAGING SOIL NUTRIENTS

Soil fertility on farms cannot be managed without knowledge of the actual fertility status, based on observations of plant growth, yield measurements, plant tissue analysis and soil fertility testing in each field. A diagnostic evaluation of soil fertility also requires information about the amount of crop residues left in the field after the harvest, how much and how often agricultural lime and other soil amendments are spread on fields, and how much fertilizer is applied each year. We should also know the crop rotation, tillage, irrigation and drainage conditions in each field. This information can explain why soil organic matter and nutrient levels are increasing, stable or declining across the farm.

Although it is tempting to believe that we should always try to increase the soil organic matter and nutrient levels, this is not always achievable or desirable. There are physical limits to the amount of organic matter that can be retained in soil, related to climatic conditions (primarily temperature and rainfall) that determine the crop residues remaining after harvest, soil microbial activity and soil texture. Increasing the concentration of water-soluble nutrients in soils with high or excessively high fertility could increase the risk of nutrient loss to the environment through biological reactions (e.g., nitrogen loss in gaseous forms such as NH_3, N_2O and N_2)

and via leaching and erosion, which transport water-soluble NO_3^- and $H_2PO_4^-$ and particulate-associated nutrients to downstream environments. Consequently, we must use fertilizers and other amendments prudently and judiciously.

The value of measuring residual soil nitrate to assess nitrogen use *efficiency*

Residual soil nitrate at the end of the growing season is N that was not used by the crop. Thus, measuring how much nitrate remains in the soil profile following harvest provides a means of assessing how effectively we have managed N sources. In semi-arid and subhumid climates, this test is used to adjust the N fertilizer requirements for the next growing season. However, in cool humid agroecosystems where there is a high potential for non-growing season loss of nitrate, the residual soil nitrate test is an indicator of the risk of nitrate loss into aquatic systems. In Agriculture and Agri-Food Canada's Agri-environmental Indicator series (Clearwater et al., 2016) the difference between N inputs and N outputs is referred to as Residual Soil Nitrogen (RSN) and has been estimated for the agricultural regions of Canada (see Figure below).

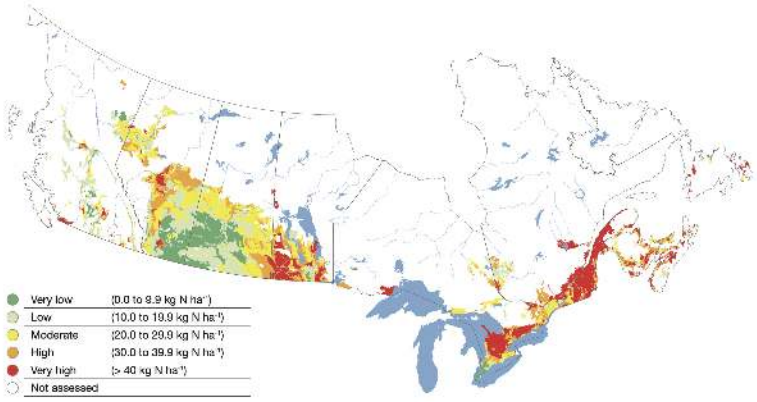


Figure 11-2). © Agriculture and Agri-Food Canada; used with permission and licensed under the Open Government Licence – Canada. https://open.canada.ca/en/open-government-licence-canada

This indicator points to potential for a large amount of nitrogen remaining in the soil in the fall after harvest in Eastern Canada with typical values exceeding 40 kg N ha^{-1}. Recent surveys of soil nitrate—measured to a depth of 45 cm—in Prince Edward Island suggest this estimate may indeed be conservative for potato production in that province with values averaging over 60 kg N ha^{-1}. The implementation of 4R practices reduced RSN in potato production in PEI by as much as 30%.

Since crops are the intended beneficiary of nutrient in agricultural lands, fertilizer should be applied in a way that optimizes nutrient acquisition by the crop. The 4R nutrient stewardship approach is a decision-making framework that guides farms in selecting the '**right**' actions related to fertilizer use, and is summarized below:

1. **Choose the right source of fertilizer** that will release water-soluble nutrients for the crop. The right source of fertilizer also depends on its availability and price. Fertilizers produced on farms, such as manure or compost, could be a readily-available, inexpensive fertilizer. Apply these fertilizers to meet crop nutrient requirements first, before purchasing commercial fertilizers that will supply the rest of the required nutrients. Is a fast-release fertilizer needed to overcome a crop nutrient deficiency? Consider a fertilizer source that dissolves in water. A controlled-release fertilizer, which is biologically or geochemically stabilized and releases water-soluble nutrients slowly, could be suitable for annual crops with a long growth period or perennial cropping systems.
2. **Choose the right rate of fertilizer** that permits your crop to reach its yield potential. When selecting the right rate, account for residual nutrients from decomposing crop residues and animal manure that were applied in previous growing seasons. Applying more nutrients than the crop can use will not improve the yield, although it could make weeds grow better. From an economic perspective, it is better to apply a little less fertilizer than you need, particularly when the growing conditions are sub-optimal due to a cold spring, flooding or drought. Applying more fertilizer during a poor growing season will not make the crop grow better and is a waste of money.
3. **Choose the right time to apply fertilizer** so that you can synchronize the supply of water-soluble nutrients with periods of high nutrient demand by the crop. Annual crops have an exponential growth phase for several weeks during their vegetative growth period, which slows down after floral initiation. In humid and irrigated soils, the best strategy is to apply fertilizer just before the exponential growth phase, so the plants will absorb nutrients efficiently. In some crops, fertilizer should be applied at

the beginning of the reproductive stage, to promote the development of high-quality seed. However, applying required fertilizers at the beginning of the growing season is a better strategy in semi-arid and subhumid regions because nutrient release is slow when precipitation occurs unevenly during the vegetative growth and initial reproductive stages. Perennial crops can be fertilized before they renew their vegetative growth each year. Perennial forages (i.e., hayfields) that are harvested two, three or four times a year can be fertilized after each successive cut, before the vegetative regrowth begins.

4. **Choose the right place to apply fertilizer** so that you can maximize uptake of water-soluble nutrients by the plant root. Precise placement is required to avoid direct contact of fertilizer salts with seeds and seedlings, but not so far that water-soluble nutrients leach from the root zone before they can be absorbed by root hairs. The most efficient methods for annual crops involve placing fertilizer in the planted row (i.e., between seeds or tubers) or beside the planted row so that transpiration aids the movement of nutrient ions through mass flow (e.g., NO_3^-, SO_4^{2-}) and diffusion (e.g., $H_2PO_4^-$, K^+) processes. Furthermore, as fertilizers dissolve or mineralize, they will release soluble nutrients that are accessible to the proliferating root system. Seeds can also be treated with fertilizer, as long as the fertilizer has a low salt content so there is no osmotic stress during germination and early seedling development. Foliar application is a rapid way to deliver small quantities of some micronutrient fertilizers, but there is a fine line between a micronutrient deficiency and a micronutrient toxicity in leaf tissues.

The 4R nutrient stewardship principles are the fundamentals of nutrient management on farms, but must be combined with soil and water conservation practices to be effective. The 4R Plus framework that is used on modern farms connects 4R nutrient management and soil and water conservation practices. The goal of the 4R Plus approach is to provide nutrients when the crop needs them, to enhance soil fertility and to improve water quality. Soil and water conservation practices that can enhance soil fertility and nutrient management include:

1. Reduce tillage to minimize soil disturbance and improve soil structure.
2. Leave crop residue on the soil surface to protect it from the impact of rainfall.
3. Grow cover crops to protect the soil and support the soil microbial community between cropping seasons.
4. Apply adequate fertilizer and manure using the 4R principles of right source, right rate, right time and right place.
5. Value manure applications, since manure can enhance soil biology and soil fertility.
6. Control field traffic to reduce soil compaction.
7. Extend crop rotations and include perennial crops to build soil organic matter.
8. Establish contour buffer strips, field buffers and stream buffers to protect sensitive environments within the field and in downstream environments.
9. Highly erodible areas can be planted with perennial crops or with trees, rather than annual crops.
10. Use prescribed or rotational grazing to minimize soil erosion.

Impact of long-term tillage and synthetic fertilization on soil parameters and crop yields under a corn–soybean rotation in eastern Canada

In eastern Canada, the corn (Zea mays L.)–soybean [Glycine max (L.) Merr.] rotation represents a dominant cereal–legume association. Although the economic benefits of this rotation are recognized, research is required to evaluate its long-term impact on soil quality and crop productivity under varying management alternatives such as reduced tillage or varying rates of synthetic fertilization. Indeed, the fundamental role of long-term field experiments in agronomic research and development has been recognized for decades by the international scientific community as essential for assessing the effects of management practices on soil fertility, soil quality, and crop productivity.

In 1992, a long-term research experiment was established in southern Quebec on a clay loam to evaluate the effect of tillage [mouldboard plow (MP) vs. no-till (NT)], synthetic N (0, 80, and 160 kg N ha^{-1}) and P fertilization (0, 17.5, and 35 kg P ha^{-1}) on soil functioning and crop yields of a grain corn–soybean crop sequence. Synthetic fertilizers were applied only to the corn phase of the rotation, while tillage was performed each year.

Results obtained 12 to 20 years after initiation of the study indicated that NT treatment enhanced organic C accumulation, NO_3-N, P and K availability, and microbial activity in the upper soil layer likely due to accumulation of crop residues at or near the soil surface. This stratification can benefit soil structure but might impact surface water quality over the long term. Conversely, the MP treatment was characterized by a more uniform distribution of organic C and nutrients to tillage depth and higher microbial activity near the bottom of the plow layer. Strong links were established between soil tests P (e.g. Mehlich-3 and Olsen P) and P balance, explaining how P can accumulate in time with continuous agricultural practices. Soil N_2O emissions were increased by N

fertilization in corn but not affected by tillage. Grain corn yield linearly responded to fertilizer N addition in all years but soybean yield was little affected by residual N. Except for soil P status, fertilizer P had limited impact on both crops on this site but lack of replenishment would eventually lead to soil P depletion. Overall, MP gave greater yields than CT in several years (see above Figure)—by an average of 10% in corn and 13% in soybean—and provided generally a better profitability. The lower performance of CT in this site compared with MP is probably because cooler and wetter soil conditions in spring due to surface crop residues.

***Note:** this site is still operating and other studies are under investigation. For more details see Ziadi et al. (2014).*

In the future, soil fertility managers will rely more on precision agriculture technologies to achieve the 4R Plus goals. The greatest advantage of precision agriculture, from the perspective of managing nutrients, is that it allows for the development of spatially-explicit maps of the soil and hydrological conditions in agricultural fields. The multiple layers of information in these maps permit us to evaluate individual parameters, such as soil pH or the soil potassium concentration across a field, or we can evaluate the interactions between multiple parameters (e.g., are changes in soil pH associated with fluctuations in the water table level across the field?). Rather than relying on an average soil fertility value for a large field, we can observe gradients in soil fertility that exist across the field. Knowing the zones of low, medium and high fertility allows us to adjust fertilizer application rates and meet the needs of the crop growing in a particular zone, rather than applying a uniform rate of fertilizer across the field. Soil and hydrological maps can also identify the vulnerable areas in fields where leaching or erosion are more likely to occur. Hence, the farm that adopts precision agriculture technologies can save money by applying fertilizers precisely, where they are needed, and establish buffer zones in the sensitive areas of the field and at the interface between fields and non-agricultural land. Precision agriculture technologies are well aligned with the 4R Plus goals and should be of great benefit to manage soil fertility in the future.

SUMMARY

1. Water-soluble nutrient ions in the soil pore water are absorbed through root hairs of plants.
2. Plant-available nutrients in the soil pore water can be transformed into unavailable forms through biological reactions, such as plant uptake, immobilization and reduction. Chemical precipitation and adsorption reactions also remove plant-available nutrients from the soil pore water. Physical loss of plant-available nutrients occurs when soils are leached and eroded.
3. Soil pore water is replenished with plant-available nutrients through mineralization, oxidation, weathering, dissolution and desorption reactions.
4. Soil fertility testing describes the soil nutrient supply and is required to make fertilizer recommendations. Crop nutrition is explained by plant tissue analysis, which determines whether the crop has sufficient nutrients or is deficient in nutrients required to achieve the economically optimal yield.
5. Inorganic and organic fertilizers can be applied to replenish the soil nutrient supply and improve crop nutrition. Dissolution of inorganic fertilizers releases water-soluble nutrient ions quickly, whereas organic fertilizers contain some organic compounds that must be mineralized to release water-soluble nutrient ions, making them a relatively slow-release fertilizer.
6. The 4R Nutrient Stewardship approach to soil nutrient management focuses on selecting the right source of fertilizer, to be applied at the right rate, at the right time and in the right place to get the greatest percentage of fertilizer nutrients into crop biomass. Achieving a high efficiency of fertilizer nutrient use in a cost-effective way is the goal.
7. Soil nutrients are susceptible to loss from the soil-plant system, so farms should be aware of the environmental implications of soil nutrient management.
8. The “Can you dig it?” boxes provide examples of soil fertility testing and soil nutrient improvement in Canada.

SUGGESTED READING

International Plant Nutrition Institute. 2012. 4R plant nutrition: a manual for improving the management of plant nutrition. North American version. IPNI, Norcross, GA, USA.

Havlin, J.L., J.D. Beaton, S.L. Tisdale, and W.L. Nelson. 2014. Soil fertility and fertilizers: an introduction to nutrient management. 8th ed. Pearson Prentice Hall, Upper Saddle River, NJ.

Munroe, J. (ed). 2016. Soil fertility handbook. 3rd edn. Ontario Ministry of Agriculture, Food and Rural Affairs Publication 611. Queen’s Printer for Ontario, Toronto. http://www.omafra.gov.on.ca/english/crops/pub611/pub611.pdf

Reid, K. 2014. Improving your soil: A practical guide to soil management for the serious home gardener. Firefly Books Ltd., Richmond Hill, Ontario, 272 p.

STUDY QUESTIONS

1. How do soil nutrients get absorbed by plants?
2. Complete the following table.

Nutrient	Macronutrient or Micronutrient?	Source in soil?	Role in plant nutrition
N			
P			
Ca			
S			
Cu			
Fe			

3. Give an example of a biological reaction that can deplete the amount of plant-available nitrogen in the soil pore water.
4. What is a chemical reaction that reduce the amount of plant-available potassium in the soil pore water?
5. Why does the application of a fertilizer containing NH_3 tend to acidify the soil?
6. Nitrogen is often the most limiting nutrient for agricultural crops to reach their yield potential, so is it a good idea to add excessively high rates of nitrogen fertilizer to agricultural land? Why or why not?
7. What is the best diagnostic tool to determine the fertilizer requirements of crops: soil fertility testing or plant tissue analysis?
8. Micronutrient fertilizers can be applied to crop foliage. What is the advantage of this fertilization method?
9. What are some considerations to help you choose the right fertilizer source for your crop?
10. What are suitable soil and water conservation practices to prevent nutrient loss through erosion?

REFERENCES

Abdi, D., Cade-Menun, B.J., Ziadi, N., and Parent, L.-É. 2014. Long-term impact of tillage practices and phosphorus fertilization on soil phosphorus forms as determined by 31P nuclear magnetic resonance spectroscopy. J. Environ. Qual. 43(4): 1431–1441.

Bender, R.R., J.W. Haegele, M.L. Ruffo, and F.E. Below. 2010. Nutrient uptake, partitioning, and remobilization in modern, transgenic insect-protected maize hybrids. Agron. J. 105(1): 161–170.

Carter, M.R. and Gregorich, E.G. 2008. Soil sampling and methods of analysis. 2nd ed. Canadian Society of Soil Science. CRC Press, Boca Raton, FL.

Centre de référence en agriculture et agroalimentaire du Québec. 2010. Guide de référence en fertilisation. 2e édn. Centre de référence en agriculture et agroalimentaire du Québec, Ste Foy, QC.

Chantigny, M.H., Angers, D.A., Morvan, T., and Pomar, C. 2004. Dynamics of pig slurry nitrogen in soil and plant as determined with N-15. Soil Sci. Soc. Am. J. 68(2): 637–643.

Deen, B., Lauzon, J., and Bruulsema, T. 2007. Corn response to intensive crop nutrition. Better Crops 91: 20–21. http://www.ipni.net/publication/bett...48874B7A852579 800081D5CE/$FILE/Better%20Crops%202007-2%20p20.pdf

Dessureault-Rompré, J., Zebarth, B.J., Chow, T.L., Burton, D.L., Sharifi, M., Georgallas, A., Porter, G., Moreau, G., Leclerc, Y., Arsenault, W.J., and Grant, C.A. 2011. Prediction of soil nitrogen supply in potato fields in a cool humid climate. Soil Sci. Soc. Am. J. 75(2): 626–637.

Drury, C.F., Yang, J., De Jong, R., Huffman, T., Reid, K., Yang, X., Bittman, S., and Desjardins, R. 2016. Residual Soil Nitrogen. Pages 114–120 In R.L. Clearwater, T. Martin, and T. Hoppe (Eds.) Environmental Sustainability of Canadian Agriculture: Agri-Environmental Indicators Report Series. Report #4. Ottawa, ON: Agriculture and Agri-Food Canada

Elbon A., and Whalen, J.K. 2015. Phosphorus supply to vegetable crops from vesicular arbuscular mycorrhizae: a review. Biol. Agric. Hort. 31(2): 73–90.

Eskin, N., Vessey, K., and Tian, L. 2014. Research progress and perspectives of nitrogen fixing bacterium, Gluconacetobacter diazotrophicus, in monocot plants. Int. J. Agron. 2014, 13 p. Article 208383. https://doi.org/10.1155/2014/208383

Evans, J.R. 1989. Photosynthesis and nitrogen relationships in leaves of C_3 plants. Oecologia 78(1): 9–19.

Fleury, D. 2020. 4R nitrogen management for wheat and canola. Top Crop Manager 19 August 2020 https://www.topcropmanager.com/4r-ni...at-and-canola/

Follett, R.H., Murphy, L.S., and Donahue, R.L. 1981. Fertilizers and soil amendments. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Havlin, J.L., Beaton, J.D., Tisdale, S.L., and Nelson, W.L. 2014. Soil fertility and fertilizers: an introduction to nutrient management. 8th ed. Pearson Prentice Hall, Upper Saddle River, NJ.

He, Z.Q., Honeycutt, C.W., Griffin, T.S., Cade-Menun, B.J., Pellechia, P.J., and Dou, Z.X. 2009. Phosphorus forms in conventional and organic dairy manure identified by solution and solid state P-31 NMR spectroscopy. J. Environ. Qual. 38(5): 1909–1918.

International Plant Nutrition Institute. 2012. 4R plant nutrition: a manual for improving the management of plant nutrition. North American version. IPNI, Norcross, GA, USA.

Jones, J.B., Jr. 2003. Agronomic handbook: management of crops, soils and their fertility. CRC Press, Boca Raton, FL.

Karamanos, R.E. 2008. Nitrogen and phosphorus recommendations for Manitoba for today's production costs. Presented at the 2008 Manitoba Agronomists Conference, Winnipeg, MB
http://umanitoba.ca/faculties/afs/MA...is%20Karamanos %20paper.pdf

Kour, D. Rana, K.L., Kaur, T., Yadav, N., Yadav, A.N., Kumar, M., Kumar, V., Dhaliwal, H.S., and Saxena, A.K. 2021. Biodiversity, current developments and potential biotechnological applications of phosphorus-solubilizing and -mobilizing microbes: A review. Pedosphere 31(1): 43–75.

Kumar, V., Vogelsang, L., Schmidt, R.R., Sharma, S.S., Seidel, T., and Dietz, K.-J. 2020. Remodeling of root growth under combined arsenic and hypoxia stress is linked to nutrient deprivation. Frontiers in Plant Science. 23 October 2020. https://doi.org/10.3389/fpls.2020.569687

Leytem, A.B., Smith, D.R., Applegate, T.J., and Thacker, P.A. 2006. The influence of manure phytic acid on phosphorus solubility in calcareous soils. Soil Sci. Soc. Am. J. 70(5): 1629–1638.

Marty, C., Lévesque, J.-A., Bradley, R.L., Lafond J., and Paré, M.C. 2019. Contrasting impacts of two weed species on lowbush blueberry fertilizer nitrogen uptake in a commercial field. PLoS ONE 14: e0215253.

Mehlich, A. 1984. Mehlich-3 soil test extractant: a modification of Mehlich-2 extractant. Commun. Soil Sci. Plant Anal. 15(12): 1409–1416.

Mortvedt, J.J., Cox, F.R., Shuman, L.M., and Welch, R.M. 1991. Micronutrients in agriculture. SSSA Book Series 4, American Society of Agronomy, Madison, WI.

Munroe, J. (ed). 2016. Soil fertility handbook. 3rd edn. Ontario Ministry of Agriculture, Food and Rural Affairs Publication 611. Queen's Printer for Ontario, Toronto. http://www.omafra.gov.on.ca/english/...611/pub611.pdf

Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA). 2006. Nutrient management planning. Best Management Practices Series, Guelph, ON.

Rascio, N., and La Rocca, N. 2013. Biological nitrogen fixation. Reference Module in Earth Systems and Environmental Sciences. Elsevier, Amsterdam. doi:10.1016/B978-0-12-409548-9.00685-0

Reid, R., and Hayes, J. 2003. Mechanisms and control of nutrient uptake in plants. Int. Rev. Cytol. 29: 73–114. doi: 10.1016/s0074-7696(03)29003-3.

Ruiz Herrera, L.F., Shane, M.W., and López-Bucio, J. 2015. Nutritional regulation of root development. WIREs Dev. Biol. 4: 431–443. doi: 10.1002/wdev.183

Sims, J.T., Maguire, R.O., Leytem, A.B., Gartley, K.L., and Pautler, M.C.. 2002. Evaluation of Mehlich 3 as an agri-environmental soil phosphorus test for the Mid-Atlantic United States of America. Soil Sci. Soc. Am. J. 66(6): 2016–2032.

Stumborg, C., Schoenau, J.J., and Malhi, S.S. 2007. Nitrogen balance and accumulation pattern in three contrasting prairie soils receiving repeated applications of liquid swine and solid cattle manure. Nutr. Cycl. Agroecosyst. 78: 15–25.

Westerman, R.L. 1990. Soil testing and plant analysis. 3rd edn. SSSA Book Series 3, American Society of Agronomy, Madison, WI.

Whalen, J.K. 2014. Managing soil biota-mediated decomposition and nutrient mineralization in sustainable agroecosystems. Advances in Agriculture 2014: 13 p. Article 384604. doi: 10.1155/2014/384604

Whalen, J.K., Thomas, B.W. and Sharifi, M. 2019. Novel practices and smart technologies to maximize the nitrogen fertilizer value of manure for crop production in cold humid temperate regions. Adv. Agron. 153: 1–85.

Zebarth, B.J., Leclerc, Y., Moreau, G., Gareau, R., and Milburn, P.H. 2003. Soil inorganic nitrogen content in commercial potato fields in New Brunswick. Can. J. Soil Sci. 83(4): 425–429.

Ziadi, N., Angers, D.A., Gagnon, B., Lalande, R., Morel, C., Rochette, P., and Chantigny, M.H. 2014. Long-term tillage and synthetic fertilization affect soil functioning and crop yields in a corn–soybean rotation in eastern Canada. Can. J. Soil Sci. 94 (3): 365–376.

Ziadi, N., Simard, R.R., Allard, G., and Lafond, J. 1999. Field evaluation of anion exchange membranes as a N soil testing method for grasslands. Can. J. Soil Sci. 79(2): 281–294.

About the Authors

Joann K. Whalen, James McGill Professor at McGill University and an Adjunct Professor at Gansu Agricultural University and with the Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences



Joann Whalen (licensed under a CC BY-NC-ND license)

Joann received her B.Sc. (Agr.) from Dalhousie University, M.Sc. from McGill University, and Ph.D. from Ohio State University (USA). She is a professional agronomist (agronome) in Quebec, Canada. Her research focuses on soil fertility and soil ecology in agroecosystems. She has published more than 230 peer-reviewed scientific publications and supervised/co-supervised more than 70 students at the M.Sc., Ph.D. and postdoctoral levels. She teaches courses in soil fertility, nutrient management planning and environmental soil chemistry.

Joann is a Chief Editor for Soil Biology and Biochemistry, an academic journal that publishes original, scientifically challenging research articles of international significance that describe and explain biological processes occurring in soil. She is a Fellow of the Soil Science Society of America and of the Canadian Society of Soil Science. Joann is the first international scientist to be elected for a 3-year term as President of the American Society of Agronomy (2022-2024). She grew up in a farming community near Sussex, New Brunswick and has always been interested in the practical aspects of growing food. In her opinion, soil fertility specialists are the most broadly-trained agricultural scientists because they have to muster knowledge from so many fields – physics, hydrology, chemistry, biochemistry, plant biology, microbiology and genetics – to understand the complex and dynamic soil processes responsible for crop growth. To make sense of soil responses, we have to dig deeper, which comes quite naturally to a soil scientist!

Noura Ziadi, Research Scientist, Agriculture and Agri-Food Canada; Adjunct Professor *Université* Laval, the Graduate University of Chinese Academy of Science, *Université du Québec en Abitibi-Témiscamingue*



Noura Ziadi (licensed under a CC BY-NC-ND license)

Noura is a research scientist with Agriculture and Agri-Food Canada in Quebec City. She holds a B.Sc. (Agronomy) from Tunisia and a M.Sc. and a Ph.D. (Soil Fertility and Plant Nutrition) from Université Laval, Quebec, Canada. Noura is also an adjunct professor at the *Université* Laval (2005-), the Graduate University of Chinese Academy of Science (2012-), and the *Université du Québec en Abitibi-Témiscamingue (2016-)*.

Noura has a distinguished career in soil science especially soil fertility and plant nutrition and she's conducting her research at the national and international scales (France, China, Switzerland, Finland, Tunisia, Iran, Saudi Arabia) within multi-disciplinary teams including government and university researchers and industry partners. Noura published more than 155 papers, five reviews and 11 book chapters including 37 papers with the Canadian Journal of Soil Science (CJSS). Noura joined the CSSS in 1994 and has served as Eastern Councillor (2008-2010), President-Elect (2017), President (2018) and Past-President (2019). Noura is an associate editor (AE) for the CJSS since 2011, has handled 107 manuscripts for this journal and was been recognized as a best AE in 2017. Noura has been recognized for her contributions to soil science and to the CSSS with a Fellow Award in 2015.

Jeff J. Schoenau, Professor and Saskatchewan Ministry of Agriculture Research Chair – Nutrient Cycling & Management, Department of Soil Science, University of Soil Science



Jeff Schoenau (licensed under a CC BY-NC-ND license)

Jeff is a professor of soil fertility and professional agrologist who works in the Department of Soil Science at the University of Saskatchewan. He holds the Saskatchewan Ministry of Agriculture Soil Nutrient Management Chair in the College of Agriculture and Bioresources, and is a fellow of the Agricultural Institute of Canada. He was born in Saskatchewan, completed his undergraduate and graduate degrees in the 1980's in the College of Agriculture at the University of Saskatchewan and has worked there since. He also farms with his spouse Lynne near Central Butte, Saskatchewan. His research, teaching and extension activities deal with soil fertility and fertilizers, nutrient cycling, and soil management practices in prairie cropping systems.

Maxime C. Paré, Professor, Université du Québec à Chicoutimi



Maxime Paré (licensed under a CC BY-NC-ND license)

I teach soil science to undergraduate students, most of whom are enrolled in the biology program at UQAC. I teach soil science as an introductory class, and we look at soil chemistry, biology, nutrient cycling, and pedology. It is always interesting how initially most students view soil science negatively—I usually do a small survey during the first class of the semester—and how their views change over the semester as the students realize the importance of soil for biology, life, and the general well-being of the planet. I

also include soil science in most of my field teaching activities. I like to show the links between spheres, such as how what we see when we look up, e.g., the traits of plants and trees, strongly influences what we find when we look down, i.e., in the soil, and vice versa. Finally, soil science allows me to do research and support many graduate students across a variety of research disciplines, from various agricultural domains (wild blueberry, haskap, potatoes, cash crops, etc.) to forestry and the environmental sciences. How ever could I be more stimulated?

David L. Burton, Professor, Department of Plant, Food, and Environmental Sciences, Faculty of Agriculture, Dalhousie University



David Burton (licensed under a CC BY-NC-ND license)

Dr. David Burton is a Soil Scientist and Professor in the Department of Plant, Food, and Environment, Dalhousie University. His research examines the role of the soil environment in influencing the nature and extent of microbial metabolism in soil. His focus has been on processes in the cycling of nitrogen in soils and their implications for soil fertility and environmental impact. It is the aim of this work to better understand the factors that control soil microbial function and to use this information to developing resilient land management systems in response to a changing climate.

Tom Bruulsema, Chief Scientist with IPNI Canada



Tom Bruulsema (licensed under a CC BY-NC-ND license)

Since July 2019, Tom Bruulsema is Chief Scientist with Plant Nutrition Canada, providing support for the nutrient stewardship programs of Fertilizer Canada, The Fertilizer Institute, and the International Fertilizer Association. Prior to the dissolution of the larger IPNI, he was vice-president for the Institute's programs in the Americas, and research globally. From 2015 to 2017 he led the Institute's Phosphorus Program. Dr. Bruulsema has been recognized as a Fellow of the American Society of Agronomy, the Soil Science Society of America, and the Canadian Society of Agronomy. He was previously a research associate at the University of Minnesota (1994), and an agronomist with the Mennonite Central Committee in Bangladesh (1986-1990).

This page titled 4.12: Soil Nutrient Cycling is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Canadian Society of Soil Science via source content that was edited to the style and standards of the LibreTexts platform.

SECTION OVERVIEW

4.13: Weathering, Erosion, and Mass Movement



Figure 4.13.1: Landslide in Venezuela (Courtesy USGS)

Movement of earth material can be so slow that it is imperceptible to the human eye, or move at tremendous speeds, covering or destroying all in its path. Here you will investigate the processes that mobilize earth materials and the surface changes that result from weathering, erosion, and mass movement.

Learning Objectives

By the end of this chapter you should be able to:

- Compare and contrast chemical and physical weathering.
- Explain how mass movement occurs and the landform features created by it.
- Explain how splash, sheet, rill and gully erosion occurs.

4.13.1: Weathering

4.13.2: Mass Movement

4.13.3: Water Erosion

4.13.4: Review and Additional Resources

This page titled 4.13: Weathering, Erosion, and Mass Movement is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Michael E. Ritter (The Physical Environment) via source content that was edited to the style and standards of the LibreTexts platform.

4.13.1: Weathering

Slopes

A **slope** is an inclined ground surface. Most slopes are mantled with unconsolidated regolith, the product of weathering. Regolith serves as parent material for soil above and grades downward into unaltered bedrock below. Loose regolith serves as a source of **colluvium**, sediment that has been eroded, transported and deposited down slope. In order to do so, the erosional forces must overcome the forces of resistance: friction, inertia (the resistance to movement), and particle cohesion.

The land surface constantly responds to endogenic and exogenic forces that shape the Earth. Being an open system, a slope seeks a state of balance between the forces of resistance and those of change. The surface attains a state of **dynamic equilibrium**, a condition where the system constantly adjusts to processes that raise the surface (e.g. tectonic uplift) and those that wear it down (e.g. erosion by water). The system fluctuates around a stable average state unless the driving forces of change exceed a geomorphic threshold, such as a massive earthquake, tsunami, volcanism. Once this occurs, a period of system adjustment occurs, finally ending in a new stable state.

Weathering

Weathering is the breakdown and decomposition of earth material, namely rocks. Weathering is an important mechanism to destabilize surface materials for their eventual removal by erosive processes. Weathering of rock-forming minerals can create new products from pre-existing rocks. The physical disintegration of rocks affects soil development and texture. Weathering releases chemical compounds that become available for biological processes. The weathering of carbonate minerals releases carbon to the atmosphere which impacts atmospheric chemistry and temperature. And the list goes on. Weathering, needless to say, is an important environmental process that bridges all elements of our physical environment and sustains the notion of a changing Earth.



Figure 4.13.1.1: Talus slopes created by physical weathering. (Courtesy USGS DDS21)

Weathering occurs in two ways. **Physical weathering**, also called mechanical weathering, involves the disintegration of rock materials. Physical weathering incurs no change in the chemistry of the material being altered. Instead, it simply breaks large pieces into smaller ones. **Chemical weathering** involves the decomposition of rocks and sediment. In this case, a chemical change occurs and a new product is created from the material that has undergone weathering. Weathering processes are determined by the climate and vegetation of a place. Dry locations tend to be dominated by physical weathering and moist places by chemical weathering.

Physical weathering

The result of physical weathering is to simply make smaller pieces out of larger ones. In so doing, physical weathering makes it easier for surface materials to chemically decompose and be eroded. When a large block of material is broken into smaller pieces additional surface area for chemical weathering to act is exposed. Examine Figure 4.13.1.2 below. On the left is a block whose length, width and height is equal to one centimeter. This means that the volume of the block is equal to 1 cubic centimeter and the total surface area is equal to 6 square centimeters. If we split the block in length and width wise (dashed lines) we create eight smaller pieces, each with a height, width, and depth of .5 centimeters. Though the volume of all eight pieces taken together is still 1 cubic centimeter, the total amount of surface area is greatly increased to 12 cm^2.

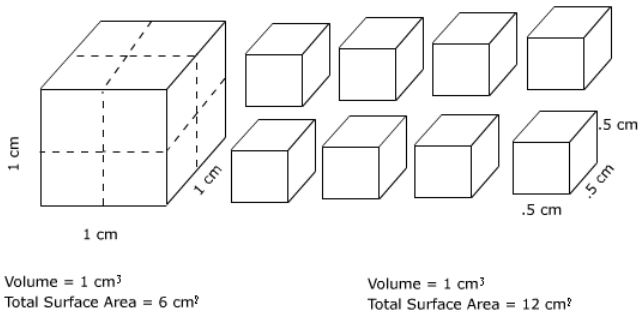


Figure 4.13.1.2: Effect of particle size and surface area.

Physical weathering processes

There are a number of physical weathering processes that break earth materials apart, a very common one is called ***root wedging***. Plant roots work their way into rock crevices called ***joints***. As they grow, roots create pressure on the sides of the crack enlarging it until the rock breaks apart. This is a common problem for home owners where trees are grow too close to a house. Tree roots can force their way into the foundation, breaking it apart, and let water seep into the owner's basement.



Figure 4.13.1.3: Tree roots growing into bedrock. (Courtesy USGS DDS21)

Frost wedging occurs when water freezes in rock fractures. As the water freezes it expands putting pressure on the sides of the crack, enlarging it until the rock breaks apart. ***Thermal expansion and contraction*** can weaken rock and cause it to disintegrate. In deserts, surface materials get exceedingly hot during the day and be exposed to cold temperatures at night. The expansion upon heating and contraction during cooling weakens rock breaking it apart. ***Alternate wetting and drying*** causes material to expand and contract, thus weakening rocks and inducing them to break as well. Regardless of process, the result is a mass of unconsolidated material.

Chemical weathering

The minerals in rocks formed beneath the surface are in equilibrium with the temperature and pressure conditions at time of their formation and thus are quite stable. However, many minerals are no longer in equilibrium with their environmental conditions when exposed at the surface and are susceptible to weathering. ***Chemical weathering*** results in the formation and retention of minerals in equilibrium with environmental conditions at the Earth's surface. The least stable minerals in igneous and metamorphic rocks are olivine and plagioclase, the most stable is quartz.

The interlocking and spacing of mineral grains controls the tendency towards weathering. Rocks with loosely interlocking mineral grains allow agents of chemical weathering to penetrate, thus speeding their decomposition. Limestone is primarily composed of calcite, a mineral that is quite soluble under surface conditions and easily dissolves in humid environments. In dry regions, the tight texture of limestone prevents it from disintegration and thus is a relatively resistant rock when found in deserts.

The process of chemical weathering tends to:

- increase bulk creating stress within rocks
- lower the density minerals
- decreased particle size resulting in increased surface area

- creates more mobile materials
- creates more stable minerals

Chemical weathering processes

Oxidation takes place when oxygen reacts with earth materials. Oxygen dissolved in water combines with atoms of metallic elements abundant in silicate minerals. Attacking metals in the soil, oxidation causes them to rust leaving the soil a brownish red to red color. When oxygen combines with iron, the reddish iron oxide *hematite* (Fe_2O_3) is formed:

$$4Fe^{+3} + 3O_2 \rightarrow 2Fe_2O_3$$

Hydrolysis is an exchange reaction involving minerals and water. Free hydrogen (H^+) and hydroxide $(OH)^-$ ions in water are able to replace mineral ions and drive them into solution. As a result, the mineral's atomic structure is changed into a new form. It is a process whereby silicate minerals like potassium feldspar are weathered and a clay mineral is formed.

$$2KAlSi_3O_8 + 2H^+ + 9H_2O \rightarrow Al_2Si_2O_5(OH)_4 + 4H_4SiO_4 + 2K^{2+}$$

Hydration involves the absorption of water like which occurs during the conversion of hematite to limonite:

$$2Fe_2O_3 + 3H_20 \rightarrow 2Fe_2O_3 \cdot 3H_20$$

Some geoscientists question whether hydration is a true chemical weathering process because the process is readily reversible and the new product is not chemically different from its precursor. Some would rather call hydration a physical weathering process.

Carbonic acid action involves combination of carbon dioxide and water. Though present in pure water, carbon dioxide dissolved in water provides ions that produces free hydrogen. Carbon dioxide in the atmosphere combines with rain water to form carbonic acid (H_2CO_3):

$$H_2O + CO_2 \rightarrow H_2CO_3$$

Though weak, when carbonic acid is combined with a mineral like calcite ($CaCO_3$) common to limestone, calcium and bicarbonate ions are released and carried off by groundwater.

$$CaCO_3 + H_2CO_3 \rightarrow Ca^{+2} + 2HCO_{-3}$$

Rock resistance to weathering

Rocks react differently to weathering due to the differences in mineral content and structure. Some minerals are unstable under surface conditions and are readily soluble. Others are stable and resist the agents of weathering. Some rock-forming minerals are physically soft, being easily crushed and split while harder minerals are less easily broken apart. The arrangement and size of mineral grains control weathering processes. Water has a difficult time penetrating intricately locked and closely spaced mineral grains to promote weathering. Larger, loosely cemented minerals disintegrate and decompose more readily. Minerals in the form of poorly joined sheets readily break apart.

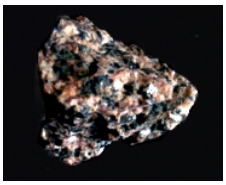


Figure 4.13.1.4: Granite, a predominate rock type that composes the continental land masses.

Granite is a coarse grained rock composed of quartz and feldspar. Both quartz and feldspar are hard minerals, but feldspar is less stable under surface conditions than quartz. The feldspar readily weathers to become clay in humid conditions. The feldspar will weather in dry climates as the somewhat porous granite allows moisture to penetrate. As the feldspar decomposes it weakens the bonds holding the rock together and it disintegrates

Karst Landscapes

The chemical weathering of carbonate-rich rocks creates a unique landscape abounding in caves, disappearing streams, and springs. **Karst**, a Yugoslavian term that comes from a narrow strip of limestone plateau noted for the assemblage of solution landforms. Karst develops in regions underlain by limestone and to a lesser extent dolomite. Chemical solution of the limestone, especially

when fractured, wears away the bedrock leaving fissures and possibly undermining the surface. Some of the most spectacular cave complexes are found in such areas. Significant karst regions are found in Jamaica, the northern Yucatan, New South Wales, northern Puerto Rico, the Great Valley Region of Pennsylvania, Maryland, Virginia, and Tennessee, central Kentucky and central Florida in the United States



Figure 4.13.1.5: Stalactites hang from the ceiling and stalagmites grow from the floor of Carlsbad Caverns, NM (Photo Credit: National Park Service)

Generally four conditions are important for karst development. First, there must be limestone at or near the surface. Though karst is found in areas underlain by dolomite, it usually far less soluble than limestone. Second, the limestone must be dense, highly jointed, and thinly bedded. If the rock is too porous, water will be rapidly absorbed throughout the entire mass and not be concentrated along restricted flow lines. Third, the existence of entrenched valleys below uplands underlain by soluble and well jointed rocks ensures downward movement of groundwater, favorable for the development of karst. Finally, at least moderate precipitation must fall in the region. Few arid or semi-arid regions exhibit karst features, though some relic features may exist from a previous moist period in the past.

Karst Landforms

One of the most common features of areas underlain by limestone are sinkholes. ***Sinkholes*** form either from beneath the surface or from the surface down. Collapse sinkholes form when the limestone is dissolved away from below, removing the support for the surface and it collapses in. Collapse sinks can be particularly dangerous because one might not know that the support for the surface under their feet is being slowly eaten way. Sinkholes can also be created where water infiltrates into the surface widen the fissure into which it flows. These sinkholes are called **dolines**.

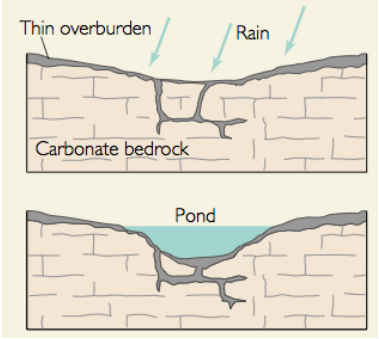


Figure 4.13.1.6: Sinkholes from dissolution (Courtesy USGS)

Sinkholes are commonly funnel-shaped and broadly open upward. They may be a few feet to more than 100 feet in depth, though usually ranging from 10 to 30 feet. Sinkhole diameter sizes range from a few square yards to several acres in area.

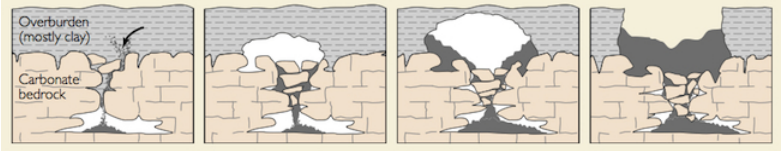


Figure 4.13.1.7: Collapse sinkhole (Courtesy USGS)

Streams flowing along the surface may enter a sinkhole as a "**disappearing stream**" and flow underground for some distance to reappear at the surface. A **karst window** forms when the roof above the underground stream collapses in. **Cavern systems** are

carved by these underground streams over time.



Figure 4.13.1.8: Sinkhole, Cape Breton, Canada (Image courtesy Geological Survey Canada)

This page titled 4.13.1: Weathering is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Michael E. Ritter (The Physical Environment) via source content that was edited to the style and standards of the LibreTexts platform.

- **17.1: Weathering** by Michael E. Ritter is licensed CC BY-SA 4.0. Original source: https://www.thephysicalenvironment.com.

4.13.2: Mass Movement

Mass movement is the down slope movement of earth materials under the influence of gravity. The detachment and movement of earth materials occurs if the stress imposed is greater than the strength of the material to hold it in place. ***Shear strength*** is a measure if the resistance of earth materials to be moved. The interlocking of soil particles increases the ability of material to stay in place. Plant roots also help bind soil particles together. ***Shear stress*** is primarily a function of the force exerted by the weight of the material under the influence of gravity acting in the down slope direction. The slope of the surface determines the amount of stress that occurs on earth materials. Water destabilizes hill slopes by creating pressure in the pore spaces of earth materials. Water infiltrating into slope materials saturates the soil particles at depth by filling the pore spaces between. The weight of water lying above creates water pressure that drives soil particles apart. This lessens the friction between them and enables them to slip past one another. Material is mobilized when the shear stress imposed on a surface exceeds the shear strength. The movement, especially in the case of slides and slumps, is along a failure plane. The failure plane may be a well-defined layer of clay or rock upon which sets the destabilized surface material. Humans induce mass movement when subjecting a slope to a load that exceeds its ability to resist movement. People building houses on scenic hill slopes often find their homes threatened by a landslide. Undercutting of hillsides during road construction commonly creates unstable slopes making them prone to failure.

Types of mass movement

Soil creep is nearly imperceptible to the naked eye as it is the slowest of all types of mass movement. Soil creep generally occurs in the top few meters of the surface and is accomplished by expansion and contraction of the soil. For instance, when water in the soil freezes the ice pushes soil particles outward perpendicular to the slope. Upon warming, the ice melts and the soil is pulled down slope under the influence of gravity. Over many freeze-thaw cycles soil moves slowly down slope. In many cases one might not be able to tell that soil creep is occurring by just examining the surface. However, trees growing on surfaces undergoing creep will have curved trunks or roots that are curved. Broken retaining walls and curved railroad tracks also indicate creep in action.

A ***slide*** is a sheet of material that slips over a failure plane ending anywhere from a meter to a kilometer down slope. Slides produce concave scars while slumps tend to produce a scarp or cliff exposure. Trees are broken and bent and the slide can bury the soil down slope. Digging into buried soils and analyzing their contents can tell us about the age and what the environment was like when the slide occurred.



Figure 4.13.2.1: La Conchita, CA 2005 landslide (Courtesy USGS; Source)

La Conchita, California has experienced devastating landslides in recent years. Unstable slopes mobilized by rain water caused a landslide and debris flow seen in Figure 4.13.2.1. The city lies on a narrow strip of coast 250m (800ft) wide between the shoreline and a 180 m (600ft) bluff above it. Extraordinary rains and rising groundwater levels caused the slope to fail, fortunately no one was killed. However, in 2005, another year of abnormally high rain fall caused the slope to fail again, this time burying structures and killing 10 people.



Video: "Landslides" (Courtesy National Geographic)

Slumps are characterized by a backward rotation of the earth material as it moves along a curved failure plane resulting in a reverse slope. Slumps take place as an intermittent movement of earth or rock material, often as several independent units, creating a number of step-like "terracettes." Undercutting of slopes by stream erosion, waves, and road building are common causes of slumping.



Figure 4.13.2.2: Slump on hill side. Notice the step-like appearance of terracettes (Courtesy USGS DDS21) (Click image to enlarge)

Solifluction is the down slope movement of soil over a permanently frozen subsurface. Solifluction is common on slopes underlain by permafrost. During the summer when the upper permafrost is activated, the waterlogged soil mass slowly moves down slope to form solifluction lobes or terraces.



Figure 4.13.2.3: Solifluction terraces, Niwot Ridge, Colorado

A **flow** is the down slope movement of water-saturated soil, regolith, weak shale, or weak clay layers. *Earth flows* are fairly slow, occurring over a few hours or so slow that they are almost imperceptible. Earth flows are accompanied with slumping, but unlike slumping, there is no backward rotation. Earth flows differ from mudflows in that they (1) tend to be slower, (2) are not confined to channels, (3) are more common in humid areas than dry, and (4) have a lower water content.



Figure 4.13.2.4: Mudflow, Pacific Palisades, CA. (Courtesy USGS DDS21) Click image to enlarge

A *mudflow* is the rapid down slope movement of water-saturated water- saturated soil, regolith. The higher water content creates a flow rapid enough to be perceptible to the eye. Conditions favorable for the development of mudflows are: (1) unconsolidated surface materials, (2) steep slopes abundant but intermittent precipitation, and (3) sparse cover of vegetation. Mudflows tend to be more prevalent in dry regions where vegetation is sparse and heavy rains may form. When set in motion, they occupy stream-cut channels rushing along in a torrential flow of mud.



Figure 4.13.2.5: Talus slope, Isabelle Valley, Colorado

Rock fall is one of the most sudden forms of mass movement. Rock fall occurs when blocks of rock shed from a cliff face and collect at the base. **Talus** is a term that is applied to an accumulation of rock by rock fall.



Video:"Riding the Storm - Landslide Danger in San Francisco" (Courtesy USGS)

This page titled 4.13.2: Mass Movement is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Michael E. Ritter (The Physical Environment) via source content that was edited to the style and standards of the LibreTexts platform.

- **17.2: Mass Movement** by Michael E. Ritter is licensed CC BY-SA 4.0. Original source: https://www.thephysicalenvironment.com.

4.13.3: Water Erosion

Erosion is the detachment of earth material from the surface. Once detached, agents like water or wind transport the material to a new location where it is deposited. The most ubiquitous form of erosion is that done by water.

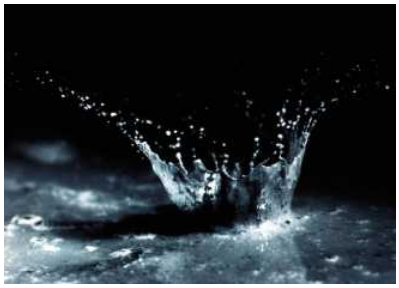


Figure 4.13.3.1: Rain drop impact causing splash erosion (Image courtesy NRCS)

Rain splash erosion is caused by the impact of water striking the surface. Rain splash erosion generally takes place in two steps. As precipitation is absorbed by the surface it fills the pore spaces, loosening soil particles and driving them apart. The impact of subsequent rain drops hitting the surface splash the particle away from the point of impact. The effect is to give the surface a dimpled-like appearance.



Figure 4.13.3.2: Severe sheet erosion on a field (Image courtesy NRCS)

Surface runoff forms when the rainfall intensity of a storm exceeds the infiltration capacity of the soil. ***Sheet erosion*** is caused by the unconfined flow of water running across the surface. The effects of sheet erosion are often hard to distinguish because such thin layers of soil are being removed. It isn't until several years later that significant degradation is perceived.



Figure 4.13.3.3: NRCS personnel inspecting rill erosion on a field (Image courtesy NRCS)

Rill erosion is caused by water concentrating into innumerable, closely-spaced small channels. Left unchecked, rills can cut vertically and horizontally and when joined form gullies.



Figure 4.13.3.4: Severe gully erosion on a field in Iowa (Image courtesy NRCS)

Gullies are steep-sided trenches formed by the coalescence of many rills. Once started they are difficult to stop.

This page titled 4.13.3: Water Erosion is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Michael E. Ritter (The Physical Environment) via source content that was edited to the style and standards of the LibreTexts platform.

- **17.3: Water Erosion** by Michael E. Ritter is licensed CC BY-SA 4.0. Original source: https://www.thephysicalenvironment.com.

4.13.4: Review and Additional Resources

Review

Review and assess your learning. Start with the "Important Terms and Concepts" to ensure you know the terminology related to the topic of the module and concepts discussed. Move on to the "Review Questions" to answer critical thinking questions about concepts and processes discussed in the module. Finally, test your overall understanding by taking the "Self-assessment quiz".

Important Terms and Concepts

- **weathering**
  the breakdown and decomposition of earth material, namely rocks
- **physical weathering**
  involves the disintegration of rock materials; breaks large pieces into smaller ones
- **chemical weathering**
  decomposition of rocks and sediment.a chemical change occurs and a new product is created from the material that has undergone weathering
- **root wedging**
  Plant roots work their way into joints; as they grow, roots create pressure on the sides of the crack enlarging it until the rock breaks apart
- **joints**
  rock crevices
- **frost wedging**
  occurs when water freezes in rock fractures. As the water freezes it expands putting pressure on the sides of the crack, enlarging it until the rock breaks apart
- **thermal expansion and contraction**
  can weaken rock and cause it to disintegrate; expansion upon heating and contraction during cooling weakens rock breaking it apart
- **regolith**
  the product of weathering
- **oxidation**
  takes place when oxygen reacts with earth materials. Oxygen dissolved in water combines with atoms of metallic elements abundant in silicate minerals.
- **hydrolysis**
  an exchange reaction involving minerals and water. Free hydrogen (H^+) and hydroxide $(OH)^-$ ions in water are able to replace mineral ions and drive them into solution
- **karst**
  a Yugoslavian term that comes from a narrow strip of limestone plateau noted for the assemblage of solution landforms
- **sinkhole**
  form either from beneath the surface or from the surface down

- **karst window**
  forms when the roof above the underground stream collapses in
- **mass movement**
  the down slope movement of earth materials under the influence of gravity
- **shear stress**
  primarily a function of the force exerted by the weight of the material under the influence of gravity acting in the down slope direction
- **shear strength**
  a measure of the resistance of earth materials to being moved
- **creep**
  occurs in the top few meters of the surface and is accomplished by expansion and contraction of the soil; slowest of all types of mass movement
- **slide**
  a sheet of material that slips over a failure plane ending anywhere from a meter to a kilometer down slope
- **slump**
  characterized by a backward rotation of the earth material as it moves along a curved failure plane resulting in a reverse slope
- **earth flow**
  fairly slow, occurring over a few hours or so slow that they are almost imperceptible. Accompanied with slumping; no backward rotation; are not confined to channels; are more common in humid areas
- **mud flow**
  is the rapid down slope movement of water-saturated water- saturated soil, regolith. The higher water content creates a flow rapid enough to be perceptible to the eye.
- **rock fall**
  occurs when blocks of rock shed from a cliff face and collect at the base
- **erosion**
  the detachment of earth material from the surface
- **rain splash erosion**
  caused by the impact of water striking the surface
- **sheet erosion**
  caused by the unconfined flow of water running across the surface
- **rill erosion**
  caused by water concentrating into innumerable, closely-spaced small channels
- **gully erosion**
  steep-sided trenches formed by the coalescence of many rills

Review Questions 4.13.4.1

What is the difference between physical and chemical weathering?

Answer

Physical weathering processes are those that break rocks into smaller pieces (disintegration). Chemical weathering chemically alters rock-forming minerals into a product (decomposition).

List the effects of chemical weathering on minerals and rocks.

Answer

- increase bulk creating stress within rocks
- lower the density minerals
- decreased particle size resulting in increased surface area
- creates more mobile materials
- creates more stable minerals

Compare and contrast hydration and hydrolysis as it relates to rock weathering.

Answer

Hydration and hydrolysis involve the absorption of water into the molecular structure of a mineral. Hydration is a reversible process and hydrolysis is not.

What is soil creep and what features of the landscape can be used to tell creep has taken place?

Answer

Soil creep is the slow down slope movement of earth material. The process is imperceptible to the human eye. Terracettes, broken retaining walls, curved tree roots, deformed railroad tracks result from creep.

Compare and contrast slumps and slides.

Answer

Slumps take place on well-defined failure planes, exhibit rotation of the failed mass, and often form terraces. Slides occur along irregular failure planes with the mass sliding down slope burying the overlying surface with debris.

Explain the process of mass movement.

Answer

Mass movement is the down slope movement of earth material under the influence of gravity. Movement occurs when the shearing stress imposed on material is greater than the shearing strength.

What does erosion mean?

Answer

Erosion is the detachment of material from the surface.

Describe the process of rain drop erosion.

Answer

Rain drop erosion occurs in a series of steps. Initial rain drops soak the surface driving soil particles apart. Subsequent rain drops detach soil particles upon impact. A dimpled surface often results from rain splash erosion.

What is particularly harmful about sheet erosion?

Answer

Sheet erosion occurs when the unconfined flow of water across the surface strips thin layers of top soil. It can be imperceptible to the human eye.

Compare and contrast rill and gully erosion.

Answer

Rill erosion occurs when moving water is confined to small channels. A gulley is a much larger channel formed from the coalescence of rills

Self-Assessment Quiz 4.13.4.1

1. Backward rotation of a mass of earth material when mobilized is typical of
   A. slides
   B. slumps
   C. rock fall
   D. creep
2. Another term that describes chemical weathering is
   A. disintegration
   B. decomposition
   C. rotting
   D. fragmentation
3. Curved tree trunks, fractured retaining walls, curved roots are typical features of
   A. slides
   B. slumps
   C. rock fall
   D. creep
4. Talus is a deposit associated with
   A. slides
   B. slumps
   C. rock fall
   D. creep
5. Physical weathering
   A. is typical of warm and wet environments
   B. results in new mineral products being formed
   C. increases surface area
   D. all the above
6. Solifluction is a mass movement process in
   A. the rain forest
   B. the desert
   C. the tundra
   D. the steppe
7. _____ erosion occurs when water is confined to small channels.
   A. rain splash
   B. sheet
   C. rill
   D. gully
8. Sheet erosion

A. occurs as channeled flow
B. in some cases is nearly imperceptible
C. occurs when the infiltration capacity of the soil exceeds the rainfall intensity
D. all the above

9. Karst landscapes develop over _____ bedrock

A. sandstone
B. limestone
C. granite
D. basalt

10. Conditions favorable for mudflows include

A. unconsolidated surface materials
B. steep slopes abundant but intermittent precipitation
C. sparse cover of vegetation
D. all the above

Answer

1. B
2. B
3. D
4. C
5. C
6. C
7. C
8. B
9. B
10. D

Additional Resources

Use these resources to further explore the world of geography

Focus on The Physical Environment: "Killer Landslides" (NOVA PBS). (29:00)

Connections: When Land Slides (*NASA Earth Observatory*)

Physical Geography Today: "Real-Time" Monitoring of an Active Landslide above Highway 50, California - USGS

Readings

When Land Slides (NASA EOS)

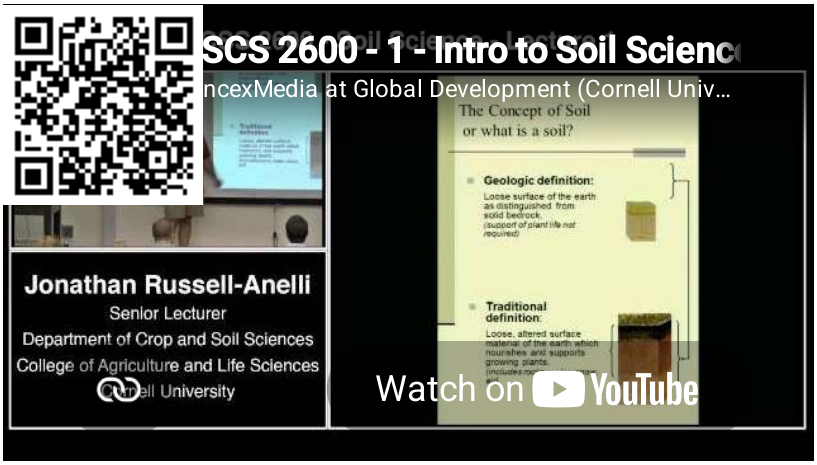
This page titled 4.13.4: Review and Additional Resources is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Michael E. Ritter (The Physical Environment) via source content that was edited to the style and standards of the LibreTexts platform.

- **17.4: Review and Additional Resources** by Michael E. Ritter is licensed CC BY-SA 4.0. Original source: https://www.thephysicalenvironment.com.

4.14: Summary



Under Construction



4.14: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

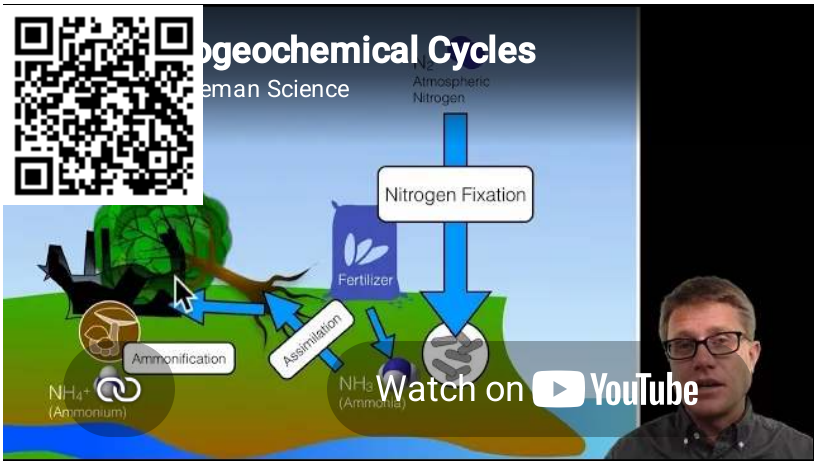
5: The Carbon Cycle of Terrestrial Ecosystems

- 5.1: Introduction
- 5.2: Photosynthesis
- 5.3: Respiration
- 5.4: Net Primary Production
  - 5.4.1: Marine Primary Production
    - 5.4.1.1: Primary production
    - 5.4.1.2: The Producers
    - 5.4.1.3: Factors Influencing Production
    - 5.4.1.4: Patterns of Primary Production
  - 5.4.2: Surface Primary Production
- 5.5: Net ecosystem production and Eddy-covariance studies
- 5.6: Remote sensing of primary production and biomass
- 5.7: Global estimates of net primary production and biomass
- 5.8: The fate of net primary production
  - 5.8.1: Introduction to and Components of Food Webs
  - 5.8.2: Quantifying Food Webs
  - 5.8.3: Trophic Cascades
  - 5.8.4: Sources and Attributions
- 5.9: Net primary production and global change
- 5.10: Detritus
- 5.11: Soil organic matter and global change
- 5.12: Summary

5: The Carbon Cycle of Terrestrial Ecosystems is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5.1: Introduction





5.1: Introduction is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

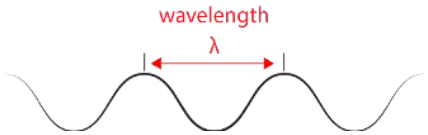
5.2: Photosynthesis

Light Energy

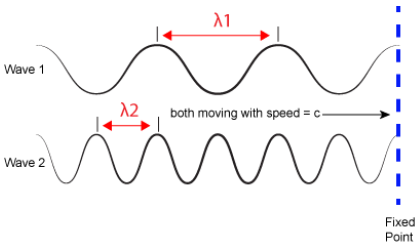
The sun emits an enormous amount of electromagnetic radiation (solar energy) that spans a broad swath of the **electromagnetic spectrum**, the range of all possible radiation frequencies. When solar radiation reaches Earth, a fraction of this energy interacts with and may be transferred to the matter on the planet. This energy transfer results in a wide variety of different phenomena from influencing weather patterns to driving a myriad of biological processes. In Bis2a we are largely concerned with the latter and we discuss some very basic concepts related light and its interaction with biology below.

First, however, we need to review a couple of key properties of light.

Light has properties of waves. A specific "color" of light has a characteristic wavelength.

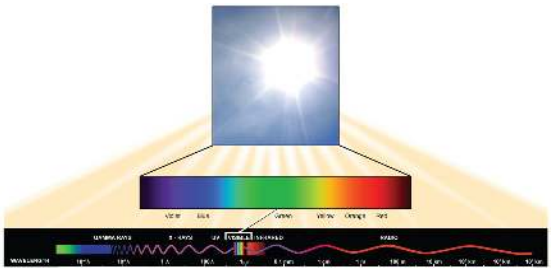


The distance between peaks in a wave is referred to as the wavelength and is abbreviated with the Greek letter lambda (λ). Attribution: Marc T. Facciotti (original work)



The inverse proportionality of frequency and wavelength. Wave 1 has a wavelength that is 2X that of wave 2 (λ1 > λ2). If the two waves are traveling at the same speed (c) - imagine that both of the the whole line that are drawn are dragged past the fixed vertical line at the same speed - then the number of times a wave peak passes a fixed point is greater for wave 2 than wave 1 (f2 > f1).
Attribution: Marc T. Facciotti (original work)

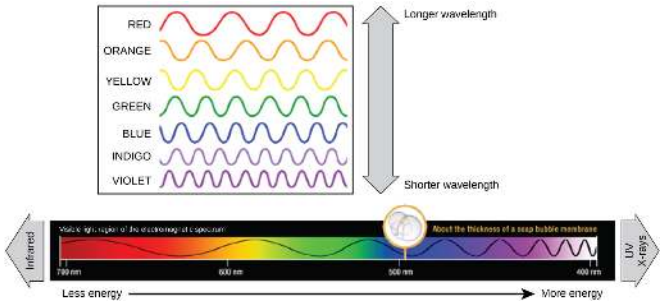
Each frequency (or wavelength) of light is associated with a specific energy. The higher the frequency (or shorter the wavelength) the more energy is associated with a specific "color". Wave 2 in the figure above has greater energy than wave 1.



The sun emits energy in the form of electromagnetic radiation. All electromagnetic radiation, including visible light, is characterized by its wavelength. The longer the wavelength, the less energy it carries. The shorter the wavelength the more energy is associated with that band of the electromagnetic spectrum.

The Light We See

The visible light seen by humans as white light is composed of a rainbow of colors, each with a characteristic wavelength. Certain objects, such as a prism or a drop of water, disperse white light to reveal the colors to the human eye. In the visible spectrum, violet and blue light have shorter (higher energy) wavelengths while the orange and red light have longer (lower energy) wavelengths.



The colors of visible light do not carry the same amount of energy. Violet has the shortest wavelength and therefore carries the most energy, whereas red has the longest wavelength and carries the least amount of energy. (credit: modification of work by NASA)

Absorption by Pigments

The interaction between light and biological systems occurs by several different mechanisms, some of which you may learn about in upper-division courses in cellular physiology or biophysical chemistry. These interactions can initiate a variety of light dependent biological processes that can be grossly grouped into two functional categories: cellular signaling and energy harvesting. Signaling interactions are largely responsible for perceiving changes in the environment (in this case changes in light). An example of a signaling interaction might be the interaction between light and the pigments expressed in an eye. By contrast, light/pigment interactions that are involved in energy harvesting are used for - not surprisingly - capturing the energy in the light and transferring it to the cell to fuel biological processes. In Bis2a we are mostly concerned about this type of interaction of light and biological pigments. Photosynthesis is one example of an energy harvesting interaction.

At the center of the biological interactions with light are groups of molecules we call organic pigments. Whether in the human retina, chloroplast thylakoid, or microbial membrane, organic pigments often have specific ranges of energy or wavelengths that they can absorb. The sensitivity of these molecules for different wavelengths of light is due to their unique chemical makeups and structures. A range of the electromagnetic spectrum is given a couple of special names because of the sensitivity of some key biological pigments: The retinal pigment in our eyes, when coupled with an opsin sensor protein, “sees” (absorbs) light predominantly between the wavelengths between of 700 nm and 400 nm. Because this range defines the physical limits of the electromagnetic spectrum that we can actually see with our eyes, we refer to this wavelength range as the "visible range". However, the sunlight reaching the Earth's surface ranges from the UV into the far infrared. Interestingly, some animals perceive wavelengths that we cannot see. Plants are also very good at perceiving light, and use this information to direct growth (out from under the shade of a taller plant, for example). However, other plant pigments absorb light in order to capture energy. Depending on the plant (or photosynthetic bacterium) the spectrum absorbed by energy-collecting pigments can range from approximately 800 nm (infrared) to approximately 300 nm (the ultraviolet). Green plants carry combinations of pigments that allow them to absorb a wide range of wavelengths, though with varying efficiencies (green plants appear green because they reflect, rather than absorb, most of the photons at a combination of wavelengths that we perceive as green).

Quick question

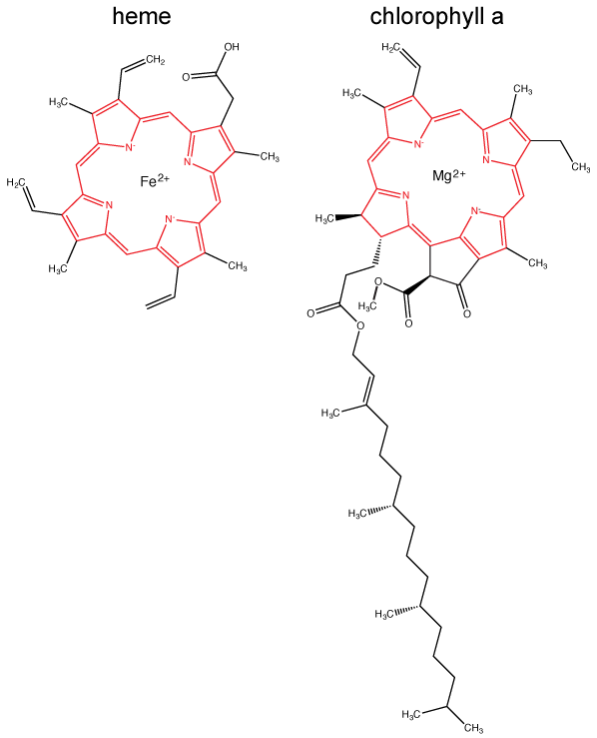
If a plant were able to absorb 100% of incident photons, what color would it be? What if a variant of the same species of plant lacked a pigment required to absorb red light?

Three Key Types of Pigments We Discuss in Bis2a

Chlorophylls

Chlorophylls (including bacteriochlorophylls) are part of a large family of pigment molecules. There are five major chlorophyll pigments named: *a*, *b*, *c*, *d*, and *f*. Chlorophyll *a* is related to a class of more ancient molecules found in bacteria called **bacteriochlorophylls**. Chlorophylls are structurally characterized by ring-like porphyrin group that coordinates a metal ion. This ring structure is chemically related to the structure of heme compounds that also coordinate a metal and are involved in oxygen

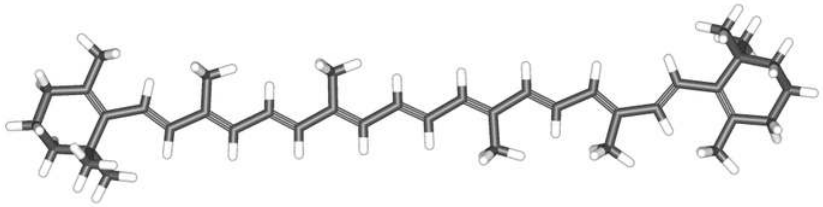
binding and/or transport in many organisms. Different chlorophylls are distinguished from one another by different "decorations"/chemical groups on the porphyrin ring.



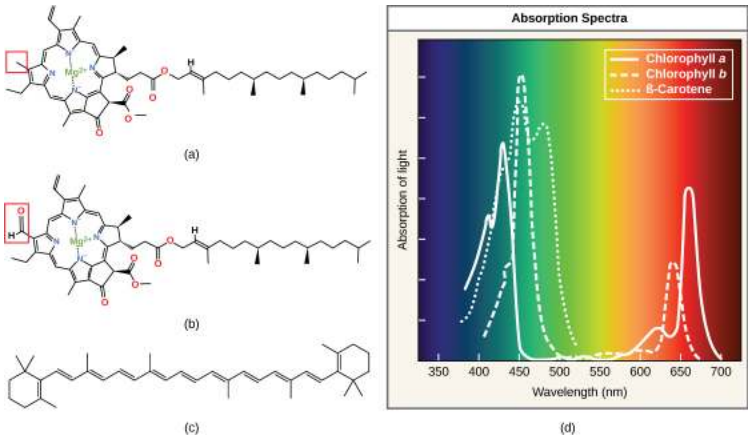
The structure of heme and chlorophyll a molecules. The common porphyrin ring is colored in red. Attribution: Marc T. Facciotti (original work)

Carotenoids

Carotenoids are the red/orange/yellow pigments found in nature. They are found in fruit — such as the red of tomato (lycopene), the yellow of corn seeds (zeaxanthin), or the orange of an orange peel (β-carotene, below) — and can be used as biological "advertisements" to attract seed dispersers (animals or insects that may carry seeds elsewhere). In photosynthesis, various carotenoids can function as either light-harvesting or protective (energy-dispersing) pigments in photosynthetic reaction centers.



Each type of pigment can be identified by the specific pattern of wavelengths it absorbs from visible light. This characteristic is known as the pigment's **absorption spectrum**. The graph in the figure below shows the absorption spectra for chlorophyll *a*, chlorophyll *b*, and a carotenoid pigment, called β-carotene. Notice how each pigment has a distinct set of peaks and troughs, revealing a highly specific pattern of absorption. These differences in absorbance are due to differences in chemical structure (some are highlighted in the figure). Chlorophyll *a* absorbs wavelengths from either end of the visible spectrum (blue and red), but not green. Because green is reflected or transmitted, chlorophyll appears green. Carotenoids absorb in the short-wavelength blue region, and reflect the longer yellow, red, and orange wavelengths.



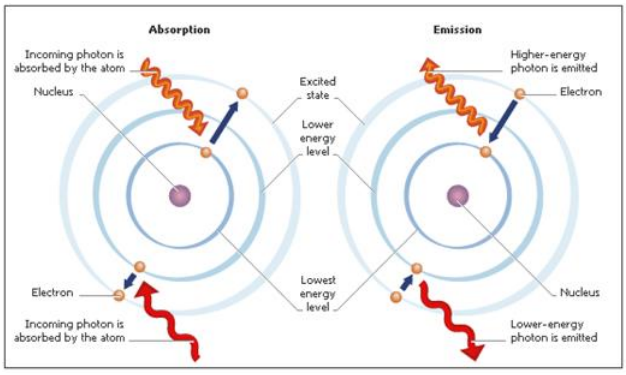
(a) Chlorophyll a, (b) chlorophyll b, and (c) β-carotene are hydrophobic organic pigments found in the thylakoid membrane. Chlorophyll a and b, which are identical except for the part indicated in the red box, are responsible for the green color of leaves. Note how the small amount of difference in chemical composition between different chlorophylls leads to different absorption spectra. β-carotene is responsible for the orange color in carrots. Each pigment has (d) a unique absorbance spectrum.

Importance of having multiple different pigments

Not all photosynthetic organisms have full access to sunlight. Some organisms grow underwater where light intensity and available wavelengths decrease and change, respectively, with depth. Other organisms grow in competition for light. Plants on the rainforest floor for instance must be able to absorb any bit of light that comes through, because the taller trees absorb most of the sunlight and scatter the remaining solar radiation. To account for these variable light conditions, many photosynthetic organisms have a mixture of pigments whose expression can be tuned to improve the organism's ability to absorb energy from a wider range of wavelengths than would be possible with one pigment alone.

What happens when an atom absorbs a photon?

When an atom absorbs a photon of light, an electron acquires that energy, leaving its ground (lowest potential energy) orbital and moving up to a higher energy orbital. This is an unstable situation.

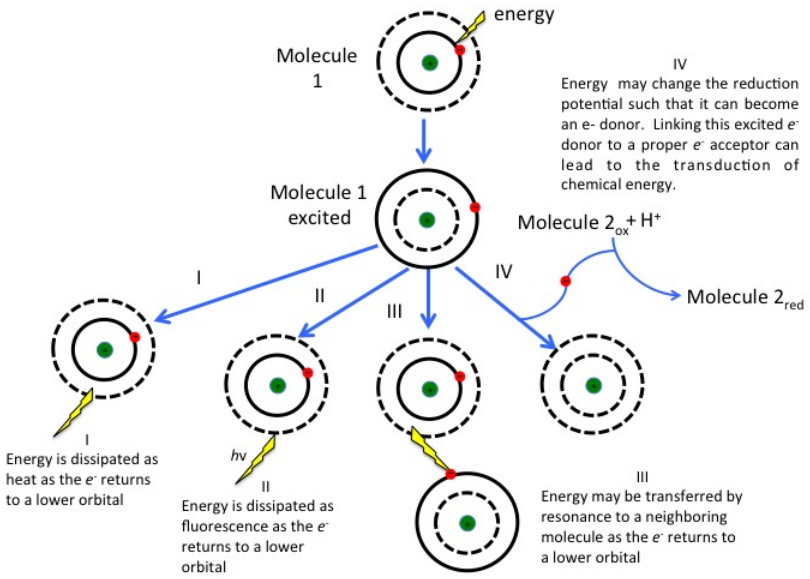


A diagram depicting what happens to a molecule that absorbs a photon of light. Please that electron hasn't left its atom, it has just moved to a higher orbital.

What are the fates of the "excited" electron? There are a few possible outcomes, which are schematically diagrammed in the figure below. These options are:

1. (I and II below) The electron can relax to a lower quantum state (back to the ground state), transferring energy as heat or light (a photon of less energy than the incident photon; the wavelength of this "fluoresced" photon is determined by the specific difference in energy between the excited state orbital and the ground state orbital).

2. (III below) The energy can be transferred by resonance to a neighboring molecule as the e^- returns to a lower quantum state. There is virtually no loss of energy in this transfer.
3. (IV below) Because the energy changed the reduction potential such that the molecule is now a stronger e^- donor, this high-energy e- can be transferred exergonically to an appropriate e^- acceptor. In other words, the excited state can be involved in a redox reactions. This is a **photochemical** reaction.



What can happen to the energy absorbed by a molecule

Photophosphorylation: an Overview

Photophosphorylation is the process of transferring the energy from light into ATP. The evolutionary roots of photophosphorylation are likely in the anaerobic world, between 3 billion and 1.5 billion years ago, when life was abundant in spite of the absence of molecular oxygen. Photophosphorylation probably evolved relatively shortly after electron transport chains and **anaerobic respiration**. The first step of the process involves the absorption of a photon by a pigment molecule. Light energy is transferred to the pigment and promotes electrons into an excited state. (Note the use of anthropomorphism here, the electrons are not suddenly hopping all over or celebrating their promotion. They are simply in a higher energy orbital. This is in contrast to the pigment in its "unexcited", or "**ground state**".) While in the excited state, the pigment has a much lower reduction potential (E°', it moves upward on our electron tower, it becomes a stronger reducing agent) and can donate these unstable, high potential energy electrons to carriers with greater E°' - carriers that the ground state pigment would not be capable of reducing. We have now gotten the ball rolling- we have a highly reducing electron that we can pass from carrier to carrier (as in respiration's electron transport chain).

A quick example: For example, as you can see in the Table below, the ground state pigment at the reaction center of PSII (the "chlorophyll a" in P680) cannot reduce anything listed in the table- it is the weakest reducing agent described there (even weaker than H_2O!). However, once it is has absorbed a photon of sufficient energy to boost P680 into its excited (*) state, it becomes a strong reducing agent, capable of reducing Pheophytin.

As electrons pass from one electron carrier to another via redox reactions, these exergonic transfers could be coupled to the endergonic transport (or "pumping") of protons across a membrane to create an electrochemical gradient (as we saw in respiration). This electrochemical gradient generates a proton motive force whose concentration gradient can then be coupled to the endergonic

production of ATP, via ATP synthase (again, just as in respiration). Electrons that have been used to generate the proton gradient are returned to an oxidized chlorophyll reaction center, restoring its missing e-.

However, the electron energized by light might have an alternative fate: it might descend through a different series of carriers (without pumping protons) and instead be deposited onto a close relative of NAD^+ called NADP. Addition of 2 e- generates NADPH, which is going to be used to build sugars from CO_2. Since the e- is not recycled to a pigment, the original oxidized pigment must regain an electron from somewhere else. This electron must come from an external source with a lower reduction potential than the (ground state) pigment and depending on the reduction potential of that pigment there are different possible sources that might be employed, including H_2O, reduced sulfur compounds such as SH_2, and even elemental S^0.

This is all very complicated, and was introduced rather quickly, so this provides you with an ideal opportunity to attempt to sketch this process, just to see what you do and don't know. Write down your "issues". This may be very challenging. While you read the text below, you can see how good your guess was. It's always easier to learn when you know what exactly you don't know!

If your drawing is done, let's go over this more specifically and carefully.

TABLE OF REDUCTION POTENTIALS

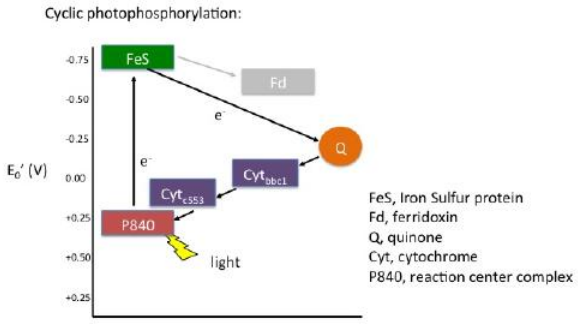
Reduction half-reaction	E'o (V)
[ ] / [ ] *reduced or oxidized?* *reduced or oxidized?*	
Phylloquinone + e⁻ / Phylloquinone	-0.8
SO_4^{2-} + 2e⁻ / HSO^-	−0.52
Pheophytin + e⁻ / Pheophytin	-0.50
Bacteriorhodopsin⁻ + e⁻ / Bacteriorhodopsin (ground state)	0.48
CO_2 + 24e⁻ / $C_6H_{12}O_6$ (glucose)	0.43
$2H^+$ + 2e⁻ / H_2	−0.42
NAD^+ + $2H^+$ + 2e⁻ / NADH + H^+	0.32
NADP⁺ + H⁺ + e⁻ / NADPH	−0.32
CO_2 + 8e⁻ / $C_3H_3O_2^-$ (acetate)	0.28
S^0 + 2e⁻ / H_2S	−0.28
CO_2 + 8e⁻ / CH_4 (methane)	−0.24
C_2H_4O (acetaldehyde) + 2H⁺ + 2e⁻ / C_2H_6O (ethanol)	0.197
$C_3H_3O_3^-$ (pyruvate) + 2e⁻ / $C_3H_4O_3$ (lactate or lactic acid)	−0.19
$C_4H_2O_4$ (fumarate) + 2e⁻ / $C_4H_4O_4$ (succinate)	+0.03
Cytochrome b_{ferric} + e⁻	+0.035
Ubiquinone$_{ox}$ + 2 e⁻	+0.11
Au^{3+} (aqueous) + 3e⁻ / Au (solid)	+0.17
Fe^{3+} + e⁻ / Fe^{2+} (pH 7)	+0.2
Cytochrome c_{ferric} + e⁻	+0.25
Bacteriochlorophyll* + e⁻ / Bacteriochlorophyll (ground state)	+0.33
Chlorophyll + e⁻ / Chlorophyll (ground state) PSI	+0.375
NO_3^- + 2e⁻ / NO_2^-	+0.42
Fe^{3+} + e⁻ / Fe^{2+} (pH 2)	+0.76
$½O_2$ + 2H⁺ + 2e⁻ / H_2O	+0.82
Chlorophyll a^+ + e⁻ / Chlorophyll a (ground state) PSII	+1.110

Table: Electron tower that includes some components of the ETC associated with photophosphorylation. PSI and PSII refer to Photosystems I (containing P700) and II (containing P680) of the oxygenic photophosphorylation pathways. Note the reducing potential of the activated forms of PSI and PSII are hard to describe accurately, as they are so short-lived. However, we know that PSI can reduce Phylloquinone^{ox}, and that PSII* can reduce Pheophytin^{ox}, so we do know the upper limit of their reduction potential.*

Cyclic Photophosphorylation (an example)

Although Green sulfur bacteria live in the absence of visible light, they can perform photosynthesis using remarkably low-energy infrared photons. In the cyclic photophosphorylation depicted below, bacteriochlorophyll$_{red}$ at the reaction center (called P840, because it has an absorption maximum at a wavelength of 840 nm) absorbs a photon, energizing an electron to the extent that it can now be transferred to the next electron carrier in an electron transport chain (here an iron-sulfur protein). In an electron transport

chain analogous to that of respiration, the electron is passed exergonically from carrier to carrier. The redox reaction at one of the carriers powers a proton pump, pushing protons into a higher concentration compartment. Eventually the electron is used to reduce bacteriochlorophyll$_{ox}$ (making a complete loop) and the whole process can start again. This flow of electrons is cyclic and is therefore said to drive "cyclic photophosphorylation". The "phosphorylation" occurs when ATP is produced from ADP when the highly concentrated protons are allowed to reenter the cytosol via ATP synthase (as in respiration).



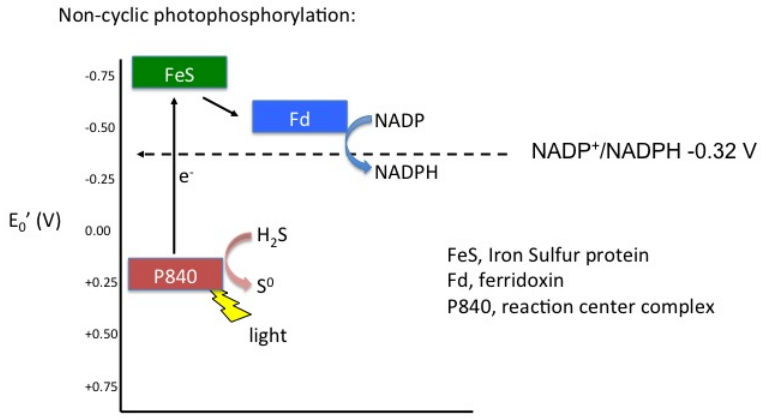
Cyclic electron flow. The reaction center P840 absorbs light energy and becomes excited, (the excited form is not illustrated here- where would you place it on this axis?). The excited electron is ejected and used to reduce an FeS protein leaving an oxidized reaction center. The electron is transferred to a quinone, then to a series of cytochromes which in turn reduce the P840 reaction center. The process is cyclical. Note the gray arrow coming from the FeS protein going to a ferridoxin (Fd), also in gray. This represents an alternative pathway the electron can take and will be discussed below in non-cyclic photophosphorylation.

Suggested discussion

The figure of cyclic photophosphorylation above depicts the flow of electrons in a respiratory chain. How does this process help generate ATP? The schematic in the figure above demonstrates how cyclic electron flow works, but does not include P840* (excited state P840). Nor does it illustrate proton pumping (by the cytochrome bc1 complex) or the synthesis of ATP. Try drawing a schematic that includes these processes also. You might want to draw two figures- one with an E°' axis (as drawn above), and another that shows the locations of these activities in the bacterial inner membrane. As always, try doing this without looking to other resources first. This will help you pinpoint any knowledge "potholes"- make a list of those, and then fix them.

Non-cyclic photophosphorylation (another example from Green sulfur bacteria)

In cyclic photophosphorylation electrons cycle from reduced bacteriochlorophyll through a series of electron carriers and eventually back to oxidized bacteriochlorophyll. There is, theoretically, no net loss of electrons; they stay in the system. In non-cyclic photophosphorylation electrons are removed from the photosystem, as they end up in NADPH after moving through an electron transport chain (which does not pump protons- there's only so much you can ask from the energy of a single photon). This NADPH will be used for carbon fixation. Thus in order to keep generating NADPH there's needs to be a source of electrons to refill the "electron hole" in bacteriochlorophyll. This source must a lower (= higher up on our chart) reduction potential than bacteriochlorophyll itself. An electron tower is provided above so you can see what compounds might *potentially* be used to reduce the oxidized form of bacteriochlorophyll. You can see that water is not an option as a source of electrons for P840ox, as water is too weak a reducing agent to donate an electron to that molecule. Whether an organism can actually employ a particular reducing agent will depend on what catalysts (enzymes) it can produce, as well as the energetics of the metabolites. Green sulfur bacteria have evolved to live in H_2S rich environments, and can utilize H_2S as an electron donor. Green plants, in contrast, can't do this. Thus for photosynthesis to occur, green sulfur bacteria must find this metabolizable source of electrons.



Non-cyclic electron flow. In this example, the P840 reaction center absorbs light energy and becomes energized, the emitted electron reduces an FeS protein and in turn reduces ferredoxin. Reduced ferredoxin (Fd_{red}) can now reduce NADP to form NADPH. The electrons are now removed from the system as NADPH, finding their way to carbon fixation or other anabolic reactions. The electrons need to be replaced on P840, which requires an external electron donor. In this case, H_2S can serve as the external electron donor.

It should be noted that for these bacterial photophosphorylation pathways, for each electron donated from a reaction center the resulting output from that electron transport chain is either the formation of NADPH (which requires 2 electrons) or ATP, not both. In other words, there are two possible paths that the electrons could take "downstream" of P840*. The cell will choose to run one or the other pathway depending on its immediate needs.

Oxygenic Photophosphorylation

Generation of NADPH and ATP

The overall function of "light-dependent" reactions of photosynthesis is to transform solar energy into chemical compounds, in the form of NADPH and ATP. This energy supports the "light-independent" reactions and fuels the assembly of sugar molecules. As illustrated above for green sulfur bacteria, protein complexes and pigment molecules will work together to produce NADPH and ATP. To quickly review what we discovered above, the absorbance of a photon makes all the difference. Normally, a good oxidizing agent ($P840_{ox}$, capable of oxidizing H_2S) would, in its reduced form, be a relatively poor reducing agent (incapable of reducing $NADP^+$). The energy added by absorbance of a photon- even, for green sulfur bacteria, a relatively weak infrared photon- turns this weak reducing agent into a strong one, at least until that energy is lost. The temporary addition of energy to $P840_{red}$ makes it a good electron donor, loss of that energy (with the electron) allows it to be a good oxidizing agent, and the net result is that electrons are transferred from H_2S to $NADP^+$, provided light is available.

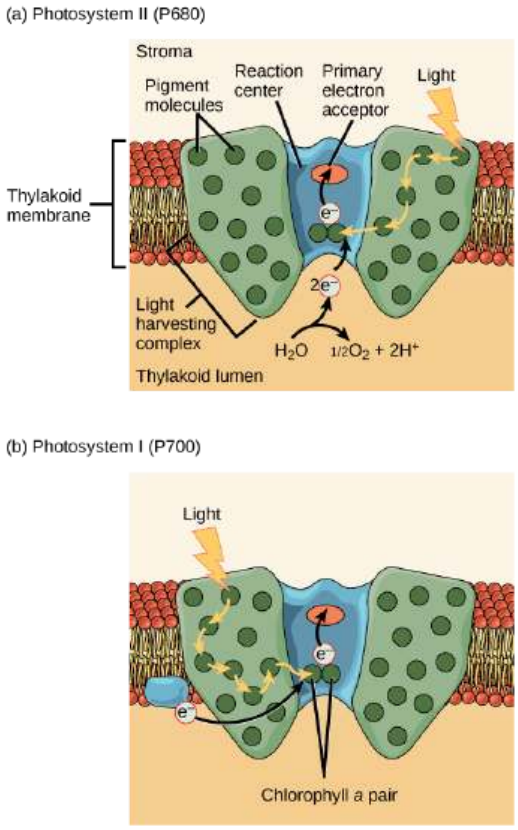
As you will see, **oxygenic photosynthesis** is more complex than the "sulfur-genic" photosynthesis described above, requiring two different reaction centers, with different reduction potentials. While this complexity is inconvenient for students (and instructors) it is extremely convenient for the bacteria and plants that employ it as they can derive electrons from an unlimited supply: water. The complexity of oxygenic photosynthesis was the only evolutionary solution **available to the Grand Design Challenge**: "What biological pigment can both oxidize water and, when activated by light, reduce $NADP^+$?" As you'll see, Nature's answer was "no can do- why not use two?".

Parts of the discussion below should sound extremely familiar if you have followed the discussion above. Try to see of you can determine where there are differences between photosynthesis as performed by green sulfur bacteria and as performed by cyanobacteria and green plants.

Suggested discussion

Step back a little. Why is it a reasonable goal to want to make NADPH and ATP? In the discussion of glycolysis and the TCA cycle the goal was also to make ATP and NADH. What is the key difference between these two processes (no, it's not the P).

Why is it a convenient to be able pull electrons from water to make NADPH?



A photosystem consists of a light-harvesting complex and a reaction center. Pigments in the light-harvesting complex pass light energy to two special chlorophyll a molecules in the reaction center. The light excites an electron from the chlorophyll a pair, which passes to the primary electron acceptor in the ETC. The excited electron must then be replaced. In (a) photosystem II, the electron is replaced by the splitting of water, which releases oxygen as a waste product. In (b) photosystem I, the electron comes from the photosynthetic electron transport chain discussed below.

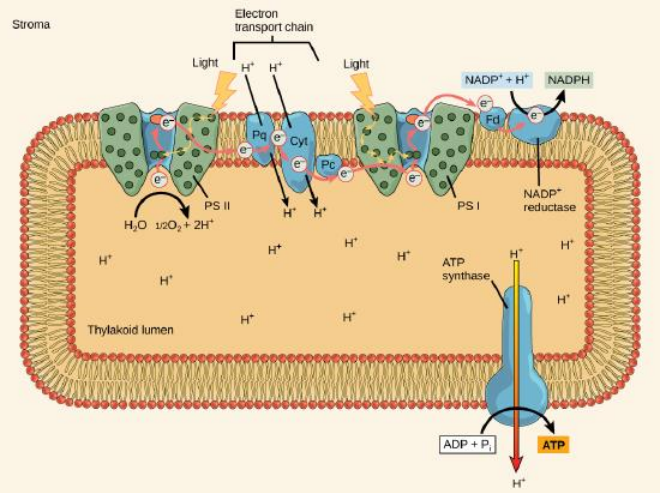
As in the green sulfur bacteria example above, the step that transfers light energy into the biomolecule takes place in a multiprotein, multipigment complex called a **photosystem.** In the green sulfur bacterium, an electron from a single type of photosystem, carrying the pigment P840, could be used to power either the formation of NADPH or the formation of ATP, depending on ther outhe of the electron transport chain, which is regulated by the needs of the bacterium.

In oxygenic photosynthesis, two types of pigment are found embedded in the thylakoid membrane (in plants) or the bacterial inner membrane (in cyanobacteria). These two types are simply called **photosystem II** (PSII, carrying P680) and **photosystem I** (PSI, carrying P700), and were named (confusingly) in the order of their discovery. The two complexes differ on the basis of what they oxidize (that is, their source of electrons) and what they reduce (the place to which they deliver their energized electrons). Working in tandem, these two photosystems can power the production of both NADPH and ATP.

Both photosystems have the same basic structure; a number of **antenna proteins** to which chlorophyll molecules are bound surround the **reaction center** where the photochemistry takes place. Each photosystem is serviced by this **light-harvesting**

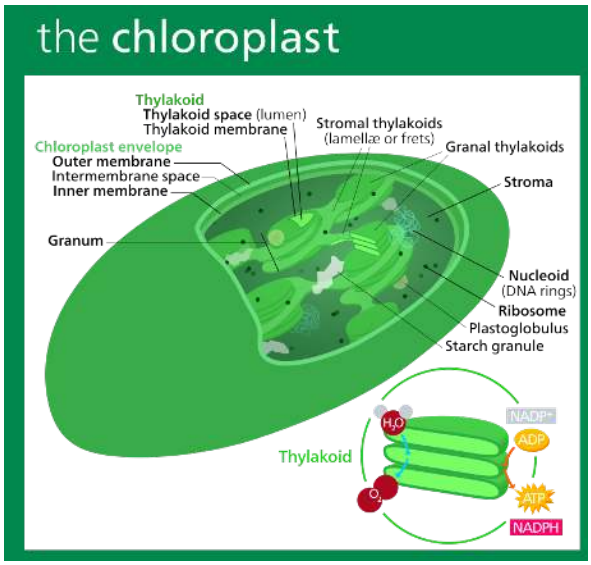
complex, which passes energy from sunlight to the reaction center; it consists of multiple antenna proteins that contain a mixture of 300–400 chlorophyll *a* and *b* molecules as well as other pigments like carotenoids. The absorption of a single **photon** by any of the chlorophylls pushes that molecule into an excited state. In short, the light energy has now been captured by biological molecules but is extremely unstable and not yet stored in any useful form. The captured energy is transferred from chlorophyll to chlorophyll until ...eventually... (after about a millionth of a second), it is delivered to the reaction center. Up to this point, only *energy* has been transferred between molecules, not *electrons*. In other words, no new bonds have formed.

Recall/consider: Both the reaction center and the antennae pigments contain chlorophyll a. What is the essential difference between and antennae chlorophyll a and reaction center chlorophyll a?



*In oxygenic photosynthesis, energy from sunlight is used to extract electrons from water. This example is from green plants, and depicts a **thylakoid**, which is located in a **chloroplast**, which is a subcellular organelle in green plants (see Figure below). Electrons from activated PSII travel through the thylakoid membrane's electron transport chain, providing energy for proton pumping into the thylakoid lumen (interior space). The electrons from PSII* eventually reach and reduce PSI_{ox}. When the now-reduced PSI is energized by light, its energized (more reducing) e- is employed to generate NADPH. While the electron transport chain moves protons across the thylakoid membrane into the lumen, the oxidation of water also adds protons to the lumen, and reduction of NADP removes protons from the stroma (exterior of the thylakoids). The net result is a low pH (high [H+]) in the thylakoid lumen, and a high pH in the stroma. ATP synthase uses this electrochemical gradient to make ATP.*

The reaction center contains a pair of chlorophyll *a* molecules with a special property. Those two chlorophylls are located close to an oxidizing agent and so can undergo oxidation (upon excitation). It is at this step that light energy is transformed into the more stable form of chemical energy. All of the subsequent redox reactions are involved in pumping protons or in delivering that e- to NADP. Once reduced, this "reducing power" carrier, as well that the newly constructed ATP, will participate in the Calvin cycle where the electron can be deposited onto CO_2 for long-term storage in the form of a carbohydrate. The Calvin cycle takes place in the stroma, conveniently located *vis a vis* the ATP- and NADPH-producing thylakoids



The **chloroplast** (image by Kelvinsong) looks like a bacterium but is actually an intracellular organelle descended from cyanobacteria. It has an outer membrane and and inner membrane, which enclose the smaller stacks of membrane-bound thylakoids. The photosynthetic pigments and ATP synthetic apparatus are embedded in the thylakoid membranes; carbon fixation (the Calvin Cycle) occurs in the stroma: the aqueous compartment that surrounds the thylakoids.

The electron transport chain

Refer to the diagram below of the "Z scheme" (you will have to turn it sideways to see the "Z"). If you have already drawn your own diagram, fantastic- you can check it against this one.

The reaction center of PSII (called **P680**) delivers its high-energy electrons, one at a time, to a **primary electron acceptor**, and these pass through the electron transport chain (Plastoquinones to cytochrome complex to plastocyanin) to the oxidized chlorophyll in $P700_{ox}$. Along the way, the cytochrome complex pumps protons across the thylakoid membrane (into the lumen) to generate a proton gradient, the energy of which will eventually be harvested for ATP synthesis.

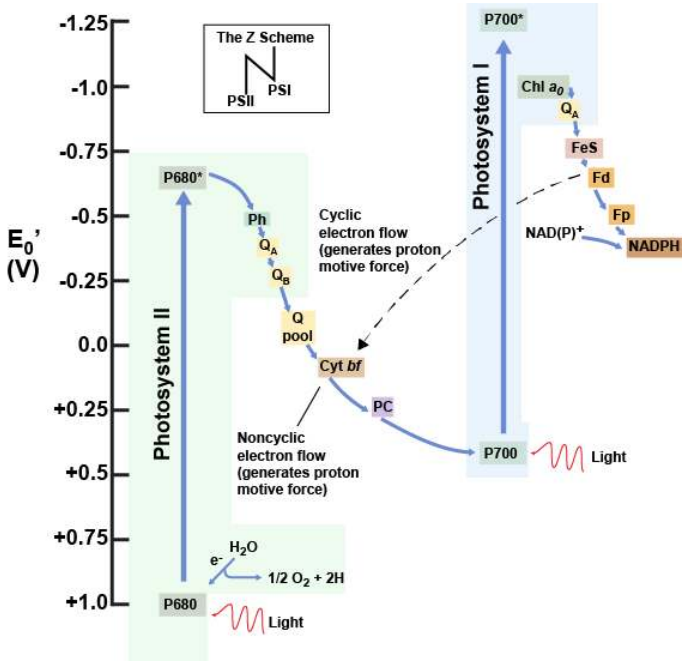
P680ox's missing electron is replaced by extracting an electron from water. Splitting one H_2O molecule releases two electrons, two hydrogen atoms, and one atom of oxygen. Splitting two molecules of water is required to form one molecule of diatomic O_2 gas. For this reason, the water-splitting apparatus of PSII carries a manganese core that gradually releases electrons, one by one, to refill the electron "hole" in P680. PSII becomes repeatedly oxidized, the Mn atoms continuously "reload" the pigment. When these metal ions are depleted of a total of 4 electrons, they become oxidizing enough to actually rip electrons from 2 molecules of water, releasing 4 protons and an O_2. In plants, about 10 percent of that O_2 is used by mitochondria in the leaf to support oxidative phosphorylation (respiration). The remainder escapes to the atmosphere as waste, where it is picked up by aerobic organisms to support their own aerobic respiration.

As electrons move through the proteins that reside between PSII and PSI, they take part in exergonic redox transfers. The free energy associated with the exergonic redox reaction is coupled to the endergonic transport of protons from the stromal side of the membrane to the thylakoid lumen. Those hydrogen ions, plus the ones produced by splitting water, accumulate in the thylakoid lumen and create a proton motive force that will be used to drive the synthesis of ATP in a later step.

The electrons, reaching the end of that electron transport chain are then deposited on $PSI(P700)^{ox}$, in their ground state. But there's no rest for the weary- they are then re-energized by light, reloaded onto another ETC, and are sufficiently energetic to be donated to $NADP^+$ at the terminus of that chain. As described above for green sulfur bacteria, the NAPDH is employed for other reductive reactions. Therefore, to complete this process another electron is delivered to PSI via the ETC that originates with PSII. That energy is transferred to the PSI reaction center (called **P700**). P700 is oxidized and sends an electron through several intermediate redox steps to $NADP^+$ to form NADPH. Thus, PSII captures the energy in light and couples its transfer via redox reactions to the creation of a proton gradient. As already noted, the exergonic and controlled relaxation of this gradient can be coupled to the

synthesis of ATP. PSI captures energy in light and couples that, through a series of redox reactions, to reduce $NADP^+$ into NADPH. The two photosystems work in concert to simultaneously produce ATP and NADPH. However, if the cell needs more ATP vs. NADPH, photophosphorylation can be decoupled from NADPH production by routing PSI's electron through the proton-pumping cyt bf complex, rather than directing it to NADP.

What if: the electron from the PSII to PSI ETC arrived at PSI and found that the molecule was already reduced?



A diagram depicting the flow of electrons and the reduction potentials of their carriers in oxygenic photosynthetic systems expressing both photosystem I (contains P700) and photosystem II (contains P680). The dotted line indicates a cyclic alternative route, which does not produce NADPH, but does pump protons.
Attribution: Marc T. Facciotti (own work)

Suggested discussion

A major difference between the green sulfur bacterium light-dependent reactions and the green plant's "Z scheme" light-dependent reactions is that the noncyclic green sulfur bacterium's system only makes NADPH, while the Z scheme's noncyclic light reactions make *both* NADP and ATP. Discuss what the differences there might be between the two systems, that allow the green plant system to make both types of energy-rich molecules?

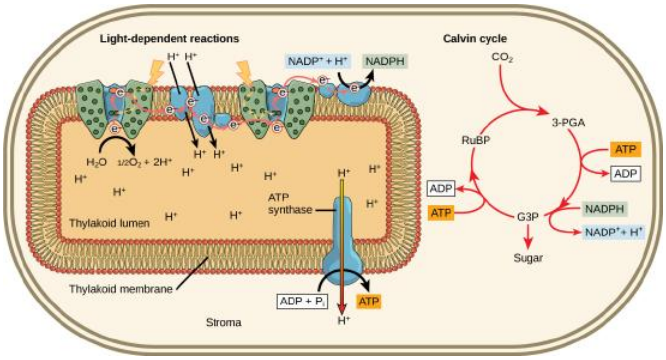
Light Independent Reactions and Carbon Fixation

A short introduction

The general principle of carbon fixation is that some cells under certain conditions can take CO_2 and reduce it to a usable cellular form. Most of us are aware that green plants can take up CO_2 and produce O_2 in during photosynthesis. We have just discussed the ability of a cell to transfer light energy onto chemicals and ultimately to produce the energy carriers ATP and NADPH in a process known as the light reactions. In photosynthesis, the plant cells use the ATP and NADPH formed during photophosphorylation to reduce CO_2 to sugar, (as we will see, specifically G3P, an intermediate in glycolysis also) in what are called the dark reactions. While we appreciate that this process happens in green plants, photosynthesis had its evolutionary origins in the bacterial world. In this module we will go over the general reactions of the Calvin Cycle, a reductive pathway that incorporates CO_2 into cellular material.

In photosynthetic bacteria such as Cyanobacteria and purple non-sulfur bacteria, as well as in plants, the energy (ATP) and **reducing power** (NADPH) obtained from the light reactions is coupled to "**Carbon Fixation**", the incorporation of inorganic carbon (CO_2) into organic molecules; initially as glyceraldehyde-3-phosphate (G3P) and eventually into glucose. Organisms that can obtain all of their required carbon from an inorganic source (CO_2) are referred to as **autotrophs**, while those organisms that require organic sources of carbon, such as glucose or amino acids, are referred to as **heterotrophs**. The biological pathway that leads to carbon fixation in green plants and cyanobacteria is called the **Calvin Cycle** and is a reductive pathway (consumes energy and electrons) which leads to the reduction of CO_2 to G3P. There are at least five other pathways for carbon fixation, used by other autotrophic prokaryotes.

The Calvin Cycle: the reduction of CO_2 to Glyceraldehyde 3-Phosphate



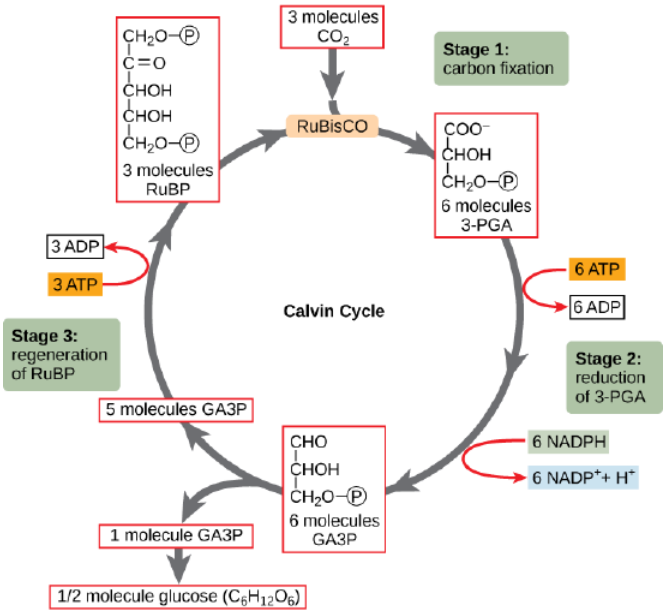
Light reactions harness energy from the sun to produce chemical bonds, ATP, and NADPH. A stylized chloroplast is shown above. These energy-carrying molecules are made in the stroma where carbon fixation (here the Calvin Cycle, drawn in an extremely simplified form on the right) takes place.

In plant cells, the Calvin cycle takes place in the stroma (outside of the thylakoids) of the chloroplasts. While the process is similar in cyanobacteria, there are no specific organelles that house the Calvin Cycle and the reactions occur in the cytoplasm around a complex intracellular membrane system derived from the plasma membrane. There is strong evidence that supports the hypothesis that the origin (actually, several independent origins) of *chloroplasts* was from a symbiosis between cyanobacteria and nonphotosynthetic cells.

> *Note that in Dr. Britt's class you will not have to memorize the Calvin cycle; you only need to know what goes in, the first step, and what comes out. Dr. Britt suggests this video, which may lack style, but it's clear on the basics.*

Stage 1: Carbon Fixation

In the stroma of plant chloroplasts, in addition to CO_2,two other components are present to initiate the light-independent reactions: an enzyme called ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), and ribulose bisphosphate (RuBP), as shown in the figure below. Ribulose-1,5-bisphosphate (RuBP) is composed of five carbon atoms and includes two phosphates.



The Calvin cycle has three stages. In stage 1, the enzyme RuBisCO incorporates carbon dioxide into an organic molecule, 3-PGA. In stage 2, the organic molecule is reduced using electrons supplied by NADPH. In stage 3, RuBP, the molecule that starts the cycle, is regenerated so that the cycle can continue. Only one carbon dioxide molecule is incorporated at a time, so the cycle must be completed three times to produce a single three-carbon GA3P molecule, and six times to produce a six-carbon glucose molecule. Here the Calvin cycle is presented in a (slightly less) symplified form.

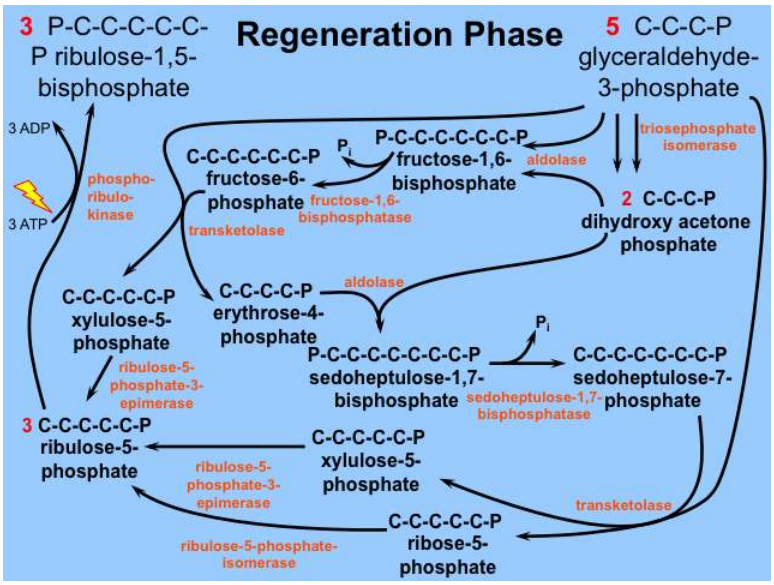
RuBisCO (ribulose bis phosphate carboxylase) catalyzes a reaction between CO_2 and RuBP. For each CO_2 molecule that reacts with one RuBP, a 6-Carbon molecules is formed. The resulting 6 C molecule will immediately split to form two molecules of another 3-carbon compound (3-PGA). Note that we have added only one C (Carbon) to an existing 5 C chain to make these 2 3-C molecules! Thus we will have to run this initial phase three times to "bleed off" a 3-C and regenerate the 3 RuBP's that were employed to capture each CO_2.

Stage 2: Reduction

ATP and NADPH are used to convert the six molecules of 3-PGA into six molecules of a 3C molecule: glyceraldehyde 3-phosphate (G3P) - a carbon compound that you might or might not remember from glycolysis. Six molecules of both ATP and NADPH are used up in the process, helping to drive the reactions and produce the electrons required to reduce the incoming CO_2. The "spent" molecules (ADP and $NADP^+$) return to the nearby thylakoids to be recycled back into ATP and NADPH.

Stage 3: Regeneration

Remember that we have consumed 3 molecules of RuBP to bleed off one G3P. Thus we need to somehow regenerate these 3 RuBP from the remaining 5 G3Ps. This regeneration phase might be politely described as "interesting", or more colloquially as "A Frigging Nightmare". Three more molecules of ATP are used in these regeneration reactions. It is included for your enjoyment below. Again, Dr. Britt will only quiz you on what goes into the Calvin cycle, its first step, and what comes out.



Additional Links of Interest

Khan Academy Links

- Calvin Cycle

Chemwiki links

- Calvin Cycle

YouTube Videos

- 3D animation of photosynthesis in plants

- Calvin Cycle

5.2: Photosynthesis is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Anne Britt, UC Davis.

- **1.16: Photosynthesis** has no license indicated.

Respiration Overview

In the next few lectures we start to learn about the process of respiration and the roles that electron transport chains play in this process. A definition of the word "respiration" that most people are familiar with is "the act of breathing". When we breath, air- including molecular oxygen- is brought into our lungs from outside of the body, the oxygen then becomes reduced, and waste products, including the reduced oxygen in the form of water, are exhaled. Simply put, some reactant comes into the organism and then gets reduced and leaves the body as a waste product. This generic idea can be generally applied across biology and oxygen need not always be the compound that brought in, reduced, and dumped as waste. These compounds where the electron is finally "dumped" are more generally known as "**terminal electron acceptors**". The molecules from which the electrons originate can vary greatly across biology (we have looked at one common source - the reduced carbon-based molecule glucose). We will often refer to these molecules as "fuel", and the analogy with the burning of gasoline, propane, or coal is quite apt; these are all simply sources of electrons, and the waste product is CO_2 and water. But once again, keep in mind that many organisms can use non-carbon based sources of electrons, and transfer the electrons to other kinds of oxidizing agents.

In between the original electron source and the terminal electron acceptor are a series of biochemical reactions involving at least one redox reaction (during the course of glycolysis, pyruvate oxidation, and the TCA cycle we saw many!). These redox reactions harvest energy for the cell by coupling exergonic redox reactions to an energy-requiring reactions in the cell. In respiration, a special set of enzymes carry out a linked series of redox reactions that ultimately transfer electrons to the terminal electron acceptor. These "chains" of redox enzymes and electron carriers are called electron transport chains (ETC). ETCs are therefore the portion of respiration that use an electron acceptor (brought in from outside of the cell) as the final/terminal acceptor for the electrons that were removed from the intermediate compounds in catabolism. In aerobically respiring eukaryotic cells the ETC is composed of four large, multiprotein complexes embedded in the inner mitochondrial membrane and two small diffusible electron carriers shuttling electrons between them. The electrons are passed from enzyme to enzyme through a series of redox reactions. These reactions couple the exergonic redox transfers to the endergonic transport of hydrogen ions across the membrane. This process contributes to the creation of a transmembrane electrochemical gradient. The electrons passing through the ETC gradually lose potential energy, and are finally deposited on the terminal electron acceptor, which is removed as waste from the cell. When oxygen is the final electron acceptor, the free energy difference of this multistep redox process- from the 1st step in the ETC to the last- is ~ -60 kcal/mol when NADH donates electrons or 45 kcal/mol when $FADH_2$ donates electrons.

Introduction to redox, oxidative phosphorylation and Electron Transport Chains

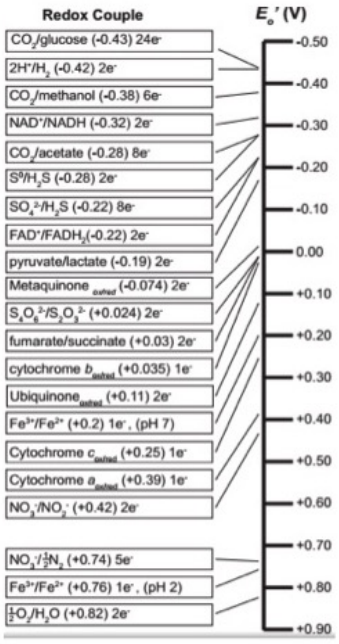
In a prior module (Cashing in on Redox) we discussed the general concept of redox reactions in biology and introduced the Electron Tower, a tool to help you understand redox chemistry and to estimate the direction and magnitude of potential energy differences for various redox couples. In earlier modules, substrate level phosphorylation and fermentation were discussed and we saw how exergonic redox reactions could be directly coupled by enzymes to the endergonic synthesis of ATP. These processes are hypothesized to be one of the oldest forms of energy production used by cells. In this section we discuss the next evolutionary advancement in cellular energy metabolism, **oxidative phosphorylation** of ADP to form ATP. First and foremost, oxidative phosphorylation *does not necessarily* imply the use of oxygen as the terminal electron acceptor. Oxidative phosphorylation evolved before O_2 was available, and has continued to evolve since then. As "obligate aerobes" (organisms that can't live without O_2) we ourselves can only use O_2 as the terminal acceptor. Some organisms are far more flexible, metabolically, and others (obligate anaerobes) can't use O_2 at all. The process of generating ATP from ADP is called oxidative phosphorylation (vs. the substrate level phosphorylation seen earlier, in glycolysis and the citric acid cycle) because it relies on redox reactions to generate a electrochemical *transmembrane potential* that can then be used by the cell to add a phosphate group to ADP (or to do other kinds of work). If you think about it, these are not the most appropriate names, as redox reactions occur in glycolysis and the citric acid cycle too.

A quick summary of Electron Transport Chains

The ETC begins with the donation of electrons from NADH, $FADH_2$ or other reduced compounds to a first acceptor in the chain. The electron moves through a series of electron transporters, carriers (mostly proteins) that undergo a series of redox reactions. The free energy transferred from these exergonic redox reactions is, in some steps, coupled to the endergonic movement of protons across a membrane. Since the membrane is an effective barrier to ions, this pumping results in an unequal accumulation of protons across the membrane. This in turn "polarizes" or "charges" the membrane, with a net positive (protons) on one side of the membrane and a negative charge on the other side of the membrane. The separation of charge creates an *electrical potential*. In addition, the accumulation of protons also causes a pH gradient known as a *chemical potential* across the membrane (the protons are being pumped from a low concentration to a higher $[H^+]$ environment). Together these two gradients (electrical and chemical) are called an *electro-chemical gradient*. The -ΔG of the electrons moving down the ETC is harvested in the form of this gradient.

Review: The Electron Tower

Since redox chemistry is so central to the topic we begin with a quick review of the table of reduction potentials - sometimes called the "redox tower" or "electron tower". You may hear your instructors use these terms interchangeably. As we discussed in previous modules, all kinds of compounds can participate in biological redox reactions. Making sense of all of this information and ranking potential redox pairs can be confusing. A tool has been developed to rate redox half reactions based on their **reduction potentials** or E_0' values. Whether a particular compound can act as an electron donor (reductant) or electron acceptor (oxidant) depends on what other compound it is interacting with. The redox tower ranks a variety of common compounds (their half reactions) from most negative E_0', compounds that *readily get rid of electrons*, to the most positive E_0', compounds *most likely to accept electrons*. The tower organizes these half reactions based on the ability of comopunds to accept electrons. In many redox towers each half reaction is written by convention with the oxidized form on the left followed by the reduced form to its right. The two forms may be either separated by a slash, for example the half reaction for the reduction of NAD^+ to NADH is written: NAD^+/NADH ($2e^-$), or by separate columns. An electron tower is shown below.



A common biological "redox tower". Orient yourself by finding the strongest electron acceptor (= strongest oxidizing agent). Hopefully it's familiar to you.

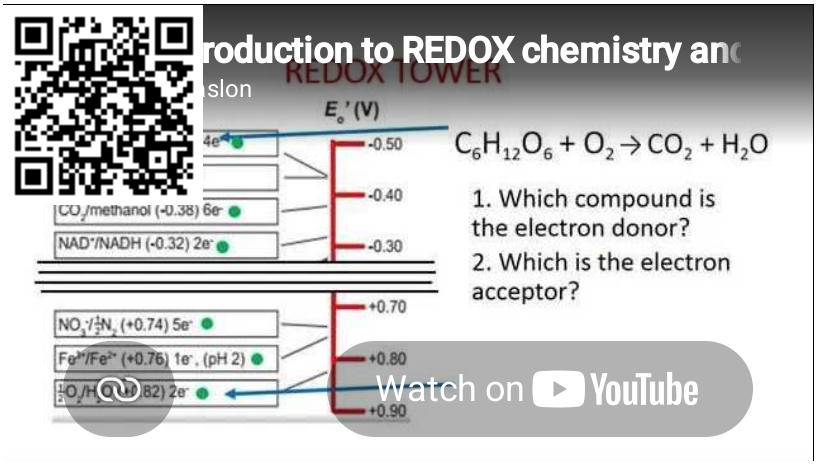
Note

Use the redox tower above as a reference guide to orient you as to the reduction potential of the various compounds. Redox reactions may be either exergonic or endergonic depending on the relative redox potentials of the donor and acceptor (represented by two different half reactions on the tower). Also remember there are many different ways of looking at this conceptually; this type of redox tower is just one way.

Note: **Language shortcuts reappear**

In the redox table above some entries seem to be written in unconventional ways. For instance Cytochrome $c_{ox/red}$. There only appears to be one form listed. Why? This is another example of a language shortcut (probably employed so it fits in the column) that can be confusing. The notation above could be rewritten as Cytochrome c_{ox}/Cytochrome c_{red} to indicate that the cytochrome c protein can exist in either an oxidized state (Cytochrome c_{ox}) or reduced state (Cytochrome c_{red}).

Review Redox Tower Video



For a short video on how to use the redox tower in redox problems click here. This video was made by Dr. Easlon for Bis2A students.

Using the redox tower: A tool to help understand electron transport chains

By convention the tower half reactions are written with the oxidized form of the compound on the left and the reduced form on the right. Notice that compounds such as glucose and hydrogen gas are excellent electron donors and have very low reduction potentials E_0'. Compounds, such as oxygen and nitrite, whose half reactions have relatively high positive reduction potentials (E_0') generally make good electron acceptors and are found at the opposite end of the table.

Example: **Menaquinone**

Let's look at menaquinone$_{ox/red}$. This compound sits in the middle of the redox tower with an half-reaction E_0' value of -0.074 eV. Menaquinone$_{ox}$ can spontaneously ($\Delta G<0$) accept electrons from reduced forms of compounds with lower half-reaction E_0'. Such transfers form menaquinone$_{red}$ and the oxidized form of the original electron donor. In the table above, examples of compounds that could act as electron donors to menaquinone include $FADH_2$, an E_0' value of -0.22, or NADH, with an E_0' value of -0.32 eV. Remember the reduced forms are on the right hand side of the red/ox pair.

Once menaquinone has been reduced to menaquinone(red), it can now spontaneously ($\Delta G<0$) donate electrons to any compound with a higher half-reaction E_0' value. Possible electron acceptors include cytochrome b_{ox} with an E_0' value of 0.035 eV; or ubiquinone$_{ox}$ with an E_0' of 0.11 eV. Remember that the oxidized forms lie on the left side of the half reaction.

Electron Transport Chains

An **electron transport chain**, or **ETC**, is composed of a group of protein complexes in and around a membrane that help energetically couple a series of exergonic/spontaneous redox reactions to the endergonic pumping of protons across the membrane to generate a an electrochemical gradient. This electrochemical gradient creates a free energy potential that is termed a **proton motive force** whose energetically "downhill" exergonic flow can later be coupled to a variety of cellular processes.

ETC Overview

Step 1. Electrons enter the ETC from an electron donor, such as NADH or $FADH_2$ which are generated during a variety of catabolic reactions, including those associated with glucose oxidation. Depending on the complexity (number and types of electron carriers) of the ETC being used by an organism, electrons can enter at a variety of places in the electron transport chain - this depends upon the respective reduction potentials of the proposed electron donors and acceptors. In some metabolically flexible organisms, the chain could also terminate at a variety of external electron acceptors.

Step 2. After the first redox reaction, the initial electron donor will become oxidized and the electron acceptor will become reduced. The difference in redox potential between the electron acceptor an donor is related to ΔG by the relationship $\Delta G = -nF\Delta E$, where n = the number of electrons transferred and F = Faraday's constant. The larger a positive ΔE the more exergonic a reaction.

Step 3. If sufficient energy transferred during an exergonic redox step the electron carrier may couple this negative change in free energy to the endergonic process of transporting a proton from one side of the membrane to the other.

Step 4. After multiple redox transfers, the electron is delivered to a molecule known as the terminal electron acceptor. In the case of humans, the terminal electron acceptor is oxygen. However, there are many, many, many, other possible electron acceptors, see below.

Note

Electrons entering the ETC do not have to come from NADH or $FADH_2$. Many other compounds can serve as electron donors, the only requirements are that there exists an enzyme that can oxidize the electron donor and then reduce another compound and that the E_0' be positive (e.g. $\Delta G<0$). Even a small amounts of free energy transfers can add up. For example there are bacteria that use H_2 as an electron donor. This is not too difficult to believe because the half reaction $2H^+ + 2\ e^-/H_2$ has a reduction potential (E_0') of -0.42 V. If these electrons are eventually delivered to oxygen then the $\Delta E_0'$ of the reaction is 1.24 V which corresponds to a large negative ΔG (-ΔG). Alternatively, there are some bacteria that can oxidize iron, Fe_2+ at pH 7 to Fe_3^+ with a reduction potential (E_0') of + 0.2 V. These bacteria use oxygen as their terminal electron acceptor and in this case, the $\Delta E_0'$ of the reaction is approximately 0.62 V. This still produces a -ΔG. The bottom line is that depending on the electron donor and acceptor that the organism uses, a little or a lot of energy (per electron) can be transferred and used by the cell as electrons are donated to the electron transport chain.

How do you calculate ΔE? It's pretty simple using the redox table- look for the **final state (in V) of the electron** (say, water, from reduction of O2) and subtract the **initial state (say, as glucose)**. Using this chart your calculation ΔE would be 0.82 - (-0.43) = 1.25 volts. You do not need to worry (yet) about the number of electrons as Δ volts have units of Δ unit energy/unit charge (for example, Joules per Coulomb).

What are the complexes of the ETC?

ETCs are made up of a series (or at least one) of membrane associated redox proteins or (some are integral) protein complexes (complex = more than one protein arranged in a quaternary structure) that move electrons from a donor source, such as NADH, to a final terminal electron acceptor, such as oxygen. This particular donor/terminal acceptor pair is the primary one used in human mitochondria. Each electron transfer in the ETC requires a reduced substrate as an electron donor and an oxidized substrate as the electron acceptor. In most cases the electron acceptor is a member of the enzyme complex. Once the complex is reduced, the complex can serve as an electron donor for the next reaction.

How do ETC complexes transfer electrons?

As previously mentioned the ETC is composed of a series of protein complexes that undergo a series of linked redox reactions. These complexes are in fact multiprotein enzyme complexes referred to as **oxidoreductases:** the enzymes perform both oxidation (of the step in the chain "above" them, energetically) and reduction (of the complex in the complex in the chain below them) or **reductases** (again, in spite of the fact that the enzyme both oxidizes an electron donor and passes that electron to an electron acceptor). The one exception to this naming convention is the terminal complex in aerobic respiration that uses molecular oxygen as the terminal electron acceptor. That enzyme complex is referred to as an **oxidase** (even though it both oxidizes and reduces something). Redox reactions within these complexes are typically carried out by a non-protein component called a **prosthetic group**. This is true for all of the electron carriers with the exception of quinones, which are a class of lipids that can directly be reduced or oxidized by the oxidoreductases. In this case, both the $Quinone^{red}$ and the $Quinone^{ox}$ is soluble within the membrane and can move from complex to complex. The prosthetic groups are directly involved in the redox reactions being catalyzed by their associated oxidoreductases. In general these prosthetic groups can be divided into two general types: those that carry both electrons and protons and those that only carry electrons.

Question:

If the redox reaction is carried out by a non-proteinaecious "prosthetic group" (see list below)- why bother with the protein part of the complex?

The Electron and Proton carriers in the ETC

- **Flavoproteins (Fp)**, these proteins contain an organic prosthetic group called a **flavin**, which is the actual moiety that undergoes the oxidation/reduction reaction. $FADH_2$ is an example of a Fp.
- **Quinones**, are a family of lipids which means they are soluble within the membrane.
- NADH and NADPH are electron (2e-) and proton ($2\ H^+$) carriers, of course. The $FADH_2$ we know from the citric acid cycle is actually bound to the enzyme succinate dehydrogenase.
- **Cytochromes** are proteins that contain a heme prosthetic group. The Heme is capable of carrying a single electron.
- **Iron-Sulfur proteins** contain a non-heme iron-sulfur clusters that can carry an electron. The prosthetic group is often abbreviated as **Fe-S**

Aerobic versus Anaerobic respiration

In the world we live in, most of the organisms we see around us breath air, which is approximately 20% oxygen. Oxygen is our **terminal electron acceptor**. We call this process respiration, specifically aerobic respiration. We breath in oxygen, it diffuses into our cells and to our mitochondria where it is used as the final acceptor of electrons from our electron transport chains. That is **aerobic respiration**: the process of using oxygen as a terminal electron acceptor in an electron transport chain.

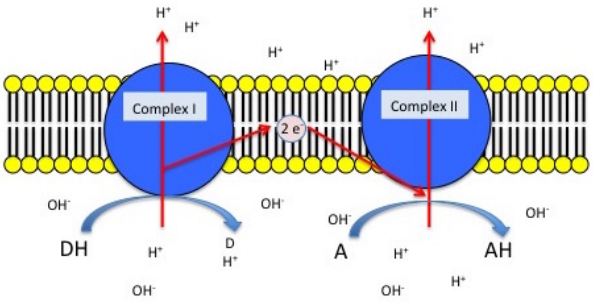
While we may use oxygen as the terminal electron acceptor for our respiratory chains, the more general process of respiration evolved at time when oxygen was not a major component of the atmosphere. Respiration or **oxidative phosphorylation** does not require oxygen at all; it simply requires a compound with a high reduction potential to act as a terminal electron acceptor. Many organisms can use a variety of compounds including nitrate (NO_3^-), nitrite (NO_2^-), even iron (Fe^{+++}) as terminal electron acceptors. When oxygen is **NOT** the terminal electron acceptor, the process is referred to as **anaerobic respiration**. The ability of organisms to vary their terminal electron acceptor provides metabolic flexibility and can ensure better survival if any given terminal acceptor is in limited supply. Think about this: in the absence of oxygen we die; but an organism that can use a different terminal electron acceptor can survive.

Question: What's the difference between anaerobic respiration and fermentation?

Here is a fun video in which a nice young man COMPLETELY FAILS TO UNDERSTAND what anaerobic respiration is, teaching us instead about fermentation. Beware of the internet! I'd say "you get what you pay for" but this is actually a site you have to pay for!

A generic example of a simple, 2 complex ETC

The figure below depicts a generic electron transport chain, composed of two integral membrane complexes; Complex I_{ox} and Complex II_{ox}. A reduced electron donor, designated DH (such as NADH or $FADH_2$) reduces Complex I_{ox} giving rise to the oxidized form D (such as NAD^+ or FAD). Simultaneously, a prosthetic group within complex I is now reduced (accepts the electrons). In this example the redox reaction is exergonic and the free energy difference is coupled by the enzymes in Complex I to the endergonic translocation of a proton from one side of the membrane to the other. The net result is that one surface of the membrane becomes more negatively charged, due to an excess of hydroxyl ions (OH^-) and the other side becomes positively charged due to an increase in protons on the other side. Complex I_{red} can now reduce the prosthetic group in Complex II_{red} while simultaneously becoming oxidized by Complex II_{ox}. Electrons pass from Complex I to Complex II via thermodynamically spontaneous redox reactions, regenerating Complex I_{ox} which can repeat the previous process. Complex II_{red} reduces A, the terminal electron acceptor to regenerate Complex II_{ox} and create the reduced form of the terminal electron acceptor. In this case, Complex II can also translocate a proton during the process. If A is molecular oxygen, water (AH) will be produced. When A is oxygen the reaction scheme would be considered a model of an aerobic ETC. However, if A is nitrate, NO_3^- then Nitrite, NO_2^- is produced (AH) and this would be an example of an anaerobic ETC.



Generic 2 complex electron transport chain. In the figure, DH is the electron donor (donor reduced) and D is the donor oxidized. A is the oxidized terminal electron acceptor and AH is the final product, the reduced form of the acceptor. As DH is oxidized to D, protons are translocated across the membrane, leaving an excess of hydroxyl ions (negatively charged) on one side of the membrane and protons (positively charged) on the other side of the membrane. The same reaction occurs in Complex II as the terminal electron acceptor is reduced to AH.

Exercise 1

Thought question

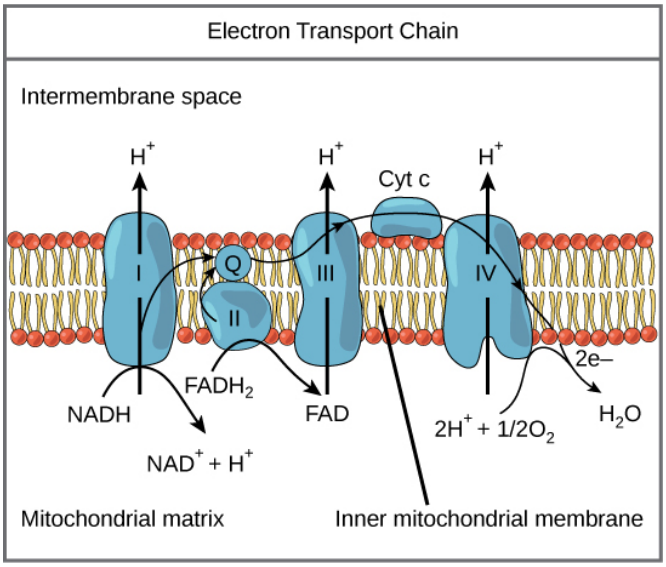
Based on the figure above and using an electron tower, what is the difference in the electrical potential if (A) DH is NADH and A is O_2 and (B) DH is NADH and A is NO_3^-. Which pairs (A or B) provide the most amount of usable energy?

In the nitrate vs nitrite example above, both the oxidized and the reduced forms have a charge of minus 1. So... how to you know which form is more oxidized?

The figure above has a problem- it depicts an electron floating within the membrane. This wouldn't happen- the membrane is not a "wire", it is an "insulator". You might draw your own corrected version.

Detailed look at aerobic respiration

The eukaryotic mitochondria has evolved a very efficient ETC. There are four complexes composed of multiple proteins and their prosthetic groups. These complexes labeled I through IV depicted in the figure below. The aggregation of these four complexes, together with associated mobile, accessory electron carriers, makes up the electron transport chain. This type of electron transport chain is present in multiple copies in the inner mitochondrial membrane of eukaryotes.



The electron transport chain is a series of electron transporters embedded in the inner mitochondrial membrane that shuttles electrons from NADH and $FADH_2$ to molecular oxygen. In the process, protons are pumped from the mitochondrial matrix to the intermembrane space, and oxygen is reduced to form water.

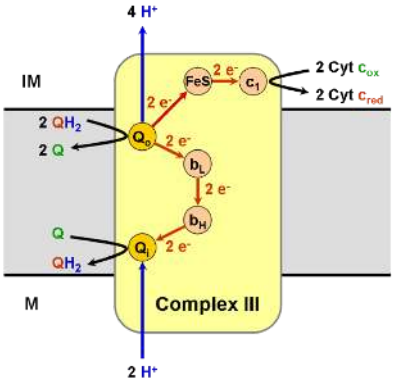
Complex I

To start, two electrons are carried to the first complex aboard NADH. This complex, labeled I, is composed of flavin mononucleotide (FMN) and an iron-sulfur (Fe-S)-containing protein. FMN, which is derived from vitamin B_2, also called riboflavin, is one of several prosthetic groups or co-factors in the electron transport chain. These **prosthetic groups** (non-protein molecule required for the activity of a protein) are organic or inorganic non-peptide molecules bound to a protein, covalently or noncovalently; prosthetic groups include co-enzymes, which are the prosthetic groups of enzymes. The enzyme in complex I is NADH dehydrogenase (aka NADH Q reductase) and is a very large protein, containing 45 amino acid chains. Complex I can pump four hydrogen ions across the membrane from the matrix into the intermembrane space, and it is in this way that the hydrogen ion gradient is established and maintained between two aqueous compartments separated by the inner mitochondrial membrane.

Q and Complex II

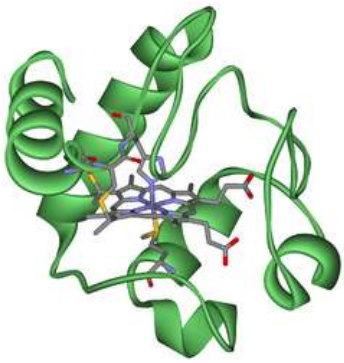
Complex II directly receives $FADH_2$, $FADH_2$'s electrons do not pass through complex I. The compound connecting the first and second complexes to the third is ubiquinone (Q). The Q molecule is lipid-soluble and freely moves through the hydrophobic core of the membrane. Once it is reduced, (QH_2), ubiquinone delivers its electrons to the next complex in the electron transport chain. Q receives the electrons derived from NADH from complex I and the electrons derived from $FADH_2$ from complex II, including succinate dehydrogenase. This enzyme and $FADH_2$ form a small complex that delivers electrons directly to the electron transport chain, bypassing the first complex. Since these electrons bypass and thus do not energize the proton pump in the first complex, fewer ATP molecules are made from the $FADH_2$ electrons. As we will see in the following section, the number of ATP molecules ultimately obtained is proportional to the number of protons pumped across the inner mitochondrial membrane.

Complex III



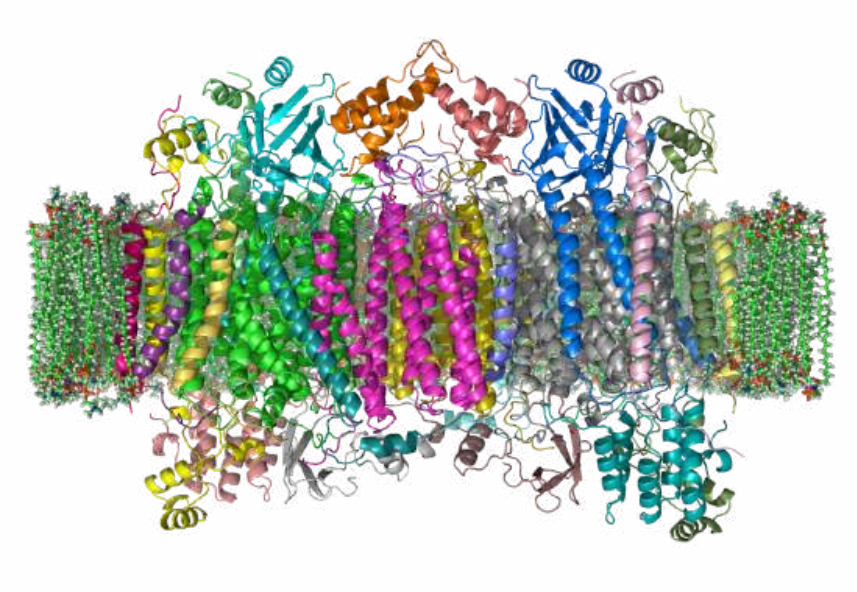
There's a lot of redox going on in this picture. Which is the upstream donor of electrons coming into Complex III? Which complex/molecule does Complex III reduce?

The third complex is composed of proteins that carry a number of prosthetic groups: a B-type cytochrome, a c-type cytochrome, and an Iron-sulfur cluster; this complex is called cytochrome bc_1 complex. Cytochrome (= "cellular pigment") proteins have a prosthetic group that includes a metal bound by heme. The heme molecule is similar to the heme in hemoglobin, but it carries electrons, not oxygen. As a result, the iron ion at its core is reduced and oxidized as it passes the electrons, fluctuating between different oxidation states: Fe^{++} (reduced) and Fe^{+++} (oxidized). The heme molecules in the cytochromes have slightly different characteristics due to the effects of the different proteins binding them, giving slightly different characteristics to each complex. Complex III pumps protons through the membrane and passes its electrons to cytochrome c for transport to the fourth complex of proteins and enzymes (cytochrome c is the acceptor of electrons from Q; however, whereas Q carries pairs of electrons, cytochrome c can accept only one at a time).



Cytochrome C

The **cytochrome complex**, or **cyt *c*** is a small hemeprotein found loosely associated with the inner membrane of the mitochondrion. It belongs to the cytochrome c family of proteins. Cytochrome c is an essential component of the electron transport chain, where it carries one electron. It is capable of undergoing oxidation and reduction as its iron atom converts between the ferrous and ferric forms, but does not bind oxygen. It transfers electrons between Complexes III (Coenzyme Q – Cyt C reductase) and IV (Cyt C oxidase).



Complex IV, cytochrome C oxidase

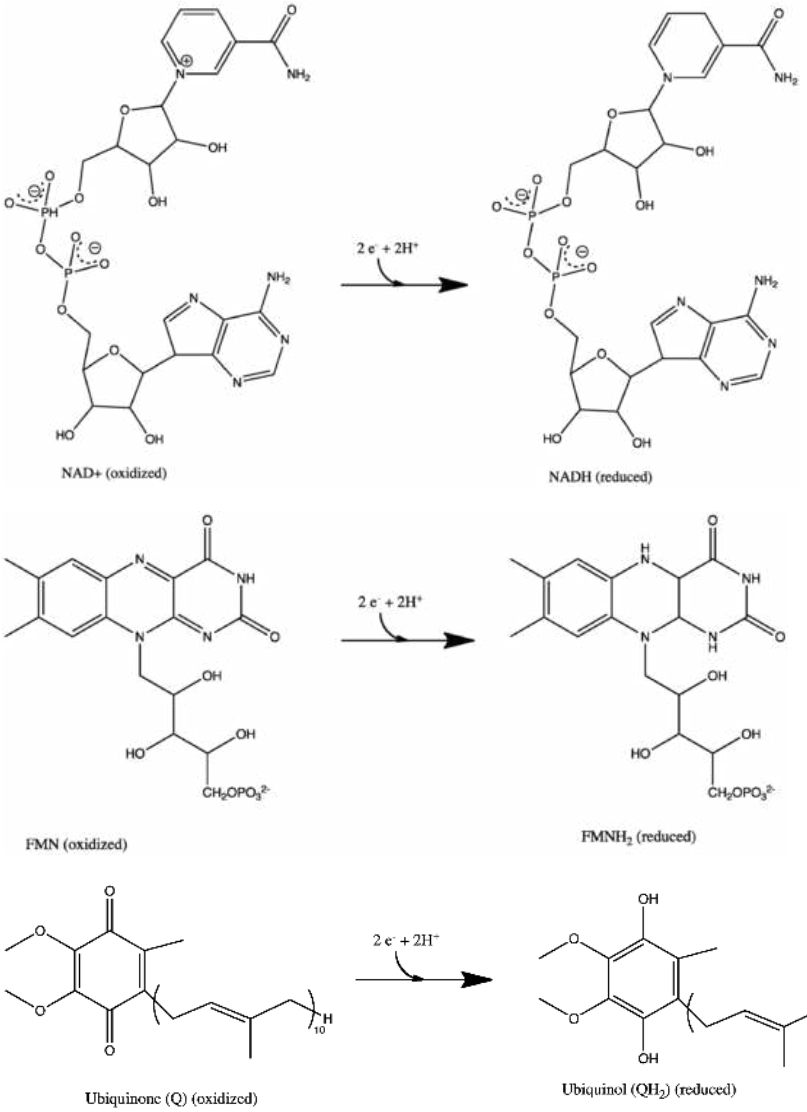
This complex contains two heme groups (one in each of the two cytochromes, a, and a3) and three copper ions (a pair of CuA and one CuB in cytochrome a3). The cytochromes hold an oxygen molecule very tightly between the iron and copper ions until the oxygen is completely reduced (the donor, cytochrome C, can only deliver one electron at a time). The reduced oxygen then picks up two hydrogen ions from the surrounding medium to make water (H_2O). The removal of the hydrogen ions from the system contributes to the ion gradient used in the process of oxidative phosphorylation (as well as providing the -ΔG required to drive other endergonic processes).

Here is a very nice video that includes the electron transport chain- I think it's very helpful, check it out! It starts with ATP synthase, which we haven't discussed yet, but that's fine, it'll explain why the cell is building that proton gradient!

I have one problem with this animation:There's really no discussion of how proton pumping works- the discussion's extremely vague- one might even come away with the notion that a gas forms within the matrix-domain of NADH dehydrogenase (complex I). In fact, the proton pump of complex I is entirely embedded within the membrane and isn't illustrated here at all. Presumably this is because the mechanism of pumping is a subject of debate. For the purposes of this class, you simply need to know- as mentioned in the video- that the endergonic formation of a proton gradient is powered by the exergonic redox reactions. In NADH dehyrogenase, this includes the oxidation of NADH by a flavin within the enzyme, some internal transfers of electrons within the enzyme, and then the transfer of the electron to a membrane-soluble quinone (coenzyme Q), to form a quinol. Each of these 2-electron transfers is illustrated below, with the oxidized molecules shown first.

Question:

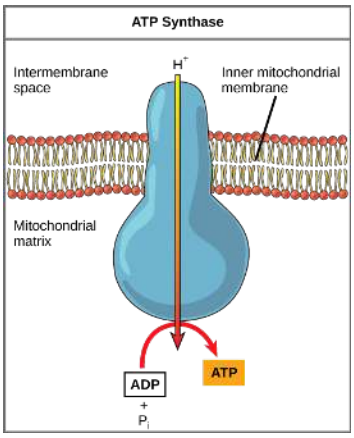
Could NADH dehyrodrogenase reduce NAD^+ as well as oxidize NADH? Under what circumstances might this happen?



Chemiosmosis

The free energy from the series of redox reactions just described is used to pump hydrogen ions (protons) across the membrane, resulting in a concentration of protons on one side of the membrane. The uneven distribution of H^+ ions across the membrane establishes both concentration and electrical gradients (thus, an "electrochemical" gradient), owing to the hydrogen ions' positive charge and their aggregation on one side of the membrane.

If the membrane were open to diffusion by the hydrogen ions, the ions would tend to diffuse back across into the matrix, driven by their electrochemical gradient. Many ions cannot diffuse through the nonpolar regions of phospholipid membranes without the aid of ion channels. Similarly, hydrogen ions in the matrix space can only pass through the inner mitochondrial membrane via an integral membrane protein called ATP synthase (depicted below, and illustrated in the video recommended above). This complex acts as a tiny generator, turned by the protons moving down their electrochemical gradient. The movement of this molecular machine serves to both lower the activation energy of the ADP + Pi -> ATP reaction and couple the exergonic transfer of energy associated with the movement of protons down their electrochemical gradient to the endergonic addition of a phosphate to ADP, forming ATP.



ATP synthase is a complex, molecular machine that uses a proton (H^+) gradient to form ATP from ADP and inorganic phosphate (Pi). (Credit: modification of work by Klaus Hoffmeier). If you have had some O-chem and would like to dig a little deeper into the mechanism through which the mechanical flexing of this complex promotes ATP formation, check out this video.

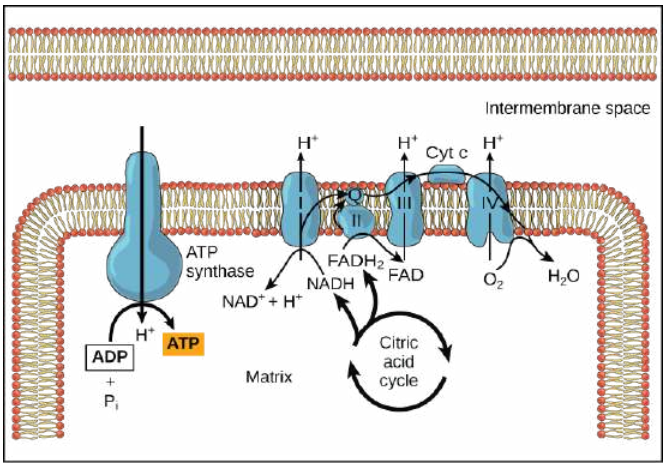
Suggested Discussion

Dinitrophenol (DNP) is a small chemical that makes the membrane leaky to protons. It is sometimes used as a weight-loss drug, and can have lethal consequences. What effect would you expect DNP to have on the difference in pH across both sides of the inner mitochondrial membrane? Why do you think this might be an effective weight-loss drug? Why might it be dangerous?

In our cells, chemiosmosis is used to generate 90 percent of the ATP made during aerobic glucose catabolism; it is also the method used in the light reactions of photosynthesis to harness the energy of sunlight in the process of photophosphorylation (vs. oxidative phosphorylation via the ATPase of mitochondria). The overall result of these reactions is the production of ATP from the energy of the electrons removed from hydrogen atoms. These atoms were originally part of a glucose molecule. At the end of the pathway, the electrons are used to reduce an oxygen molecule to water.

Super-helpful video:

How ATP is made by ATP synthase



In oxidative phosphorylation, the pH gradient formed by the electron transport chain is used by ATP synthase to form ATP in a Gram- bacteria.

Suggested Discussion

Cyanide inhibits cytochrome C oxidase, a component of the electron transport chain. If cyanide poisoning occurs, would you expect the concentration of protons of the intermembrane space to increase or decrease? What effect would cyanide have on

ATP synthesis?

5.3: Respiration is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Anne Britt, UC Davis.

- **1.15: Respiration** has no license indicated.

SECTION OVERVIEW

5.4: Net Primary Production

5.4.1: Marine Primary Production

5.4.1.1: Primary production

5.4.1.2: The Producers

5.4.1.3: Factors Influencing Production

5.4.1.4: Patterns of Primary Production

5.4.2: Surface Primary Production

5.4: Net Primary Production is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

SECTION OVERVIEW

5.4.1: Marine Primary Production

5.4.1.1: Primary production

5.4.1.2: The Producers

5.4.1.3: Factors Influencing Production

5.4.1.4: Patterns of Primary Production

5.4.1: Marine Primary Production is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

5.4.1.1: Primary production

Primary production is the creation of new organic matter from inorganic substrates, and it is this organic matter that serves as the base of the food web for most marine consumers. Primary production generally refers to the process of **photosynthesis,** or the utilization of light energy to produce chemical fuels that is undertaken by plants and algae according to the reaction:

$$6\,CO_2 + 6\,H_2O \longrightarrow C_6H_{12}O_6 + 6\,O_2 \tag{5.4.1.1.1}$$

Here, powered by light energy, carbon dioxide and water combine to produce glucose and oxygen. However, primary production is also carried out by bacteria in the absence of light through **chemosynthesis**. Instead of light providing the energy for the reaction, the energy comes from the oxidation of inorganic materials, such as hydrogen sulfide (see section 4.11 on hydrothermal vents). Here we will concentrate on photosynthesis because it plays a much larger role in total oceanic productivity than chemosynthesis.

The organisms responsible for oceanic primary production include a wide diversity of marine plants and algae. While many people may be more familiar with the larger seagrasses and macroalgae (seaweeds), by far the greatest amount of photosynthesis in the ocean comes from microscopic algae, the **phytoplankton**. The term “plankton” refers to organisms that drift with the currents, and the phytoplankton are the free-floating algae that undergo photosynthesis (contrast this with the **zooplankton**, who are the drifting animals). In section 7.2 we will take a closer look at the organisms responsible for oceanic primary production.

The total amount of organic material created by the producers is called the **gross primary productivity**, or **total production**. However, the primary producers consume a portion of this organic matter themselves through respiration, so the total amount that is left to support the consumers in the ocean is called **net production** (gross productivity – respiration). Gross production can be divided into two components, new production and regenerated production. **New production** is supported by nutrients brought in from outside of the local ecosystem through processes such as upwelling or ocean currents. **Regenerated production** results from the recycling of nutrients within an ecosystem.

Overall, marine productivity is similar to terrestrial production. Marine net production is about 35-50 billion metric tons per year, while terrestrial production reaches 50-70 billion tons per year. However, the biomass responsible for that production in the ocean is about 1-2 billion metric tons, compared to 600-1000 billion metric tons of biomass in terrestrial systems. So the oceans are producing almost as much organic material as terrestrial producers, but are doing it from only a fraction of the amount of producer biomass. One reason for this discrepancy is that the phytoplankton are constantly being consumed, while much of the terrestrial biomass is much longer-lived than the plankton.

5.4.1.1: Primary production is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **7.1: Primary Production** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

5.4.1.2: The Producers

Although the phytoplankton are microscopic in size compared to marine plants and macroalgae like seaweeds and seagrasses, they account for by far the greatest amount of photosynthesis in the oceans; about 95% of all marine primary productivity. Most of the production by phytoplankton comes from three groups, the diatoms, dinoflagellates, and coccolithophores.

Diatoms are single-celled algae consisting of cellular material inside a shell, or test, made of silica, a component of glass. Diatoms are relatively large, reaching up to about 1 mm in diameter, and come in a range of shapes, from circular disks to elongated or triangular forms (Figure 5.4.1.2.1). In some species of diatoms individual cells link together into multicellular chains. Diatoms are very efficient producers, with up to 55% of the absorbed sunlight energy incorporated into carbohydrate formation; this is one of the highest photosynthetic efficiencies known. Diatoms are most abundant in coastal and cold, nutrient-rich waters. Where diatoms are bountiful, the underlying sediments are rich in their silica shells, creating siliceous sediment and diatomaceous earth (see section 12.3).

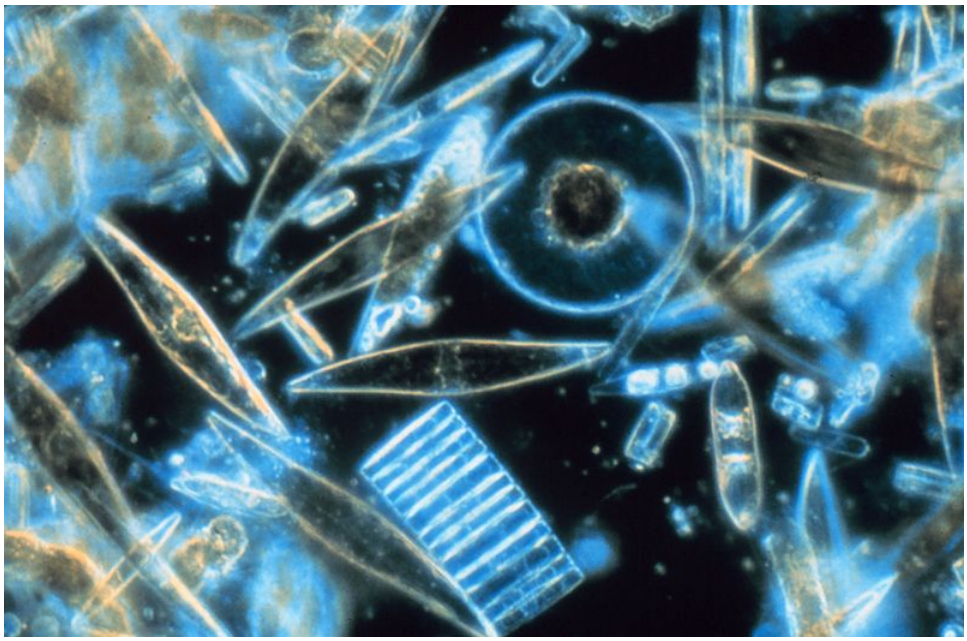


Figure 5.4.1.2.1 A variety of diatom species from the Southern Ocean, seen under a light microscope (Prof. Gordon T. Taylor, Stony Brook University, [Public domain], via Wikimedia Commons).

Dinoflagellates are another form of single-shelled photosynthetic algae that are generally smaller than diatoms, most with sizes in the 0.015 – 0.04 mm range. Most dinoflagellates have a characteristic pair of flagella (hence the name); small whip-like “tails” that they use for locomotion. Usually there is one flagellum that trails from the body to provide forward movement, and another that encircles the cell to make it spin as it moves. Unlike diatoms, dinoflagellates do not have a mineralized shell. Instead, many are covered by cellulose, which easily decomposes in seawater, so their shells do not really contribute to sediment formation (Figure 5.4.1.2.2). While most dinoflagellates undergo photosynthesis, some species will also ingest prey.

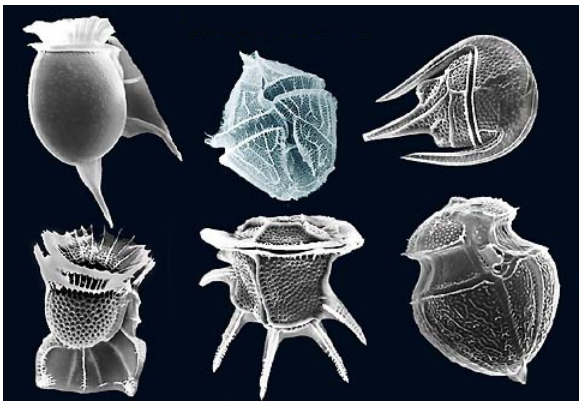


Figure 5.4.1.2.2 Electron microscope image of several dinoflagellate species. In some of the species, a groove can be seen running around the circumference of the organism; this groove is usually home to one of the flagella (fickleandfreckled, https://www.flickr.com/photos/fickle...led/6939384773, CC-BY 2.0).

A third, much smaller type of phytoplankton includes the **coccolithophores**, which range from about 5-100 micrometers wide. As with diatoms and dinoflagellates, these are single-celled photosynthetic algae that contribute significantly to oceanic primary production, but their cellular material is encased in a very different kind of shell. The test (shell) is made up of a number of interlocking circular plates composed of calcium carbonate that link together to form a sphere (Figure 5.4.1.2.3). Coccolithophores are most abundant in warm, open ocean waters, and their sinking tests can lead to calcium carbonate sediments in some parts of the ocean (see section 12.3).

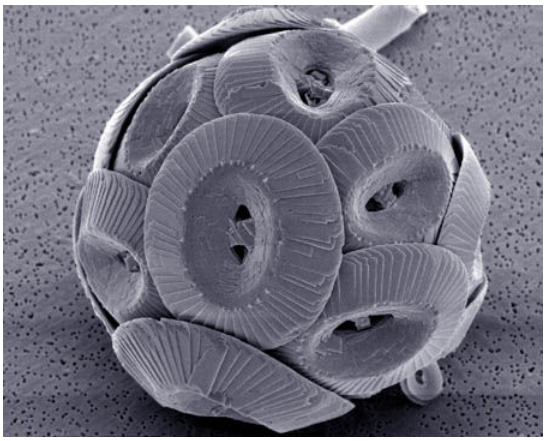


Figure 5.4.1.2.3 Electron microscope image of the North Atlantic coccolithophore *Coccolithus pelagicus* (Richard Lampitt, Jeremy Young, The Natural History Museum, London (http://planktonnet.awi.de/) [CC BY 2.5], via Wikimedia Commons).

Recent evidence suggests that another group of organisms, the **bacteria**, or **picoplankton**, may be very important primary producers. Although they are very small, on the order of 0.2-2 micrometers long, they can be found in very high concentrations, and may be responsible for up to 70% of all productivity in some parts of the ocean.

Harmful algal blooms

Primary production provides plentiful food resources for ocean consumers, so a high abundance of phytoplankton is a good thing, right? As in many other cases, too much of a good thing can sometimes be dangerous, and an overabundance of dinoflagellates or diatoms can often create serious concerns. These events are referred to as **harmful algal blooms**, or HABs. HABs can occur for a number of reasons, although a common one is an overabundance of nutrients, which is often due to excessive terrestrial runoff of fertilizers or other nitrogen- and phosphate-containing materials. These conditions lead to an explosion in algal populations that can change the color of the water if the cells are in high enough concentrations. Figure 5.4.1.2.4 shows a massive bloom that contained so many dinoflagellate cells that it turned the water reddish-brown, a so-called "**red tide**." (It has been suggested that Biblical references to seas being "turned to blood" may have actually been describing red tide events).



Figure 5.4.1.2.4 A red tide caused by dinoflagellates near the Scripps Institution of Oceanography Pier, La Jolla California (Alejandro Díaz, Public Domain via Wikimedia Commons).

These massive algal blooms can have some serious consequences. For one, when all of this algae eventually dies off and sinks, their decomposition uses up the dissolved oxygen in the water, leaving anoxic or hypoxic conditions that can lead to the mass die-off of fish and invertebrates. Dinoflagellates and diatoms are also capable of producing toxins themselves. These phytoplankton get eaten by fish, shellfish and other organisms, and in high abundances the toxins become concentrated in the tissues of the consumers. When humans or other higher-level consumers then eat these organisms, the toxins are concentrated enough to cause sickness or even death. For example, some dinoflagellates produce a toxin that causes **paralytic shellfish poisoning**, which can occur in humans as soon as 30 minutes after eating infected shellfish. This toxin attacks the nervous system, producing symptoms of dizziness, nausea, slurred speech, loss of feeling, and uncoordinated movements, and can ultimately be fatal. Diatoms produce a toxin called domoic acid that causes **amnesic shellfish poisoning**, leading to memory loss, seizures, and potentially death. Domoic acid poisoning also affects marine animals; it is thought to have been responsible for an event in Capitola, California in 1961 where flocks of seabirds acted crazily, even attacking humans. This event inspired the Alfred Hitchcock movie “The Birds.”

Additional links for more information:

- For more on harmful algal blooms, visit NOAA’s HAB site: http://www.noaa.gov/what-is-harmful-algal-bloom

---

This page titled 5.4.1.2: The Producers is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Paul Webb via source content that was edited to the style and standards of the LibreTexts platform.

- **7.2: The Producers** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

5.4.1.3: Factors Influencing Production

For terrestrial plants, many factors affect productivity, including light, temperature, nutrients, soil, and water. For phytoplankton, soil is obviously not needed, and water availability is not an issue. Temperature is generally more stable in the ocean than on land, so for phytoplankton, productivity comes down to the availability of light and nutrients.

Light

Since light is vital for photosynthesis, phytoplankton and other primary producers are limited to the uppermost layers of the ocean where light is abundant enough to sustain the reaction. As depth increases, light intensity decreases until there reaches a depth where photosynthesis can no longer occur (Figure 5.4.1.3.1). The region through which sufficient light for photosynthesis can penetrate is called the photic or euphotic zone, which extends down to about 200 m (section 6.5).

In addition to undergoing photosynthesis, phytoplankton also respire, consuming some of the organic compounds they produce. Rates of respiration are not light dependent, and respiration occurs at all depths and light levels. Therefore, as depth increases the rate of photosynthesis declines as light is diminished, until a point is reached where the rate of photosynthesis equals the respiration rate (Figure 5.4.1.3.1). This depth is the **compensation depth**, and it marks the lower level of the photic zone, and represents the depth where net primary production ends. Below this depth, there is net respiration.

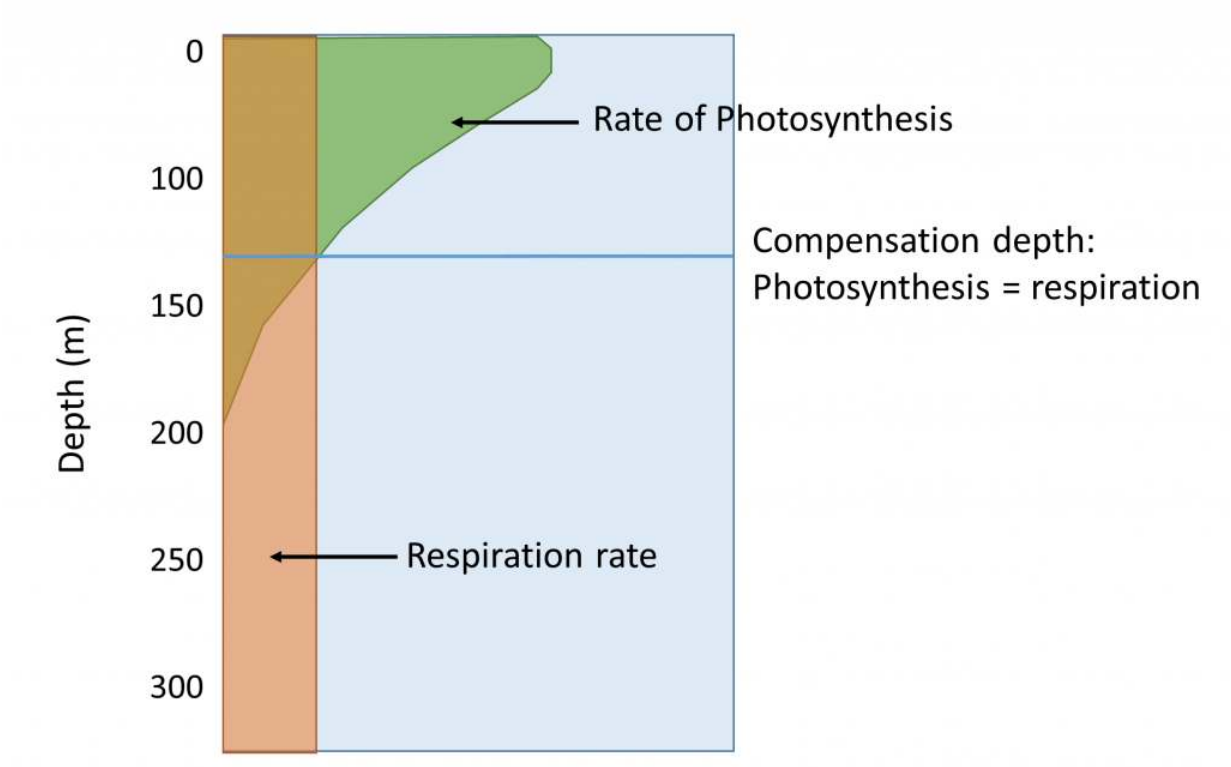


Figure 5.4.1.3.1 As depth increases the rate of photosynthesis declines as light becomes limited. The rate of respiration remains consistent at all depths. The depth where photosynthesis equals respiration is the compensation depth (PW).

Nutrients

Nutrients are required by all of the marine primary producers. The major nutrients required by phytoplankton are nitrogen and phosphorus, in the forms of nitrate NO_3^-, nitrite NO_2^{2-}, ammonium NH_4^+, and phosphate PO_4^{3-}. Many phytoplankton, particularly the diatoms, also require silica, SiO_2, for shell formation. All of these nutrients occur in very small amounts in seawater, so they are often the limiting factors for phytoplankton growth in most situations, particularly the nitrogen compounds. For example, agricultural soil contains 0.5% nitrogen in the upper meter of soil, while surface ocean water contains about 0.00005% nitrogen, 1/10,000 the amount in soil.

As we saw in section 5.6, nutrients are not distributed evenly throughout the water column (Figure 5.4.1.3.2). Near the surface nutrients are quickly utilized by phytoplankton as they become available, so surface waters are nutrient-poor. But as the phytoplankton are consumed or die they are recycled into particles of organic matter, such as fecal pellets or carcasses, that sink into deeper water. Once in deep water, decomposition of these materials releases the nutrients back into the water column. Because there are no producers to utilize them at depth, nutrients are more abundant in deeper water.

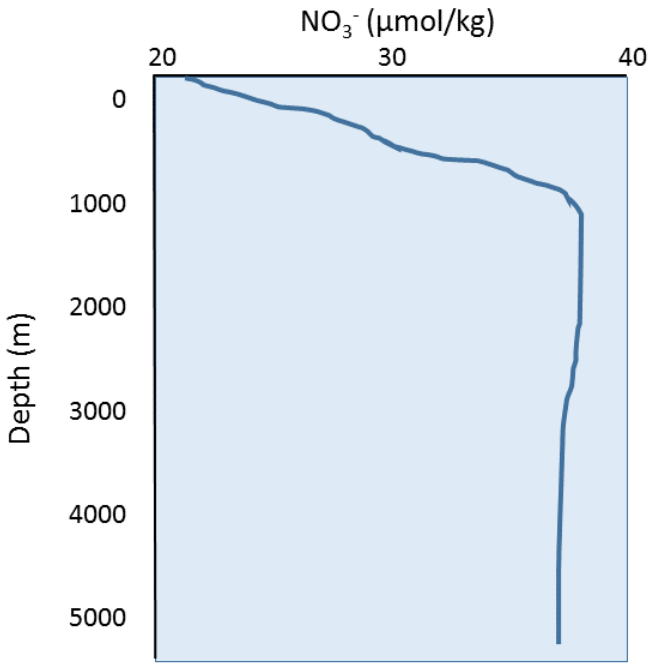


Figure 5.4.1.3.2 Representative nutrient profile in the open ocean. While this profile shows nitrate abundance, the profile is similar for other nutrients such as phosphate and silica (PW).

These deep water nutrients are out of reach of the phytoplankton at the surface. The thermocline and density stratification of the water column generally prevents the nutrient-rich deep water from mixing with the surface water. However, under certain conditions this nutrient-rich deep water may be brought to the surface through the process of upwelling (see section 9.5). Where upwelling occurs there is usually high productivity as the phytoplankton can take advantage of the input of nutrients.

This page titled 5.4.1.3: Factors Influencing Production is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Paul Webb via source content that was edited to the style and standards of the LibreTexts platform.

- **7.3: Factors Influencing Production** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

5.4.1.4: Patterns of Primary Production

Primary productivity varies both geographically and seasonally. Geographically, phytoplankton abundance generally decreases as you move from coastal to oceanic waters (Figure 5.4.1.4.1). Coastal waters are more productive than the central ocean for two main reasons. First, runoff from land often contains a high abundance of nutrients which get deposited in coastal waters and stimulate production. Second, the shallower bottom along the continental shelf can trap nutrients and prevent them from sinking to greater depths. It is easier for these nutrients to be brought back to the surface when they remain trapped in the shallows. Conversely, the central ocean generally has very low primary production, as these areas are far removed from any terrestrial sources of nutrients, and the great depth prevents the deep nutrients from returning to the surface.

Global averages for ocean surface primary production are about 75-150 g C/m^2/yr. Some highly productive areas include the California coast (200-300 g C/m^2/year), the Southern Ocean (200-400 g C/m^2/year), and the coast of Peru (200-400 g C/m^2/year), all regions with significant upwelling. The central ocean, by contrast, produces less than 50 g C/m^2/year.

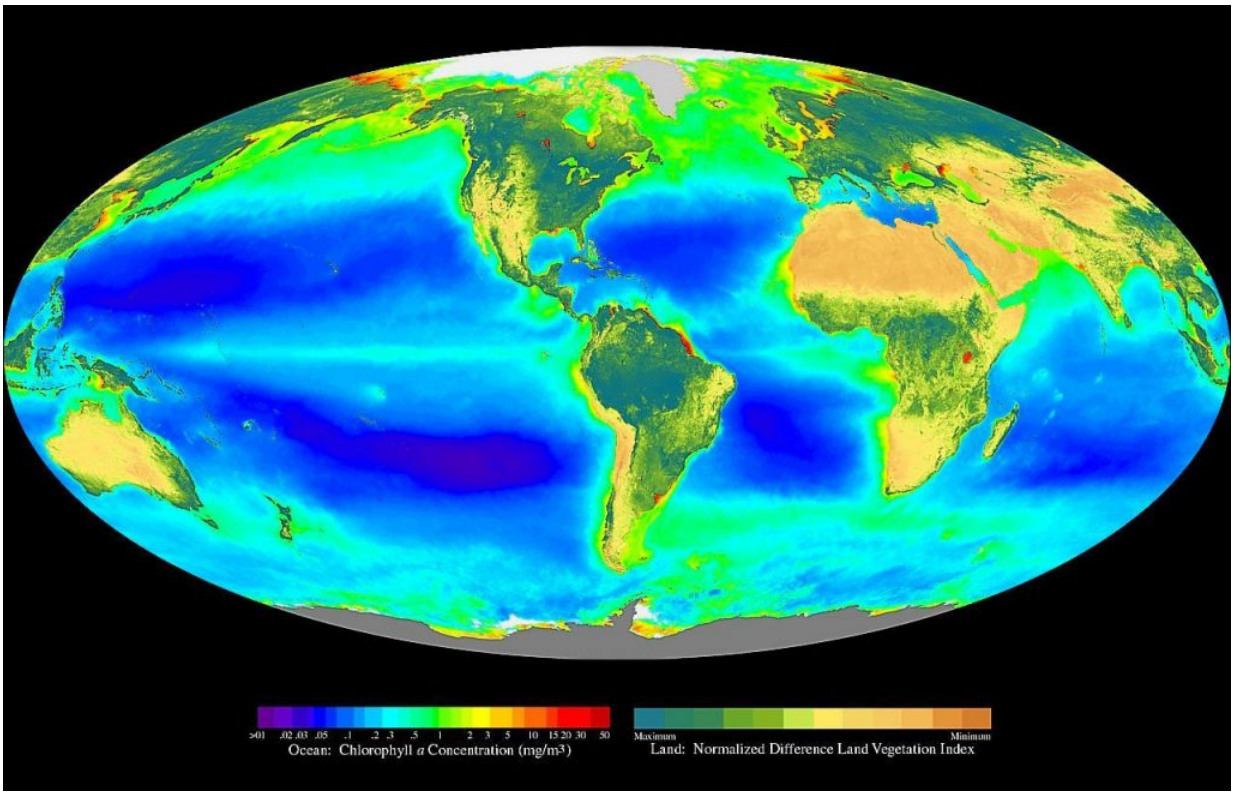


Figure 5.4.1.4.1 Global surface ocean primary productivity, as measured by chlorophyll concentration (Provided by the SeaWiFS Project, Goddard Space Flight Center and ORBIMAGE [Public domain], via Wikimedia Commons).

Regional and seasonal changes in primary production are due to a combination of the availability of light and the amount of nutrients provided by water mixing above the thermocline. In tropical regions sunlight is plentiful throughout the year, so light is not a limiting factor. The surface water is always warm and there is always a pronounced thermocline, leading to highly stratified water that prevents the nutrient-rich bottom water from reaching the surface (section 6.2). Thus productivity in tropical water is always nutrient-limited, and productivity is low throughout the year (Figure 5.4.1.4.2). Because tropical water is nutrient-poor with little phytoplankton production, the water is very clear, as is the case with water in the central ocean.

At the poles, the water is uniformly cold at all depths, so there is not much of a thermocline and little stratification, allowing mixing to occur year-round (section 6.2). This mixing distributes nutrients throughout the water column, so that for much of the year productivity will not be nutrient-limited. However, the polar regions may experience several months with little or no light during the winter, and the fluctuation in light levels leads to variation is seasonal productivity. In the winter months, mixing is occurring and nutrients are abundant, but there is no light, so there is no productivity. By late spring the sunlight returns, and combined with the abundance of nutrients, a spring/summer bloom of phytoplankton occurs (Figure 5.4.1.4.2). By late summer, the

nutrients have been depleted and zooplankton have been grazing on the phytoplankton, so the bloom begins to decline. In the autumn, light levels decline and prevent further production throughout the winter. But during the winter the mixing is distributing nutrients throughout the water, ready for the sun to return and stimulate a bloom in the following summer.

In temperate regions there is much more seasonal variation in the depth and intensity of the thermocline (see section 6.2). The thermocline is shallower and stronger in the summer, and is deeper and weaker in the winter, so mixing of deep and surface water is more pronounced in the winter months. As with the poles, this winter mixing creates nutrient-rich water during the winter, but the lack of light limits winter productivity. When light levels increase in the spring, the combination of abundant light and nutrients creates a spring bloom of productivity. By late summer there is still plenty of light, but the nutrients have been depleted by the phytoplankton bloom, and the summer thermocline has prevented further mixing, so productivity declines. In the autumn, cooler temperatures weaken the thermocline, and increasing storms cause a deeper mixed layer to form, bringing nutrients back to the surface. At the same time, there is still sufficient light available that a smaller autumn bloom occurs (Figure 5.4.1.4.2). But this bloom is short-lived, as light declines throughout the autumn and into the winter. Again, there is little production during the winter due to light limitations, but the winter storms and deep thermocline recharge the water with nutrients for the next spring bloom.

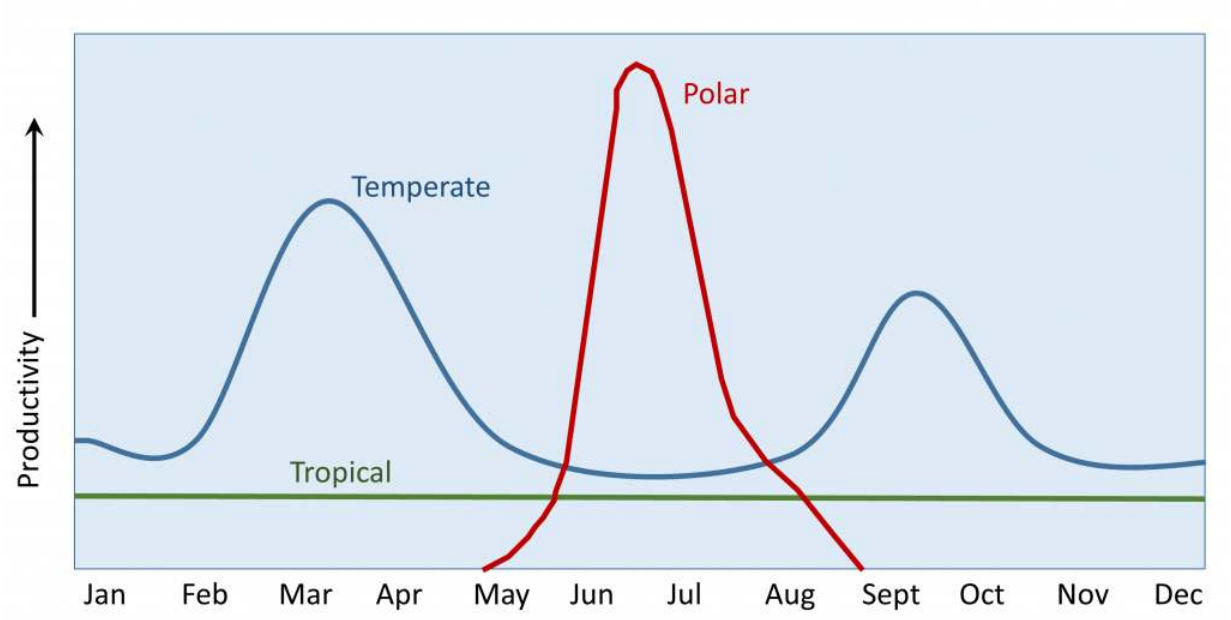


Figure 5.4.1.4.2 Seasonal patterns of productivity in the Northern Hemisphere. Tropical regions are always nutrient-limited and show low productivity. Polar regions are light limited in the winter and only display production during the late spring and summer months when light is available. Northern temperate regions have a spring bloom, and a smaller autumn bloom (PW).

This page titled 5.4.1.4: Patterns of Primary Production is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Paul Webb via source content that was edited to the style and standards of the LibreTexts platform.

- **7.4: Patterns of Primary Production** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

5.4.2: Surface Primary Production

Net primary production (NPP) is the amount of biomass or carbon produced by primary producers per unit area and time, obtained by subtracting plant respiratory costs (Rp) from gross primary productivity (GPP) or total photosynthesis. The term net ecosystem production (NEP) is used to express net carbon accumulation by ecosystems, which is obtained by subtracting the respiratory costs of all organisms (Rall), including plants, grazers and microbes, from GPP. When NEP is positive (i.e. GPP > Rall), there is a net gain of carbon in the ecosystem.



Figure 5.4.2.1: Erosion on hillsides. Extreme topsoil erosion in Iceland. An extreme example of near total loss of topsoil having huge impacts on NPP. In Iceland, thick Andosoils have been removed by the forces of wind and water leaving shallow and poor soils with limited vegetation cover.
Source: Zinneke. Wikimedia Commons CC-BY.

Terrestrial ecosystems are dependent on the sun's energy to support growth and maintenance. Plants are primary producers that, through photosynthesis, manufacture organic molecules such as carbohydrates and lipids from raw inorganic materials (CO_2, water, mineral nutrients). Primary productivity is thus a fundamental determinant of both the structure and functioning of terrestrial biomes. The energy and carbon of primary production supplies consumers, including humans, with the necessary fuel to support their metabolism while providing essential carbon compounds that form the bricks and mortar of living cells. In addition to solar radiation, the main abiotic factors that affect rates of photosynthesis and NPP are water, temperature, carbon dioxide concentration, and nutrients. Globally, there is broad equilibrium relationship between NPP, temperature and precipitation that is strongly impacted by nutrient limitations and, hence, knowledge of the global distribution of NPP is important for understanding vegetation dynamics in biomes, patterns of biodiversity, potential agricultural yield, and predicting global climatic changes.

Some of the key factors that lead to long-term reductions in NPP are associated with various forms of land degradation. This includes soil erosion (wind, water), nutrient depletion, salinisation, soil compaction and crusting, topsoil losses and nutrient depletion, alterations of vegetation composition and structure, and water depletion. In fact, most types of soil disturbance can have devastating impacts on a region’s productivity. One of the most extreme examples of soil loss’s impacts on productivity is Iceland. Icelandic agriculture is primarily based on sheep farming, dairy and poultry production. Due to numerous factors, but especially poor grazing management and wood harvesting, the entire topsoil has been lost over vast areas of the island. Numerous biogeochemical processes affect the carbon balance of terrestrial biomes, including photosynthesis, plant respiration, microbial respiration, leaching losses, erosion, herbivory, fire, and rates of rock weathering. Human appropriation of NPP and modification of the Earth's surface over the past several centuries has altered many of these processes. Net biome production (NBP), which applies to changes in carbon stocks over large spatial areas and long time periods, is helpful to describe changes in carbon balances after losses due to natural or anthropogenic disturbances. NBP balances carbon emissions with non-respiratory losses such as fire,

deforestation, insect infestation, droughts, agricultural harvests, and ecosystem respiration fluxes with NPP, and indicates the carbon source-sink strength and possible positive feedback flux to atmospheric CO2 concentrations. Climate change and human-induced degradation may well result in an increased or decreased NBP. Global environmental change is rapidly altering the dynamics of terrestrial biomes. This has major consequences for the functioning and structure of the Earth system, including the provision of ecosystem services. Long-term satellite measurements have identified a widespread greening of the Earth. The four key drivers of this change are the fertilisation effects of atmospheric CO2 (this explains 70 % of the observed greening trend), nitrogen deposition (9 %), climate change (8 %) and land cover change (4 %). Global CO2 concentrations have risen from about 280 ppm at the start of the industrial revolution to about 406 ppm in 2017.

Elevated carbon dioxide concentrations have numerous effects on plants, such as acting as a fertiliser that stimulates increases in photosynthesis. Elevated CO2 concentrations also tend to reduce water loss in plants, which may be more important than the direct effect of increased photosynthesis rates due to global trends in changing aridity. There is evidence that the CO2 fertilisation effect can change plant-species mix (such as enhancing woody plant growth over grass growth), lower the carbon-to-nitrogen ratio of plants (making grazing and browse material less palatable), and evoke long-term evolutionary responses. It is therefore likely that increased NPP due to increased CO2 concentrations may be changing useful ecosystem services in some areas, despite the higher plant productivity. However, generalisations are difficult to make because there are many feedbacks and interactions with other variables, such as temperature, nutrients, water availability, and plant-plant competition.

The direct CO2 effect on plants should be most strongly expressed in warm, arid environments where water is the dominant limitation to vegetation growth and where land degradation is widespread. Indeed, it has been shown that the 14 % increase in atmospheric CO2 (covering the period 1982–2010) led to a 5-10 % increase in green foliage in warm, arid environments. Can global increases in NPP mask the impacts of degradation? Long-term change in NPP is potentially a useful indicator of land degradation, but interpreting short-term changes in NPP as degradation can be misleading because it can be a reflection of climatic fluctuations. Scientists have attempted to use rain use efficiency (NPP per unit of water) to better understand degradation tends, which have been most successful over long time periods. Also, in combination with climate impact, anthropogenic land use can increase NPP but mask other forms of degradation. Highly productive cultivation systems may increase NPP but affect other ecosystem services, such as water and nutrient supply. Grazing pressure can cause a species change to, for example, woody shrubs showing an increase in NPP but reduce palatability and biodiversity. Hence, it is important to include NPP changes in land degradation assessment but these changes can only be interpreted correctly when considering the dynamics of other ecosystem services and socio-economic situations.



Figure 5.4.2.1: A flux tower in Kruger National Park South Africa, measures the “breathing” of an African savanna ecosystem. During summer days, CO2 is taken up by the vegetation leading to an increase in overall biomass. Some of this biomass is lost to herbivory and microbial decay, but most is lost to the frequent fires that are an integral component of the ecosystem. An important question in the context of climate change is whether this system is in long-term equilibrium, or if the carbon store is increasing or decreasing. The flux tower helps scientists understand the dynamics of the system, and trends in carbon dynamics, as well as being a useful tool in calibrating satellite-based NPP products.
Source: photo from and text based on: https://www.csir.co.za/eddy-covariance-flux-towers

Biomass production is the most important process of the biosphere. It directly impacts many ecosystem services, such as the global carbon cycle, which in turn affects the water cycle and climate. In fact, many ecosystem services are positively correlated with net primary production (NPP), including food production, climate regulation, purification of water, maintenance of nutrients, healthy soils, carbon sinks, biodiversity, and aesthetic landscape function. NPP dictates the amount of carbon synthesised within an ecosystem, which is ultimately available to consumers, including humans. In fact, associated with increased population growth over the last millennium, a disproportionate amount of the world’s NPP is now consumed by humans.

Humans have major impacts on NPP through the use of irrigation and fertilisers. It is the loss of NPP through actions such as increased soil erosion, deforestation and soil salinisation that forms the basis for many forms of land degradation. There are also forms of degradation where NPP may stay constant, or even increase, but where important ecosystem services change. Examples include plant species, compositional changes in response to grazing pressure where palatable grasses are replaced by less palatable ones, or in some cases where palatable grasses are replaced by unpalatable woody shrubs. In such circumstances the grazing capacity of the rangeland may be greatly reduced, biodiversity can be lost, but carbon sequestration and other regulating ecosystem services may be maintained or even enhanced. CO2 fertilisation is having an impact on Leaf Area Index (LAI) (and hence NPP) over vast areas of the Earth. 70 % of the observed impacts on LAI to CO2 fertilisation effects, with factors such as nitrogen

deposition, climate change-induced rainfall and temperature as well as land-cover change being responsible for the remaining observed changes.

Source:

World Atlas of Desertification, European Commission, Net Primary Production Accessed December 2023 from https://wad.jrc.ec.europa.eu/primaryproduction

---

5.4.2: Surface Primary Production is shared under a Public Domain license and was authored, remixed, and/or curated by World Atlas of Desertification, European Commission.

5.5: Net ecosystem production and Eddy-covariance studies

Net ecosystem production (NEP) in ecology, limnology, and oceanography, is the difference between gross primary production (GPP) and net ecosystem respiration.[1] Net ecosystem production represents all the carbon produced by plants in water through photosynthesis that does not get respired by animals, other heterotrophs, or the plants themselves.

Net ecosystem production describes the total carbon in an ecosystem that can be stored, exported, or oxidized back into carbon dioxide gas. NEP is written in units of mass of carbon per unit area per time, for example, grams carbon per square meter per year (g C m−2 yr−1). In a given ecosystem, carbon quantified as net ecosystem production can eventually end up: oxidized by fire or ultraviolet radiation, accumulated as biomass, exported as organic carbon to another system, or accumulated in sediments or soils. Carbon classified as NEP can be in the form of particles in the particulate organic carbon (POC) pool such as phytoplankton cells (living) and detritus (non-living), or it can be in the form of dissolved substances that have not yet been decomposed in the dissolved organic carbon (DOC) pool.[2] In any form, if the carbon gets respired or decomposed by any living organism (plant, animal, bacteria, or other microscopic organism) to release carbon dioxide, that carbon no longer counts as NEP.[1]

- NEP = GPP - respiration [by plants] - respiration [by animals and other heterotrophs]

Net ecosystem production is all the carbon not respired, including respiration by plants and heterotrophic organisms such as animals and microbes. In contrast, net primary production (NPP) is all the carbon taken up by plants (autotrophs) minus the carbon that the plants themselves respire through cellular respiration.

- NPP = GPP - respiration [by plants]

1. *Lovett, Gary M.; Cole, Jonathan J.; Pace, Michael L. (2006-02-01). "Is Net Ecosystem Production Equal to Ecosystem Carbon Accumulation?". Ecosystems. **9** (1): 152–155. doi:10.1007/s10021-005-0036-3. ISSN 1435-0629. S2CID 5890190.*
2. *^ Chester, Roy; Jickells, Tim (2012). Marine Geochemistry. John Wiley & Sons. doi:10.1002/9781118349083. ISBN 9781118349083.*

Eddy Covariance

The **eddy covariance** (also known as **eddy correlation** and **eddy flux**) is a key atmospheric measurement technique to measure and calculate vertical turbulent fluxes within atmospheric boundary layers. The method analyses high-frequency wind and scalar atmospheric data series, gas, energy, and momentum,[1] which yields values of fluxes of these properties. It is a statistical method used in meteorology and other applications (micrometeorology, oceanography, hydrology, agricultural sciences, industrial and regulatory applications, etc.) to determine exchange rates of trace gases over natural ecosystems and agricultural fields, and to quantify gas emissions rates from other land and water areas. It is frequently used to estimate momentum, heat, water vapour, carbon dioxide and methane fluxes.[2][3][4][5][6][7]

The technique is also used extensively for verification and tuning of global climate models, mesoscale and weather models, complex biogeochemical and ecological models, and remote sensing estimates from satellites and aircraft. The technique is mathematically complex, and requires significant care in setting up and processing data. To date,[*when?*] there is no uniform terminology or a single methodology for the eddy covariance technique, but much effort is being made by flux measurement networks (e.g., FluxNet, Ameriflux, ICOS, CarboEurope, Fluxnet Canada, OzFlux, NEON, and iLEAPS) to unify the various approaches.



Figure 5.5.1: An eddy correlation instrument measuring oxygen fluxes in benthic environments.

The technique has additionally proven applicable under water to the benthic zone for measuring oxygen fluxes between the sea floor and overlying water.[8] In these environments, the technique is generally known as the eddy correlation technique, or just eddy correlation. Oxygen fluxes are extracted from raw measurements largely following the same principles as used in the atmosphere, and they are typically used as a proxy for carbon exchange, which is important for local and global carbon budgets. For most

benthic ecosystems, eddy correlation is the most accurate technique for measuring *in-situ* fluxes. The technique's development and its applications under water remains a fruitful area of research.[9][10][11][12][13]

General principles

Representation of the air flow in the atmospheric boundary layer

Air flow can be imagined as a horizontal flow of numerous rotating eddies, that is, turbulent vortices of various sizes, with each eddy having horizontal and vertical components. The situation looks chaotic, but vertical movement of the components can be measured from the tower.



Figure 5.5.2:

Physical meaning

At one physical point on the tower, at time 1, eddy 1 moves parcel of air c_1 down at speed w_1. Then, at time 2, eddy 2 moves parcel c_2 up at speed w_2. Each parcel has gas concentration, pressure, temperature, and humidity. If these factors, along with the speed are known, we can determine the flux. For example, if one knew how many molecules of water went down with eddies at time 1, and how many molecules went up with eddies at time 2, at the same point, one could calculate the vertical flux of water at this point over this time. So, vertical flux can be presented as a covariance of the vertical wind velocity and the concentration of the entity of interest.

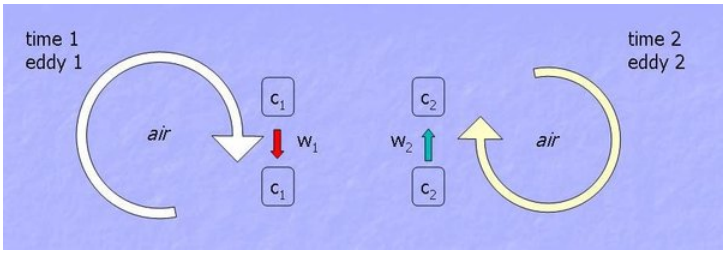


Figure 5.5.3:

Summary

The 3D wind and another variable (usually gas concentration, temperature or momentum) are decomposed into mean and fluctuating components. The covariance is calculated between the fluctuating component of the vertical wind and the fluctuating component of gas concentration. The measured flux is proportional to the covariance.

The area from which the detected eddies originate is described probabilistically and called a flux footprint. The flux footprint area is dynamic in size and shape, changing with wind direction, thermal stability and measurements height, and has a gradual border.

The effect of sensor separation, finite sampling length, sonic path averaging, as well as other instrumental limitations, affect frequency response of the measurement system and may need a co-spectral correction, especially noticeable with closed-path instruments and at low heights below 1 to 1.5 m.

Major assumptions

- Measurements at a point can represent an upwind area
- Measurements are done inside the boundary layer of interest
- Fetch/flux footprint is adequate – fluxes are measured only at area of interest
- Flux is fully turbulent – most of the net vertical transfer is done by eddies
- Terrain is horizontal and uniformed: average of fluctuations is zero; density fluctuations negligible; flow convergence & divergence negligible
- Instruments can detect very small changes at high frequency, ranging from minimum of 5 Hz and to 40 Hz for tower-based measurements

1. *Liang, Shunlin; Li, Xiaowen; Wang, Jindi, eds. (2012-01-01), "Chapter 16 - Vegetation Production in Terrestrial Ecosystems", Advanced Remote Sensing, Academic Press, pp. 501–531, doi:10.1016/b978-0-12-385954-9.00016-2, ISBN 978-0-12-385954-9, retrieved 2020-03-12*
2. Baldocchi, D., B. Hicks, and T. Meyers. 1988. Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. Ecology 69, 1331-1340
3. Verma, S.B.: 1990, Micrometeorological methods for measuring surface fluxes of mass and energy, Remote Sensing Reviews 5(1): 99-115
4. Lee, X., W. Massman, and B. Law. 2004. Handbook of Micrometeorology. Kluwer Academic Publishers, The Netherlands, 250 pp.
5. Burba, G., 2013. Eddy Covariance Method for Scientific, Industrial, Agricultural and Regulatory Applications: a Field Book on Measuring Ecosystem Gas Exchange and Areal Emission Rates. LI-COR Biosciences, Lincoln, USA, 331 pp.
6. Aubinet, M., T. Vesala, D. Papale (Eds.), 2012. Eddy Covariance: A Practical Guide to Measurement and Data Analysis. Springer Atmospheric Sciences, Springer Verlag, 438 pp.
7. *Burba, George (2022-09-06). Eddy Covariance Method For Scientific, Regulatory, and Commercial Applications. LI-COR Biosciences. ISBN 978-0-578-97714-0.*
8. Berg, P., H. Røy, F. Janssen, V. Meyer, B. B. Jørgensen, M. Hüttel, and D. de Beer. 2003. Oxygen uptake by aquatic sediments measured with a novel non-invasive eddy correlation technique. Marine Ecology Progress Series. 261:75-83.
9. University of Virginia. Aquatic Eddy Covariance Research Lab. Retrieved: 22 June 2015.
10. The Florida State University. Eddy Correlation - Further Development and Studies of Flow and Light driven dynamics of Benthic Oxygen Exchange Archived 2014-04-18 at the Wayback Machine. Retrieved: 22 June 2015.
11. Leibniz-Institute of Freshwater Ecology and Inland Fisheries. Eddy Correlation in Natural Waters. Retrieved: 22 June 2015.
12. Max Planck Institute for Marine Microbiology. Eddy Correlation System (ECS). Retrieved: 22 June 2015.
13. Centre for Coastal Biogeochemistry Research. Eddy Correlation Archived 2013-12-13 at the Wayback Machine. Retrieved: 22 June 2015.

Excerpted from:

Wikipedia, https://en.Wikipedia.org/wiki/Eddy_covariance, November 2023

Wikipedia, https://en.Wikipedia.org/wiki/Net_ec...tem_production, November 2023

5.5: Net ecosystem production and Eddy-covariance studies is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Wikipedia.

5.6: Remote sensing of primary production and biomass

Understanding regional variability in carbon cycle processes requires a dramatically more spatially detailed analysis of global land surface processes than was available to climate scientists before 2000. In March 2000, the Moderate Resolution Imaging Spectroradiometer (MODIS), aboard NASA's Terra and Aqua satellites, began producing a regular global estimate of near-weekly gross photosynthesis (GPP) and annual net primary production (NPP) of the entire terrestrial Earth surface at up to 1 km spatial resolution. However, ecosystem respiration, and therefore the full terrestrial carbon balance, cannot be estimated from satellite measurements, but only from models or ground-based measurements.

The GPP and NPP products are designed to provide an accurate, regular measure of the production activity or growth of terrestrial vegetation. (Note: only NPP is available in the NEO site.) These products have both theoretical and practical utility. The theoretical use is primarily for defining the seasonally dynamic terrestrial surface carbon dioxide uptake for global carbon cycle studies such as answering the "missing sink question" of carbon. The spatial and seasonal dynamics of carbon dioxide flux are also of high interest in global climate modeling because carbon dioxide is an important greenhouse gas. Both GPP and RE, and therefore the terrestrial carbon balance, can be modeled. Currently, global carbon cycle models are being integrated with climate models with the goal of creating integrated Earth Systems Models that will represent the dynamic interactions between the atmosphere, biosphere, and oceans. The weekly and monthly MODIS GPP and NPP products are useful to constrain models that are used to answer carbon dioxide flux questions.

The practical utility of these GPP/ NPP products includes routine measurements of crop yield, range forage, and forest production, and other economically and socially significant products of vegetation growth. The value of an unbiased, regular source of crop, range, and forest production estimates for global political and economic decision-making is immense. Moreover, these products are freely available for all users worldwide. This daily computed GPP more correctly defines terrestrial carbon dioxide fluxes than simple Normalized Difference Vegetation Index (NDVI) correlations currently done to increase understanding on how the seasonal fluxes of net photosynthesis are related to seasonal variations of driving factors such as atmospheric carbon dioxide, temperature, light, and precipitation.

Source:

NASA Earth Observatory, Net Primary Production, Accessed December 2023, https://neo.gsfc.nasa.gov/view.php?datasetId=MOD17A3H_Y_NPP

5.6: Remote sensing of primary production and biomass is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5.7: Global estimates of net primary production and biomass

Probably the single most fundamental measure of "global change" of practical interest to humankind is change in terrestrial biological productivity. Biological productivity is the source of all the food, fiber, and fuel that humans survive on, so it defines most fundamentally the habitability of Earth. The spatial variability of net primary productivity (NPP) over the globe is enormous, from about 1000 grams of carbon per cubic meter per year for evergreen tropical rain forests to less than 30 grams of carbon per cubic meter per year for deserts. Primary productivity is largely balanced by ecosystem respiration (RE). The difference between GPP and RE is the terrestrial carbon balance, called net ecosystem productivity (NEP). With increased atmospheric abundance of carbon dioxide, NPP is expected to increase but ongoing global warming is likely to increase RE as well, so that the effect of predicted climate change on the terrestrial carbon uptake or release is largely a matter of ongoing research.

Below are a series of maps showing global net primary production and biomass. Some are linked to interactive maps on other sites.

Terrestrial Net Primary Production

The World Atlas of Desertification has an interactive map of net primary production a snapshot of which is shown in Figure 5.7.1 and is based on data from the NASA Earth Observatory.

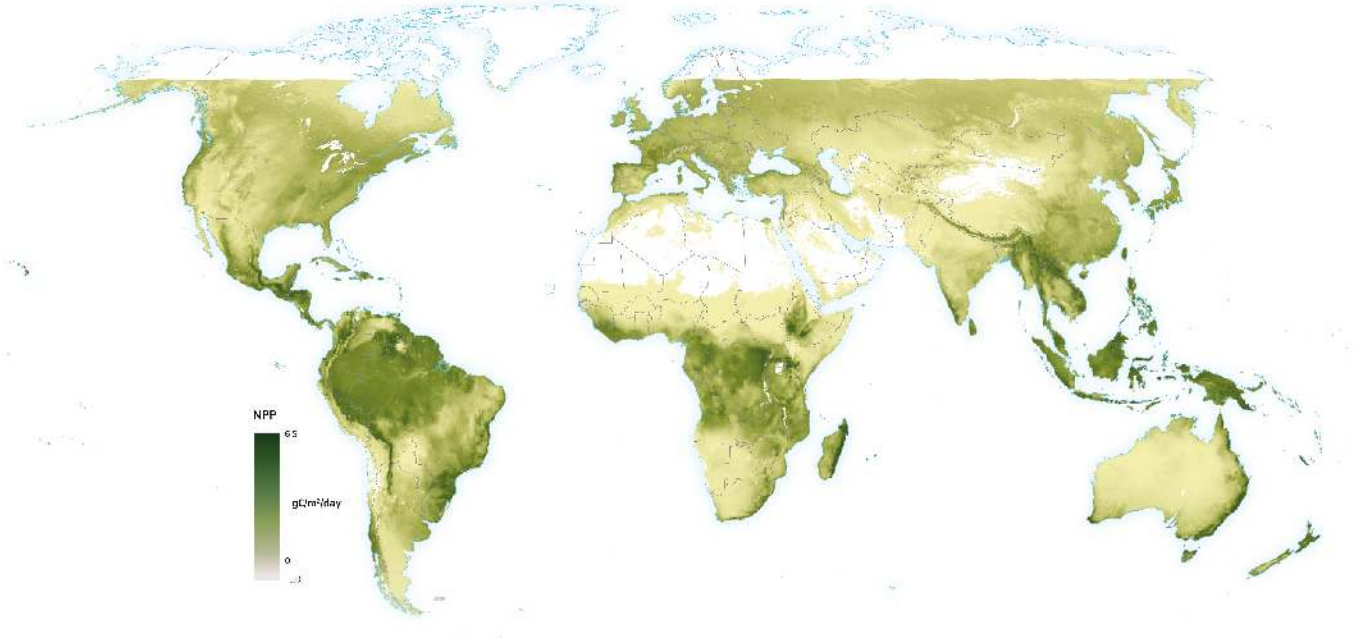


Figure 5.7.1: Net Primary Production. The colours on these maps indicate how fast carbon was taken in for every square metre of land, during the year 2015. Values range from -1.0 grammes of carbon per square metre per day (tan) to 6.5 grammes per square metre per day (dark green). A negative value means decomposition or respiration overpowered carbon absorption; more carbon was released to the atmosphere than the plants took in. These NASA images were made by Reto Stockli, NASA's Earth Observatory Team, using data provided by the MODIS Land Science Team. *Source: NASA images compiled by Reto Stockli, NASA's Earth Observatory Team, using data provided by the MODIS Land Science Team. Public Domain*

Understanding regional variability in carbon cycle processes requires a dramatically more spatially detailed analysis of global land surface processes than was available to climate scientists before 2000. In March 2000, the Moderate Resolution Imaging Spectroradiometer (MODIS), aboard NASA's Terra and Aqua satellites, began producing a regular global estimate of near-weekly gross photosynthesis (GPP) and annual net primary production (NPP) of the entire terrestrial Earth surface at up to 1 km spatial resolution. However, ecosystem respiration, and therefore the full terrestrial carbon balance, cannot be estimated from satellite measurements, but only from models or ground-based measurements.

The GPP and NPP products are designed to provide an accurate, regular measure of the production activity or growth of terrestrial vegetation. (Note: only NPP is available in the NEO site.) These products have both theoretical and practical utility. The theoretical use is primarily for defining the seasonally dynamic terrestrial surface carbon dioxide uptake for global carbon cycle studies such as answering the "missing sink question" of carbon. The spatial and seasonal dynamics of carbon dioxide flux are also of high

interest in global climate modeling because carbon dioxide is an important greenhouse gas. Both GPP and RE, and therefore the terrestrial carbon balance, can be modeled. Currently, global carbon cycle models are being integrated with climate models with the goal of creating integrated Earth Systems Models that will represent the dynamic interactions between the atmosphere, biosphere, and oceans. The weekly and monthly MODIS GPP and NPP products are useful to constrain models that are used to answer carbon dioxide flux questions.

The practical utility of these GPP/ NPP products includes routine measurements of crop yield, range forage, and forest production, and other economically and socially significant products of vegetation growth. The value of an unbiased, regular source of crop, range, and forest production estimates for global political and economic decision-making is immense. Moreover, these products are freely available for all users worldwide. This daily computed GPP more correctly defines terrestrial carbon dioxide fluxes than simple Normalized Difference Vegetation Index (NDVI) correlations currently done to increase understanding on how the seasonal fluxes of net photosynthesis are related to seasonal variations of driving factors such as atmospheric carbon dioxide, temperature, light, and precipitation.

Marine Net Primary Production

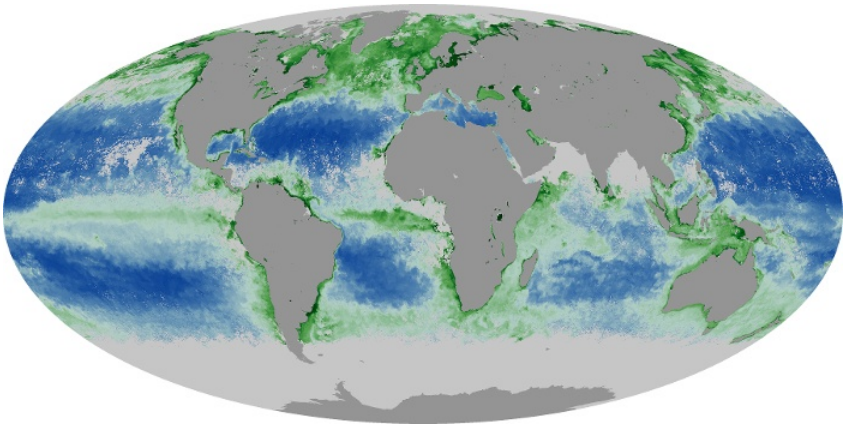


Figure 5.7.2: Map of cholophyll from space via the MODIS insturment. Chlorophyll is a tracer for phytoplankton, the base of the marine food web and an indicator of net primary production. *Source: NASA Earth Observatory Team, using data provided by the MODIS Ocean Science Team. Public Domain*

At the base of the ocean food web are single-celled algae and other plant-like organisms known as phytoplankton. Like plants on land, phytoplankton use chlorophyll and other light-harvesting pigments to carry out photosynthesis, absorbing atmospheric carbon dioxide to produce sugars for fuel. Chlorophyll in the water changes the way it reflects and absorbs sunlight, allowing scientists to map the amount and location of phytoplankton. These measurements give scientists valuable insights into the health of the ocean environment, and help scientists study the ocean carbon cycle.

These chlorophyll maps show milligrams of chlorophyll per cubic meter of seawater each month. Places where chlorophyll amounts were very low, indicating very low numbers of phytoplankton are blue. Places where chlorophyll concentrations were high, meaning many phytoplankton were growing, are dark green. The observations come from the Moderate Resolution Imaging Spectroradiometer (MODIS) on NASA's Aqua satellite. Land is dark gray, and places where MODIS could not collect data because of sea ice, polar darkness, or clouds are light gray.

The highest chlorophyll concentrations, where tiny surface-dwelling ocean plants are thriving, are in cold polar waters or in places where ocean currents bring cold water to the surface, such as around the equator and along the shores of continents. It is not the cold water itself that stimulates the phytoplankton. Instead, the cool temperatures are often a sign that the water has welled up to the surface from deeper in the ocean, carrying nutrients that have built up over time. In polar waters, nutrients accumulate in surface waters during the dark winter months when plants can't grow. When sunlight returns in the spring and summer, the plants flourish in high concentrations.

A band of cool, plant-rich waters circles the globe at the Equator, with the strongest signal in the Atlantic Ocean and the open waters of the Pacific Ocean. This zone of enhanced phytoplankton growth comes from the frequent upwelling of cooler, deeper water as a result of the dominant easterly trade winds blowing across the ocean surface. In many coastal areas, the rising slope of the sea floor pushes cold water from the lowest layers of the ocean to the surface. The rising, or upwelling water carries iron and

other nutrients from the ocean floor. Cold coastal upwelling and subsequent phytoplankton growth are most evident along the west coasts of North and South America and southern Africa.

Biomass Maps

NASA's Oak Ridge National Laboratory Distributed Active Archive Center has a map of above ground and below ground biomass. Details are provided in the open article at Nature, Scientific Data by Spawn, Sullivan, and Lark.

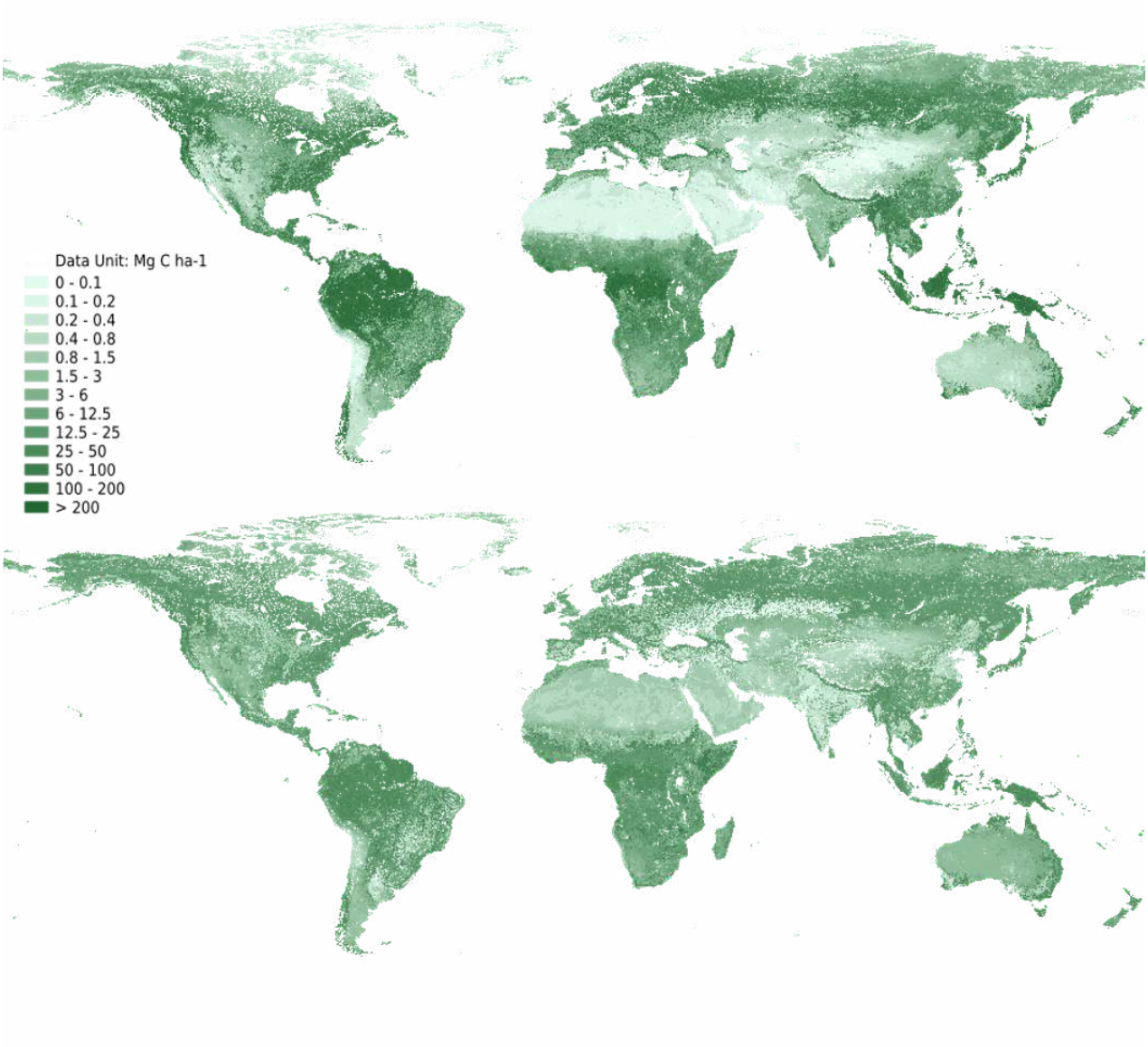


Figure 5.7.3: New maps of above (top) and belowground biomass (bottom) for 2010. Raster values in the maps are in units of megagrams carbon per hectare, with darker green indicating higher carbon density. Image credit: ORNL DAAC Public Domain

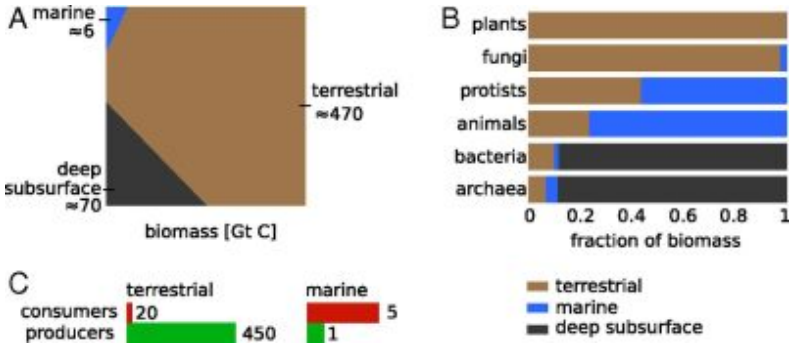


Figure 5.7.1. Biomass distribution of organisms living in different environments (marine -blue-, terrestrial -brown- and deep subsurfaces -black). (A) Absolute biomass is represented using a Voronoi diagram, with the area of each cell proportional to the overall biomass in each environment. (B) Fraction of the biomass of each group concentrated in the terrestrial, marine, or deep subsurface environment. Numbers are expressed in gigatons of carbon. [Source: diagram after Bar-On et al. ref [4]; Open Access article distributed under CC BY-NC-ND 4.0 license]

The differences in global biomass between terrestrial and marine environments are highly marked (Figure 7 and Table 2). The ocean covers 71% of the Earth's surface and occupies a much larger volume than the terrestrial environment, yet the terrestrial biomass, at ≈470 Gt C, is about two orders of magnitude higher than the ≈6 Gt C of the marine biomass, as shown in Figure 7A. In contrast, the primary productivity of the two environments is roughly equal [23].

- For plants, most biomass is concentrated in terrestrial environments (plants are only a small fraction of marine biomass, <1 Gt C, in the form of algae -green and red- and seagrass;).
- For animals, most biomass is concentrated in the marine environment, from fish (≈0.7 Gt C), marine arthropods (≈1 Gt C), molluscs, and annelids.
- For bacteria and archaea, most of the biomass is concentrated in deep subsurface environments [24],[25], such as deep aquifers and the ocean's crust, which may hold the largest aquifer on Earth [26].

Table 5.7.1. *Global biomass of taxa in terrestrial, marine, or deep subsurface environments. [Table based on data from Bar-On et al. ref [4]; Open access article distributed under CC BY-NC-ND 4.0 license]*

Terrestrial		Marine		Deep subsurface	
Taxon	Biomass [Gt C]	Taxon	Biomass [Gt C]	Taxon	Biomass [Gt C]
Plants	450	Fish	0.7		
Soil fungi	12	Marine fungi	0.3		
Soil bacteria	7	Marine bacteria	1.3	Terrestrial deep subsurface bacteria	60
				Marine deep subsurface bacteria	7
Terrestrial protists	1.6	Marine protists	2		
Soil archaea	0.5	Marine archaea	0.3	Terrestrial deep subsurface archaea	4
				Marine deep subsurface archaea	3
Terrestrial arthropods	0.2	Marine arthropods	1		
Wild mammals	0.003	Marine mammals	0,004		
Terrestrial nematodes	0.006	Marine nematodes	0.01		
Annelids	0.2				
Livestock	0.1				
Humans	0.06				
Wild birds	0.002				
		Marine molluscs	0.2		
		Cnidaria	0.1		
Sum	470	Sum	6	Sum	70

However, several of the results in The table should be interpreted with caution due to the large uncertainty associated with some of the estimates, primarily those for total terrestrial protists, marine fungi, and -more broadly- contributions from deep subsurface environments (See Focus How to estimate global biomass?).

The Global Climate Observing System has a map of above ground biomass in global forests

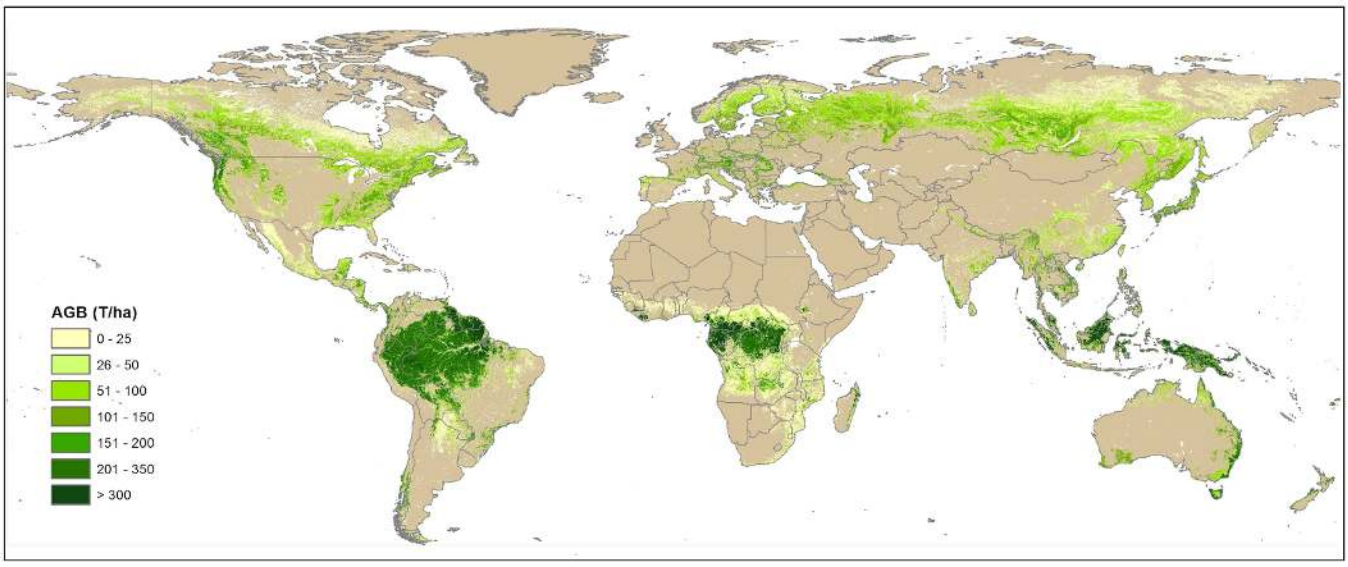


Figure 5.7.5: GEOCARBON global forest above-ground biomass map for 2010 at 0.01° (lucid.wur.nl). Forest areas according to the GLC2000 map (lucid.wur.nl).

References

Spawn, S.A., Sullivan, C.C., Lark, T.J. *et al*. Harmonized global maps of above and below ground biomass carbon density in the year 2010. *Sci Data* **7**, 112 (2020). https://doi.org/10.1038/s41597-020-0444-4

Chlorophyll, *NASA Earth Observatory Team, using data provided by the MODIS Ocean Science Team. Public Domain Accessed December 2023 from https://earthobservatory.nasa.gov/global-maps/MY1DMM_CHLORA*

Net Primary Production, *NASA images compiled by Reto Stockli, NASA's Earth Observatory Team, using data provided by the MODIS Land Science Team. Public Domain. Accessed December 2023 from* https://neo.gsfc.nasa.gov/view.php?datasetId=MOD17A3H_Y_NPP

5.7: Global estimates of net primary production and biomass is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

- **3.1: Biomass of organisms living in different environments** is licensed CC BY-NC-SA 4.0.

SECTION OVERVIEW

5.8: The fate of net primary production

The fate of net primary production is to be eaten, and then, of course, the eaters become the eaten for the next level on the food web, and if not eaten their wastes fertilize plant growth.

5.8.1: Introduction to and Components of Food Webs

5.8.2: Quantifying Food Webs

5.8.3: Trophic Cascades

5.8.4: Sources and Attributions

5.8: The fate of net primary production is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5.8.1: Introduction to and Components of Food Webs

All living things require energy in one form or another. Energy is required by most complex metabolic pathways (often in the form of adenosine triphosphate, ATP), especially those responsible for building large molecules from smaller compounds, and life itself is an energy-driven process. Living organisms would not be able to assemble macromolecules (proteins, lipids, nucleic acids, and complex carbohydrates) from their monomeric subunits without a constant energy input.

Food webs illustrate how energy flows directionally through ecosystems, including how efficiently organisms acquire it, use it, and how much remains for use by other organisms of the food web.

Food Chains and Food Webs

In ecology, a **food chain** is a linear sequence of organisms through which nutrients and energy pass: primary producers, primary consumers, and higher-level consumers are used to describe ecosystem structure and dynamics. There is a single path through the chain.

Food chains do not accurately describe most ecosystems. Even when all organisms are grouped into appropriate trophic levels, some of these organisms can feed on species from more than one trophic level; likewise, some of these organisms can be eaten by species from multiple trophic levels. In other words, the linear model of ecosystems, the food chain, is not completely descriptive of ecosystem structure. A holistic model—which accounts for all the interactions between different species and their complex interconnected relationships with each other and with the environment—is a more accurate and descriptive model for ecosystems. A food web is a graphic representation of a holistic, nonlinear web of primary producers, primary consumers, and higher-level consumers used to describe ecosystem structure and dynamics (Figure 5.8.1.1).

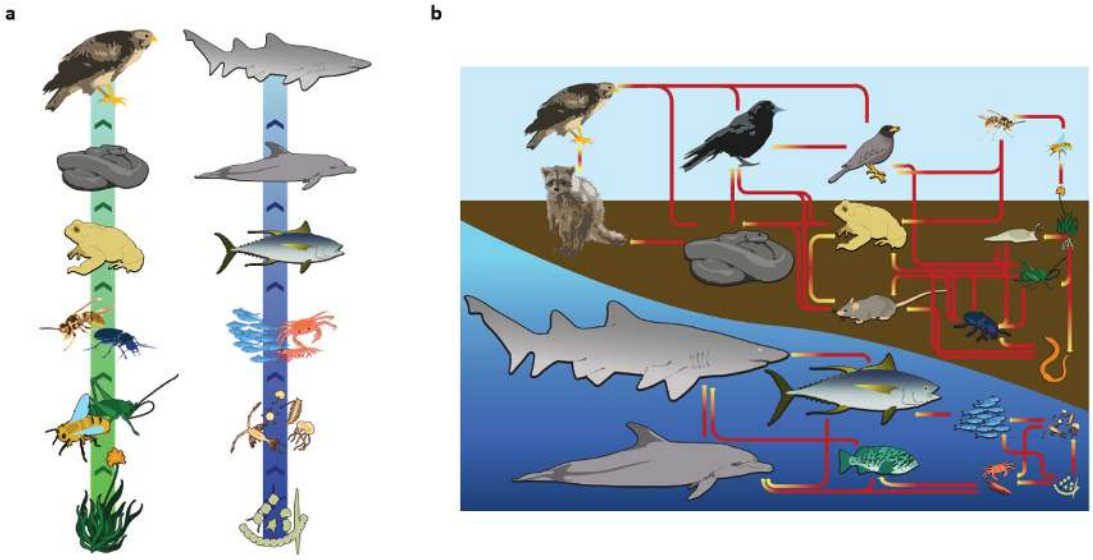


Figure 5.8.1.1: Example of simplified food chains (a) and food webs (b) of terrestrial and marine ecosystems. Developed by LadyofHats and licensed under CC0.

Though more complex than a food chain, a food web remains a simplified illustration of the direct and indirect trophic interactions among species in an ecosystem. Food webs often aggregate many species into trophic groups, which are functional groups of species that have the same predators and prey in a food web. Software can be used to model more complex interactions (Figure 5.8.1.2), but no food web model can capture all of the complexity found within a natural ecosystem.

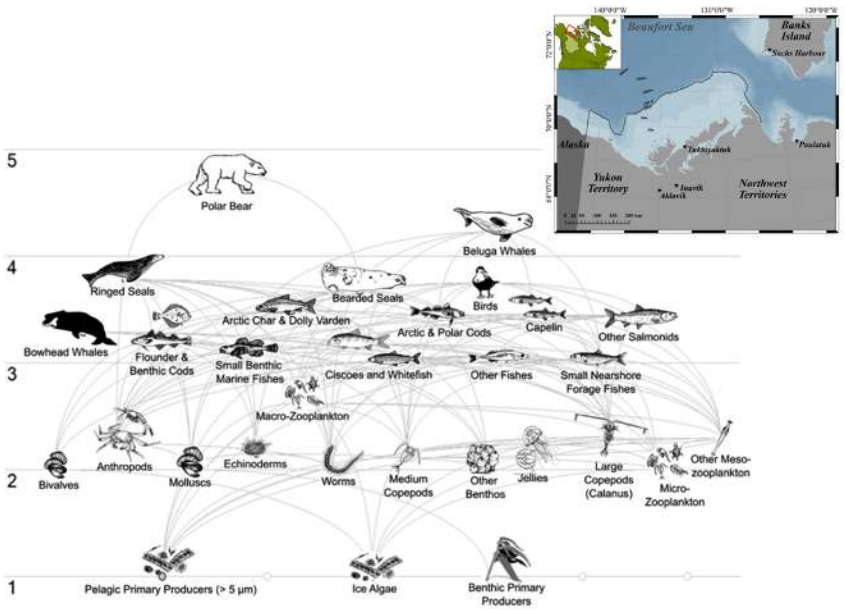


Figure 5.8.1.2**:** An example of a more complex food web developed by Hoover et al. (2021) using a program called Ecopath. This food web depicts trophic relationships among species in the Canadian Beaufort Sea. Horizontal lines represent trophic level. Image licensed under CC-BY 4.0.

Components of a Food Web

The three basic ways in which organisms get food are as producers, consumers, and decomposers.

- Producers (autotrophs) are typically plants or algae. Plants and algae do not usually eat other organisms, but pull nutrients from the soil or the ocean and manufacture their own food using photosynthesis. For this reason, they are called primary producers. In this way, it is energy from the sun that usually powers the base of the food chain (Cengage Learning, 2002). An exception occurs in deep-sea hydrothermal ecosystems, where there is no sunlight. Here primary producers manufacture food through a process called chemosynthesis (van Dover, 2000).
- Consumers (heterotrophs) are species that cannot manufacture their own food and need to consume other organisms. Animals that eat primary producers (like plants) are called herbivores. Animals that eat other animals are called carnivores, and animals that eat both plants and other animals are called omnivores.
- Decomposers (detritivores) break down dead plant and animal material and wastes and release it again as energy and nutrients into the ecosystem for recycling. Decomposers, such as bacteria and fungi (mushrooms), feed on waste and dead matter, converting it into inorganic chemicals that can be recycled as mineral nutrients for plants to use again.

Autotrophs

Energy is acquired by living things in three ways: photosynthesis, chemosynthesis, and the consumption and digestion of other living or previously living organisms by heterotrophs.

Photosynthetic and chemosynthetic organisms are both grouped into a category known as **autotrophs**: organisms capable of synthesizing their own food (more specifically, capable of using inorganic carbon as a carbon source). Photosynthetic autotrophs (**photoautotrophs**) use sunlight as an energy source, whereas chemosynthetic autotrophs (**chemoautotrophs**) use inorganic molecules as an energy source. Autotrophs are critical for all ecosystems. Without these organisms, energy would not be available to other living organisms and life itself would not be possible.

Photoautotrophs, such as plants, algae, and photosynthetic bacteria, serve as the energy source for a majority of the world's ecosystems. Photoautotrophs harness the solar energy of the sun by converting it to chemical energy in the form of ATP (and NADP). The energy stored in ATP is used to synthesize complex organic molecules, such as glucose.

Chemoautotrophs are primarily bacteria that are found in rare ecosystems where sunlight is not available, such as in those associated with dark caves or hydrothermal vents at the bottom of the ocean (Figure 5.8.1.3). Many chemoautotrophs in hydrothermal vents use hydrogen sulfide (H2S), which is released from the vents as a source of chemical energy. This allows chemoautotrophs to synthesize complex organic molecules, such as glucose, for their own energy and in turn supplies energy to the rest of the ecosystem.



Figure 5.8.1.3**:** Swimming shrimp, a few squat lobsters, and hundreds of vent mussels are seen at a hydrothermal vent at the bottom of the ocean. As no sunlight penetrates to this depth, the ecosystem is supported by chemoautotrophic bacteria and organic material that sinks from the ocean's surface. This picture was taken in 2006 at the submerged NW Eifuku volcano off the coast of Japan by the National Oceanic and Atmospheric Administration (NOAA). The summit of this highly active volcano lies 1535 m below the surface.

Not Your Average Food Web: Deep Sea 5.8.1.1

Food webs in the deep sea vary depending on proximity to seamount, hydrothermal vents, and trenches. In areas near hydrothermal vents, chemosynthetic bacteria are the major primary producers. These chemoautotrophs are what provides energy for the rest of the trophic levels in this system.

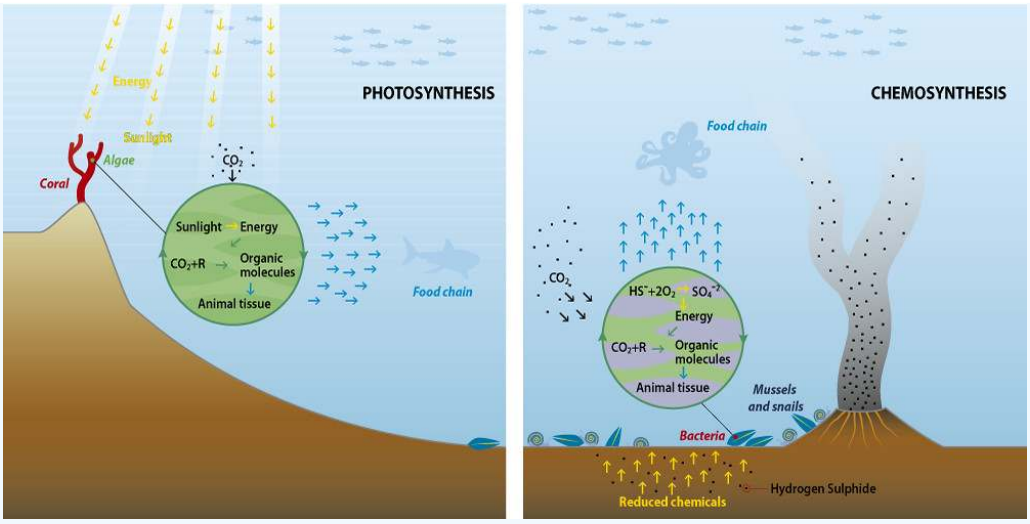


Figure 5.8.1.4**:** A comparison of photosynthetic (left) vs. chemosynthetic (right) food webs. Diagram developed by GRID-Arendal and licensed under CC-SA-NC.

Species in deep-sea ecosystems have adapted to interact with each other in many ways. One key interaction is the symbiosis between many species and chemosynthetic bacteria in hydrothermal vent systems. These bacteria live within the body of species like tubeworms, which are dependent on the bacteria to survive, similar to the relationship between zooxanthellae and coral. Another important type of deep sea community develops when a dead whale (or other large marine organism) carcass sinks to the ocean floor and provides an influx of nutrients. The communities support scavengers like hagfish, opportunists like bristle worms, and eventually enter a sulfophilic stage that appears similar to a hydrothermal vent community.



Figure 5.8.1.5**:**Whale falls serve as an extremely important influx of nutrients to the sun-starved deep ocean. This photo shows a Whale skeleton submerged in Monterey Bay National Marine Sanctuary, covered in octopuses and several other species. Photo by National Marine Sanctuaries is licensed under CC 2.0.

Heterotrophs

Unlike autotrophs, **heterotrophs** consume rather than produce biomass energy as they metabolize, grow, and add to levels of secondary production. A food web depicts a collection of polyphagous heterotrophic consumers that network and cycle the flow of energy and nutrients from a productive base of self-feeding autotrophs (Pimm et al., 1991; Odum & Barrett, 2005; Benke, 2010).

Autotrophs and heterotrophs come in all sizes, from microscopic to many tonnes - from cyanobacteria to giant redwoods, and from viruses to blue whales.

A gradient exists between trophic levels running from complete autotrophs that obtain their sole source of carbon from the atmosphere, to mixotrophs (such as carnivorous plants) that are autotrophic organisms that partially obtain organic matter from sources other than the atmosphere, and complete heterotrophs that must feed to obtain organic matter.

There are different kinds of feeding relations that can be roughly divided into herbivory, carnivory, scavenging and parasitism. Some of the organic matter eaten by heterotrophs, such as sugars, provides energy. An often overlooked but key component of food webs are the decomposers.

Not Your Average Food Web: Wasp-Waist Ecosystems 5.8.1.2

Food webs can be controlled by top-down mechanisms (predator abundance determines the abundance of lower trophic levels), bottom-up mechanisms (primary producer abundance determines the abundance of higher trophic levels), or a combination of both. In wasp-waist food webs, population dynamics are controlled by planktivorous lower trophic level species such as sardine, anchovy, and small squids rather than the bottom or the top (Cury et al., 2011). These lower trophic level species often have high abundance but low diversity. The term “wasp-waist” describes the shape of these food webs, with many species existing at lower trophic levels (i.e., the plankton) and at higher trophic level (i.e., the predators), but very few lower trophic level species linking the plankton and the predators. These lower trophic level species exert top-down control on zooplankton and bottom-up control on top predators, with environmental factors largely affecting their abundance (Cury et al., 2000; Cury et al., 2003). Wasp-waist ecosystems are highly vulnerable to collapse when forage fish decline due to the critical energetic links that they provide between highly available zooplankton and larger predators (Shannon, 2000).

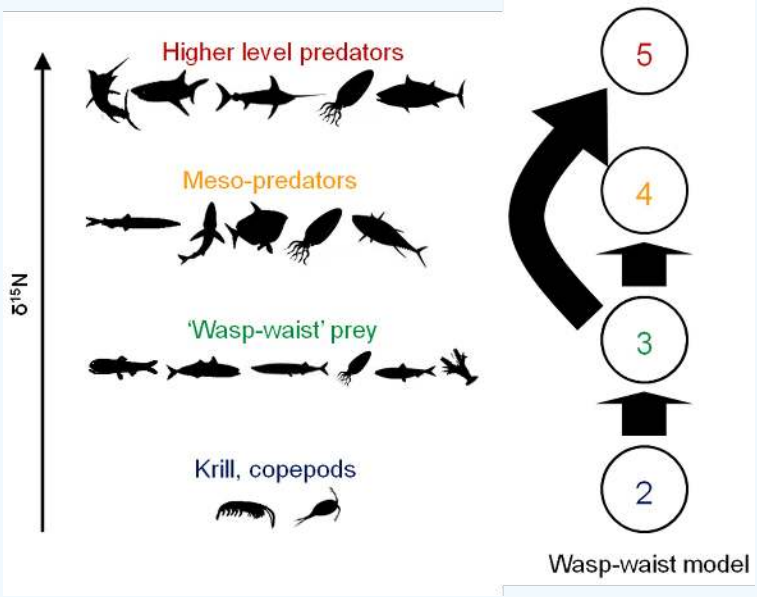


Figure 5.8.1.6**:** A diagram showing the structure of a wasp-waist model for the California Current Large Marine Ecosystem. Arrows indicate inputs of a trophic group to another (credit: modified from Madigan et al., 2012).

Scavengers are animals that consume dead organisms that have died from causes other than predation or have been killed by other predators (Tan & Corlett, 2011). While scavenging generally refers to carnivores feeding on carrion, it is also a herbivorous feeding behavior (Getz, 2011). Scavengers play a fundamental role in the environment through the removal of decaying organisms, serving as a natural sanitation service (Ogada et al., 2011). While microscopic and invertebrate decomposers break down dead organisms into simple organic matter which are used by nearby autotrophs, scavengers help conserve energy and nutrients obtained from

carrion within the upper trophic levels, and are able to disperse the energy and nutrients farther away from the site of the carrion than decomposers (Olson et al., 2016). Decomposers and detritivores complete this process, by consuming the remains left by scavengers. Scavengers are not typically thought to be detritivores, as they generally eat large quantities of organic matter.

Decomposers are often left off food webs, but if included, they mark the end of a food chain (Hutchinson, 2013). Thus food chains start with primary producers and end with decay and decomposers. Since decomposers recycle nutrients, leaving them so they can be reused by primary producers, they are sometimes regarded as occupying their own trophic level (Kane et al., 2016; Pahl & Ruedas, 2021).

Detritivores (also known as **detrivores**, **detritophages**, **detritus feeders**, or **detritus eaters**) are heterotrophs that obtain nutrients by consuming detritus (decomposing plant and animal parts as well as feces) (Wetzel, 2001). By doing so, detritivores contribute to decomposition and to nutrient cycles. Plant tissues are made up of resilient molecules (cellulose, chitin, lignin and xylan) that decay at a much lower rate than other organic molecules. The activity of detritivores is the reason why we do not see an accumulation of plant litter in nature (Keddy, 2017; Sagi et al., 2019). **Coprophagy** is a specific case of detritivory used to describe animals that eat feces (their own, from another individual of their species, or from another species).

Detritivores are an important aspect of many ecosystems. They can live on any type of soil with an organic component, including marine ecosystems, where they are termed interchangeably with bottom feeders. Typical detritivorous animals include millipedes, springtails, woodlice, dung flies, slugs, many terrestrial worms, sea stars, sea cucumbers, fiddler crabs, and some sedentary polychaete worms.

Decomposers are organisms that break down dead or decaying organisms; they carry out decomposition, a process possible by only certain kingdoms, such as fungi (NOAA, 2014). Like herbivores and predators, decomposers are heterotrophic, meaning that they use organic substrates to get their energy, carbon and nutrients for growth and development. While the terms decomposer and detritivore are often interchangeably used, detritivores *ingest* and digest dead matter internally, while decomposers *directly absorb* nutrients through external chemical and biological processes (Keddy, 2017). Thus, invertebrates such as earthworms, woodlice, and sea cucumbers are technically detritivores, not decomposers, since they must ingest nutrients - they are unable to absorb them externally (Sagi et al., 2019).



Figure 5.8.1.7**:**Fungi are the primary decomposers in most environments, illustrated here *Mycena interrupta*. Only fungi produce the enzymes necessary to decompose lignin, a chemically complex substance found in wood.

Not Your Average Food Web: Detrital Web 5.8.1.3

Detritus is dead particulate organic material, as distinguished from dissolved organic material. Detritus typically includes the bodies or fragments of bodies of dead organisms, and fecal material. Detritus typically hosts communities of microorganisms that colonize and decompose (i.e. remineralize) it. In terrestrial ecosystems it is present as leaf litter and other organic matter that is intermixed with soil, which is denominated "soil organic matter". The detritus of aquatic ecosystems is organic material that is suspended in the water and accumulates in depositions on the floor of the body of water; when this floor is a seabed, such a deposition is denominated "marine snow".



Figure 5.8.1.8**:**Earthworms are soil-dwelling detritivores.



Figure 5.8.1.9**:** Two Adonis blue butterflies lap at a small lump of feces lying on a rock.

In a detrital web, plant and animal matter is broken down by decomposers, e.g., bacteria and fungi, and moves to detritivores and then carnivores (Gönenç et al., 2007). There are often relationships between the detrital web and the grazing web. Mushrooms produced by decomposers in the detrital web become a food source for deer, squirrels, and mice in the grazing web. Earthworms are detritivores that consume decaying leaves and are then consumed by a variety of wildlife, especially birds.

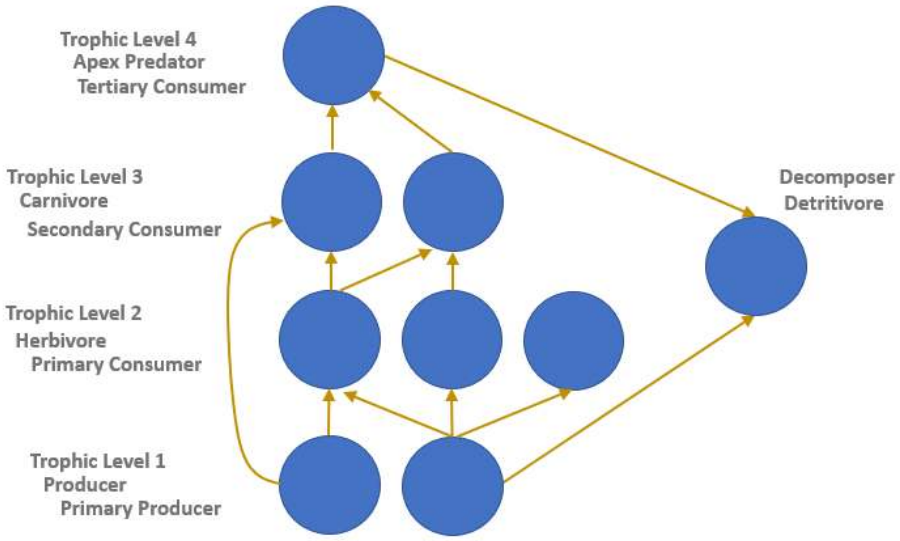


Figure 5.8.1.10**:** Food web diagram showing the various ways in which organism roles can be differentiated. Developed by N. Gownaris.

The **trophic level** of an organism is the position it occupies in a food web. The trophic level of an organism is the number of steps it is from the start of a food chain. A food web starts at trophic level 1 with primary producers, followed by herbivores at level 2, carnivores at level 3 or higher, and apex predators at level 4 or 5. The path along the chain can form either a one-way flow or a food "web". Ecological communities with higher biodiversity form more complex trophic paths. The word trophic derives from the Greek τροφή (trophē) referring to food or nourishment (merriam-webster.com, 2017).

- Level 1: Plants and algae make their own food and are called producers.
- Level 2: Herbivores eat plants and are called primary consumers.
- Level 3: Carnivores that eat herbivores are called secondary consumers.
- Level 4: Carnivores that eat other carnivores are called tertiary consumers.
- Apex predators by definition have no predators and are at the top of their food web.

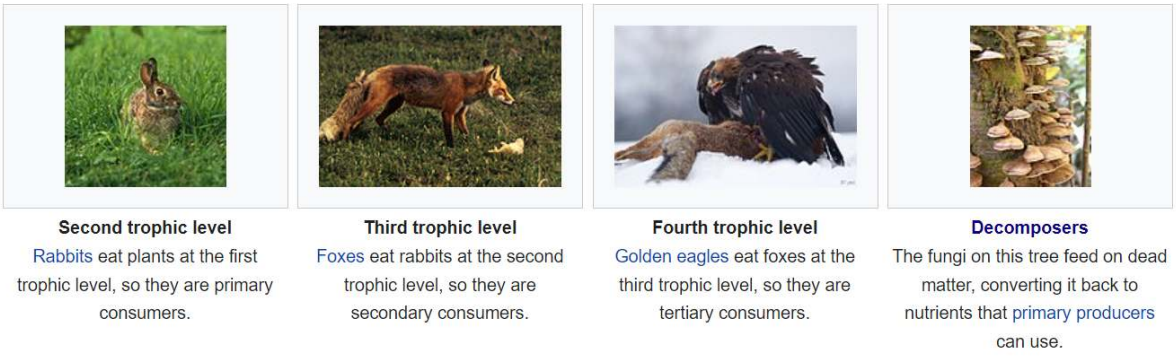


Figure 5.8.1.11**:**Examples of species found at each trophic level of a terrestrial ecosystem.

The trophic level concept was introduced in a historical landmark paper on trophic dynamics in 1942 by Raymond L. Lindeman. The basis of trophic dynamics is the transfer of energy from one part of the ecosystem to another (Odum & Heald, 1975; Cortés, 1999). The trophic dynamic concept has served as a useful quantitative heuristic, but it has several major limitations including the

precision by which an organism can be allocated to a specific trophic level. Omnivores, for example, are not restricted to any single level. Nonetheless, recent research has found that discrete trophic levels do exist, but "above the herbivore trophic level, food webs are better characterized as a tangled web of omnivores." (Pauly et al., 1998).



Figure 5.8.1.12**:**Killer whales (orca) are apex predators but they are divided into separate populations that hunt specific prey, such as tuna, small sharks, and seals.

The fisheries scientist Daniel Pauly sets the values of trophic levels to one in plants and detritus, two in herbivores and detritivores (primary consumers), three in secondary consumers, and so on. The definition of the trophic level, TL, for any consumer species is (Pauly & Palomares, 2005)

$$TL_i = 1 + \sum_j (TL_j \cdot DC_{ij})$$

where

$$TL_j$$

is the fractional trophic level of the prey j, and

$$DC_{ij}$$

represents the fraction of j in the diet of i. That is, the consumer trophic level is one plus the weighted average of how much different trophic levels contribute to its food.

In the case of marine ecosystems, the trophic level of most fish and other marine consumers takes a value between 2.0 and 5.0. The upper value, 5.0, is unusual, even for large fish (Cortés, 1999), though it occurs in apex predators of marine mammals, such as polar bears and orcas (Pauly et al., 1998).

Not Your Average Food Web: Microbial Loop 5.8.1.4

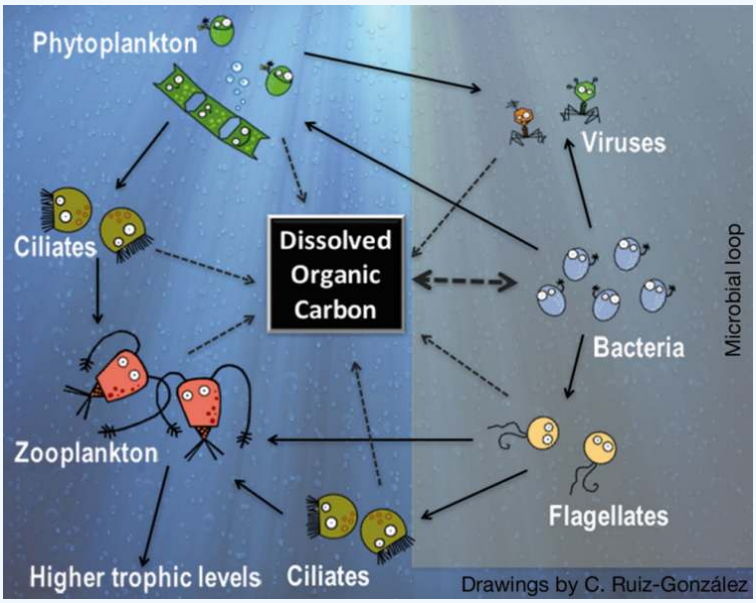


Figure 5.8.1.13**:** Simplified microbial food web in the sunlit ocean by Anders et al. is licensed under CC-BY-SA 4.0. Left side: classic description of the carbon flow from photosynthetic algae to grazers and higher trophic levels in the food chain. Right side: microbial loop, with bacteria using dissolved organic carbon to gain biomass, which then re-enters the classic carbon flow through protists. Based on DeLong & Karl (2005).

The microbial food web refers to the combined trophic interactions among microbes in aquatic environments. These microbes include viruses, bacteria, algae, heterotrophic protists (such as ciliates and flagellates) (Mostajir et al., 2015). Scientists have relatively recently begun to appreciate the importance of this microscopic food web to the functioning of higher trophic levels.

In aquatic environments, microbes constitute the base of the food web. Single celled photosynthetic organisms such as diatoms and cyanobacteria are generally the most important primary producers in the open ocean. Many of these cells, especially cyanobacteria, are too small to be captured and consumed by small crustaceans and planktonic larvae. Instead, these cells are consumed by phagotrophic protists which are readily consumed by larger organisms. Viruses can infect and break open bacterial cells and (to a lesser extent), planktonic algae (a.k.a. phytoplankton). Therefore, viruses in the microbial food web act to reduce the population of bacteria and, by lysing bacterial cells, release particulate and dissolved organic carbon (DOC). DOC may also be released into the environment by algal cells. The microbial loop describes a pathway in the microbial food web where DOC is returned to higher trophic levels via the incorporation into bacterial biomass.

Ecological pyramids

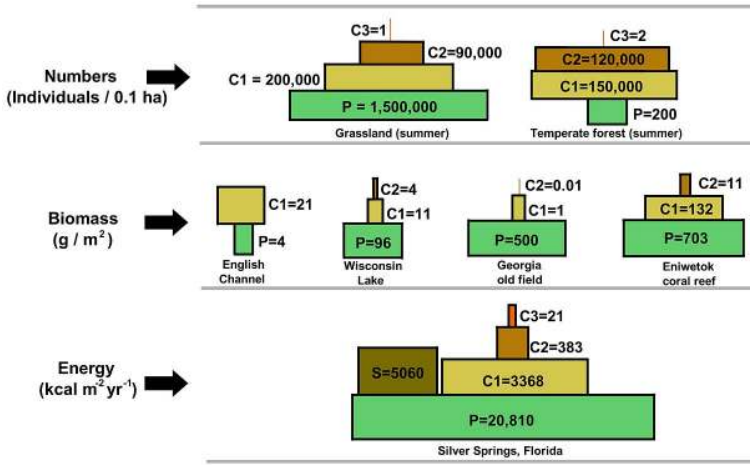


Figure 5.8.1.14**:** Illustration of a range of ecological pyramids, including **top** pyramid of numbers, **middle** pyramid of biomass, and **bottom** pyramid of energy. The terrestrial forest (summer) and the English Channel ecosystems exhibit inverted pyramids.*Note:* trophic levels are not drawn to scale and the pyramid of numbers excludes microorganisms and soil animals. *Abbreviations:* P=Producers, C1=Primary consumers, C2=Secondary consumers, C3=Tertiary consumers, S=Saprotrophs (Odum & Barrett, 2005).

Ecological trophic pyramids are typically one of three kinds: 1) pyramid of numbers, 2) pyramid of biomass, or 3) pyramid of energy (Odum & Barrett, 2005). In a pyramid of numbers, the number of consumers at each level decreases significantly, so that a single top consumer, (e.g., a polar bear or a human), will be supported by a much larger number of separate producers. There is usually a maximum of four or five links in a food chain, although food chains in aquatic ecosystems are more often longer than those on land. Eventually, all the energy in a food chain is dispersed as heat (Odum & Barrett, 2005).

Pyramid structure can vary across ecosystems and across time. In some instances biomass pyramids can be inverted. This pattern is often identified in aquatic and coral reef ecosystems. The pattern of biomass inversion is attributed to different sizes of producers. Aquatic communities are often dominated by producers that are smaller than the consumers that have high growth rates. Aquatic producers, such as planktonic algae or aquatic plants, lack the large accumulation of secondary growth as exists in the woody trees of terrestrial ecosystems. However, they are able to reproduce quickly enough to support a larger biomass of grazers. This inverts the pyramid. Primary consumers have longer lifespans and slower growth rates that accumulate more biomass than the producers they consume. Phytoplankton live just a few days, whereas the zooplankton eating the phytoplankton live for several weeks and the fish eating the zooplankton live for several consecutive years (Spellman, 2008). Aquatic predators also tend to have a lower death rate than the smaller consumers, which contributes to the inverted pyramidal pattern. Population structure, migration rates, and environmental refuge for prey are other possible causes for pyramids with biomass inverted. Energy pyramids, however, will always have an upright pyramid shape if all sources of food energy are included and this is dictated by the second law of thermodynamics (Odum & Barrett, 2005; Wang et al., 2009).

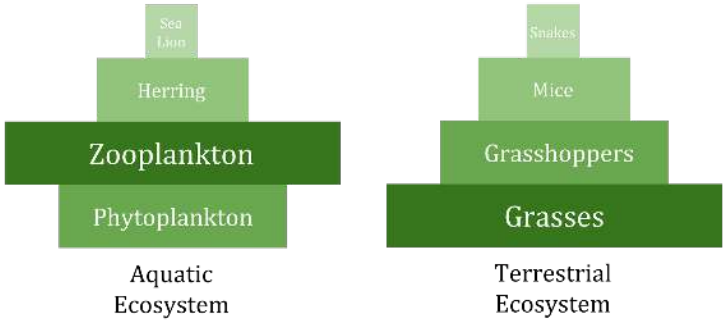


Figure 5.8.1.15**:** A pyramid of biomass shows the total biomass of the organisms involved at each trophic level of an ecosystem. These pyramids are not necessarily upright. There can be lower amounts of biomass at the bottom of the pyramid if the rate of primary production per unit biomass is high.

5.8.1: Introduction to and Components of Food Webs is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

5.8.2: Quantifying Food Webs

Ecologists collect data on trophic levels and food webs to statistically model and mathematically calculate parameters, such as those used in other kinds of network analysis (e.g., graph theory), to study emergent patterns and properties shared among ecosystems. There are different ecological dimensions that can be mapped to create more complicated food webs, including: species composition (type of species), richness (number of species), biomass (the dry weight of plants and animals), productivity (rates of conversion of energy and nutrients into growth), and stability (food webs over time). A food web diagram illustrating species composition shows how change in a single species can directly and indirectly influence many others. The sections below describe three metrics that ecologists use to describe and quantify food webs (Transfer Efficiency, Food Chain Length, Connectance).

Transfer Efficiency

The biomass of each trophic level decreases from the base of the chain to the top. This is because energy is lost to the environment with each transfer as entropy increases. About eighty to ninety percent of the energy is expended for the organism's life processes or is lost as heat or waste. Only about ten to twenty percent of the organism's energy is generally passed to the next organism (Spellman, 2008). The amount can be less than one percent in animals consuming less digestible plants, and it can be as high as forty percent in zooplankton consuming phytoplankton (Kent, 2000). Graphic representations of the biomass or productivity at each trophic level are called ecological pyramids or trophic pyramids. The proportion of energy that is transferred from one trophic level to the next is known as the **transfer efficiency** of a food web.

The **ten percent law** of transfer of energy from one trophic level to the next can be attributed to Raymond Lindeman (1942) (Odum & Barrett, 2005). However, Lindeman did not call it a "law" and cited transfer efficiencies ranging from 0.1% to 37.5%. According to this law, during the transfer of organic food energy from one trophic level to the next higher level, only about ten percent of the transferred energy is stored as flesh. The remaining is lost during transfer, broken down in respiration, or lost to incomplete digestion by higher trophic level.

When organisms are consumed, approximately 10% of the energy in the food is fixed into their flesh and is available for the next trophic level (carnivores or omnivores). When a carnivore or an omnivore in turn consumes that animal, only about 10% of energy is fixed in its flesh for the higher level. Again, it is important to remember that the 10% value represents an estimate and that, in reality, transfer efficiency varies highly among ecosystems, time periods, and trophic levels.

Example: The Sun releases 10,000 J of energy, then plants take only 100 J of energy from sunlight (this is an exception to the 10% rule, since only 1% of energy is taken up by plants from sun); thereafter, a deer would take 10 J (10% of energy) from the plant. A wolf eating the deer would only take 1 J (10% of energy from deer). A human eating the wolf would take 0.1J (10% of energy from wolf), etc.

The ten percent law provides a basic understanding on the cycling of food chains. Furthermore, the ten percent law shows the inefficiency of energy capture at each successive trophic level. The rational conclusion is that energy efficiency is best preserved by sourcing food as close to the initial energy source as possible.

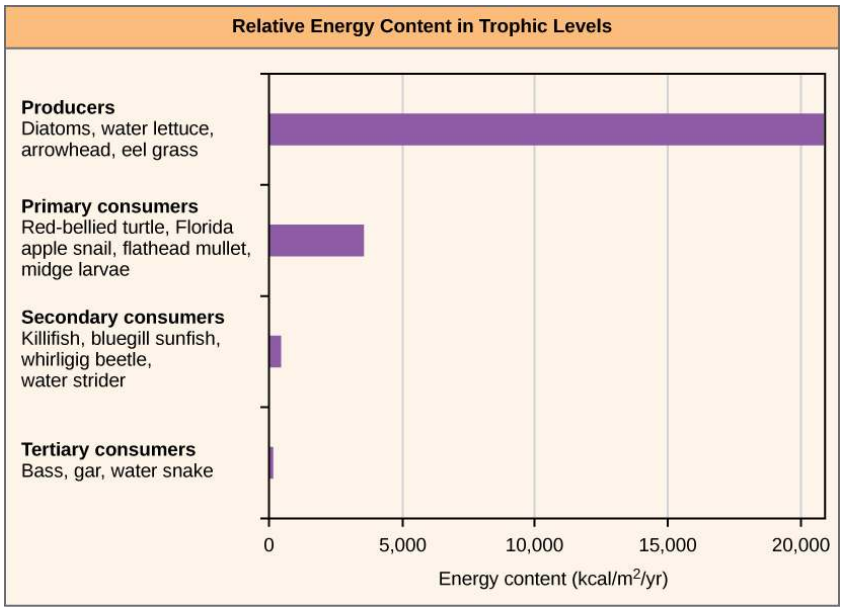


Figure 5.8.2.1**:** The relative energy in trophic levels in a Silver Springs, Florida, ecosystem is shown. Each trophic level has less energy available, and usually, but not always, supports a smaller mass of organisms at the next level.

Food Chain Length

A common metric used to quantify food web trophic structure is food chain length. Food chain length is another way of describing food webs as a measure of the number of species encountered as energy or nutrients move from the plants to top predators (Post, 1993). In its simplest form, the length of a chain is the number of links between a trophic consumer and the base of the web or the maximum trophic level of a species in that food web. The mean chain length of an entire web is the arithmetic average of the lengths of all chains in a food web (Pimm, 1979; Odum & Barrett, 2005). In a simple predator-prey example, a deer is one step removed from the plants it eats (chain length = 1) and a wolf that eats the deer is two steps removed from the plants (chain length = 2). Because energy is lost at each transfer from one trophic level to the next (see Transfer Efficiency), the maximum number of trophic levels is generally constrained to only four or five levels. The number of trophic levels is ultimately determined by the productivity at the base of the food web and by how efficiently this energy is transferred up the food chain. Greater productivity and higher efficiency can support a greater number of trophic levels.

Connectance

Food webs are extremely complex. Complexity is a measure of an increasing number of permutations and it is also a metaphorical term that conveys the mental intractability or limits concerning unlimited algorithmic possibilities. In food web terminology, complexity is a product of the number of species and connectance (Neutel et al., 2002; Leveque, 2003; Proctor et al., 2005). Connectance is "the fraction of all possible links that are realized in a network" (Dunne et al., 2002). These concepts were derived and stimulated through the suggestion that complexity leads to stability in food webs, such as increasing the number of trophic levels in more species rich ecosystems. This hypothesis was challenged through mathematical models suggesting otherwise, but subsequent studies have shown that the premise holds in real systems (Neutel et al., 2002; Banasek-Richter et al., 2009).

While the complexity of real food webs connections are difficult to decipher, ecologists have found mathematical models on networks an invaluable tool for gaining insight into the structure, stability, and laws of food web behaviors relative to observable outcomes. Quantitative formulas simplify the complexity of food web structure. The number of trophic links (tL), for example, is converted into a connectance value:

$$C = L/[S(S-1)/2]$$

where, S(S-1)/2 is the maximum number of binary connections among S species (Paine 1988). "Connectance (C) is the fraction of all possible links that are realized (L/S2) and represents a standard measure of food web complexity..." (Williams et al., 2002).

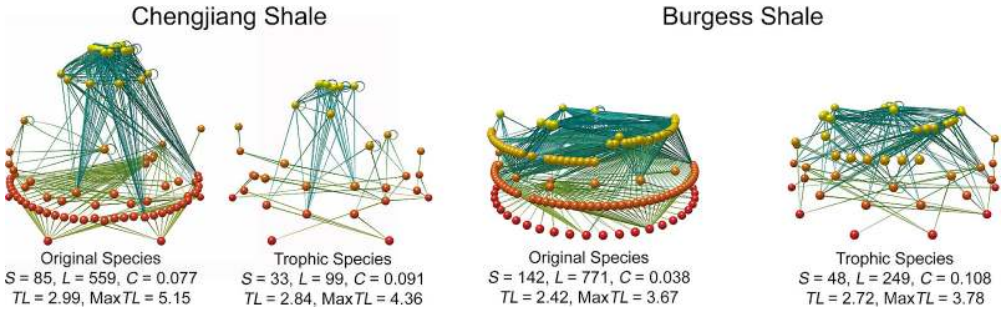


Figure 5.8.2.2**:** Food web and trophic level of the Chengjiang and Burgess Shale. S: number of species (nodes). L: number of trophic links. C: connectance; L/S2. MaxTL: maximum trophic level of a species in the web. Paleoecological studies can reconstruct fossil food-webs and trophic levels. Primary producers form the base (red spheres), predators at top (yellow spheres), the lines represent feeding links. Original food-webs (left) are simplified (right) by aggregating groups feeding on common prey into coarser grained trophic species (Dunne et al., 2008). Figure developed by Dunne et al. 2008 is licensed under CC-BY 4.0.

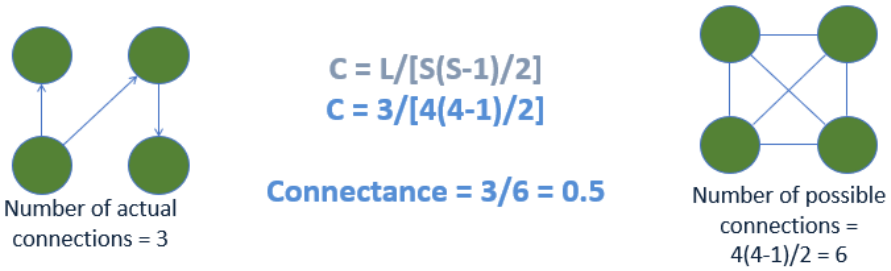


Figure 5.8.2.3**:** Worked example of calculating connectance for a food web. Diagram by N. Gownaris and A. Wilson.

Techniques for Studying Food Webs 5.8.2.1

Food web linkages, or the feeding relationships between species inhabiting a shared ecosystem, are an ecological lens through which ecosystem structure and function can be assessed, and thus are fundamental to informing sustainable resource management. Empirical feeding datasets have traditionally been painstakingly generated from stomach content analysis, direct observations and from biochemical trophic markers (stable isotopes, fatty acids, molecular tools). Each approach carries inherent biases and limitations, as well as advantages.

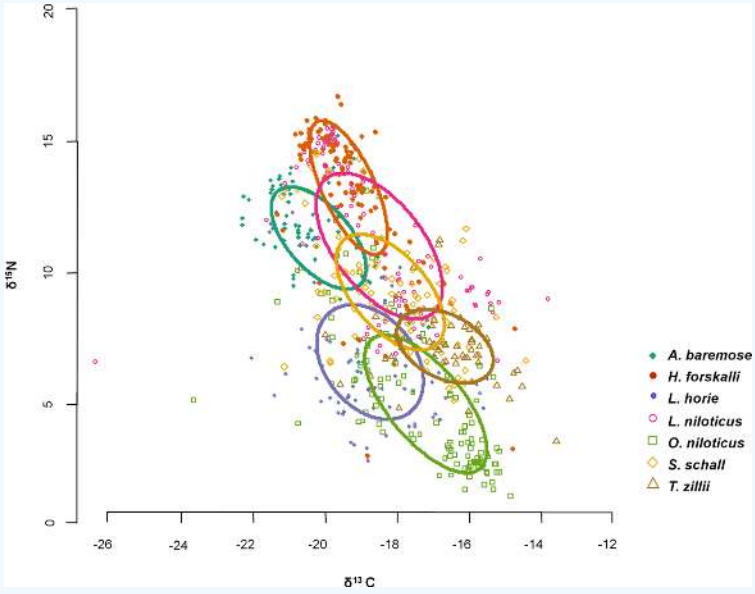
Traditional Approaches to Studying Food Webs

Traditional approaches to food web analysis include gut and scat content analysis. In these analyses, the diet of species is determined based on the remains of prey found in their stomachs or in their feces. These approaches are limited because they only allow researchers to examine the hard parts or otherwise undigested components of prey (e.g., otoliths, or ear bones, of fish and teeth of mice). Over the past two decades, stable isotope analysis has become an increasingly common approach to studying food webs. Even more recently, two new methods to studying food webs have emerged: the use of remotely operated

vehicles and the use of DNA in feces and gut contents. Below, we provide an example of a recent study using remotely operated vehicles to study food webs and diet in the deep sea.



Figure 5.8.2.4**:** Traditional food web analysis has often entailed gut content analysis or fecal content analysis. Scientists use the bits of prey that remain to calculate estimates of variables like percent occurrence for each prey type. The top photo (taken by N. Gownaris, licenced CC-BY) shows a tilapia in the stomach of a Nile perch; very rarely is gut content analysis this straightforward! The bottom photo (inside the owl pellet by Art Siegel and licensed under CC-BY-NC 2.0) shows prey remains removed from an owl pellet, including teeth, claws, and bones of owl prey.



https://geo.libretexts.org/@go/page/25906

Figure 5.8.2.5**:** An example of a food web study using stable isotopes of seven key fish species in Lake Turkana, Kenya. In stable isotope studies, δ13C provides information on the foraging habitat of each species, while δ15N provides information on the trophic level. Species foraging more offshore or more pelagically have a lower δ13C signature and species feeding at lower trophic levels have a lower δ15N signature. In this example, *H. forskalii* is a pelagic fish (i.e., it feeds in the lake's open waters) and has the highest trophic level among the species studied. Figure from Gownaris et al., 2015.

The Use of Remotely Operated Vehicles to Understand Understudied Food Webs

Within the deep sea, Earth's largest ecosystem, the challenge of gathering empirical feeding data for food webs is particularly formidable due to logistical access and sampling constraints (Robison, 2004). Analyzing the contents of a consumer's stomach (gut or stomach content analysis) is the common and most directly quantitative way of inferring diet, and is an irreplaceable approach for determining the taxonomic identity of food web components. However, for deeper-dwelling fishes with internal gas spaces, stomach eversion can confound this approach (Drazen & Sutton, 2017). These analyses may also fail to quantify gelatinous prey that are readily digested and become quickly unrecognizable (Hyslop, 1980; Choy, 2013).

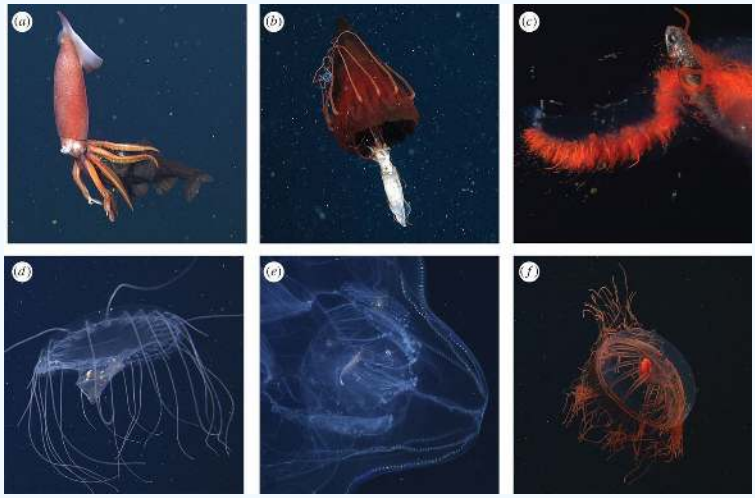


Figure 5.8.2.6**:** A suite of six illustrative ROV frame grabs of pelagic predators and their prey included in Choy et al. (2017). No scale bars for size are available from these sequences. From left to right, top to bottom: (a) *Gonatus sp.* (squid) feeding on a bathylagid fish (Bathylagidae); (b) *Periphylla periphylla*, the helmet jellyfish, feeding on a gonatid squid (Gonatidae), with a small narcomedusa (*Aegina sp.*) also captured; (c) an undescribed physonect siphonophore known as 'the galaxy siphonophore' feeding on a lanternfish of the family Myctophidae; (d) a narcomedusa, *Solmissus*, ingesting a salp chain (Salpida); (e) the ctenophore *Thalassocalyce inconstans*, with a euphausiid (Euphausiacea) in its gut; and (f) the trachymedusa, *Halitrephes maasi*, with a large, red mysid (Mysidae) in its gut. Figure from Choy et al., 2017.

Choy et al. (2017) used 27 years (1991–2016) of *in situ* feeding observations collected by remotely operated vehicles (ROVs) to quantitatively characterize the deep pelagic food web of central California within the California Current, complementing existing studies of diet and trophic interactions with a unique perspective. Seven hundred and forty-three independent feeding events were observed with ROVs from near-surface waters down to depths approaching 4000 m, involving an assemblage of 84 different predators and 82 different prey types, for a total of 242 unique feeding relationships. The greatest diversity of prey was consumed by narcomedusae, followed by physonect siphonophores, ctenophores and cephalopods. This study highlighted key interactions within the poorly understood 'jelly web', showing the importance of medusae, ctenophores and siphonophores as key predators, whose ecological significance is comparable to large fish and squid species within the central California deep pelagic food web.

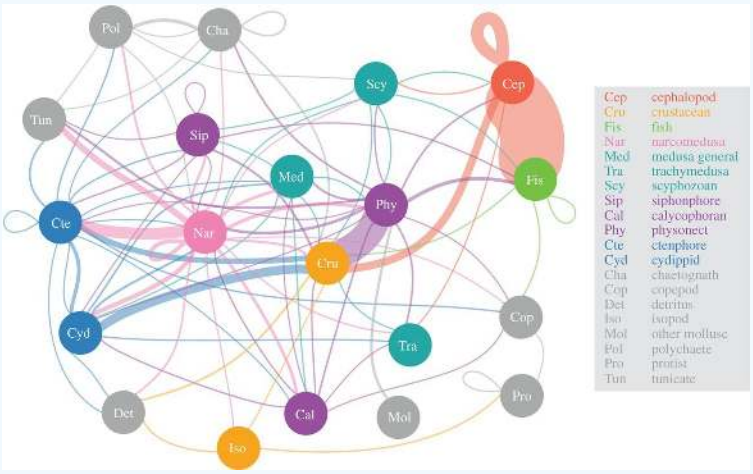


Figure 5.8.2.7**:** An *in situ* perspective of the food web derived from ROV-based observations of feeding, as represented by 20 broad taxonomic groupings. The linkages between predator to prey are coloured according to predator group origin, and loops indicate within-group feeding. The thickness of the lines or edges connecting food web components is scaled to the log of the number of unique ROV feeding observations across the years 1991–2016 between the two groups of animals. The different groups have eight colour-coded types according to main animal types as indicated by the legend and defined here: red, cephalopods; orange, crustaceans; light green, fish; dark green, medusa; purple, siphonophores; blue, ctenophores and grey, all other animals. In this plot, the vertical axis does not correspond to trophic level, because this metric is not readily estimated for all members. Note that for the Sip and Med groups, there are overlapping sub-groups (calycophoran and physonect siphonophores, and trachymedusa and scyphozoan medusa, respectively), which is attributed to varying levels of taxonomic discrimination possible from *in situ* video observations. Figure from Choy et al., 2017.

5.8.2: Quantifying Food Webs is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

5.8.3: Trophic Cascades

Trophic cascades are powerful indirect interactions that can control entire ecosystems, occurring when a trophic level in a food web is suppressed. For example, a top-down cascade will occur if predators are effective enough in predation to reduce the abundance, or alter the behavior of their prey, thereby releasing the next lower trophic level from predation (or herbivory if the intermediate trophic level is a herbivore).

The trophic cascade is an ecological concept which has stimulated new research in many areas of ecology. For example, it can be important for understanding the knock-on effects of removing top predators from food webs, as humans have done in many places through hunting and fishing.

A **top-down cascade** is a trophic cascade where the top consumer/predator controls the primary consumer population. In turn, the primary producer population thrives. The removal of the top predator can alter the food web dynamics. In this case, the primary consumers would overpopulate and exploit the primary producers. Eventually there would not be enough primary producers to sustain the consumer population. Top-down food web stability depends on competition and predation in the higher trophic levels. Invasive species can also alter this cascade by removing or becoming a top predator. This interaction may not always be negative. Studies have shown that certain invasive species have begun to shift cascades; and as a consequence, ecosystem degradation has been repaired (Kotta et al., 2018; Megrey & Werner).

These trophic interactions shape patterns of biodiversity globally. Humans and climate change have affected these cascades drastically. One example can be seen with sea otters (*Enhydra lutris*) on the Pacific coast of the United States of America. Over time, human interactions caused a removal of sea otters. One of their main prey, the pacific purple sea urchin (*Strongylocentrotus purpuratus*) eventually began to overpopulate. The overpopulation caused increased predation of giant kelp (*Macrocystis pyrifera*). As a result, there was extreme deterioration of the kelp forests along the California coast (Estes & Palmisano, 1974; Szpak et al., 2013). This is why it is important for countries to regulate marine and terrestrial ecosystems (Zhang et al., 2018).

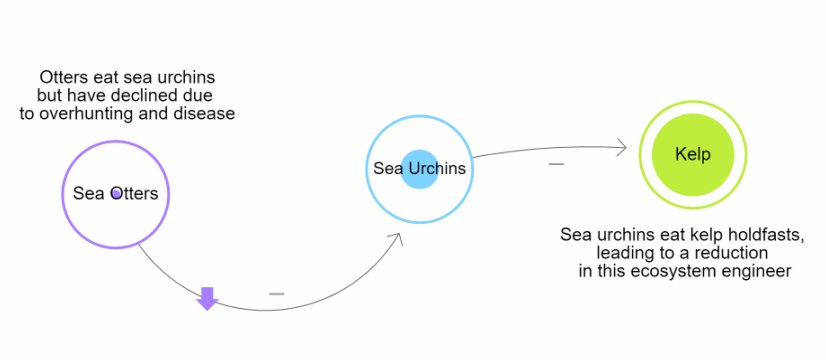


Figure 5.8.3.1**:** Diagram representing the trophic cascade that occurs when sea otters are removed from coastal marine ecosystems in California, ultimately driving a reduction in habitat due to the loss of Kelp. Developed by N. Gownaris using loopy. Create your own using LOOPY: a tool for thinking in systems (ncase.me)

Predator-induced interactions could heavily influence the flux of atmospheric carbon if managed on a global scale. For example, a study was conducted to determine the cost of potential stored carbon in living kelp biomass in sea otter (*Enhydra lutris*) enhanced ecosystems. The study valued the potential storage between $205 million and $408 million dollars (US) on the European Carbon Exchange (2012) (Wilmers et al., 2012).

One criticism of trophic cascades is that, in nature, communities rarely act as food chains with discrete trophic levels. Most communities, however, have complex food webs. In real food webs, consumers often feed at multiple trophic levels (omnivory),

organisms often change their diet as they grow larger, cannibalism occurs, and consumers are subsidized by inputs of resources from outside the local community, all of which blur the distinctions between trophic levels (Polis & Strong, 1996).

Classic Examples of Trophic Cascades

The earliest empirical demonstrations of trophic cascades came from marine and, especially, aquatic ecosystems. In addition to the sea otter example discussed above, two famous examples include:

- In North American lakes, piscivorous fish can dramatically reduce populations of zooplanktivorous fish; zooplanktivorous fish can dramatically alter freshwater zooplankton communities, and zooplankton grazing can in turn have large impacts on phytoplankton communities. Removal of piscivorous fish can change lake water from clear to green by allowing phytoplankton to flourish (Carpenter et al., 1985).
- In the Eel River, in Northern California, fish (steelhead and roach) consume fish larvae and predatory insects. These smaller predators prey on midge larvae, which feed on algae. Removal of the larger fish increases the abundance of algae (Power, 1990).

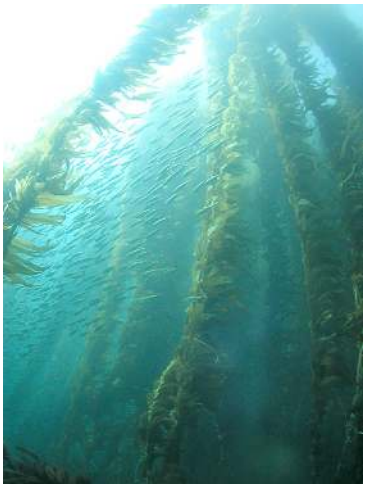


Figure 5.8.3.2**:** Healthy Pacific kelp forests, like this one at San Clemente Island of California's Channel Islands, have been shown to flourish when sea otters are present. When otters are absent, sea urchin populations can irrupt and severely degrade the kelp forest ecosystem. This image by Aquaimages is licensed under CC BY-SA 2.5.

More recent examples of trophic cascades in marine ecosystems have been identified:

- An example of a cascade in a complex, open-ocean ecosystem occurred in the northwest Atlantic during the 1980s and 1990s. The removal of Atlantic cod (*Gadus morhua*) and other ground fishes by sustained overfishing resulted in increases in the abundance of the prey species for these ground fishes, particularly smaller forage fishes and invertebrates such as the northern snow crab (*Chionoecetes opilio*) and northern shrimp (*Pandalus borealis*). The increased abundance of these prey species altered the community of zooplankton that serve as food for smaller fishes and invertebrates as an indirect effect (Frank et al., 2005).
- A similar cascade, also involving the Atlantic cod, occurred in the Baltic Sea at the end of the 1980s. After a decline in Atlantic cod, the abundance of its main prey, the sprat (*Sprattus sprattus*), increased (Alheit et al. 2005) and the Baltic Sea ecosystem shifted from being dominated by cod into being dominated by sprat. The next level of trophic cascade was a decrease in the abundance of *Pseudocalanus acuspes* (Mollmann et al., 2008), a copepod which the sprat prey on.
- On Caribbean coral reefs, several species of angelfishes and parrotfishes eat species of sponges that lack chemical defenses. Removal of these sponge-eating fish species from reefs by fish-trapping and netting has resulted in a shift in the sponge community toward fast-growing sponge species that lack chemical defenses (Loh & Pawlik, 2014). These fast-growing sponge

species are superior competitors for space, and overgrow and smother reef-building corals to a greater extent on overfished reefs (Loh et al., 2015).

The fact that the earliest documented trophic cascades all occurred in lakes and streams led a scientist to speculate that fundamental differences between aquatic and terrestrial food webs made trophic cascades primarily an aquatic phenomenon. Trophic cascades were restricted to communities with relatively low species diversity, in which a small number of species could have overwhelming influence and the food web could operate as a linear food chain. Additionally, well documented trophic cascades at that point in time all occurred in food chains with algae as the primary producer. Trophic cascades, Strong argued, may only occur in communities with fast-growing producers which lack defenses against herbivory.

Subsequent research has documented trophic cascades in terrestrial ecosystems, including:

- A classic example of a terrestrial trophic cascade is the reintroduction of gray wolves (*Canis lupus*) to Yellowstone National Park, which reduced the number, and changed the behavior, of elk (*Cervus canadensis*). This in turn released several plant species from grazing pressure and subsequently led to the transformation of riparian ecosystems. This example of a trophic cascade is vividly shown and explained in the viral video "How Wolves Change Rivers" (Sustainable Human, 2014).
- In the coastal prairie of Northern California, yellow bush lupines are fed upon by a particularly destructive herbivore, the root-boring caterpillar of the ghost moth. Entomopathogenic nematodes kill the caterpillars, and can increase the survival and seed production of lupines (Strong et al., 1999; Preisser, 2003).
- In Costa Rican rain forest, a clerid beetle specializes in eating ants. The ant *Pheidole bicornis* has a mutualistic association with *Piper* plants: the ant lives on the *Piper* and removes caterpillars and other insect herbivores. The clerid beetle, by reducing the abundance of ants, increases the leaf area removed from *Piper* plants by insect herbivores (Letourneau & Dyer, 1998).

Critics pointed out that published terrestrial trophic cascades generally involved smaller subsets of the food web (often only a single plant species). This was quite different from aquatic trophic cascades, in which the biomass of producers as a whole were reduced when predators were removed. Additionally, most terrestrial trophic cascades did not demonstrate reduced plant biomass when predators were removed, but only increased plant damage from herbivores (Polis et al., 2000). It was unclear if such damage would actually result in reduced plant biomass or abundance.

In 2002 a meta-analysis found trophic cascades to be generally *weaker* in terrestrial ecosystems, meaning that changes in predator biomass resulted in smaller changes in plant biomass (Shurin et al., 2002). In contrast, a study published in 2009 demonstrated that multiple species of trees with highly varying autecologies are in fact heavily impacted by the loss of an apex predator (Beschta & Ripple, 2009). Another study, published in 2011, demonstrated that the loss of large terrestrial predators also significantly degrades the integrity of river and stream systems, impacting their morphology, hydrology, and associated biological communities (Beschta & Ripple, 2011).

An Ecosystem-Wide Trophic Cascade: The Wolves of Yellowstone National Park

The critics' model is challenged by studies accumulating since the reintroduction of gray wolves (*Canis lupus*) to Yellowstone National Park. The gray wolf, after being extirpated in the 1920s and absent for 70 years, was reintroduced to the park in 1995 and 1996. Since then a three-tiered trophic cascade has been reestablished involving wolves, elk (*Cervus elaphus*), and woody browse species such as aspen (*Populus tremuloides*), cottonwoods (*Populus* spp.), and willows (*Salix* spp.). Mechanisms likely include actual wolf predation of elk, which reduces their numbers, and the threat of predation, which alters elk behavior and feeding habits, resulting in these plant species being released from intensive browsing pressure. Subsequently, their survival and recruitment rates have significantly increased in some places within Yellowstone's northern range. This effect is particularly noted among the range's riparian plant communities, with upland communities only recently beginning to show similar signs of recovery (Ripple & Beschta, 2012).

Examples of this phenomenon include:

- A 2–3 fold increase in deciduous woody vegetation cover, mostly of willow, in the Soda Butte Creek area between 1995 and 1999 (Groshong, 2004).
- Heights of the tallest willows in the Gallatin River valley increasing from 75 cm to 200 cm between 1998 and 2002 (Ripple & Beschta, 2004).
- Heights of the tallest willows in the Blacktail Creek area increased from less than 50 cm to more than 250 cm between 1997 and 2003. Additionally, canopy cover over streams increased significantly, from only 5% to a range of 14–73% (Beschta & Ripple, 2007).
- In the northern range, tall deciduous woody vegetation cover increased by 170% between 1991 and 2006 (Baril, 2009).
- In the Lamar and Soda Butte Valleys the number of young cottonwood trees that had been successfully recruited went from 0 to 156 between 2001 and 2010 (Ripple & Beschta, 2012).

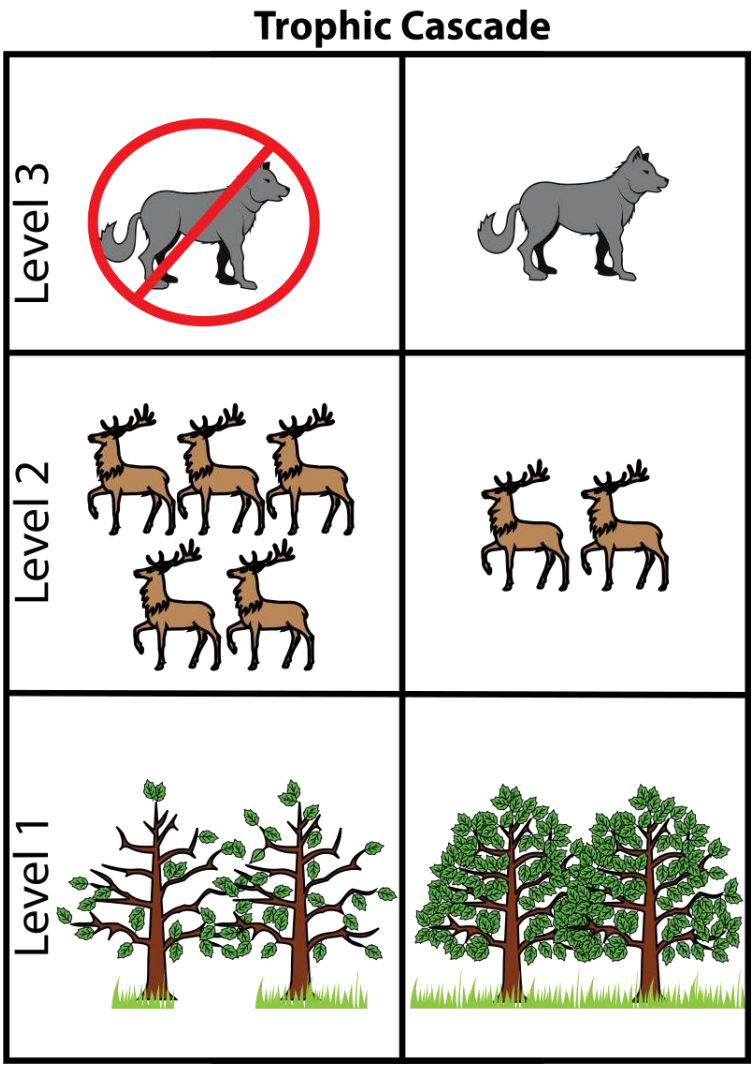


Figure 5.8.3.3**:**This diagram illustrates trophic cascade caused by removal of the top predator. When the top predator is removed the population of deer is able to grow unchecked and this causes over-consumption of the primary producers. This image by Zirguezi is licensed under CC BY 3.0.

Trophic cascades also impact the biodiversity of ecosystems, and when examined from that perspective wolves appear to be having multiple, positive cascading impacts on the biodiversity of Yellowstone National Park. These impacts include:

- Scavengers, such as ravens (*Corvus corax*), bald eagles (*Haliaeetus leucocephalus*), and even grizzly bears (*Ursus arctos horribilis*), are likely subsidized by the carcasses of wolf kills (Wilmers et al., 2003).
- In the northern range, the relative abundance of six out of seven native songbirds which utilize willow was found to be greater in areas of willow recovery as opposed to those where willows remained suppressed (Baril, 2009).
- Bison (*Bison bison*) numbers in the northern range have been steadily increasing as elk numbers have declined, presumably due to a decrease in interspecific competition between the two species (Painter & Ripple, 2012).
- Importantly, the number of beaver (*Castor canadensis*) colonies in the park has increased from one in 1996 to twelve in 2009. The recovery is likely due to the increase in willow availability, as they have been feeding almost exclusively on it. As keystone species, the resurgence of beaver is a critical event for the region. The presence of beavers has been shown to positively impact streambank erosion, sediment retention, water tables, nutrient cycling, and both the diversity and abundance of plant and animal life among riparian communities (Ripple & Beschta, 2012).

There are a number of other examples of trophic cascades involving large terrestrial mammals, including:

- In both Zion National Park and Yosemite National Park, the increase in human visitation during the first half of the 20th century was found to correspond to the decline of native cougar (*Puma concolor*) populations in at least part of their range. Soon after, native populations of mule deer (*Odocoileus hemionus*) erupted, subjecting resident communities of cottonwoods (*Populus fremontii*) in Zion and California black oak (*Quercus kelloggii*) in Yosemite to intensified browsing. This halted successful recruitment of these species except in refugia inaccessible to the deer. In Zion the suppression of cottonwoods increased stream erosion and decreased the diversity and abundance of amphibians, reptiles, butterflies, and wildflowers. In parts of the park where cougars were still common these negative impacts were not expressed and riparian communities were significantly healthier (Ripple & Beschta, 2006; Ripple & Beschta, 2008).
- In sub-Saharan Africa, the decline of lion (*Panthera leo*) and leopard (*Panthera pardus*) populations has led to a rising population of olive baboon (*Papio anubis*). This case of mesopredator release negatively impacted already declining ungulate populations and is one of the reasons for increased conflict between baboons and humans, as the primates raid crops and spread intestinal parasites (Prugh et al., 2009; Estes et al., 2011).
- In the Australian states of New South Wales and South Australia, the presence or absence of dingoes (*Canis lupus dingo*) was found to be inversely related to the abundance of invasive red foxes (*Vulpes vulpes*). In other words, the foxes were most common where the dingoes were least common. Subsequently, populations of an endangered prey species, the dusky hopping mouse (*Notomys fuscus*) were also less abundant where dingoes were absent due to the foxes, which consume the mice, no longer being held in check by the top predator (Letnic & Dworjanyn, 2011).



5.8.3: Trophic Cascades is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

5.8.4: Sources and Attributions

Contributors and Attributions

This chapter was compiled and written by N. Gownaris with text taken from the following CC-BY resources:

- **Trophic Cascade by Wikipedia, the free encyclopedia**
- **Trophic Level by Wikipedia, the free encyclopedia**
- **Food Web by Wikipedia, the free encyclopedia**
- **Ecological Pyramid by Wikipedia, the free encyclopedia**
- **Decomposer by Wikipedia, the free encyclopedia**
- **Detritivore by Wikipedia, the free encyclopedia**
- **Detritus by Wikipedia, the free encyclopedia**
- **Scavenger by Wikipedia, the free encyclopedia**
- **Microbial Food Web by Wikipedia, the free encyclopedia**
- **Ecological Efficiency by Wikipedia, the free encyclopedia**
- **Ecology of Ecosystems (Chapter 46.1) from OpenStaxx Biology 2e**
- **Energy Flow through Ecosystems (Chapter 46.2) from OpenStaxx Biology 2e**
- **Madigan, D. J., Carlisle, A. B., Dewar, H., Snodgrass, O. E., Litvin, S. Y., Micheli, F., & Block, B. A. (2012). Stable isotope analysis challenges wasp-waist food web assumptions in an upwelling pelagic ecosystem. *Scientific reports*, *2*(1), 1-10.**
- **Choy, C. A., Haddock, S. H., & Robison, B. H. (2017). Deep pelagic food web structure as revealed by in situ feeding observations. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1868), 20172116.**
- **Hoover, C., Giraldo, C., Ehrman, A., Suchy, K. D., MacPhee, S. A., Brewster, J., ... & Loseto, L. (2021). The Canadian Beaufort Shelf trophic structure: evaluating an ecosystem modelling approach by comparison with observed stable isotopic structure. *Arctic Science*, 1-21.**

Sources

Alheit, J., Möllmann, C., Dutz, J., Kornilovs, G., Loewe, P., Mohrholz, V., & Wasmund, N. (2005). Synchronous ecological regime shifts in the central Baltic and the North Sea in the late 1980s. *ICES Journal of Marine Science*, *62*(7), pp. 1205–1215. doi:10.1016/j.icesjms.2005.04.024.

Banasek-Richter, C., Bersier, L.L., Cattin, M., Baltensperger, R., Gabriel, J., Merz, Y., et al. (2009). Complexity in quantitative food webs. *Ecology*, *90*(6), pp. 1470–1477. doi:10.1890/08-2207.1. hdl:1969.1/178777. PMID 19569361.

Barbosa, P., & Castellanos, I. (Eds.). (2005). Ecology of predator-prey interactions, pp. 306, Oxford University Press.ISBN 9780199883677.

Baril, L.M. (2009). Change in deciduous woody vegetation, implications of increased willow (Salix spp.) growth for bird species diversity and willow species composition in and around Yellowstone National Park's northern range (MS). *Bozeman, USA: Montana State University.*

Benke, A.C. (2010). Secondary production. *Nature Education Knowledge*, *1*(8), pp. 5.

Beschta, R.L., & Ripple, W.J. (2011). The role of large predators in maintaining riparian plant communities and river morphology. *Geomorphology*, 157–158, 88–98. doi:10.1016/j.geomorph.2011.04.042.

Beschta, R.L., & W.J. Ripple. (2009). Large predators and trophic cascades in terrestrial ecosystems of the western. *United States Biological Conservation*, *142*, pp. 2401–2414.

Beschta, R.L., & Ripple, W.J. (2007). Increased willow heights along northern Yellowstone's Blacktail Deer Creek following wolf reintroduction. *Western North American Naturalist*, *67*(4), pp. 613–617. doi:10.3398/1527-0904(2007)67[613:iwhany]2.0.co;2.

Carpenter, S.R., Kitchell, J.F., & Hodgson, J.R. (1985). Cascading trophic interactions and lake productivity. *BioScience*, *35*(10), pp. 634–639. doi:10.2307/1309989. JSTOR 1309989.

Choy, C.A., Portner, E., Iwane, M., & Drazen, J.C. (2013). Diets of five important predatory mesopelagic fishes of the central North Pacific. *Mar. Ecol. Prog. Ser.*, *492*, pp. 169–184. (doi:10.3354/meps10518)

Choy, C.A., Haddock, S.H., & Robison, B.H. (2017). Deep pelagic food web structure as revealed by in situ feeding observations. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1868), 20172116.

Cortés, E. (1999). Standardized diet compositions and trophic levels of sharks. *ICES J. Mar. Sci.*, *56*(5), pp. 707–717. doi:10.1006/jmsc.1999.0489.

Cury, P., et al. (2000). Small pelagics in upwelling systems: patterns of interaction and structural changes in “wasp-waist” ecosystems. *ICES J. Mar. Sci.*, *57*, pp. 603–618 (2000).

Cury, P.M., et al. (2011). Global seabird response to forage fish depletion - one-third for the birds. *Science*, *334*, pp. 1703–1706.

Cury, P., Shannon, L., & Shin, Y.-J. (2003). In Sinclair, M., & Valdimarsson, G. (Eds.), *Responsible fisheries in the marine ecosystem*, (pp. 103–123). AO, Rome, & CABI Publishing, Wallingford, UK.

Definition of Trophic. www.merriam-webster.com. Retrieved 16 April 2017.

Drazen, J.C., & Sutton, T.T. (2017). Dining in the deep: The feeding ecology of deep-sea fishes. *Annu. Rev. Mar. Sci.*, *9*, pp. 337–366. (doi:10.1146/annurev-marine-010816-060543)

Dunne, J.A., Williams, R.J., Martinez, N.D., Wood, R.A., & Erwin, D.H. (2008). Compilation and Network Analyses of Cambrian Food Webs. *PLoS Biol.*, *6*(4): e102. doi:10.1371/journal.pbio.0060102

Dunne, J.A., Williams, R.J., & Martinez, N.D. (2002). Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences*, *99*(20), pp. 12917–12922. Bibcode:2002PNAS...9912917D. doi:10.1073/pnas.192407699. PMC 130560. PMID 12235364.

Dunne, J.A., Williams, R.J., Martinez, N.D., Wood, R.A., Erwin, D.H., & Dobson, Andrew P. (2008). Compilation and network analyses of cambrian food webs. *PLOS Biology*, *6*(4): e102. doi:10.1371/journal.pbio.0060102. PMC 2689700. PMID 18447582.

Eisenberg, C. (2011). The wolf's tooth: Keystone predators, trophic cascades, and biodiversity. pp. 15. Island Press. ISBN 978-1-59726-398-6.

Estes, J.A., & Palmisano, J.F. (1974). Sea otters: Their role in structuring nearshore communities. *Science*, *185*(4156), pp. 1058–1060. doi:10.1126/science.185.4156.1058. PMID 17738247. S2CID 35892592.

Estes, J.A., et al. (2011). Trophic Downgrading of Planet Earth. *Science*, *333*(6040), pp. 301–306. CiteSeerX 10.1.1.701.8043. doi:10.1126/science.1205106. PMID 21764740. S2CID 7752940.

Frank, K.T.; Petrie, B., Choi, J.S., & Leggett, W.C. (2005). Trophic cascades in a formerly cod-dominated ecosystem. *Science*, *308*(5728), pp. 1621–1623. doi:10.1126/science.1113075. ISSN 0036-8075. PMID 15947186. S2CID 45088691.

Getz, W. (2011). Biomass transformation webs provide a unified approach to consumer–resource modelling. *Ecology Letters*, *14*(2), pp. 113–124. doi:10.1111/j.1461-0248.2010.01566.x. PMC 3032891. PMID 21199247.

Santos, G.N., Danac, A.C., & Ocampo, J.P. (2003). E-biology ii. *Rex Book Store*. p. 58. ISBN 978-971-23-3563-1.

Gönenç, I.E., Koutitonsky, V.G., & Rashleigh, B. (2007). Assessment of the fate and effects of toxic agents on water resources. *Springer*. p. 279. ISBN 978-1-4020-5527-0.

Gownaris, N.J., Pikitch, E.K., Ojwang, W.O., Michener, R., & Kaufman, L. (2015). Predicting species’ vulnerability in a massively perturbed system: The fishes of Lake Turkana, Kenya. *PloS one*, *10*(5), e0127027.

Groombridge, B., & Jenkins, M. (2002). World atlas of biodiversity: Earth's living resources in the 21st century. *World Conservation Monitoring Centre, United Nations Environment Programme*. ISBN 978-0-520-23668-4.

Groshong, L.C. (2004). Mapping riparian vegetation change in Yellowstone's Northern Range using high spatial resolution imagery (MA Thesis). *Eugene, Oregon, USA: University of Oregon*.

Hutchinson, J. (2013). Tyrannosaurus rex: Predator or media hype? *What's in John's Freezer?* Retrieved August 26, 2013.

Hyslop E.J. (1980). Stomach contents analysis: A review of methods and their application. *J. Fish Biol., 17*, pp. 411–429. (doi:10.1111/j.1095-8649.1980.tb02775.x)

Kane; et al. (2016). Body size as a driver of scavenging in theropod dinosaurs. *The American Naturalist, 187*(6), pp. 706–16. doi:10.1086/686094. hdl:10023/10617. PMID 27172591. S2CID 3840870.

Keddy, P. (2017). Plant ecology, origins, processes, consequences (2nd Ed.). *New York: Cambridge University Press*. pp. 92–93. ISBN 978-1-107-11423-4.

Kent, M. (2000). Advanced biology. *Oxford University Press US*. p. 511. ISBN 978-0-19-914195-1.

Kotta, J., Wernberg, T., Jänes, H., Kotta, I., Nurkse, K., Pärnoja, M., & Orav-Kotta, H. (2018). Novel crab predator causes marine ecosystem regime shift. *Scientific Reports, 8*(1), pp. 4956. doi:10.1038/s41598-018-23282-w. PMC 5897427. PMID 29651152.

Letnic, M .,& Dworjanyn, S.A. (2011). Does a top predator reduce the predatory impact of an invasive mesopredator on an endangered rodent? *Ecography, 34*(5), pp. 827–835. doi:10.1111/j.1600-0587.2010.06516.x.

Letourneau, D.K., & Dyer, L.A. (1998). Experimental test in lowland tropical forest shows top-down effects through four trophic levels. *Ecology, 79*(5), pp. 1678–1687. doi:10.2307/176787. JSTOR 176787.

Leveque, C., ed. (2003). Ecology: From ecosystem to biosphere. Science Publishers. p. 490. ISBN 978-1-57808-294-0.

Loh, T.L., et al. (2015). Indirect effects of overfishing on Caribbean reefs: Sponges overgrow reef-building corals. PeerJ. 3: e901. doi:10.7717/peerj.901. PMC 4419544. PMID 25945305.

Loh, T.L., & Pawlik, J.R. (2014). Chemical defenses and resource trade-offs structure sponge communities on Caribbean coral reefs. *Proceedings of the National Academy of Sciences, 111*(11), pp. 4151–4156. doi:10.1073/pnas.1321626111. ISSN 0027-8424. PMC 3964098. PMID 24567392.

Madigan, D.J., Carlisle, A.B., Dewar, H., Snodgrass, O.E., Litvin, S.Y., Micheli, F., & Block, B.A. (2012). Stable isotope analysis challenges wasp-waist food web assumptions in an upwelling pelagic ecosystem. *Scientific reports, 2*(1), pp. 1-10.

Megrey, B., & Werner, F. Evaluating the role of topdown vs. bottom-up ecosystem regulation from a modeling perspective.

Mollmann, C., Muller-Karulis, B., Kornilovs, G., & St. John, M.A. (2008). Effects of climate and overfishing on zooplankton dynamics and ecosystem structure: Regime shifts, trophic cascade, and feedback loops in a simple ecosystem. *ICES Journal of Marine Science, 65*(3), pp. 302–310. doi:10.1093/icesjms/fsm197.

Mostajir, B., Amblard, C., Buffan-Dubau, E., De Wit, R., Lensi, R., & Sime-Ngando, T. (2015). Microbial food webs in aquatic and terrestrial ecosystems. In Bertrand, J.C., Caumette, P., Lebaron, P., Matheron, R., Normand, P., & Sime-Ngando, T. (Eds.), *Environmental Microbiology: Fundamentals and Applications*, (pp. 485-510). Microbial Ecology, Springer. ISBN 9789401791182.

Murdoch, W.M. (1966). Community structure, population control, and competition – a critique. *The American Naturalist, 100*(912), pp. 219–226. doi:10.1086/282415. S2CID 84354616.

Neutel, A., Heesterbeek, J.A.P., & de Ruiter, P.D. (2002). Stability in real food webs: Weak link in long loops. *Science, 295*(550), pp. 1120–1123. Bibcode:2002Sci...296.1120N. doi:10.1126/science.1068326. hdl:1874/8123. PMID 12004131. S2CID 34331654. Archived from the original (PDF) on 2011-09-28.

NOAA. ACE Basin National Estuarine Research Reserve: Decomposers. Archived from the original on 2014-10-09. Retrieved 2012-09-17.

Odum, E.P., & Barrett, G.W. (2005). Fundamentals of ecology (5th ed.). Brooks/Cole, a part of Cengage Learning. ISBN 978-0-534-42066-6. Archived from the original on 2011-08-20.

Odum, W.E., & Heald, E.J. (1975). The detritus-based food web of an estuarine mangrove community. Pages 265–286 In Cronin, L.E. (Eds.), *Estuarine research* (pp. 265-286). Academic Press, New York.

Ogada, D.L., Keesing, F., & Virani, M.Z. (2011). Dropping dead: Causes and consequences of vulture population declines worldwide. *Annals of the New York Academy of Sciences, 1249*(1), pp. 57–71. Bibcode:2012NYASA1249...57O. doi:10.1111/j.1749-6632.2011.06293.x. ISSN 0077-8923. PMID 22175274. S2CID 23734331.

Olson, Z.H., Beasley, J.C., & Rhodes, O.E. (2016). Carcass type affects local scavenger guilds more than habitat connectivity. *PLOS ONE*, *11*(2): e0147798. Bibcode:2016PLoSO..1147798O. doi:10.1371/journal.pone.0147798. ISSN 1932-6203. PMC 4757541. PMID 26886299.

Pahl and Ruedas (2021). Carnosaurs as apex scavengers: Agent-based simulations reveal possible vulture analogues in late Jurassic Dinosaurs. *Ecological Modelling*, 458. doi:10.1016/j.ecolmodel.2021.109706e.

Paine, R.T. (1988). Road maps of interactions or grist for theoretical development? *Ecology*, *69*(6), pp. 1648–1654. doi:10.2307/1941141. JSTOR 1941141. Archived from the original (PDF) on 2011-07-28.

Painter, L.E., & Ripple, W.J. (2012). Effects of bison on willow and cottonwood in northern Yellowstone National Park. *Forest Ecology and Management*, *264*, pp. 150–158. doi:10.1016/j.foreco.2011.10.010.

Pauly, D., & Palomares, M.L. (2005). Fishing down marine food webs: It is far more pervasive than we thought. *Bulletin of Marine Science*, *76*(2), pp. 197–211. Archived from the original (PDF) on 14 May 2013.

Pauly, D., Trites, A., Capuli, E., & Christensen, V. (1998). Diet composition and trophic levels of marine mammals. *ICES J. Mar. Sci.*, *55*(3), pp. 467–481. doi:10.1006/jmsc.1997.0280.

Pimm, S.L. (1979). The structure of food webs. *Theoretical Population Biology*, *16*(2), pp. 144–158. doi:10.1016/0040-5809(79)90010-8. PMID 538731. Archived from the original (PDF) on 2011-09-27.

Pimm, S.L., Lawton, J.H., & Cohen, J.E. (1991). Food web patterns and their consequences. *Nature*, *350*(6320), pp. 669–674. Bibcode:1991Natur.350..669P. doi:10.1038/350669a0. S2CID 4267587. Archived from the original (PDF) on 2010-06-10.

Polis, G.A., Sears, A.L, W., Huxel, G.R., et al. (2000). When is a trophic cascade a trophic cascade? *Trends in Ecology & Evolution*, *15*(11), pp. 473–475. doi:10.1016/s0169-5347(00)01971-6. PMID 11050351.

Polis, G.A., & Strong, D.R. (1996). Food web complexity and community dynamics. *The American Naturalist*, *147*(5), pp. 813–846. doi:10.1086/285880. S2CID 85155900.

Post, D.M. (1993). The long and short of food-chain length. *Trends in Ecology and Evolution*, *17*(6), pp. 269–277. doi:10.1016/S0169-5347(02)02455-2.

Power, M.E. (1990). Effects of fish in river food webs. *Science*, *250*(4982), pp. 811–814. doi:10.1126/science.250.4982.811. PMID 17759974. S2CID 24780727.

Preisser, E.L. (2003). Field evidence for a rapidly cascading underground food web. *Ecology*, *84*(4). pp. 869–874. doi:10.1890/0012-9658(2003)084[0869:fefarc]2.0.co;2.

Proctor, J.D., & Larson, B.M.H. (2005). Ecology, complexity, and metaphor. *BioScience*, *55*(12), pp. 1065–1068. doi:10.1641/0006-3568(2005)055[1065:ECAM]2.0.CO;2.

Proulx, S.R., Promislow, D.E.L., & Phillips, P.C. (2005). Network thinking in ecology and evolution. *Trends in Ecology and Evolution*, *20*(6), 345–353. doi:10.1016/j.tree.2005.04.004. PMID 16701391. Archived from the original (PDF) on 2011-08-15.

Prugh, L.R., et al. (2009). The Rise of the Mesopredator. *BioScience*, *59*(9), pp. 779–791. doi:10.1525/bio.2009.59.9.9. S2CID 40484905.

Raffaelli, D. (2002). From Elton to mathematics and back again. *Science*, *296*(5570), pp. 1035–1037. doi:10.1126/science.1072080. PMID 12004106. S2CID 177263265.

Rickleffs, R.E. (1996). The economy of nature. *University of Chicago Press*. p. 678. ISBN 978-0-7167-3847-3.

Ripple, W.J., & Beschta, R.L. (2012). Trophic cascades in Yellowstone: The first 15 years after wolf reintroduction. *Biological Conservation*, *145*, pp. 205–213. doi:10.1016/j.biocon.2011.11.005.

Ripple, W.J., & Beschta, R.L. (2004). Wolves, elk, willows, and trophic cascades in the upper Gallatin Range of Southwestern Montana, USA. *Forest Ecology and Management*, *200*(1–3), pp. 161–181. doi:10.1016/j.foreco.2004.06.017.

Ripple, W.J., & Beschta, R.L. (2006). Linking a cougar decline, trophic cascade, and catastrophic regime shift in Zion National Park. *Biological Conservation*, *133*(4), pp. 397–408. doi:10.1016/j.biocon.2006.07.002.

Ripple, W.J., & Beschta, R.L. (2008). Trophic cascades involving cougar, mule deer, and black oaks in Yosemite National Park. *Biological Conservation*, *141*(5), pp. 1249–1256. doi:10.1016/j.biocon.2008.02.028.

Robison, B.H. (2004). Deep pelagic biology. *J. Exp. Mar. Biol. Ecol. 300*, pp. 253–272. (doi:10.1016/j.jembe.2004.01.012)

Sagi, N., Grünzweig, J.M.,& Hawlena, D. (2019). Burrowing detritivores regulate nutrient cycling in a desert ecosystem. *Proceedings Biological Sciences, 286*(1914): 20191647. doi:10.1098/rspb.2019.1647. PMC 6842856. PMID 31662076.

Science of Earth Systems. Cengage Learning. 2002. p. 537. ISBN 978-0-7668-3391-3.

Shannon, L.J., Cury, P.M. & Jarre, A. (2000). Modelling effects of fishing in the Southern Benguela ecosystem. *ICES J. Mar. Sci*., pp. 720–72221934 .

Shurin, J.B., Borer, E.T., Seabloom, E.W. Anderson, K., Blanchette, C.A., Broitman, B, Cooper, S.D., & Halpern, B.S. (2002). A cross-ecosystem comparison of the strength of trophic cascades. *Ecology Letters, 5*(6), pp. 785–791. doi:10.1046/j.1461-0248.2002.00381.x.

Spellman, F.R. (2008). The science of water: Concepts and applications. *CRC Press*. p. 165. ISBN 978-1-4200-5544-3.

Strong, D.R., Whipple, A.V., Child, A.L., & Dennis, B. (1999). Model selection for a subterranean trophic cascade: Root-feeding caterpillars and entomopathogenic nematodes. *Ecology, 80*(8), pp. 2750–2761. doi:10.2307/177255. JSTOR 177255.

Sustainable Humans. 2014. "How Wolves Change Rivers"

Szpak, P., Orchard, T.J., Salomon, A.K., & Gröcke, D.R. (2013). Regional ecological variability and impact of the maritime fur trade on nearshore ecosystems in southern Haida Gwaii (British Columbia, Canada): Evidence from stable isotope analysis of rockfish (Sebastes spp.) bone collagen. *Archaeological and Anthropological Sciences.* doi:10.1007/s12520-013-0122-y. S2CID 84866250.

Tan, C.K.W., & Corlett, R.T. (2011). Scavenging of dead invertebrates along an urbanisation gradient in Singapore. *Insect Conservation and Diversity, 5*(2), pp. 138–145. doi:10.1111/j.1752-4598.2011.00143.x. ISSN 1752-458X. S2CID 86467187.

van Dover, C. (2000). The ecology of deep-sea hydrothermal vents. *Princeton University Press*. p. 399. ISBN 978-0-691-04929-8.

Wang, H., Morrison, W., Singh, A., & Weiss, H. (2009). Modeling inverted biomass pyramids and refuges in ecosystems. *Ecological Modelling*, *220*(11), pp. 1376–1382. doi:10.1016/j.ecolmodel.2009.03.005. Archived from the original (PDF) on 2011-10-07.

Wetzel, R.G. (2001). Limnology: Lake and river ecosystems (3rd. ed.). *Academic Press*. p. 700. ISBN 978-0-12-744760-5.

Whitman, W.B., Coleman, D.C., & Wieb, W.J. (1998). Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci. USA., 95*(12), pp. 6578–83. Bibcode:1998PNAS...95.6578W. doi:10.1073/pnas.95.12.6578. PMC 33863. PMID 9618454.

Williams, R.J., Berlow, E.L., Dunne, J.A., Barabási, A., & Martinez, N.D. (2002). Two degrees of separation in complex food webs. *Proceedings of the National Academy of Sciences*, *99*(20), pp. 12913–12916. Bibcode:2002PNAS...9912913W. doi:10.1073/pnas.192448799. PMC 130559. PMID 12235367.

Wilmers, C.C., Estes, J.A., Edwards, M., Laidre, K.L., & Konar, B. (2012). Do trophic cascades affect the storage and flux of atmospheric carbon? An analysis of sea otters and kelp forests. *Frontiers in Ecology and the Environment*, *10*(8), pp. 409–415. doi:10.1890/110176. ISSN 1540-9309. S2CID 51684842.

Wilmers, C.C., Crabtree, R.L., Smith, D.W., Murphy, K.M., & Getz, W.M. (2003). Trophic facilitation by introduced top predators: Grey wolf subsidies to scavengers in Yellowstone National Park. *Journal of Animal Ecology, 72*(6), pp. 909–916. doi:10.1046/j.1365-2656.2003.00766.x.

Zhang, J., Qian, H., Girardello, M., Pellissier, V., Nielsen, S.E., & Svenning, J.C. (2018). Trophic interactions among vertebrate guilds and plants shape global patterns in species diversity. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1883): 20180949. doi:10.1098/rspb.2018.0949. PMC 6083253. PMID 30051871.

5.8.4: Sources and Attributions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

5.9: Net primary production and global change

Probably the single most fundamental measure of "global change" of practical interest to humankind is change in terrestrial biological productivity. Biological productivity is the source of all the food, fiber, and fuel that humans survive on, so it defines most fundamentally the habitability of Earth. The spatial variability of net primary productivity (NPP) over the globe is enormous, from about 1000 grams of carbon per cubic meter per year for evergreen tropical rain forests to less than 30 grams of carbon per cubic meter per year for deserts. Primary productivity is largely balanced by ecosystem respiration (RE). The difference between GPP and RE is the terrestrial carbon balance, called net ecosystem productivity (NEP). With increased atmospheric abundance of carbon dioxide, NPP is expected to increase but ongoing global warming is likely to increase RE as well, so that the effect of predicted climate change on the terrestrial carbon uptake or release is largely a matter of ongoing research.

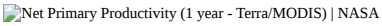


Figure 5.9.1: Net primary productivity, 2022. Figure at the link is interactive showing NPP for 2001-2022. NASA Earth Observatory, public domain.

Understanding regional variability in carbon cycle processes requires a dramatically more spatially detailed analysis of global land surface processes than was available to climate scientists before 2000. In March 2000, the Moderate Resolution Imaging Spectroradiometer (MODIS), aboard NASA's Terra and Aqua satellites, began producing a regular global estimate of near-weekly gross photosynthesis (GPP) and annual net primary production (NPP) of the entire terrestrial Earth surface at up to 1 km spatial resolution. However, ecosystem respiration, and therefore the full terrestrial carbon balance, cannot be estimated from satellite measurements, but only from models or ground-based measurements.

The GPP and NPP products are designed to provide an accurate, regular measure of the production activity or growth of terrestrial vegetation. (Note: only NPP is available in the NEO site.) These products have both theoretical and practical utility. The theoretical use is primarily for defining the seasonally dynamic terrestrial surface carbon dioxide uptake for global carbon cycle studies such as answering the "missing sink question" of carbon. The spatial and seasonal dynamics of carbon dioxide flux are also of high interest in global climate modeling because carbon dioxide is an important greenhouse gas. Both GPP and RE, and therefore the terrestrial carbon balance, can be modeled. Currently, global carbon cycle models are being integrated with climate models with the goal of creating integrated Earth Systems Models that will represent the dynamic interactions between the atmosphere, biosphere, and oceans. The weekly and monthly MODIS GPP and NPP products are useful to constrain models that are used to answer carbon dioxide flux questions.

The practical utility of these GPP/ NPP products includes routine measurements of crop yield, range forage, and forest production, and other economically and socially significant products of vegetation growth. The value of an unbiased, regular source of crop, range, and forest production estimates for global political and economic decision-making is immense. Moreover, these products are freely available for all users worldwide. This daily computed GPP more correctly defines terrestrial carbon dioxide fluxes than simple Normalized Difference Vegetation Index (NDVI) correlations currently done to increase understanding on how the seasonal fluxes of net photosynthesis are related to seasonal variations of driving factors such as atmospheric carbon dioxide, temperature, light, and precipitation.

5.9: Net primary production and global change is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

5.10: Detritus

In biology, **detritus** is dead particulate organic material, as distinguished from dissolved organic material. Detritus typically includes the bodies or fragments of bodies of dead organisms, and fecal material. Detritus typically hosts communities of microorganisms that colonize and decompose it. In terrestrial ecosystems it is present as leaf litter and other organic matter that is intermixed with soil, which is denominated "soil organic matter". The detritus of aquatic ecosystems is organic substances that is suspended in the water and accumulates in depositions on the floor of the body of water; when this floor is a seabed, such a deposition is denominated "marine snow".



Figure 5.10.1: *Sequoiadendron giganteum* (giant sequoia) cones and foliage, sugar pine and white fir foliage, and other plant litter constitute the duff layer that covers the ground of Mariposa Grove in Yosemite National Park, United States.

The corpses of dead plants or animals, material derived from animal tissues (e.g. molted skin), and fecal matter gradually lose their form due to physical processes and the action of decomposers, including grazers, bacteria, and fungi. Decomposition, the process by which organic matter is decomposed, occurs in several phases. Micro- and macro-organisms that feed on it rapidly consume and absorb materials such as proteins, lipids, and sugars that are low in molecular weight, while other compounds such as complex carbohydrates are decomposed more slowly. The decomposing microorganisms degrade the organic materials so as to gain the resources they require for their survival and reproduction. Accordingly, simultaneous to microorganisms' decomposition of the materials of dead plants and animals is their assimilation of decomposed compounds to construct more of their biomass (i.e. to grow their own bodies). When microorganisms die, fine organic particles are produced, and if small animals that feed on microorganisms eat these particles they collect inside the intestines of the consumers, and change shape into large pellets of dung. As a result of this process, most of the materials of dead organisms disappear and are not visible and recognizable in any form, but are present in the form of a combination of fine organic particles and the organisms that used them as nutrients. This combination is detritus.

In ecosystems on land, detritus is deposited on the surface of the ground, taking forms such as the humic soil beneath a layer of fallen leaves. In aquatic ecosystems, most detritus is suspended in water, and gradually settles. In particular, many different types of material are collected together by currents, and much material settles in slowly flowing areas.

A large amount of detritus is used as a source of nutrition for animals. In particular, many bottom feeding animals (benthos) living in mud flats feed in this way. In particular, since excreta are materials which other animals do not need, whatever energy value they might have, they are often unbalanced as a source of nutrients, and are not suitable as a source of nutrition on their own. However, there are many microorganisms which multiply in natural environments. These microorganisms do not simply absorb nutrients from these particles, but also shape their own bodies so that they can take the resources they lack from the area around them, and this allows them to make use of excreta as a source of nutrients. In practical terms, the most important constituents of detritus are complex carbohydrates, which are persistent (difficult to break down), and the microorganisms which multiply using these absorb carbon from the detritus, and materials such as nitrogen and phosphorus from the water in their environment to synthesise the components of their own cells.

A characteristic type of food chain called the detritus cycle takes place involving detritus feeders (detritivores), detritus and the microorganisms that multiply on it. For example, mud flats are inhabited by many univalves which are detritus feeders. When these detritus feeders take in detritus with microorganisms multiplying on it, they mainly break down and absorb the microorganisms, which are rich in proteins, and excrete the detritus, which is mostly complex carbohydrates, having hardly broken it down at all. At first, this dung is a poor source of nutrition, and so univalves pay no attention to it, but after several days, microorganisms begin to multiply on it again, its nutritional balance improves, and so they eat it again. Through this process of eating the detritus many times over and harvesting the microorganisms from it, the detritus thins out, becomes fractured and becomes easier for the microorganisms to use, and so the complex carbohydrates are also steadily broken down and disappear over time.

What is left behind by the detritivores is then further broken down and recycled by decomposers, such as bacteria and fungi.

This detritus cycle plays a large part in the so-called purification process, whereby organic materials carried in by rivers is broken down and disappears, and an extremely important part in the breeding and growth of marine resources. In ecosystems on land, far more essential material is broken down as dead material passing through the detritus chain than is broken down by being eaten by animals in a living state. In both land and aquatic ecosystems, the role played by detritus is too large to ignore.

Aquatic ecosystems

In contrast to land ecosystems, dead materials and excreta in aquatic ecosystems are typically transported by water flow; finer particles tend to be transported farther or suspended longer. In freshwater bodies organic material from plants can form a silt known as *mulm* or humus on the bottom. This material, some called *undissolved organic carbon* breaks down into dissolved organic carbon and can bond to heavy metal ions via chelation. It can also break down into colored dissolved organic matter such as tannin, a specific form of tannic acid. In saltwater bodies, organic material breaks down and forms a marine snow. This example of detritus commonly consists of organic materials such as dead phytoplankton and zooplankton, the outer walls of diatoms and coccolithophores, dead skin and scales of fish, and fecal pellets. This material will slowly sink to the seafloor, where it makes up the majority of sediment in some areas. Once settled, the material will not only contribute to sediments but will help to feed different species of detritivore, organisms which feed on detritus, such as annelid worms and sea cucumbers, to name a few. The exact composition of this detritus varies based on location and time of year, as it is very closely tied to primary production.

Terrestrial ecosystems

Detritus occurs in a variety of terrestrial habitats including forest, chaparral and grassland. In forests, the detritus is typically dominated by leaf, twig, and bacteria litter as measured by biomass dominance. This plant litter provides important cover for seedling protection as well as cover for a variety of arthropods, reptiles and amphibians. Some insect larvae feed on the detritus. Fungi and bacteria continue the decomposition process after grazers have consumed larger elements of the organic materials, and animal trampling has assisted in mechanically breaking down organic matter. At the later stages of decomposition, mesophilic micro-organisms decompose residual detritus, generating heat from exothermic processes; such heat generation is associated with the well known phenomenon of the elevated temperature of composting.

Consumers

There is an extremely large number of detritus feeders in water. After all, a large quantity of material is carried in by water currents. Even if an organism stays in a fixed position, as long as it has a system for filtering water, it will be able to obtain enough food to get by. Many immobile organisms survive in this way, using developed gills or tentacles to filter the water to take in food, a process known as filter feeding.

Another more widely used method of feeding, which also incorporates filter feeding, is a system where an organism secretes mucus to catch the detritus in lumps, and then carries these to its mouth using an area of cilia. This is called mucus feeding.

Many organisms, including sea slugs and serpent's starfish, scoop up the detritus which has settled on the water bed. Bivalves which live inside the water bed do not simply suck in water through their tubes, but also extend them to fish for detritus on the surface of the bed.

Producers

In contrast, from the point of view of organisms using photosynthesis such as plants and plankton, detritus reduces the transparency of the water and gets in the way of this process. Given that these organisms also require a supply of nutrient salts, in other words fertilizer, for photosynthesis, their relationship with detritus is a complex one.

In land ecosystems, the waste products of plants and animals collect mainly on the ground (or on the surfaces of trees), and as decomposition proceeds, plants are supplied with fertilizer in the form of inorganic salts. In water ecosystems, relatively little waste collects on the water bed, and so the progress of decomposition in water takes a more important role. Investigating the level of inorganic salts in sea ecosystems shows that unless there is an especially large supply, the quantity increases from winter to spring —but is normally extremely low in summer. As such, the quantity of seaweed present reaches a peak in early summer and then decreases. The thinking is that organisms like plants grow quickly in warm periods and thus the quantity of inorganic salts is not

enough to keep up with the demand. In other words, during winter, plant-like organisms are inactive and collect fertilizer, but if the temperature rises to some extent they will use this up in a very short period.

It is not entirely true that their productivity falls during the warmest periods. Organisms such as dinoflagellate have mobility, the ability to take in solid food, and the ability to photosynthesise. This type of micro-organism can take in substances such as detritus to grow, without waiting for it to be broken down into fertilizer.

From Wikipedia, November 2023 https://en.Wikipedia.org/wiki/Detritus

---

5.10: Detritus is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5.11: Soil organic matter and global change

Soil, land and climate change

Soil contains significant amounts of carbon and nitrogen, which can be released into the atmosphere depending on how we use the land. Clearing or planting forests, the melting of permafrost can tilt the greenhouse gas emission balance one way or the other. Climate change can also substantially alter what farmers can produce and where. The Food and Agriculture Organization of the United Nations (FAO) map below shows that the top 30 cm of the world's soil contains about twice as much carbon as the entire atmosphere. After oceans, soil is the second largest natural **carbon sink**, surpassing forests and other vegetation in its capacity to capture carbon dioxide from air. These facts remind us how important healthy soils are, not just for our food production but also for our efforts to prevent the worst effects of climate change.

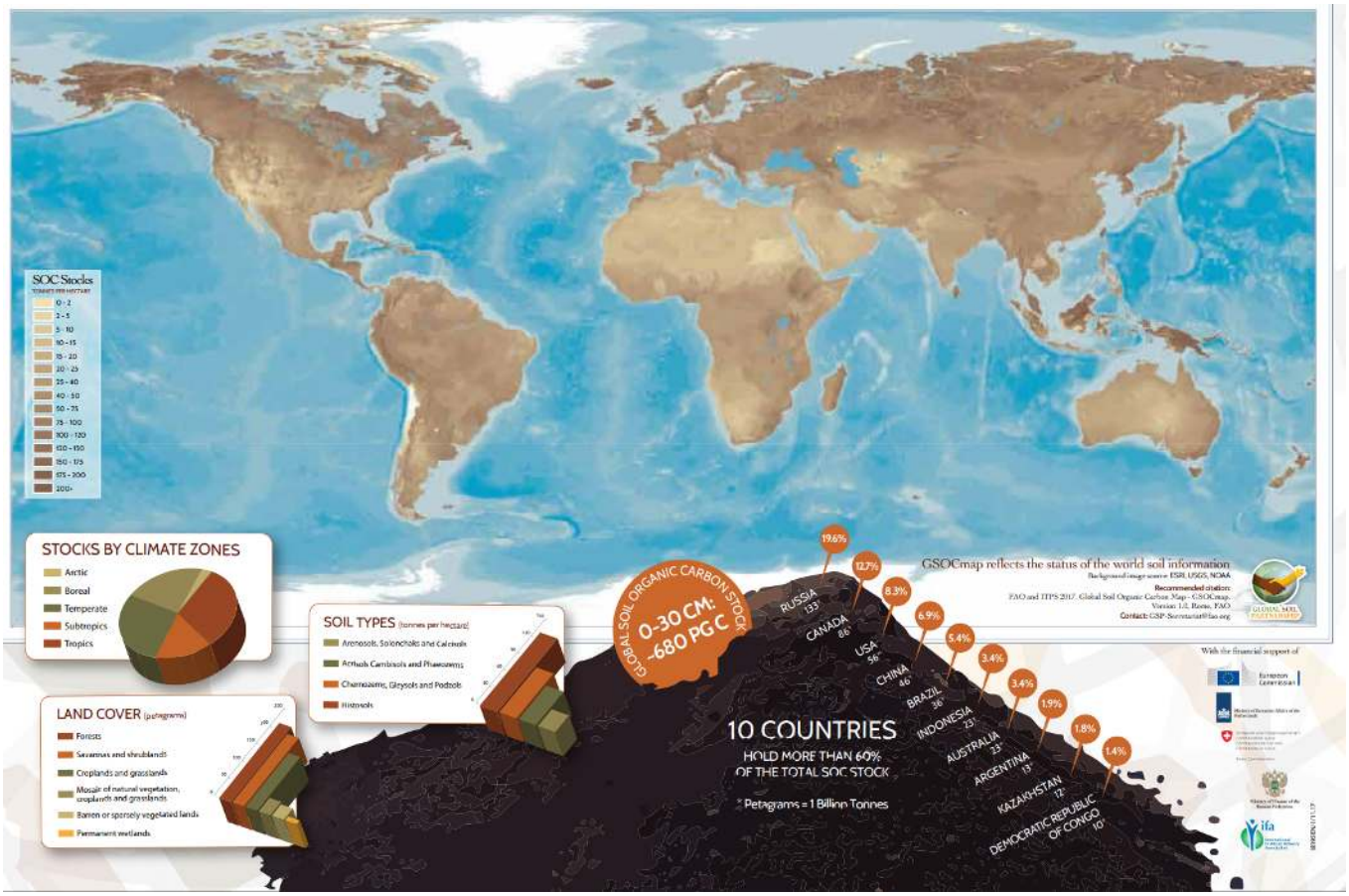


Figure 5.11.1: Copy and Paste Caption here. (Copyright; author via source)

Soil organic carbon (SOC) is the carbon that remains in the soil after partial decomposition of any material produced by living organisms. It constitutes a key element of the global carbon cycle through atmosphere, vegetation, soil, rivers and the ocean is the main component of soil organic matter (SOM) and as such constitutes the fuel of any soil. SOM supports key soil functions as it is critical for the stabilization of soil structure, retention and release of plant nutrients, and allowing water infiltration and storage in soil. It is therefore essential to ensuring soil health, fertility and food production. The loss of SOC indicates a certain degree of soil degradation. Soils represent the largest terrestrial organic carbon reservoir. Depending on local geology, climatic conditions and land use and management (amongst other environmental factors), soils hold different amounts of SOC.

The largest amounts of SOC have been estimated to be stored in the northern permafrost region with around 190 Pg C in the first 30 cm of the soil (0-30cm) [1] , mostly in peat soils. There, carbon accumulates in soils in huge quantities due to the low temperatures leading to low biological activity and slow SOM decomposition. The corresponding soil type is called Histosol and is characterized by a SOC content of 12 to 18% [2] . In contrast, in dry and hot regions such as the Sahara Desert, plant growth is naturally scarce and only very little carbon enters the soil. Arenosols, the typical soils of these areas, have mostly less than 0.6%

SOC [3]. Black soils, such as Chernozems, are inherently fertile because of their relatively high SOC content (over 1% [2]) and optimal plant growth conditions in terms of nutrient exchange capacity and a well-developed structure enabling sufficient water provision.

Unsustainable management practices such as excessive irrigation or leaving the soil bare endanger these soils, causing SOC loss and massive erosion. Caring for these soils and preserving the SOC they contain can be achieved through sustainable soil management, including mulching, planting cover crops, judicious fertilization and moderate irrigation.

Loss of SOC negatively affects not only soil health and food production, but also exacerbates climate change. When SOM is decomposed, carbon-based greenhouse gases are emitted to the atmosphere. If this occurs at too high rates, soils can contribute to warming our planet. On the flip side, many soils have the potential to increase their SOC stocks, thus mitigating climate change by reducing the atmospheric CO2 concentration.

The Global Soil Organic Carbon Map (GSOCmap), a country driven endeavour, allows the estimation of SOC stock from 0 to 30 cm. The GSOCmap represents the first ever global soil organic carbon assessment produced through a participatory approach in which countries developed their capacities and stepped up efforts to compile all the available soil information at national level. In many cases, this is paving the way to establishing national soil information systems and represents the first step toward introducing a soil monitoring program.

Changes in seasonal temperatures can also shift the annual cycles of plants and animals, resulting in lower yields. For example, spring can arrive earlier and trees can blossom before their pollinators have hatched. With the expected population growth, world food production needs to increase rather than decrease. This hinges largely on maintaining healthy soil and managing agricultural areas sustainably. At the same time, there is a growing demand for biofuels and other plant-based products, driven by the urgent need to replace fossil fuels and prevent greenhouse gas emissions.

There are other impacts on soil related to climate change, including **erosion**, which can be accelerated by extreme climate events, such as intense rain, drought, heat waves and storms. In addition to causing the loss of areas of land, **rising sea levels** may change soil in coastal areas or bring contaminants, including salt, from the sea. In relation to land use, climate change may make some agricultural areas, mainly in the south, unusable or less productive while possibly opening up new possibilities further north. In forestry, the decline in economically valuable tree species might cut the value of forest land in Europe by between 14 and 50 % by 2100. The overall impacts of climate change could produce a significant loss for the agricultural sector with large regional variations.

Yet perhaps the biggest climate concern linked to soil is the carbon dioxide and methane stored in permafrost in boreal regions, mainly in Siberia. As the global temperatures increase, the permafrost melts. This thawing causes the organic material trapped in the frozen soil to disintegrate, which can lead to the release of massive amounts of greenhouse gases into the atmosphere, which could hence lead to the accelerating of global warming far beyond people's control.

Tackling the climate crisis with soil

In April 2019, a group of highly influential scientists and activists called for 'defending, restoring and re-establishing forests, peatlands, mangroves, salt marshes, natural seabeds and other crucial ecosystems' to let nature remove carbon dioxide from the atmosphere and store it. Restoring ecosystems would also support biodiversity and enhance a wide range of ecosystem services, including cleaning air and water, and providing people with enjoyable spaces for recreation.

Despite the uncertainties, restoring ecosystems and improving soil quality could be a very cost-efficient measure in terms of **climate action** with a triple impact. First, growing plants remove carbon dioxide from the atmosphere. According to the FAO, restoring currently degraded soils could remove up to 63 billion tonnes of carbon, which would offset a small but important share of global greenhouse gas emissions. Second, healthy soils keep the carbon underground. Third, many natural and semi-natural areas act as powerful defences against the impacts of climate change.

The examples of benefits are many. For example, areas next to rivers (riparian zones) and green spaces in cities can act as cost-effective **protection against floods and heat waves**. Healthy land and soil can absorb and store excess water and alleviate floods. Parks and other natural areas in cities can also help with cooling down during heat waves, partly because of the water present in their soil. During dry seasons, healthy ecosystems can slowly release the water they have stored underground, mitigating the worst impacts of droughts.

Capturing the carbon in the air

There are also various methods for increasing land's capacity to **capture carbon dioxide** from air. A recent European research project (Caprese study) found that the conversion of arable land to grassland is the most rapid way of increasing the amount of carbon in soil. For arable land, the use of cover crops — plants such as clover grown in between harvest and sowing the next crop mainly to increase soil fertility and avoid erosion — was the most effective way of increasing carbon stocks in soil.

In contrast, decisions to use land differently can also change areas, making them sources of emissions. Notable examples of this are draining **peatlands**, burning peat from bogs for heating, ploughing up grassland and cropland, which releases previously stored carbon. For **forests**, the dynamic is the same but with a different timescale. Like soil, forests are both carbon stocks and carbon sinks, meaning that they both store carbon and capture it from the air. In many cases, young, growing forests capture carbon more rapidly than old forests but harvesting old forests removes the carbon stock from the forest. Depending on how the wood is used, the carbon may be released sooner, such as when the wood is burned for heating, or much later, when the wood is used for building houses, for example.

Healthier soils and land ecosystems could capture and store more carbon dioxide from the atmosphere than they currently do. Green spaces and natural areas could also help people and nature to adapt to the inevitable changes in our climate. Soil alone cannot fix climate change but it needs to be factored in and could be a powerful partner in our efforts.

Excerpted from:

Food and Agricultural Organization, "World's most comprehensive map showing the amount of carbon stocks in the soil launched" November 2023, Accessed November 2023
https://web.archive.org/web/20180128061450/http://www.fao.org/3/a-i8195e.pdf

European Environment Agency, 2019. Accessed November 2023
https://www.eea.europa.eu/signals-archived/signals-2019-content-list/articles/soil-land-and-climate-change#:~:text=Climate%20change%20affects%20soil&text=Continuing%20declines%20in%20soil%20moisture,dramatic%20impacts%20on%20food%20production.

5.11: Soil organic matter and global change is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by European Environment Agency, 2019 & FAO 2017.

5.12: Summary

5.12: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

6: Biogeochemical Cycling on Land

- 6.1: Biogeochemical cycling in land plants
- 6.2: Nutrient allocations and cycling in land vegetation
- 6.3: Biogeochemical cycling in the soil
  - 6.3.1: Soils and the carbon cycle
  - 6.3.2: Soils and the Water Cycle
- 6.4: Calculating landscape mass balance
- 6.5: Ecosystem models and remote sensing
- 6.6: Ecosystem Monitoring
- 6.7: Human impacts on terrestrial biogeochemistry
  - 6.7.1: Introduction
  - 6.7.2: Human-Induced Changes
  - 6.7.3: Sinks and Cycles
  - 6.7.4: Impacts and Options
  - 6.7.5: References
- 6.8: Summary

6: Biogeochemical Cycling on Land is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.1: Biogeochemical cycling in land plants

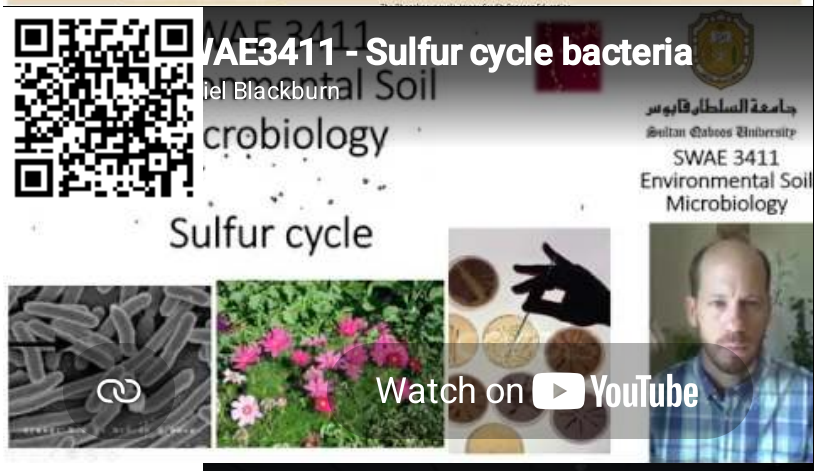


Maps take some time to build because we have to find or write matching materials. LibreTexts POV is that it is best to make available pages that we have finished rather than wait till the entire project is complete .This map is not completely finished, some pages are missing but we are workin' on it. . . If you know of some open materials which could be used here let us know info@libretexts.org (Public Domain ; Public Domain Pictures)

Till then here are some videos from Daniel Blackburn at Sultan Qaboos University that cover the subject









6.1: Biogeochemical cycling in land plants is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.2: Nutrient allocations and cycling in land vegetation

Table 6.2.1: Management actions affecting soil nutrient cycles and their impact on ecosystem services.

Management action or other driver of change	Provisioning service impact	Regulating service impact	Supporting service impact	Cultural service impact
Intensive addition of mineral fertilizers	Increased food, fibre, and feedstock production	Reduced water quality through eutrophication, reduced air quality through emission and volatilization of reactive N gases	Increased primary production; alteration of the nutrient and C cycling; possible reduction of biodiversity	
Use of organic soil amendments (e.g. manure, composts and biochar)	Increased food, fibre, and feedstock production; may increase water retention	Increase C sequestration	Increase nutrient retention	
Implementation of no-till			Increase nutrient retention	
Precision agriculture	Increase efficient production of food	Reduced GHG emissions per unit production	Reduce consumption of water and nutrient by improving use efficiency	
Prescribed use of fire for pasture management	Increase feedstock production	Increase C sequestration by conversion to BC	Reduce N recycling by storing black nitrogen	
Use of biological soil supplements	Stimulate productivity; act as fertilizers	May improve pest and disease control	Improved nutrient cycling	

Soils support primary production among other services, which in turn delivers the provisioning services of food and fibre production (Table 6.2.1). As such, soils are vital to humanity since they provide essential nutrients, such as nitrogen (N), phosphorus (P), and potassium (K) and many trace elements that support biomass production, which is essential for the supply of human and animal food, for energy and fibre production and (future) feedstock for the chemical industry (Table 6.2.1). Since the 1950s, higher biomass production and yield increases have been supported by fertilizers derived from mined minerals or industrial synthesis (Fig. 6.2.1). Intensification of agricultural practices and land use has in many regions resulted in a decline in the content of organic matter in agricultural, arable soils (Table 6.2.1; Matson et al., 1997; Smith et al., 2015). In some areas, extensive use of mineral fertilizers has led to atmospheric pollution, greenhouse gas emissions (e.g. N_2O, very important for climate regulation), water eutrophication, and human health risks (Galloway et al., 2008), thereby negatively affecting the regulating services of soil, air, and water quality (Table 6.2.1; Smith et al., 2013). During the 21st century, it is likely that the human population and demand for food, feed, and energy will rise. In order to sustain biomass production in the future, and to avoid negative environmental impacts, fertile soils need to be preserved and soil fertility needs to be restored where lost. This can be done through both the recycling and accumulation of sufficient amounts of organic matter in soils (Janzen, 2006), through a combination of plant production and targeted additions of organic and mineral amendments to soils.

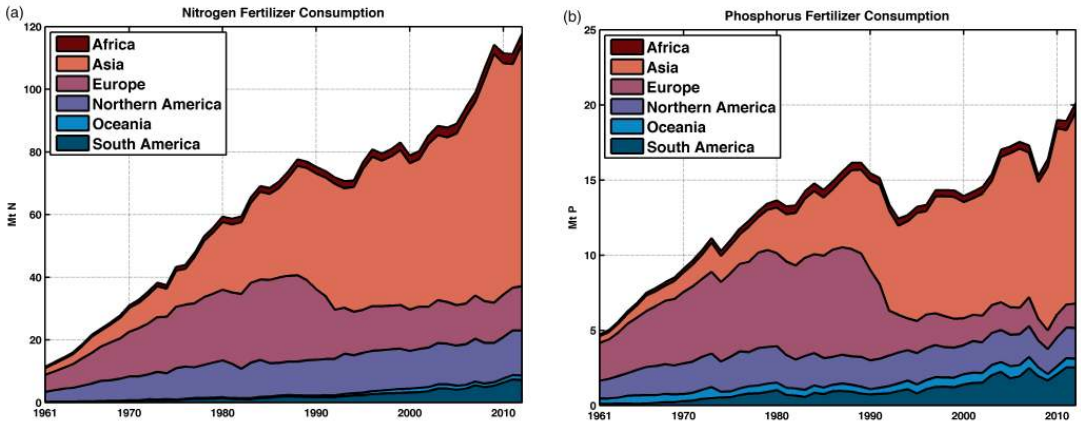


Figure 6.2.1:Global (a) nitrogen (N) and (b) phosphorus (P) fertilizer use between 1961 and 2012 split for the different continents in Mt P per year; plotted from FAOSTAT data (FAOSTAT, 2015). (Copyright: CC-BY, P. Smith, et al., SOIL https://soil.copernicus.org/articles/1/665/2015/)

The soil function "fertility" refers to the ability of soil to support and sustain plant growth, which relates to making available N, P, other nutrients, water, and oxygen for root uptake. This is facilitated by (i) their storage in soil organic matter, (ii) nutrient recycling from organic to plant available mineral forms, and (iii) physical-chemical processes that control their sorption, availability, displacement, and eventual losses to the atmosphere and

water (Table 6.2.1). Managed soils are a highly dynamic system and it is this very dynamism that makes the soil work and supply ecosystem services to humans. Overall, the fertility and functioning of soils strongly depend on interactions between the soil mineral matrix, plants, and microbes; these are responsible for both building and decomposing SOM, and therefore for the preservation and availability of nutrients in soils (Cotrufo et al., 2013). To sustain this service, the cycling of nutrients in soils must be preserved (Table 6.2.1).

After carbon, N is the most abundant nutrient in all forms of life, since it is contained in proteins, nucleic acids, and other compounds (Galloway et al., 2008). Humans and animals ultimately acquire their N from plants, which on land is mostly taken up in mineral form (i.e. NH_4^+ and NO_3^-) from the soil. The parent material of soils does not contain significant amounts of N (most other nutrients such as P largely originate from the parent material). New N mostly enters the soil through the fixation of atmospheric N_2 by a specialized group of microorganisms. However, the largest flux of N within the soils is generated through the continuous recycling of N internal to the plant-soil system: soil mineral N is taken up by the plant, is fixed into biomass, and eventually N returns in the form of plant debris to the soil. Here microorganisms decompose it, mineralizing part of the N and making it newly available for plant growth, while transforming the other part into SOM, which ultimately is the largest stock of stable N in soil. Generally, N cycles tightly in the system with minimal losses. Nitrogen is lost from the soil to the water system by leaching and to the atmosphere by gas efflux (NH_4, N_2O, and N_2). In most terrestrial natural ecosystems, N availability limits productivity. Through the cultivation of N_2 fixing crops, the production and application of mineral N fertilizer, the increasing application of animal manure from livestock and bio-wastes, and the unintentional deposition of atmospheric reactive N (ultimately derived from industrialera human activities), humans have applied twice as much reactive N to soils as the N introduced by natural processes, significantly increasing biomass production on land (Vitousek and Matson, 1993; Erisman et al., 2008). In some regions of the world, mineral fertilizer is applied in excess of plant requirement, but in other regions, in particular in Sub-Saharan Africa, where economic constraints limit the use of fertilizers, productivity is strongly limited by soil available N and other nutrients, notably P and K (N and P; Fig. 6.2.2).

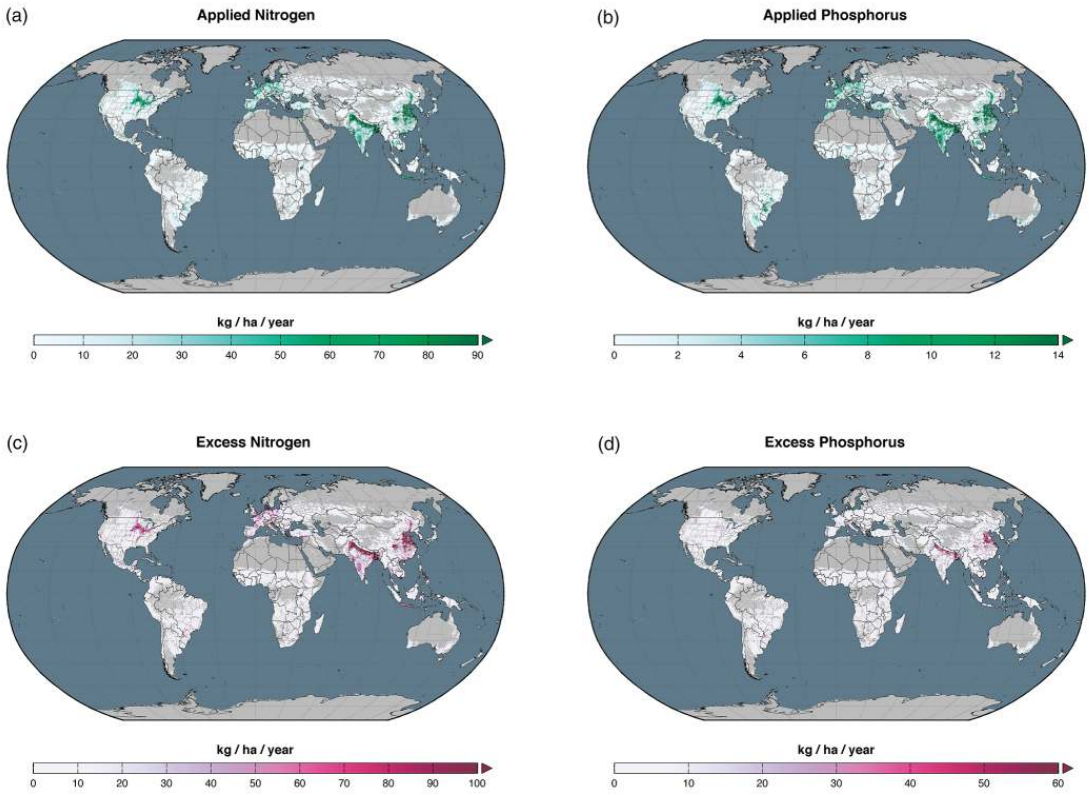


Figure 6.2.2: Applied and excess nitrogen and phosphorus in croplands. Nitrogen and phosphorus inputs and excess were calculated using a simple mass balance model (West et al., 2014), extended to include 175 crops. To account for both the rate and spatial extent of croplands, the data are presented as kilogram per hectare of the landscape. (a) Applied nitrogen, including N deposition; (b) applied phosphorus; (c) excess nitrogen; and (d) excess phosphorus. (Copyright: CC-BY, P. Smith, et al., SOIL https://soil.copernicus.org/articles/1/665/2015/)

Phosphorus derived from parent material, through weathering, cycles internally in the plant-soil system between biochemical molecules (e.g. nucleic acid, phospholipids) and mineral forms after decomposition (e.g. H_3PO_4). In soils, P is among the most limiting of nutrients, since it occurs in small amounts and is only available to plants in its dissolved ionic forms, which promptly react with calcium, iron, and aluminium cations to form highly insoluble compounds. Largely in these forms, P is lost to the aquatic system through erosion and surface runoff. Losses may also occur in dissolved form, for instance via subsurface flow and groundwater (McDowell et al., 2015). An important form of loss is in the export of organic P in agricultural products. Due to widespread agricultural P deficiencies, humans started to mine "primary" P from guano or rock phosphate deposits and added it to soils in the form of mineral fertilizer (Fig. 2). This external input has led to positive agronomic P balances (MacDonald et al., 2011) and excesses of P

and N in many regions (West et al., 2014; Fig. 6.2.2). There are large variations across the world, with high surpluses in the USA, Europe, and Asia and deficits in Russia, Africa, and South America (Fig. 3). Since plant P uptake is a relatively inefficient process with roughly 60% of the total P input to soils not taken up in the short term, a 3 -fold increase in the export of P to water bodies has been estimated, with significant impacts on water quality (Bennett et al., 2001).

Clearly, management practices need to be implemented that sustain, restore, or increase soil fertility and biomass production by promoting the accrual of SOM and nutrient recycling, applying balanced C amendments and fertilization of $\mathrm{N, P}$, and other nutrients to meet plant and soil requirements, while limiting the addition of excess fertilizer and retaining nutrients in the soil-plant system (Table 6.2.1). C, N, and P cycling in soils is coupled by tight stoichiometric relationships (e.g. relatively fixed C : N : P in plants and microorganisms; Güsewell, 2004); thus their management needs to be studied in concert. Nutrient management has been extensively studied, with the aim of identifying and proposing management practices (e.g. precision agriculture) that improve nutrient use efficiency and productivity and reduce potentially harmful losses to the environment (Table 6.2.1; van Groenigen et al., 2010; Venterea et al., 2011). Yet, our ability to predict the ecosystem response to balanced fertilization is still limited, and effectiveness and reliability would benefit from continued monitoring of efforts. Further benefits are anticipated from improved plant varieties with root morphologies that have better capacity to extract P from soils or use it more efficiently, perhaps in concert with mycorrhizal symbionts. Fertilization with nutrients other than N and P has been less well explored within the realm of understanding soil organic matter responses to agricultural C inputs and the potential to restore and increase soil organic matter (e.g. Lugato et al., 2006). Hence, we stress the importance of an integrated approach to nutrient management, which supports plant productivity while preserving or enhancing SOM stocks, and reducing nutrient losses to the atmosphere or water resources. Several issues exist where prediction and optimization of performance would benefit from relevant and continued data acquisition for the range of climate and environmental and agro-ecological conditions. Table 6.2.1 summarizes some management actions affecting soil nutrient cycles and their impacts on ecosystem services.

References

Bennett, E. M., Carpenter, S. R., and Caraco, N. F.: Human impact on erodable phosphorus and eutrophication: a global perspective. Increasing accumulation of phosphorus in soil threatens rivers, lakes, and coastal oceans with eutrophication, BioSci., 51, 227– 234, 2001.

Cotrufo, M. F., Wallenstein, M. D., Boot, C., Denef, K., and Paul, E.: The microbial efficiency-matrix stabilisation (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable organic matter?, Glob. Change Biol., 19, 988–995, 2013.

Galloway, J. N., Townsend, A. R., Erisman, J. W., Bekunda, M., Cai, Z., Freney, J. R., Martinelli, L. A., Seitzinger, S. P., and Sutton, M. A.: Transformation of the nitrogen cycle: recent trends, questions, and potential solutions, Science, 320, 889–892, 2008

Güsewell, S.: N : P ratios in terrestrial plants: variation and functional significance, New Phytol., 164, 243–266, 2004.

Janzen, H. H.: The soil carbon dilemma: Shall we hoard it or use it?, Soil Biol. Biochem., 38, 419–424, 2006

Lugato, E., Berti, A., and Giardini, L.: Soil organic carbon (SOC) dynamics with and without residue incorporation in relation to different nitrogen fertilisation rates, Geoderma, 135, 315–321, 2006.

Matson, P. A., Parton, W. J., Power, A. G., and Swift, M. J.: Agricultural intensification and ecosystem properties, Science, 277, 504–509, 1997.

McDowell, R. W., van der Weerden, T. J., and Campbell, J.: Nutrient losses associated with irrigation, intensification and management of land use: a study of large scale irrigation in North Otago, New Zealand, Agric. Water Manage., 98, 877–885, 2011.

McDowell, R. W., Cox, N., Daughney, C. J., Wheeler, D., and Moreau, M.: A national assessment of the potential linkage between soil, and surface and groundwater concentrations of phosphorus, J. Am. Water Res. Assoc., 51, 992–1002, 2015.

Smith, P., Ashmore, M., Black, H., Burgess, P. J., Evans, C., Quine, T., Thomson, A. M., Hicks, K., and Orr, H.: The role of ecosystems and their management in regulating climate, and soil, water and air quality, J. Appl. Ecol., 50, 812–829, 2013.

Smith, P., House, J. I., Bustamante, M., Sobocká, J., Harper, R., Pan, G., West, P. C., Clark, J. M., Adhya, T., Rumpel, C., Paustian, K., Kuikman, P., Cotrufo, M. F., Elliott, J. A., McDowell, R., Griffiths, R. I., Asakawa, S., Bondeau, A., Jain, A. K, Meersmans, J., and Pugh, T. A. M.: Global change pressures on soils from land use and management, Glob. Change Biol., in press, doi:10.1111/gcb.13068, 2015.

Van Groenigen, J. W., Velthof, G. L., Oenema, O., Van Groenigen, K. J., and Van Kessel, C.: Towards an agronomic assessment of N2O emissions: a case study for arable crops, Eur. J. Soil Sci., 61, 903–913, 2010.

Venterea, R. T., Maharjan, B., and Dolan M. S.: Fertilizer source and tillage effects on yield-scaled nitrous oxide emissions in a corn cropping system, J. Environ. Qual., 40, 1521–1531, 2011.

Vitousek, P. M. and Matson, P. A.: Biological diversity and terrestrial ecosystem biogeochemistry, in: The biogeochemistry of global change: radiative trace gases, edited by: Oremland, R. S., Chapman and Hall, New York, 193–208, 1993.

West, P. C., Gerber, J. S., Engstrom, P. M., Mueller, N. D., Brauman, K. A., Carlson, K. M., Cassidy, E. S., Johnston, M., MacDonald, G. K., Ray, D., and Siebert, S.: Leverage points for improving global food security and the environment, Science, 345, 325–328, 2014

Excerpted from:

Smith, P., Cotrufo, M. F., Rumpel, C., Paustian, K., Kuikman, P. J., Elliott, J. A., McDowell, R., Griffiths, R. I., Asakawa, S., Bustamante, M., House, J. I., Sobocká, J., Harper, R., Pan, G., West, P. C., Gerber, J. S., Clark, J. M., Adhya, T., Scholes, R. J., and Scholes, M. C.: Biogeochemical cycles and biodiversity as key drivers of ecosystem services provided by soils, SOIL, 1, 665–685, https://doi.org/10.5194/soil-1-665-2015, 2015. https://soil.copernicus.org/articles/1/665/2015/

6.2: Nutrient allocations and cycling in land vegetation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

SECTION OVERVIEW

6.3: Biogeochemical cycling in the soil

Soils play a pivotal role in major global biogeochemical cycles (carbon, nutrient, and water), while hosting the largest diversity of organisms on land. Because of this, soils deliver fundamental ecosystem services, and management to change a soil process in support of one ecosystem service can either provide co-benefits to other services or result in trade-offs.

6.3.1: Soils and the carbon cycle

6.3.2: Soils and the Water Cycle

Excerpted from:

Smith, P., Cotrufo, M. F., Rumpel, C., Paustian, K., Kuikman, P. J., Elliott, J. A., McDowell, R., Griffiths, R. I., Asakawa, S., Bustamante, M., House, J. I., Sobocká, J., Harper, R., Pan, G., West, P. C., Gerber, J. S., Clark, J. M., Adhya, T., Scholes, R. J., and Scholes, M. C.: Biogeochemical cycles and biodiversity as key drivers of ecosystem services provided by soils, SOIL, 1, 665–685, https://doi.org/10.5194/soil-1-665-2015, 2015.

6.3: Biogeochemical cycling in the soil is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.3.1: Soils and the carbon cycle

Soil C stocks:

Carbon (C) storage is an important ecosystem function of soils that has gained increasing attention in recent years. Changes in soil C impacts on, and feedbacks to, the Earth's climate system through emissions of CO_2 and CH_4 as well as storage of carbon removed from the atmosphere during photosynthesis (climate regulation; Table 6.3.1.1). Soil organic matter itself also confers multiple benefits for human society, e.g. enhancing water purification and water holding capacity, protecting against erosion risk, and enhancing food and fibre provision through improved soil fertility (Table 6.3.1.1; Pan et al., 2013, 2014).

Table 6.3.1.1: Management actions affecting the soil carbon cycle and their impact on ecosystem services

Management action or other driver of change	Provisioning service impact	Regulating service impact	Supporting service impact	Cultural service impact
Land use change (conversion of for-est/grassland/wetland to cropland)	Increased production of food, fibre, and energy crops; reduced availability of natural raw materials; potential change in hydrology/water availability	Decreased soil C sequestration and storage - increased GHG flux; increased erosion and sediment yield - reduced regulations of water flow and quality	Primary production may be changed; nutrient recycling reduced if no inputs, increased if there are inputs	Lower recreation value; may have impact on cultural value in recreating diverse landscapes
Land use change (establishment of forest or grassland on agricultural land)	Raw material provision may be increased; agricultural production likely decreased (but not always, e.g. agroforestry)	Increased C sequestration; increased regulation of water flow and quality	Primary production may be changed; increased water recycling	Increased recreation value; may have impact on cultural value in recreating diverse landscapes
Intensified nutrient management through fertilization and liming	Increased production of food and other raw materials	Effect on net soil C sequestration uncertain; increased GHG flux from fertilizer production and use; water and air pollution	Increased primary production; increased nutrient recycling	
Soil amelioration using organic amendments such as compost and biochar	Increased food production; more raw materials; more water available for plant growth	Increased C sequestration; increased water purification value	Increased primary production; increased nutrient cycling; improved water infiltration and retention	
Diversification of crop production systems (i.e. more perennials, reduced bare fallow)	Potential impact on agricultural production (±); more diverse products	Increased C sequestration; increased purification value	Changed primary production; increased nutrient retention; improved water infiltration and retention	Improved cultural value from more diverse landscapes
Replacement of hay forage production with pasture use on grasslands	No impact	Effect on C sequestration uncertain		Increased recreation value; may have impact on cultural value in recreating diverse landscapes
Improved grazing management	Increased food production; reduced runoff and improved water use	Increased C sequestration; increased purification value; water flow regulation	Increased primary production; improved water infiltration and retention	

Soil is an important C reservoir that contains more C (at least $1500 - 2400 \mathrm{PgC}$) than the atmosphere ($590 \mathrm{PgC}$) and terrestrial vegetation ($350 - 550 \mathrm{PgC}$) combined (Schlesinger and Bernhardt, 2013; Ciais et al., 2013), and an increase in soil C storage can reduce atmospheric CO_2 concentrations (Table 6.3.1.1; Whitmore et al., 2014). All three reservoirs of C are in constant exchange but with various turnover times, with soil as the largest active terrestrial reservoir in the global C cycle (Lal, 2008). Carbon storage in soils occur in both organic and inorganic form. Organic C stocks in the world's soils have been estimated to comprise $1500 \mathrm{Pg}$ of C to $1 \mathrm{m}$ depth and $2500 \mathrm{Pg}$ to $2 \mathrm{m}$ (Batjes, 1996). Recent studies have shown that the soil C pool to $1 \mathrm{m}$ depth may be even greater

and could account for as much as $2000\mathrm{Pg}$. These higher values are mainly based on increased estimates of the C stored in boreal soils under permafrost conditions (Tarnocai et al., 2009), in which decomposition is inhibited by low temperature, lack of oxygen, and low pH in waterlogged soils, e.g. peats (Smith et al., 2010). Although the highest C concentrations are found in the top $30\ \mathrm{cm}$ of soil, the major proportion of total C stock is present below $30\ \mathrm{cm}$ depth (Batjes, 1996). In the northern circumpolar permafrost region, at least 61% of the total soil C is stored below $30\ \mathrm{cm}$ depth (Tarnocai et al., 2009). Peatlands are particularly important components of the global soil carbon store, covering only 3% of the land area but containing around $500\mathrm{PgC}$ in organic-rich deposits ranging from 0.5 to $8\ \mathrm{m}$ deep (Gorham, 1991; Yu, 2012), with storage in deeper layers as yet unquantified.

In arid and semi-arid soils, significant inorganic C can be present as carbonate minerals (typically $\mathrm{Ca/MgCO_3}$, called "calcrete" or "caliche" in various parts of the world), formed from the reaction of bicarbonate (derived from $\mathrm{CO_2}$ in the soil) with free base cations, which can then be precipitated in subsoil layers (Nordt et al., 2000). Soils derived from carbonate-containing parent material (e.g. limestone) can also have significant amounts of inorganic C. The inorganic C pool globally is large, estimated to be $\sim 750\mathrm{Pg\ C}$ to a depth of $1\ \mathrm{m}$ (Batjes, 1996). However, in most cases, changes in inorganic C stocks are slow, are not amenable to traditional soil management practices, and do not play a significant role in terms of most ecosystem services (though a major exception is the geoengineering proposal to add finely ground silicate minerals to soils, which will then weather to carbonates, taking up CO2 in the process; Köhler et al., 2010).

The net balance of soil C depends on the inputs of C to soils relative to C losses. Losses can occur via mineralization (i.e. decomposition), leaching of dissolved C, and carbonate weathering (Smith, 2012; Schlesinger and Bernhardt, 2013). Thus, the soil organic C stock may either increase or decrease in response to changes in climate and land use practices (Smith et al., 2015). Furthermore, rates of SOC stock change in different parts of the profile can vary for different

soils and types of perturbation, because some portion of the C stored in soil, mainly in topsoil, turns over rapidly, while other soil C fractions can have a long residence time (von Lützow et al., 2008; Rumpel and Kögel-Knabner, 2011). The accumulation of stabilized C with long residence times in deep soil horizons may be due to continuous transport, temporary immobilization and microbial processing of dissolved organic matter within the soil profile (Kalbitz and Kaiser, 2012), and/or efficient stabilization of root-derived organic matter within the soil matrix (Rasse et al., 2005). The process of soil formation - i.e. the development of depth, horizons, and specific properties - is itself a supporting service (Table $6.3.1.1$).

High SOC content also improves other chemical and physical soil properties, such as nutrient storage (supporting service), water holding capacity (supporting and regulating service), aggregation, and sorption of organic or inorganic pollutants (regulating service). Carbon sequestration in soils may therefore be a cost-effective and environmentally friendly way to not only store C for climate regulation but also enhance other ecosystem services derived from soil, such as agricultural production, clean water supply, and biodiversity (Table $6.3.1.1$; Pan et al., 2013) by improving soil organic matter (SOM) content and thereby soil quality (Lal, 2004). Moreover, processes which improve SOM may themselves provide services, e.g. use of cover crops, which can provide provisioning or water regulation services while improving soil C (Table $6.3.1.1$). SOM or soil carbon are widely used proxy variables for soil health (e.g. Kibblewhite et al., 2008).

C cycling:

Carbon enters the soil as aboveground or belowground plant litter and exudates. C input is not homogenous within the soil profile. Whereas topsoil receives higher amounts of aboveground litter, subsoil C originates from root C as well as dissolved C, transported down the soil profile. Root C has a greater likelihood of being preserved in soil compared to shoot C, and was therefore hypothesized to account for most of the SOC (Rasse et al., 2005). The majority of plant litter compounds pass through and are modified by the soil biota. Thus, SOM is composed of plant litter compounds as well as microbial and, to a smaller extent, faunal decomposition products (Paul, 2014). It is a complex biogeochemical mixture comprising molecules derived from organic material in all stages of decomposition. Some organic matter compounds, including microbial decomposition products, may be stabilized for centuries to millennia by binding to soil minerals or by physical occlusion into micro-aggregates (von Lützow et al., 2008), for example with iron oxyhydrates (Zhou et al., 2009), or through protection by occlusion within soil aggregates (Dungait et al., 2012). The inherent chemical recalcitrance of some plant litter compounds (e.g. lignin) has a minor influence on their longevity in soil (Thévenot et al., 2010), whereas the location of SOM within the soil matrix has a much stronger control on its turnover (Chabbi et al., 2009; Dungait et al., 2012). Mineralassociated SOM is predominantly composed of microbial

products (Miltner et al., 2012). Therefore, microbial use efficiency of plant inputs largely determines SOM stabilization through interaction with the mineral phase (Cotrufo et al., 2013), in addition to the environmental controls discussed elsewhere in this

section. In peatlands, organic matter is stabilized by high water tables that slow down biological activity and decomposition. SOM is mineralized to carbon dioxide (CO_2) in aerobic environments, or reduced to methane (CH_4) in anaerobic environments. Soil CO_2 efflux, resulting from SOM mineralization, and from rhizosphere respiration and inorganic C weathering, is the largest terrestrial flux of CO_2 to the atmosphere ($\sim 60 PgC$; the sink of carbon on the other hand contributes to the climate regulation service; Smith, 2004). This flux is an order of magnitude larger than anthropogenic CO_2 emissions due to fossil fuel burning and land use change ($1.1\ PgCyr^{-1}$; Ciais et al., 2013). Under anaerobic conditions, CH_4 is formed by methanogenic microorganisms. A proportion of this CH_4 is oxidized to CO_2 by methanotrophic microorganisms, but a proportion can be emitted from the soil surface (Reay et al., 2010). Since CH_4 is many times more potent as a greenhouse gas than CO_2 on a per-molecule or per-mass basis (Ciais et al., 2013), soil CH_4 emissions and their mitigation play an important role in climate regulation (Table 6.3.1.1).

Fire may affect many ecosystem services, including C sequestration. For fires in natural ecosystems, a decrease in soil C storage is often observed initially, but through positive effects on plant growth, as well as input of very stable pyrogenic C, C storage may increase at longer timescales (Knicker, 2007). An additional long-term C pool in many soils is pyrogenic carbon (PyC), formed from partially combusted (i.e. pyrolysed) biomass during wildfires or other combustion processes (Schmidt and Noack, 2000). Globally, soils are estimated to contain between 54 and $109 Pg$ PyC (Bird et al., 2015). Some of this PyC has a highly condensed aromatic structure that retards microbial decay, and can thus persist in soils for relatively long periods (Singh et al., 2012). Soil amended with industrially produced PyC (biochar) as a climate mitigation technique often shows no increase in soil respiration despite the additional carbon, the reduced ecosystem carbon turnover results in increased soil carbon storage (Stewart et al., 2013). PyC additions to soil affect regulating ecosystem services, such as C sequestration, nutrient cycling, and adsorption of contaminants. However, PyC properties, and as result their effect on ecosystem services, may be strongly dependent on fire conditions.

Factors influencing soil C storage: Fundamentally, the amount of C stored in a given soil is determined by the balance of C entering the soil, mainly via plant production but also through manures or amendments such as organic sludge or biochar, and C leaving the soil through mineralization (as CO_2), driven by microbial processes, and to a lesser extent leaching out of the soil of dissolved carbon and carbonate weathering. Locally, C can be lost or gained through soil erosion or deposition, leading to a redistribution of soil C, at landscape and regional scales (van Oost et al., 2007).

Consequently, the main controls on soil C storage are the amount and type of organic matter inputs, the efficiency by which this is used by microbes, and the capacity of the soil to retain it by physical or chemical stabilization (Cotrufo et al., 2013). In most natural and agricultural ecosystems, plant productivity and subsequent death and senescence of biomass provide the input of organic C to the soil system (Table 6.3.1.1). Thus, higher levels of plant residue inputs will tend to support higher soil organic carbon stocks, and vice versa (Paustian et al., 1997), though this does not continue indefinitely (Zvomuya et al., 2008). Plants also affect soil C cycling by their specific mycorrhizal associations (Brzostek et al., 2015). Shifts in specific mycorrhizal associations affect SOM storage by contributing to both SOM formation and decomposition. Ectomychorrizhal turnover is a dominant process of SOM formation (Godbold et al., 2006), possibly due to the more recalcitrant nature of the chitin in fungal tissues, compared to the cellulose and lignin in plant residues. In arbuscular mycorrhizal fungi, it has been suggested that glomalin, a highly resistant glycoprotein, has an active role in aggregate formation and SOM stocks (Rillig, 2004). Symbiotic mycorrhizal fungi can also directly impact the turnover of organic matter by the production of exo-enzymes (Averill et al., 2014; Finzi et al., 2015).

In many regions of the world, SOM accumulates because of inhibition of microbial SOM decomposition, due to cold, dry, or anoxic conditions (Trumbore, 2009). In general, when water is not limiting, higher soil temperatures increase the rate of microbial decomposition of organic matter. Thus soil temperature is a major control of SOM storage in soil C cycle models (Peltoniemi et al., 2007). The temperature sensitivity of SOM decomposition is not, however, as straightforward as represented in most models but varies between the many different forms of chemical and physical protection of organic matter in soil (Conant et al., 2011; Zheng et al., 2012). Water influences soil C storage through several processes. Moist, but well-aerated, soils are optimal for microbial activity and decomposition rates decrease as soils become drier. However, flooded (saturated) soils have lower rates of organic matter decay due to restricted aeration and thus often have very high amounts of soil C (e.g. peat soils). High precipitation may also lead to C transport down the soil profile as dissolved and/or particulate organic matter, as well as lateral transport through soil erosion and deposition. During dry periods, SOM decomposition is decreased, but after rewetting there may be an accelerated pulse of CO_2 emission in aerobic soils (Borken and Matzner, 2008), whereas drought and lowering water tables may increase decomposition in naturally anaerobic peats (Freeman et al., 2001; Clark et al., 2012). However, the effect of drought is not only

direct via soil microbial activity. There are feedback loops concerning drought and C storage via plant activities, such as litter input and rhizodeposition. Drought was found to affect

plant litter composition (Sanaullah et al., 2014), plant C flow and root exudation (Sanaullah et al., 2012), as well as the resulting enzyme activities in the rhizosphere (Sanaullah et al., 2011).

C cycling in soils is strongly linked to the cycling of N and P. Since the $C:N:P$ stoichiometry in SOM is generally lower than in plant material - i.e. there is more N and P per unit $C-C$ generally accumulates in aerobic soil where nutrients are not limiting (Alberti et al., 2014). Nevertheless, an increase in organic C is often accompanied by increased N resource use efficiency in croplands (Pan et al., 2009), especially when SOC is increased with biochar (Huang et al., 2013). In nutrient-limited peatlands, inputs of nitrogen and/or phosphorus within the tolerance levels of sensitive plant species have increased rates of carbon accumulation (Aerts et al., 1992; Turunen et al., 2004; Olid et al., 2014). The relationship between nutrients and C cycling is not straightforward, since nutrients are also needed by soil microbes to degrade SOM. Thus, nutrient addition can either decrease or increase C storage, depending on the initial SOM stoichiometry, the ability of the soil minerals to stabilize microbial products of decomposition, and the simultaneous effects on plant productivity and organic matter inputs to soils.

The amount and type of clay particles (and to a lesser extent silt particles) are the major factors controlling the quantity and composition of soil C (Sollins et al., 1996; von Lützow et al., 2006). Clays are mainly sheet-like crystals of silicon and aluminium, known as phyllosilicates, often located as skins coating soil aggregates. In clay-rich soils, higher organic matter content and a greater concentration of O-alkyl C derived from polysaccharides may be expected compared to sandy soil, which are characterized by lower C contents and high concentrations of alkyl C (Rumpel and Kögel-Knabner, 2011). Aliphatic material may be responsible for the hydrophobicity of soils, which can lead to reduced microbial accessibility and therefore increased C storage (Lorenz et al., 2007). Many of the OM-matrix interactions are driven by expandable and non-expandable phyllosilicates, which interact with organic compounds through their large surface areas, micropores, and micro-aggregation, particularly in acid soils. In neutral and calcareous soils, polyvalent cations (especially Ca^{2+}) predominate in the interaction mechanism, forming bridges between the largely negatively charged SOM and negatively charged phyllosilicates (Cotrufo et al., 2013). Short-order silicates, like allophane, provide some of the strongest organo-mineral interactions and stabilize both proteins and carbohydrate monomers, though their occurrence is very geographically restricted (Buurman et al., 2007; Dümig et al., 2012: Mikutta and Kaiser, 2011). Pedogenic oxides (for example iron oxyhydrates in rice paddies) usually act as a coating of soil mineral particles and stabilize carbon, contributing to a higher C storage and stability compared to other soils (Song et al., 2012).

Bioturbation (the mixing of soil by organisms) may further influence the amount as well as the chemical nature of soil C. It greatly influences the heterogeneity of soils by creating hotspots of carbon and biological activity. On biologically active sites, incorporation and transformation of organic compounds into soil is usually enhanced, leading to more organo-mineral interactions and increased C storage (Wilkinson et al., 2009).

Microbial decomposition of SOM may be stimulated by the input of labile (easily decomposed) organic matter through the priming effect (Jenkinson et al., 1971). Positive priming refers to greater mineralization of otherwise stable C through shifts in microbial community composition and activity (Fontaine et al., 2003). However, in some cases, the addition of organic matter to soil may also impede mineralization of native SOM (negative priming effect), thereby protecting SOM from its decomposition. Plant communities (Table 6.3.1.1) are the main controlling factors of these processes because they influence organic matter input and microbial activity by their effects on soil water, labile C input, pH, and nutrient cycling (Kuzyakov et al., 2000).

By storing and cycling C, nutrients, and water, soils provide supporting services like soil formation and nutrient and water retention, which underpin both primary production and landscape hydrology (the processes which deliver provisioning services such as food, fibre, and water; Table 6.3.1.1), in addition to the regulating services such as climate regulation already discussed (Fig. 1). To ensure that soils continue to provide these key services, soil will require to be managed for C preservation - thus mitigating climate change - while simultaneously permitting continued SOM recycling (Table 6.3.1.1). Janzen (2006) pointed to this dilemma, that there is a tradeoff between improved soil fertility to support the provisioning services of food/timber production and the regulating service of soil carbon sequestration aiding climate regulation. Despite knowledge on which practices are likely to lead to improved SOC status, a better understanding of the controls on SOM distribution, stabilization, and turnover will help to better target these practices. This will be an important contribution to the mitigation of greenhouse gases, while assuring decomposition and, with it, the cycling of nutrients necessary to support food production. Table 6.3.1.1 summarizes management actions affecting the soil carbon cycle and their impacts on ecosystem services.

References:

Aerts, R., Wallen, B., and Malmer, N.: Growth-limiting nutrients in sphagnum-dominated bogs subject to low and high atmospheric nitrogen supply, J. Ecol., 80, 131–140, 1992

Alberti, G., Vicca, S., Inglima, I., Belelli-Marchesini, L., Genesio, L., Miglietta, F., Marjanovic, H., Martinez, C., Matteucci, G., D'Andrea E., Peressotti, A., Petrella, F., Rodeghiero, M., and Cotrufo, M. F.: Soil C : N stoichiometry controls carbon sink partitioning between above-ground tree biomass and soil organic matter in high fertility forests, iForest, 8, 195–206, doi:10.3832/ifor1196-008, 2014.

Averill, C., Turner, B. L., and Finzi, A. C.: Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage, Nature, 505, 543–545, 2014.

Batjes, N. H.: Total carbon and nitrogen in the soils of the world, European J. Soil Sci., 47, 151–163, 1996.

Bird, M. I., Wynn, J. G., Saiz, G., Wurster, C. M., and McBeath, A.: The Pyrogenic Carbon Cycle, Ann. Rev. Earth Planet. Sci., 43, 9.1–9.26, doi:10.1146/annurev-earth-060614-105038, 2015.

Borken, W. and Matzner, E.: Reappraisal of drying and wetting effects on C and N mineralization and fluxes in soils, Glob. Change Biol., 15, 808–824, 2008.

Buurman, P., Bartoli, F., Basile, A., Füleky, G., Garcia Rodeja, E., Hernandez Moreno, J., and Madeira, M.: The physico-chemical database, in: Soils of Volcanic Regions in Europe, edited by: Arnalds, Ó., Óskarsson, H., Bartoli, F., Buurman, P., Stoops, G., and García-Rodeja, E., Springer, Berlin, 271–287, 2007.

Brzostek, E. R., Dragoni, D., Brown, Z. A., and Phillips, R. P.: Mycorrhizal type determines the magnitude and direction of rootinduced changes in decomposition in a temperate forest, New Phytol., 206, 1274–1282, 2015.

Chabbi, A., Kögel-Knabner, I., and Rumpel, C.: Stabilised carbon in subsoil horizons is located in spatially distinct parts of the soil profile, Soil Biol. Biochem., 41, 256–271, 2009.

Ciais, P. and Sabine, C.: Carbon and other biogeochemical cycles, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Edenhofer, O., Pichs-Madruga, R., and Sokonaet, Y., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013

Clark, J. M., Heinemeyer, A., Martin, P., and Bottrell, S. H.: Processes controlling DOC in pore water during simulated drought cycles in six different UK peats, Biogeochem., 109, 253–270, 2012.

Conant, R. T., Ryan, M. G., Ågren, G. I., Birge, H. E., Davidson, E. A., Eliasson, P. E., Evans, S.E., Frey, S. D., Giardina, C. P., Hopkins, F. M., Hyvönen, R., Kirschbaum, M. U. F., Lavallee, J. M., Leifeld, J., Parton, W. J., Steinweg, J. M., Wallenstein, M. D., Wetterstedt, J. A. M., and Bradford, M. A.: Temperature and soil organic matter decomposition rates – synthesis of current knowledge and a way forward, Glob. Change Biol., 17, 3392– 3404, 2011.

Cotrufo, M. F., Wallenstein, M. D., Boot, C., Denef, K., and Paul, E.: The microbial efficiency-matrix stabilisation (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable organic matter?, Glob. Change Biol., 19, 988–995, 2013.

Dümig, A., Häusler, W., Steffens, M., and Kögel-Knabner, I.: Clay fractions from a soil chronosequence after glacier retreat reveal the initial evolution of organo–mineral associations, Geochim. Cosmochim. Acta, 85, 1–18, 2012.

Dungait, J. A. J., Hopkins, D. W., Gregory, A. S., and Whitmore, A. P.: Soil organic matter turnover is governed by accessibility not recalcitrance, Glob. Change Biol., 18, 1781–1796, 2012.

Finzi, A. C., Abramoff, R. Z., Spiller, K. S., Brzostek, E. R., Darby, B. A., Kramer, M. A., and Phillips, R. P.: Rhizosphere processes are quantitatively important components of terrestrial carbon and nutrient cycles, Glob. Change Biol., 21, 2082–2094, 2015

Fontaine, S., Mariotti, A., and Abbadie, L.: The priming effect of organic matter: a question of microbial competition, Soil Biol. Biochem., 35, 837–843, 2003.

Freeman, C., Ostle, N., and Kang, H.: An enzymic “latch” on a global carbon store, Nature, 409, 149, 2001.

Godbold, D. L., Hoosbeek, M. R., Lukac, M. Cotrufo, M. F., Janssens, I. A., Ceulemans, R., Polle, A., Velthorst, E. J., Scarascia-Mugnozza, G., De Angelis, P., Miglietta, F., and Peressotti, A.: Mycorrhizal hyphal turnover as a dominant process for carbon input into soil organic matter, Plant Soil 281, 15–24, 2006.

Gorham, E.: Northern peatlands: Role in the carbon cycle and probable responses to climatic warming, Ecol. Appl., 1, 182–195, 1991.

Huang, M., Yang, L., Qin, H. D., Jiang, L. G., and Zou, Y. B.: Quantifying the effect of biochar amendment on soil quality and crop productivity in Chinese rice paddies, Field Crops Res., 154, 172– 177, 2013.

Kalbitz, K. and Kaiser, K.: Cycling downwards – dissolved organic matter in soils, Soil Biol. Biochem., 52, 29–32, 2012.

Kibblewhite, M., Ritz, K., and Swift, M.: Soil health in agricultural systems, Phil. Trans. R. Soc. B, 363, 685–701, 2008.

Knicker, H.: How does fire affect the nature and stability of soil organic nitrogen and carbon? A review, Biogeochem., 85, 91– 118, 2007.

Kuzyakov, Y., Freidel, J. K., and Stahr, K.: Review of mechanisms and quantification of priming effects, Soil Biol. Biochem., 32, 1485–1498, 2000.

Lal, R.: Sequestration of atmospheric CO2 in global carbon pools, Energy Env. Sci., 1, 86–100, 2008.

Lorenz, K., Lal, R., Preston, C. M., and Nierop, K. G. I. M: Strengthening the soil organic carbon pool by increasing contributions from recalcitrant aliphatic bio(macro)molecules, Geoderma, 142, 1–10, 2007.

Mikutta, R. and Kaiser, K.: Organic matter bound to mineral surfaces: Resistance to chemical and biological oxidation, Soil Biol. Biochem., 43, 1738–1741, 2011.

Miltner, A., Bombach, P., Schmidt-Brücken, B., and Kästner, M.: SOM genesis: microbial biomass as a significant source, Biogeochem., 111, 41–55, 2012

Olid, C., Nilsson, M. B., Eriksson, T., and Klaminder, J.: The effects of temperature and nitrogen and sulfur additions on carbon accumulation in a nutrient-poor boreal mire: Decadal effects assessed using 210 Pb peat chronologies, J. Geophys. Res.-Biogeosci., 119, 392–403, 2014.

Pan, G., Zhou, P., Li, Z. P., Smith, P., Li, L. Q., Qiu, D. S., Zhang, X. H., Xu, X. B., Shen, S. Y., and Chen, X. M.: Combined inorganic/organic fertilization enhances N efficiency and increases rice productivity through organic carbon accumulation in a rice paddy from the Tai Lake region, China, Agr. Ecosys. Env., 131, 274–280, 2009

Pan, G., Huang, Z. Q., Wang, J. K., Li, H., Chabbi, A., Paustian, K., and Smith, P.: Soil organic matter dynamics: beyond carbon; A report of the 4th International Symposium on Soil Organic Matter Dynamics, Carbon Manage., 4, 485–489, 2013.

Paul, E. A.: Soil microbiology, ecology and biochemistry, Academic Press, Burlington, MA, USA and Oxford, UK, 2014.

Paustian, K., Andrén, O., Janzen, H. H., Lal, R., Smith, P., Tian, G., and Woomer, P. L.: Agricultural soils as a sink to mitigate CO2

Peltoniemi, M., Thürig, E., Ogle, S., Palosuo, T., Shrumpf, M., Wützler, T., Butterbach-Bahl, K., Chertov, O., Komarov, A., Mikhailov, A., Gärdenäs, A., Perry, C., Liski, J., Smith, P., and Mäkipää, R.: Models in country scale carbon accounting of forest soils, Silva Fennica, 41, 575–602, 2007.

Rasse, D. P., Rumpel, C., and Dignac, M.-F.: Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation, Plant Soil, 269, 341–356, 2005

Reay, D., Smith, P., and van Amstel, A. (Eds.): Methane and Climate Change, Earthscan, London, 272 pp., 2010.

Rillig, M. C.: Arbuscolar mycorrhizae, glomalin and soil aggregation, Can. J. Soil Sci., 84, 355–363, 2004.

Rumpel, C. and Kögel-Knabner, I.: Deep soil organic matter – a key but poorly understood component of terrestrial C cycle, Plant Soil, 338, 143–158, 2011.

Sanaullah, M., Blagodatskaya, E., Chabbi, A., Rumpel, C., and Kuzyakov, Y.: Drought effects on microbial biomass and enzyme activities in the rhizosphere of grasses depend on plant community composition, Appl. Soil Ecol., 48, 38–44, 2011.

Sanaullah, M., Chabbi, A., Rumpel, C., and Kuzyakov, Y.: Carbon allocation in grassland communities under drought stress followed by 14C pulse labelling, Soil Biol. Biochem., 55, 132–139, 2012.

Sanaullah, M., Chabbi, A., Girardin, C., Durand, J. L., Poirier, M., and Rumpel, C.: Effects of drought and elevated temperature on biochemical composition of forage plants and their impact on carbon storage in grassland soil, Plant Soil, 374, 767–778, 2014.

Schlesinger, W. M. and Bernhardt, E. S.: Biogeochemstry. An analysis of global change, 3rd Edition, Academic Press, NY, USA, 2013.

Schmidt, M. W. I. and Noack, A. G.: Black carbon in soils and sediments: Analysis, distribution, implications, and current challenges, Glob. Biogeochem. Cycles, 14, 777–793, 2000.

Singh, N., Abiven, S., Torn, M. S., and Schmidt, M. W. I.: Fire-derived organic carbon in soil turns over on a centennial scale, Biogeosciences, 9, 2847–2857, doi:10.5194/bg-9-2847- 2012, 2012.

Smith, P.: Soils as carbon sinks – the global context, Soil Use Manage., 20, 212–218, 2004.

Smith, P.: Soils and climate change, Curr. Op. Env. Sust., 4, 539– 544, 2012.

Smith, P., House, J. I., Bustamante, M., Sobocká, J., Harper, R., Pan, G., West, P. C., Clark, J. M., Adhya, T., Rumpel, C., Paustian, K., Kuikman, P., Cotrufo, M. F., Elliott, J. A., McDowell, R., Griffiths, R. I., Asakawa, S., Bondeau, A., Jain, A. K, Meersmans, J., and Pugh, T. A. M.: Global change pressures on soils from land use and management, Glob. Change Biol., in press, doi:10.1111/gcb.13068, 2015.

Sollins, P., Homann, P., and Caldwell, B. A.: Stabilization and destabilization of soil organic matter: mechanisms and controls, Geoderma, 74, 65–105, 1996.

Song, X. Y., Li, L. Q., Zheng, J. F., Pan, G., Zhang, X. H., Zheng, J. W., Hussain, Q., Han, X. J., and Yu, X. Y.: Sequestration of maize crop straw C in different soils: role of oxyhydrates in chemical binding and stabilization as recalcitrance, Chemosphere, 87, 649–654, 2012

Stewart, C. E., Zheng, J. Y., Botte, J., and Cotrufo, M. F.: Cogenerated fast pyrolysis biochar mitigates greenhouse gas emissions and increases carbon sequestration in temperate soils, Glob. Change Biol. Bioenergy, 5, 153–164, 2013.

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region, Glob. Biogeochem. Cycles, 23, GB2023, doi:10.1029/2008GB003327, 2009

Thévenot, M., Dignac, M.-F., and Rumpel, C.: Fate of lignins in soils: a review, Soil Biol. Biochem., 42, 1200–1211, 2010.

Trumbore, S.: Radiocarbon and soil carbon dynamics, Ann. Rev. Earth Planet. Sci., 37, 47–66, 2009.

Turunen, J., Roulet, N. T., Moore, T. R., and Richard, P. J. H.: Nitrogen deposition and increased carbon accumulation in ombrotrophic peatlands in eastern Canada, Glob. Biogeochem. Cycles, 18, GB3002, doi:10.1029/2003GB002154, 2004.

Van Oost, K., Quine, T. A., Govers, G., De Gryze, S., Six, J., Harden, J. W., Ritchie, J. C., McCarty, G. W., Heckrath, G., Kosmas, C., Giraldez, J. V., da Silva, J. R. M., and Merckx, R.: The impact of agricultural soil erosion on the global carbon cycle, Science, 318, 626–629, 2007

von Lützow, M., Kögel-Knabner, I., Ludwig, B., Matzner, E., Flessa, H., Ekschmitt, K. Guggenberger, G., Marschner, B., and Kalbitz, K.: Stabilization mechanisms of organic matter in four temperate soils: Development and application of a conceptual model, J. Plant Nut. Soil Sci., 171, 111–124, 2008.

Whitmore, A. P., Kirk, G. J. D., and Rawlins, B. G.: Technologies for increasing carbon storage in soil to mitigate climate change, Soil Use Manage., 30, 10–22, 2014.

Wilkinson, M. T., Richards, P. J., and Humphreys, G. S.: Breaking ground: Pedological, geological and ecological implications of soil biotrubation, Earth Sci. Rev., 97, 257–272, 2009.

Yu, Z. C.: Northern peatland carbon stocks and dynamics: a review, Biogeosci. 9, 4071–4085, 2012.

Zheng, J. F., Li, L. Q., Pan, G., Zhang, X. H., Smith, P., and Hussain, Q.: Potential aerobic C mineralization of a red earth paddy soil and its temperature dependence under long-term fertilizer treatments, Soil Use Manage., 28, 185–193, 2012.

Zhou, P., Song, G. H., Pan, G., Li, L. Q., and Zhang, X. H.: Role of chemical protection by binding to oxyhydrates in SOC sequestration in three typical paddy soils under long-term agro-ecosystem experiments from South China, Geoderma, 153, 52–60, 2009.

Zvomuya, F., Janzen, H. H., Larney, F. J., and Olson, B. M.: A longterm field bioassay of soil quality indicators in a semiarid environment, Soil Sci. Soc. Am. J., 72, 683–692, 2008.

Excerpted from:

Smith, P., Cotrufo, M. F., Rumpel, C., Paustian, K., Kuikman, P. J., Elliott, J. A., McDowell, R., Griffiths, R. I., Asakawa, S., Bustamante, M., House, J. I., Sobocká, J., Harper, R., Pan, G., West, P. C., Gerber, J. S., Clark, J. M., Adhya, T., Scholes, R. J., and Scholes, M. C.: Biogeochemical cycles and biodiversity as key drivers of ecosystem services provided by soils, SOIL, 1, 665–685, https://doi.org/10.5194/soil-1-665-2015, 2015. https://soil.copernicus.org/articles/1/665/2015/

6.3.1: Soils and the carbon cycle is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

6.3.2: Soils and the Water Cycle

Soils provide important ecosystem services through their control on the water cycle. These services include provisioning services of food and water security, regulating services associated with moderation, and purification of water flows, and they contribute to the cultural services of landscapes/water bodies that meet recreation and aesthetic values (Table 6.3.2.1; Dymond, 2014). At the pedon to hill slope scale, water stored in soil is used for evapotranspiration and plant growth that supplies food, stabilizes the land surface to prevent erosion, and regulates nutrient and contaminant flow. At a catchment and basin scale, the capacity of the soil to infiltrate water attenuates stream and river flows and can prevent flooding, while water that percolates through soil can replenish groundwater that can maintain water supplies and sustain surface water ecosystems while promoting a continued flow during periods of reduced precipitation (Guswa et al., 2014).

Table 6.3.2.1: Soil functions related to the water cycle and ecosystem services

Soil function	Mechanism	Consequence	Ecosystem service
Stores (storage)	Water held in soil pores supports plant and microbial communities	Biomass production Surface protection	Food Aesthetics Erosion control
Accepts (sorptivity)	Incident water infiltrates into soil with excess lost as runoff	Storm runoff reduction	Erosion control Flood protection
Transmits (hydraulic conductivity)	Water entering the soil is redistributed and excess is lost as deep percolation	Percolation to ground- water	Groundwater recharge Stream flow maintenance
Cleans (filtering)	Water passing through the soil matrix interacts with soil particles and biota	Contaminants removed by biological degradation/retention on sorption sites	Water quality

The soil functions of accepting, storing, transmitting and cleaning of water shown in Table 6.3.2.1 are inter-related. Soil water storage depends on the rate of infiltration into the soil relative to the rate of precipitation. Soil hydraulic conductivity redistributes water within and through the soil profile. The infiltration rate and hydraulic conductivity both depend on the water stored in the soil. The initially high rate of infiltration into dry soil declines as the soil water content increases and water replaces air in the pore space. Conversely, hydraulic conductivity increases with soil moisture content as a greater proportion of the pores are transmitting water. Water content and transmission times are also important to the filtering function of soil because contact with soil surfaces and residence time in soil are important controls on contaminant supply and removal (McDowell and Srinivasan, 2009).

The quantity of water which a soil can store depends on the thickness of the soil layer, its porosity, and soil matrixwater physical interactions. The latter are expressed as a water retention curve, the relationship between the soil water content and the forces holding it in place. The porosity and water retention curve are in turn influenced primarily by the particle size distribution and the soil bulk density, but also by the amount of SOM and the macropores created by biotic activity (Kirkham, 2014).

Optimum growth of most plants occurs when roots can access both oxygen and water in the soil. The soil must therefore infiltrate water, drain quickly from saturation to allow air to reach plant roots, and retain and redistribute water for plant use. An ideal soil for plant production depends on the climatic conditions. Soil structural stability and porosity are also important for the infiltration of water into soil. In addition to soil texture, organic matter improves soil aggregate stability (Das et al., 2014). While plant growth and surface mulches can help protect the soil surface, a stable, well aggregated soil structure that resists surface sealing and continues to infiltrate water during intense rainfall events will decrease the potential for downstream flooding resulting from rapid overland flow. Porosity (especially macropores of a diameter $\geq 75\mu\mathrm{m}$) controls transmission of water through the soil. In addition to total porosity, the continuity and structure of the pore network are as important to these functions as they are in filtering out contaminants in flow. Furthermore, the soil must support biota that will degrade the compounds of interest or have sorption sites available to retain the chemical species. Soil organic matter is important for these roles and, together with mineral soil (especially the clay fraction), provides sorption sites (Bolan et al., 2011). Flow through macropores, which bypass the soil matrix, where biota and sorption sites are generally located, can quickly transmit water and contaminants through the soil to groundwater or artificial drains, but for filtering purposes, a more tortuous route through the soil matrix is more effective (McDowell et al., 2008). There are multiple other links between soil biota and soil water, with water potential in particular having a pivotal role in the structure, growth, and activity of the soil microbial community (Parr et al., 1981).

Table 6.3.2.2:

Management action or other driver of change	Provisioning service impact	Regulating service impact	Supporting service impact	Cultural service impact
Land use change (increase change of agricultural to urban)	Decreased biomass; decreased availability of water for agricultural use	Increased impervious surface; decreased infiltration, storage, soil-mediated water regulation	Decreased genetic diversity; reduction of rainfall recycling, e.g. in the tropics	Decreased natural environment
Land use change (increase change of arable to intensive grassland)	Increased yield of animal over vegetable protein	Increased C sequestration; greater requirement of water; stress on ecosystem health of downstream waterways	Increased genetic diversity associated with mixed pastures	Change from traditional values and aesthetic value
Irrigation (increase)	Increased biomass over dryland agriculture; decreased availability of water for urban use	Increased C sequestration, but decreased filtration potential	Improved habitat for plant species	Infrastructure alters landscape decreasing spiritual connection with catchment
Drainage (increasing in marginal land)	Decreased soil saturation; increased biomass; removal of wetlands	Decreased C sequestration, denitrification, and flood attenuation	Better habitat for productive grassland plants, but loss of genetic diversity	Decreased recreational potential (e.g. eco- tourism)

Management of soil alters the ecosystem services provided by water (Table 6.3.2.2). Soil conservation and sustainable management practices to combat desertification help to retain soil organic matter, structural stability, infiltration, and profile water holding capacity. The promotion of soil as a C sink to offset greenhouse gas emissions generally helps to maintain or improve soil hydrological functions as well. Deforestation, overgrazing, and excessive tillage of fragile lands, however, will lead to soil structural deterioration and a loss of infiltration, water retention, and surface water quality (Table 6.3.2.2; Steinfeld et al., 2006). Anthropogenic modifications to the water cycle can aid soil function. In dry regimes, inadequate soil moisture can be mitigated through supplementary irrigation, and where waterlogging occurs it can be relieved by land drainage. However, irrigation and drainage can have consequences for water regulation services. Irrigation that enables a shift to intensive land use can increase the contaminant load of runoff and drainage (Table 6.3.2.2; McDowell et al., 2011). Furthermore, drainage of wetland soils has been shown to reduce water and contaminant storage capacity in the landscape and can increase the potential for downstream flooding, as well as increasing the potential for GHG emissions due to the rapid decomposition of SOC in soil and dissolved organic C in drainage water (IPCC, 2013). The removal of surface or groundwater for irrigation disrupts the natural water cycle and may stress downstream ecosystems and communities. Irrigation of agricultural lands accounts for about 70% of ground and surface water withdrawals, and in some regions competition for water resources is forcing irrigators to tap unsustainable sources. Irrigation with wastewater may conserve fresh water resources, but the fate of waterborne contaminants in soil and crops is a potential concern (Sato et al., 2013).

References:

Bolan, N. S., Adriano, D. C., Kunhikrishnan, A., James, T., McDowell, R. W., and Senesi, N.: Dissolved organic matter: biogeochemistry, dynamics and environmental significance in soils, Adv. Agron., 110, 1–75, 2011.

Das, B., Chakraborty, D., Singh, V. K., Aggarwal, P., Singh, R., Dwivedi, B. S., and Mishra, R. P.: Effect of integrated nutrient management practice on soil aggregate properties, its stability and aggregate-associated carbon content in an intensive ricewheat system, Soil Till. Res., 136, 9–18, 2014

Dymond, J.: Ecosystem services in New Zealand, Manaaki Whenua Press, Lincoln, New Zealand, 540 p., 2014.

Guswa, A. J., Brauman, K. A., Brown, C., Hamel, P., Keeler, B. L., and Sayre, S. S.: Ecosystem services: Challenges and opportunities for hydrologic modelling to support decision making, Wat. Resour. Res., 50, 4535–4544, 2014

IPCC: Supplement to the 2006 Guidelines for National Greenhouse Gas Inventories: Wetlands, Cambridge University Press, Cambridge, UK, 2013.

Kirkham, M. B.: Principles of soil and plant water relations, Academic Press, San Diego, CA, 2014.

McDowell, R. W. and Srinivasan, M. S.: Identifying critical source areas for water quality: 2. Validating the approach for phosphorus and sediment losses in grazed headwater catchments, J. Hydrol., 379, 68–80, 2009.

McDowell, R. W., Houlbrooke, D. J., Muirhead, R. W., Müller, K., Shepherd, M., and Cuttle, S. P.: Grazed Pastures and Surface Water Quality, Nova Science Publishers, New York, NY, 2008.

McDowell, R. W., van der Weerden, T. J., and Campbell, J.: Nutrient losses associated with irrigation, intensification and management of land use: a study of large scale irrigation in North Otago, New Zealand, Agric. Water Manage., 98, 877–885, 2011.

Parr, J. F., Gardner, W. R., and Elliot, L. F.: Water potential relations in soil microbiology: proceedings of a symposium, SSSA Special Publication number 9, Soil Science Society of America, Madison, WI, 151 pp., 1981.

Sato, T., Qadir, M., Yamamoto, S., Endo, T., and Zahoor, A.: Global, regional, and country level need for data on wastewater generation, treatment, and use, Agric. Water Manage., 130, 1–13, 2013.

Steinfeld, H., Gerber, P., Wassenaar, C. V., Rosales, M., and de Haan, C.: Livestock's long shadow, Food and Agriculture Organization of the United Nations, Rome, Italy, available at: http://www.fao.org/docrep/010/a0701e/a0701e00.htm (last access: August 2014), 2006.

Excepted from:

Smith, P., Cotrufo, M. F., Rumpel, C., Paustian, K., Kuikman, P. J., Elliott, J. A., McDowell, R., Griffiths, R. I., Asakawa, S., Bustamante, M., House, J. I., Sobocká, J., Harper, R., Pan, G., West, P. C., Gerber, J. S., Clark, J. M., Adhya, T., Scholes, R. J., and Scholes, M. C.: Biogeochemical cycles and biodiversity as key drivers of ecosystem services provided by soils, SOIL, 1, 665–685, https://doi.org/10.5194/soil-1-665-2015, 2015. https://soil.copernicus.org/articles/1/665/2015/

6.3.2: Soils and the Water Cycle is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

6.4: Calculating landscape mass balance



Maps take some time to build because we have to find or write matching materials. LibreTexts POV is that it is best to make available pages that we have finished rather than wait till the entire project is complete .This map is not completely finished, some pages are missing but we are workin' on it. . . If you know of some open materials which could be used here let us know info@libretexts.org (Public Domain ; Public Domain Pictures)

Till then here are two videos from about land modeling in Earth System Models from the Community Earth System Model symposia. ESMs are the most sophisticated climate system models, integrating the behavior of the surface, the oceans and the atmopshere.



6.4: Calculating landscape mass balance is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.5: Ecosystem models and remote sensing

The study of the changes in ecosystem structure caused by changes in the environment (disturbances) or by internal forces is called ecosystem dynamics. Ecosystems are characterized using a variety of research methodologies. Some ecologists study ecosystems using controlled experimental systems, while some study entire ecosystems in their natural state, and others use both approaches.

A holistic ecosystem model attempts to quantify the composition, interaction, and dynamics of entire ecosystems; it is the most representative of the ecosystem in its natural state. A food web is an example of a holistic ecosystem model. However, this type of study is limited by time and expense, as well as the fact that it is neither feasible nor ethical to do experiments on large natural ecosystems. To quantify all different species in an ecosystem and the dynamics in their habitat is difficult, especially when studying large habitats such as the Amazon Rainforest, which covers 1.4 billion acres (5.5 million km^2) of the Earth's surface.

For these reasons, scientists study ecosystems under more controlled conditions. Experimental systems usually involve either partitioning a part of a natural ecosystem that can be used for experiments, termed a mesocosm, or by re-creating an ecosystem entirely in an indoor or outdoor laboratory environment, which is referred to as a microcosm. A major limitation to these approaches is that removing individual organisms from their natural ecosystem or altering a natural ecosystem through partitioning may change the dynamics of the ecosystem. These changes are often due to differences in species numbers and diversity and also to environment alterations caused by partitioning (mesocosm) or re-creating (microcosm) the natural habitat. Thus, these types of experiments are not totally predictive of changes that would occur in the ecosystem from which they were gathered.

As both of these approaches have their limitations, some ecologists suggest that results from these experimental systems should be used only in conjunction with holistic ecosystem studies to obtain the most representative data about ecosystem structure, function, and dynamics.

Scientists use the data generated by these experimental studies to develop ecosystem models that demonstrate the structure and dynamics of ecosystems. Three basic types of ecosystem modeling are routinely used in research and ecosystem management: a conceptual model, an analytical model, and a simulation model. A conceptual model is an ecosystem model that consists of flow charts to show interactions of different compartments of the living and nonliving components of the ecosystem. A conceptual model describes ecosystem structure and dynamics and shows how environmental disturbances affect the ecosystem; however, its ability to predict the effects of these disturbances is limited. Analytical and simulation models, in contrast, are mathematical methods of describing ecosystems that are indeed capable of predicting the effects of potential environmental changes without direct experimentation, although with some limitations as to accuracy. An analytical model is an ecosystem model that is created using simple mathematical formulas to predict the effects of environmental disturbances on ecosystem structure and dynamics. A simulation model is an ecosystem model that is created using complex computer algorithms to holistically model ecosystems and to predict the effects of environmental disturbances on ecosystem structure and dynamics. Ideally, these models are accurate enough to determine which components of the ecosystem are particularly sensitive to disturbances, and they can serve as a guide to ecosystem managers (such as conservation ecologists or fisheries biologists) in the practical maintenance of ecosystem health.

Conceptual Models

Conceptual models are useful for describing ecosystem structure and dynamics and for demonstrating the relationships between different organisms in a community and their environment. Conceptual models are usually depicted graphically as flow charts. The organisms and their resources are grouped into specific compartments with arrows showing the relationship and transfer of energy or nutrients between them. Thus, these diagrams are sometimes called compartment models.

To model the cycling of mineral nutrients, organic and inorganic nutrients are subdivided into those that are bioavailable (ready to be incorporated into biological macromolecules) and those that are not. For example, in a terrestrial ecosystem near a deposit of coal, carbon will be available to the plants of this ecosystem as carbon dioxide gas in a short-term period, not from the carbon-rich coal itself. However, over a longer period, microorganisms capable of digesting coal will incorporate its carbon or release it as natural gas (methane, CH_4), changing this unavailable organic source into an available one. This conversion is greatly accelerated by the combustion of fossil fuels by humans, which releases large amounts of carbon dioxide into the atmosphere. This is thought to be a major factor in the rise of the atmospheric carbon dioxide levels in the industrial age. The carbon dioxide released from burning fossil fuels is produced faster than photosynthetic organisms can use it. This process is intensified by the reduction of

photosynthetic trees because of worldwide deforestation. Most scientists agree that high atmospheric carbon dioxide is a major cause of global climate change.

Conceptual models are also used to show the flow of energy through particular ecosystems. Figure 6.5.1 is based on Howard T. Odum's classical study of the Silver Springs, Florida, holistic ecosystem in the mid-twentieth century.[2] This study shows the energy content and transfer between various ecosystem compartments.

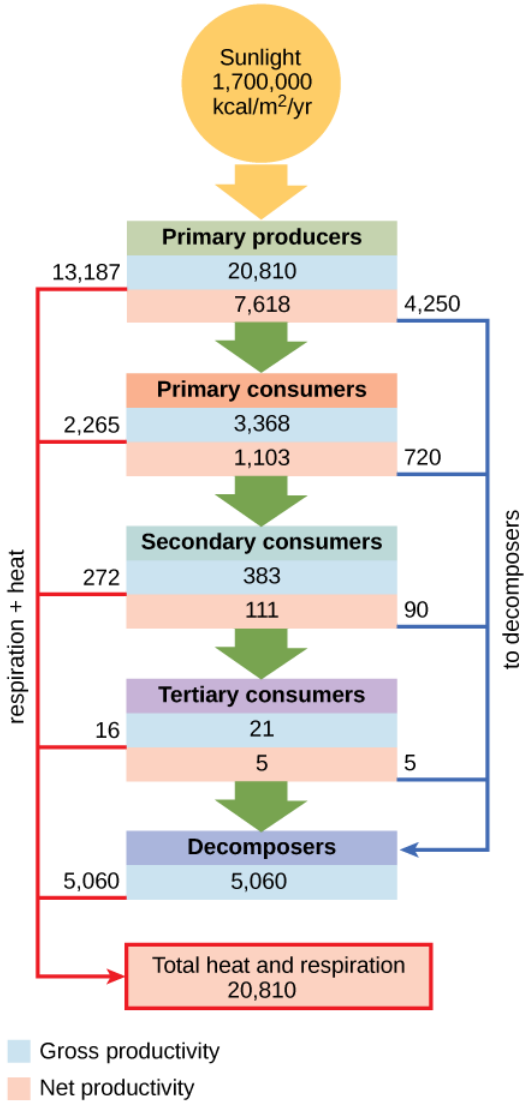


Figure 6.5.1: This conceptual model shows the flow of energy through a spring ecosystem in Silver Springs, Florida. Notice that the energy decreases with each increase in trophic level.

Exercise 6.5.1

Why do you think the value for gross productivity of the primary producers is the same as the value for total heat and respiration (20,810 kcal/m^2/yr)?

Answer

According to the first law of thermodynamics, energy can neither be created nor destroyed. Eventually, all energy consumed by living systems is lost as heat or used for respiration, and the total energy output of the system must equal the energy that went into it.

Analytical and Simulation Models

The major limitation of conceptual models is their inability to predict the consequences of changes in ecosystem species and/or environment. Ecosystems are dynamic entities and subject to a variety of abiotic and biotic disturbances caused by natural forces and/or human activity. Ecosystems altered from their initial equilibrium state can often recover from such disturbances and return to a state of equilibrium. As most ecosystems are subject to periodic disturbances and are often in a state of change, they are usually either moving toward or away from their equilibrium state. There are many of these equilibrium states among the various components of an ecosystem, which affects the ecosystem overall. Furthermore, as humans have the ability to greatly and rapidly alter the species content and habitat of an ecosystem, the need for predictive models that enable understanding of how ecosystems respond to these changes becomes more crucial.

Analytical models often use simple, linear components of ecosystems, such as food chains, and are known to be complex mathematically; therefore, they require a significant amount of mathematical knowledge and expertise. Although analytical models have great potential, their simplification of complex ecosystems is thought to limit their accuracy. Simulation models that use computer programs are better able to deal with the complexities of ecosystem structure.

A recent development in simulation modeling uses supercomputers to create and run individual-based simulations, which accounts for the behavior of individual organisms and their effects on the ecosystem as a whole. These simulations are considered to be the most accurate and predictive of the complex responses of ecosystems to disturbances.

Note

Visit Climatic & Ecological Modelling for Adaptive Forest Applications by Tongli Wang to learn more about ecological modeling

Visit The Darwin Project to view a variety of ecosystem models.

Modeling Ecosystems Energy Flow: Ecological Pyramids

The structure of ecosystems can be visualized with ecological pyramids, which were first described by the pioneering studies of Charles Elton in the 1920s. Ecological pyramids show the relative amounts of various parameters (such as number of organisms, energy, and biomass) across trophic levels.

Pyramids of numbers can be either upright or inverted, depending on the ecosystem. As shown in Figure 6.5.2, typical grassland during the summer has a base of many plants and the numbers of organisms decrease at each trophic level. However, during the summer in a temperate forest, the base of the pyramid consists of few trees compared with the number of primary consumers, mostly insects. Because trees are large, they have great photosynthetic capability, and dominate other plants in this ecosystem to obtain sunlight. Even in smaller numbers, primary producers in forests are still capable of supporting other trophic levels.

Another way to visualize ecosystem structure is with pyramids of biomass. This pyramid measures the amount of energy converted into living tissue at the different trophic levels. Using the Silver Springs ecosystem example, this data exhibits an upright biomass pyramid (Figure 6.5.2), whereas the pyramid from the English Channel example is inverted. The plants (primary producers) of the Silver Springs ecosystem make up a large percentage of the biomass found there. However, the phytoplankton in the English Channel example make up less biomass than the primary consumers, the zooplankton. As with inverted pyramids of numbers, this inverted pyramid is not due to a lack of productivity from the primary producers, but results from the high turnover rate of the phytoplankton. The phytoplankton are consumed rapidly by the primary consumers, thus, minimizing their biomass at any particular point in time. However, phytoplankton reproduce quickly, thus they are able to support the rest of the ecosystem.

Pyramid ecosystem modeling can also be used to show energy flow through the trophic levels. Notice that these numbers are the same as those used in the energy flow compartment diagram in Figure 6.5.2. Pyramids of energy are always upright, and an ecosystem without sufficient primary productivity cannot be supported. All types of ecological pyramids are useful for characterizing ecosystem structure. However, in the study of energy flow through the ecosystem, pyramids of energy are the most consistent and representative models of ecosystem structure (Figure 6.5.2).

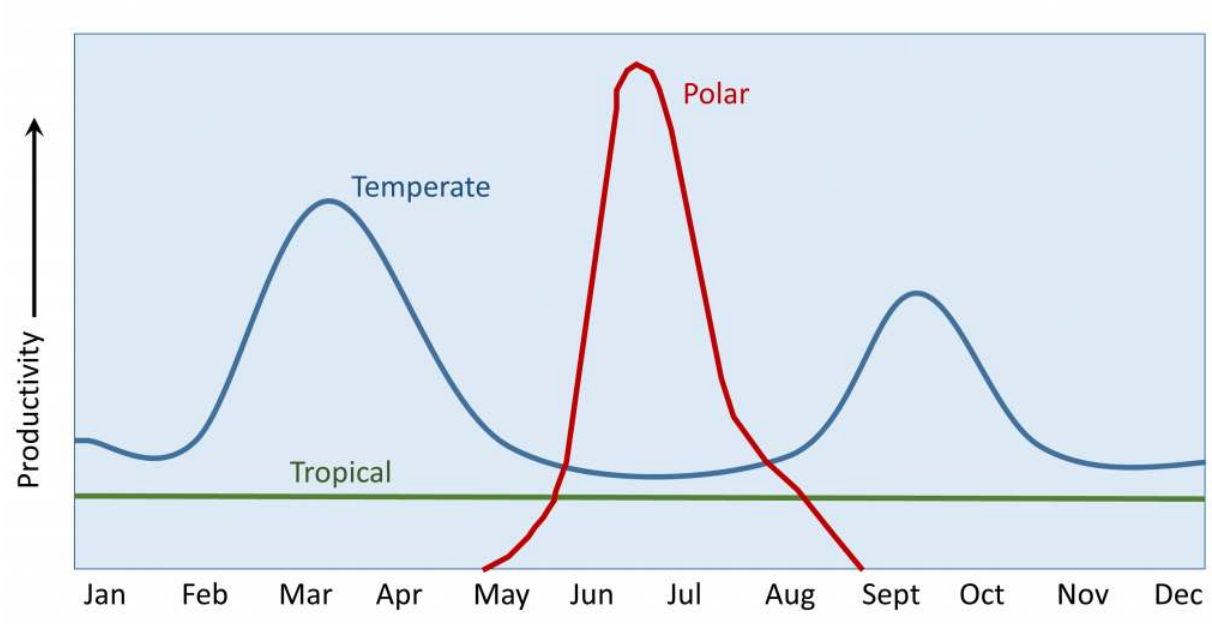


Figure 6.5.2: Ecological pyramids depict the (a) biomass, (b) number of organisms, and (c) energy in each trophic level.

? Exercise 6.5.1

Pyramids depicting the number of organisms or biomass may be inverted, upright, or even diamond-shaped. Energy pyramids, however, are always upright. Why?

Answer

Pyramids of organisms may be inverted or diamond-shaped because a large organism, such as a tree, can sustain many smaller organisms. Likewise, a low biomass of organisms can sustain a larger biomass at the next trophic level because the organisms reproduce rapidly and thus supply continuous nourishment. Energy pyramids, however, must always be upright because of the laws of thermodynamics. The first law of thermodynamics states that energy can neither be created nor destroyed; thus, each trophic level must acquire energy from the trophic level below. The second law of thermodynamics states that, during the transfer of energy, some energy is always lost as heat; thus, less energy is available at each higher trophic level.

Summary

Ecosystems exist on land, at sea, in the air, and underground. Different ways of modeling ecosystems are necessary to understand how environmental disturbances will affect ecosystem structure and dynamics. Conceptual models are useful to show the general relationships between organisms and the flow of materials or energy between them. Analytical models are used to describe linear food chains, and simulation models work best with holistic food webs.

Footnotes

1. 1 *Nature* (Vol. 458, April 1, 2009)
2. 2 Howard T. Odum, “Trophic Structure and Productivity of Silver Springs, Florida,” *Ecological Monographs* 27, no. 1 (1957): 47–112.

6.5: Ecosystem models and remote sensing is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.6: Ecosystem Monitoring

A complex and adaptive ecosystem in which all the chemical, physical, and biological components, functions, and processes are intact and functioning normally is considered a healthy ecosystem (but see Cumming and Cumming (2015), for a discussion on this value-based term). In contrast, disturbing any of an ecosystem's components, functions, and/or processes will, by definition, alter them to some degree (Table 10.1). In many cases, ecosystems that have been exposed to certain forms and levels of disturbances remain healthy because there is redundancy in the roles performed by different ecosystem components (Section 4.2.1). This ability of an ecosystem to withstand certain forms and levels of disturbances is referred to as ecosystem stability. Ecosystem stability could be the result of one or both of two qualities: resistance and resilience. Resistance is the ability of an ecosystem to retain the same characteristic communities and natural cycles throughout and after a disturbance event, while resilience is the ability of an ecosystem to rapidly recover or adapt after a disturbance event. For example, if the number of native aquatic insect species decline after non-native fishes are introduced to previously fish-free ponds, the pond's ecosystem has low resistance. But if the native insect community recovers rapidly after the non-native fishes were removed, the ecosystem is resilient.

Table 10.1 Three ways how humans have changed the natural world.

Natural function	Changes attributed to human activities
Land surface	As much as half of the world's ice-free land surface has been transformed to cater to people's need for natural resources. Much of these changes are driven by agricultural activities.
Nitrogen cycle	Human activities release massive amounts of nitrogen into natural ecosystems on a daily basis. Much of this occurs through the use of nitrogen fertilisers, burning fossil fuels, and cultivating nitrogen-fixing crops.
Atmospheric carbon cycle	Scientists estimate that humans would have doubled levels of carbon dioxide in the Earth's atmosphere by the middle of this century. This is primarily the result of fossil fuel use and deforestation.
Wildlife populations	Between 1970 and 2014, Sub-Saharan Africa have lost three quarters of its freshwater vertebrates; the rate of these declines shows no sign of reducing.
Pollutants	Pollution from human activities have become so omnipresent that it is hard to escape its impacts. Microplastics have been found in drinking water and the food we eat (Chapter 7).

Sources: MEA, 2005; Kulkarni et al., 2008; http://www.livingplanetindex.org

Many forms of ecosystem disturbances are easy to observe. Consequently, monitoring these visible forms of disturbances—such as the outright destruction of a forest or plastic pollution on a beach—focuses less on detection and more on developing systematic survey protocols (Section 9.1) that can provide information on whether a disturbance is spreading and increasing in intensity, or whether conservation action is successful in containing the threat. However, some disturbances are subtler, unobtrusive, and thus difficult to detect; examples include pesticide drift and agricultural runoff (Section 7.1). Adopting a "wait and see" approach to detecting these invisible forms of disturbances can be particularly damaging, since that approach generally ends at a point where the harm will either be impossible to reverse or will require significantly more resources and time than would have been the case if the problem was addressed earlier. In this way, there are many similarities between monitoring ecosystem health and human health —some ailments are easier to diagnose than others, but we avoid the worst-case scenarios by screening regularly for diseases and treating the threatening ones promptly.



Figure 10.2 Biologists sampling aquatic macroinvertebrates in the Okavango Delta, Botswana, as part of a biomonitoring project of the Freshwater Research Centre. Photograph by Helen Dallas/FRCSA, CC BY 4.0.

Perhaps the most popular method conservation biologists use to monitor ecosystem health is known as biomonitoring. By monitoring the abundance and/or fitness of sensitive species (Box 10.1), biologists can sometimes detect ecosystem degradation before it becomes apparent to the human eye or escalates to a point where it starts impacting human lives (Bornman and Bouwman, 2012). Monitoring environmental indicators such as macroinvertebrates (Figure 10.2) is particularly popular when examining the ecological condition of aquatic ecosystems; mayflies, caddisflies, and stoneflies—specialists of undisturbed streams—are often replaced by flies and midges in polluted and disturbed environments. Sometimes however, when plants or animals are not easily monitored, certain aspects of those species can still be monitored. One example is monitoring total plant biomass as a proxy for soil nutrients or intensity of herbivory. Another option is to perform a bioassay, during which a sensitive organism (typically water fleas or plankton) is released into a potentially contaminated environment to see if death or declining health occurs.

By monitoring the abundance and/or fitness of sensitive species, biologists can detect threats to biodiversity before it becomes apparent to the human eye.

Box 10.1 Using Insects to Monitor Environmental Health

Rosina Kyerematen

Department of Animal Biology and Conservation Science,

University of Ghana,

Legon, Ghana.

rkyerematen@ug.edu.gh

Insects are important to nearly every terrestrial food web in the world and serve a multitude of different purposes: some insects are responsible for pollination of plants while others are scavengers that clean up dead plant and animal material. In some cases, our understanding of an insect species' ecological role can make it suitable as an indicator of environmental health. Biomonitoring looks at the presence and abundance of organisms within their natural communities to assess the impact of environmental disturbances; this knowledge can then be used to guide ecosystem management. An indicator taxon is one whose impact can be specifically and precisely measured; its abundance serves as a measure of the overall health of an ecosystem. Understanding how the presence and abundance of indicator species, and the relative abundance of tolerant and intolerant species, reflects the relative health of an environment can allow for rapid surveys of impaired ecosystems to assess trends as well as to track changes following remediation and restoration efforts.

Over the past few years, biomonitoring with insects as indicator taxa has become increasingly popular in Ghana. Butterflies are especially popular because they show varying relative sensitivities to environmental change; the abundance of certain butterfly species can for example be used to study the impact of habitat loss, fragmentation, and climate change (Kyerematen et al.,

2018). The presence and abundance of butterflies more characteristic of open and disturbed ecosystems (Figure 10.A) can, for example, be used as an indicator of forest degradation.



Figure 10.A Because the green-banded swallowtail (*Papilio nireus*) prefers open woodlands, researchers in Ghana are using their abundances in forests as a measure of habitat degradation. Photograph by Celesta von Chamier, CC BY 4.0.

Aquatic insects, particularly benthic macroinvertebrates, are also useful bioindicators. Freshwater resources, such as lakes and rivers, provide water for drinking and washing to local people, and a home for economically important taxa, such as fish and shellfish. Protecting these water sources is therefore important for safeguarding people's health and livelihoods. The presence, absence, and diversity of certain benthic macroinvertebrates, even at the order level, can provide valuable information about whether a waterbody is being degraded or not (Kyerematen et al., 2014; Nnoli et al., 2019). A recent study showed that dragonfly and damselfly diversities and populations along the coastal Densu River in Ghana vary widely depending on the physical condition of the river and surrounding area (Acquah-Lamptey et al., 2013).

With their high diversity and varying tolerances for ecosystem conditions, insects are extraordinarily suited as ecological indicators in environmental monitoring. Each insect species is also part of a wider biological community with important ecological roles. If lost, not only will an abundance of other life be affected, such a loss may also hint at a looming crisis facing people living in those compromised ecosystems.

At times, conservation biologists may need to measure the physical environment to assess environmental health. This approach is particularly common when tracking pollution, for example by monitoring for changes in biochemical indicators. For example, measuring total phosphorus, nitrogen, and dissolved oxygen load in streams and other surface water can help scientists track eutrophication (Section 7.1.1). Measuring these and other biochemical indicators is usually accomplished directly via chemical analysis of environmental samples, such as soil and water. Sometimes however, biochemicals indicators are tracked indirectly via biological samples obtained from plants and animals. Because they bio-accumulate heavy metals and other pollutants, filter feeders such as clams and mussels (e.g. Bodin et al., 2013) are particularly useful in this regard as they can be used to detect very low concentrations of harmful chemicals in the environment.

Monitoring ecosystems with geospatial analysis

A persistent challenge facing biologists who monitor ecosystems—and other aspects of biodiversity—is achieving consistency across space and time. Consider a survey of a sensitive bird community to track ecosystem change; not only will different observers have varying levels of experience, but they will almost certainly see different birds during an early morning census compared to one later in the afternoon due to differences in biology and behavior between species. These factors introduce error into monitoring data, which in turn can mask the effect a biologist tries to measure (Buckland and Johnston, 2017). Laboratory scientists' control for these confounding factors by making multiple measurements under strictly controlled conditions. But for conservation biologists working outside in the wind and rain, repeated observations under similar conditions can be near impossible.

Geospatial analysis offers a variety of tools that allow biologists to overcome some of these traditional field monitoring challenges. These tools use geographic information systems (GIS) computer software packages to store, display, and manipulate a wide variety

of data representing the natural environment, biodiversity, and human land-use patterns as they relate to one another on Earth's surface. GIS thereby allows biologists to easily visualise and analyse spatial relationships between mapped data, which may include aspects such as vegetation types, climate, soils, topography, geology, water availability, species distributions, existing protected areas, human settlements, and human resource use (Figure 10.3). Understanding such relationships helps conservation biologists to prioritise their actions, for example by identifying areas where data are lacking, where an environmental change requires further investigation, or where gaps in regional protected areas network exist.

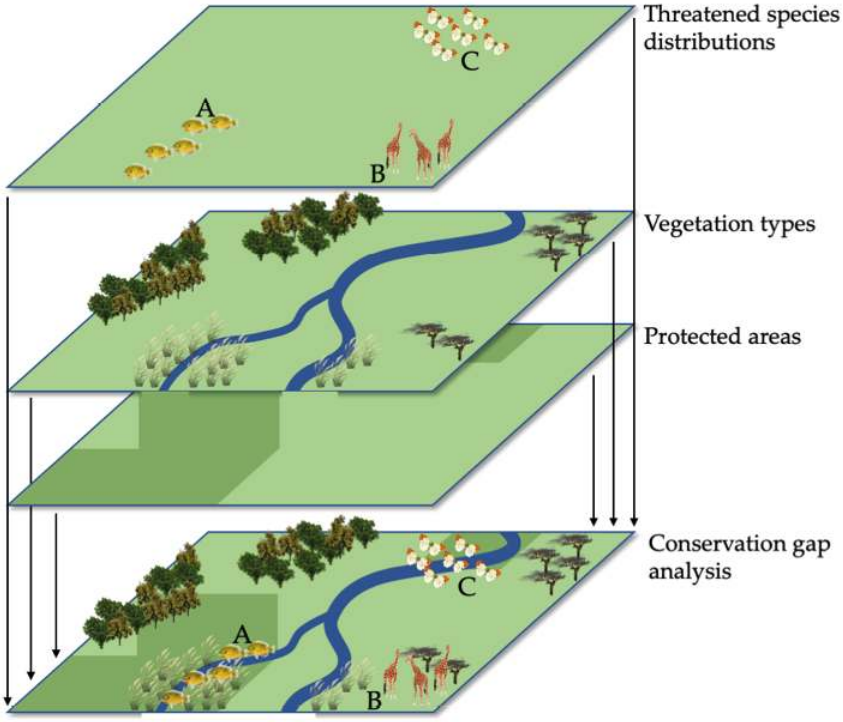


Figure 10.3 Distribution maps for three threatened species are overlaid onto maps of protected areas and vegetation types to identify potential conservation gaps (discussed in Chapter 13). This GIS analysis shows that Species A is well protected, Species B is protected to some extent, and Species C is not protected at all. Species C would thus be considered a high conservation priority. After Scott et al., 1991, CC BY 4.0.

Remote sensing is a special branch of geospatial analysis directed at obtaining ecosystem data without making physical contact (i.e. boots on the ground) with the observation site. Before the turn of the 20th century, the most popular form of remote sensing was aerial photography from airplanes. These aerial photographs facilitated geographers' ability to draw maps of landscape features, including human infrastructure and natural vegetation patterns. Remote sensing opportunities greatly expanded from 1960 onward, with the launch of the National Aeronautical Space Administration's (NASA) first Earth observation satellites, to take photographs of Earth from space for weather forecasting. Subsequent satellite programs expanded their scope to also collect additional data of Earth's surface and atmosphere. While much of this data would have been useful to conservation, early satellite data products were very expensive and thus largely out of reach of the larger conservation community. This all changed in 2008 when NASA started distributing their Earth observation products for free to the public, heralding an era in which remote sensing became a standard tool in the conservation field.

Remote sensing offers a variety of tools that allow biologists to monitor biodiversity beyond the abilities of traditional field monitoring techniques.

Today, hundreds of Earth observation satellites circle the planet, offering near real-time access to unbiased and consistent environmental datasets of nearly all terrestrial surfaces, oceanic surfaces and floor depths, and the atmosphere, all from the comfort of a computer connected to the internet (Wilson et al., 2013). Scientists use these products in ecosystem monitoring efforts, including monitoring water quality (Dube et al. 2015), forest loss (Laporte et al., 2007), coral reef health (McClanahan et al., 2011), desertification (Symeonakis et al., 2004), and fire regimes (Archbald et al., 2010). Linking the information obtained by Earth observation satellites to biological information collected on the ground has proved invaluable in monitoring species' threat statuses

(Di Marco et al., 2014), ecosystem connectivity (Wegmann et al., 2014), and habitat suitability (Torres et al., 2010), as well as understanding how biodiversity responds to environmental changes (Box 10.2).

Box 10.2 Remote Sensing and Spatial Analysis for African Conservation

Barend F. N. Erasmus

Global Change Institute (GCI), University of the Witwatersrand,

Johannesburg, South Africa.

Barend.Erasmus@wits.ac.za

Remote sensing is the art and science of observing objects or landscapes from a distance, without being in direct contact with the environment. Although you can think of wildlife photography as a type of remote sensing, the term usually refers to aerial photography (Figure 10.B), or images taken via satellite. For both cases, there is a trade-off between the area covered by each photo (the footprint), how much digital storage space is available, and how often a satellite takes a picture of the same area. For satellites that take pictures of the earth systematically, along a pre-defined path, cloud cover determines how often you can get a usable image. Recent improvements in technology, now allow for the deployment of constellations of satellites that have a collective point-and-track capability to observe an area almost continuosly.

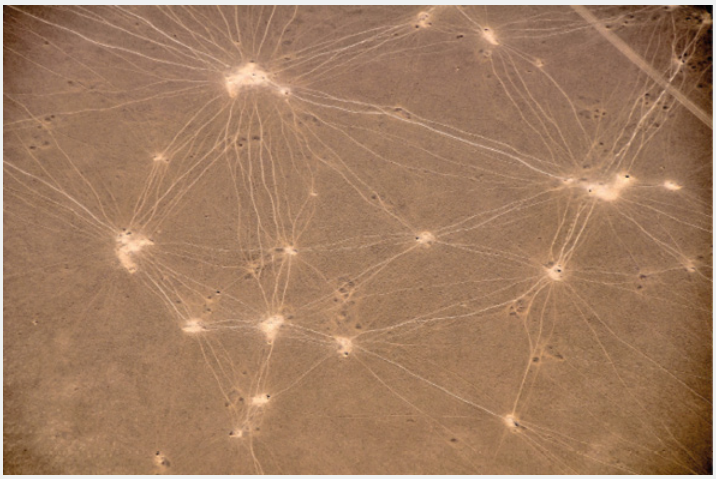


Figure 10.B Aerial photograph of Kaa pan in the southwestern Kalahari region of Botswana. Gemsbok (Oryx gazella, LC), red hartebeest (Alcelaphus buselaphus caama, LC), and eland (Tragelaphus oryx, LC) excavate holes for lekking (eating soil with high salt content), and the footpaths between the different lekking sites clearly show preferences for certain sites. In the same way, on a much broader scales, ungulates move across landscapes to access other resources, such as water and grass, whilst avoiding human hunters and predators. Photograph by B.F.N. Erasmus, CC BY 4.0.

Our eyes are sensitive to the colours red, green, and blue, together called the visible part of the electromagnetic spectrum. This is actually a very small part of the entire spectrum, and scientists have found, for example, that vegetation seen as uniformly green shows a lot of variation in the infrared part of the spectrum. For this reason, many satellites carry cameras that can “see” infrared light, and by proxy, measure vegetation health, biomass, and sometimes even structure. This capability, together with frequent revisit rates, allows for a unique view of how African landscapes change, whether through habitat loss, seasonal changes, or drought. Analyses of 13 years of remotely sensed data show that the most arid parts of southwest Botswana now experience typical summer vegetation conditions later in the year (Dubovyk et al., 2015), so herbivores must cope with a much longer dry season now than they experienced in 2000.

African landscapes are always changing, and sometimes in unpredictable ways—localised thunderstorms at the beginning of the rainy season can rapidly change a dry dustbowl landscape into green grazing. These “wet footprints” can cover areas as small as 1 km × 3 km; in contrast, a large frontal weather system can cover dozens or even hundreds of kilometres. An ungulate looking for green grazing not too far from drinking water, while trying to avoid predators and hunters, needs the ability to detect such green patches, and the strength and knowledge to move there. Obtaining this knowledge from an ungulate’s perspective of the landscape is no small task. Hopcraft et al. (2014) show how common wildebeest and plains zebra

> in the Serengeti have different migration strategies: wildebeest move to green grass as quickly as they can, with little effort to avoid predators, whereas zebra gauge predation risk and forage quality concurrent to their trek.

Our ability to understand ungulate movements and other ecosystem patterns has been greatly enhanced over the last few years, as animal-tracking technology (where GPS positions are logged and stored using radio-frequency tracking, mobile phone networks, satellite systems, or any combination) and the resulting analyses became more sophisticated. It is now possible to distinguish between locations where animals often spend a little time (for example, a preferred shady tree for resting during the heat of the day), or the same location where they infrequently spend a lot of time (for example, once during an oestrus cycle, a lactating lioness will spend a lot of time at her den, with cubs). It is also possible to "see" when such an animal changes "mode" of movement. For example, researchers have found that springbok (Antidorcas marsupialis, LC) rams around the Etosha pan in Namibia show sedentary behavior (small movements around a specific area) when grazing is good, which transitions to searching behavior during dry seasons (long, relatively fast movements in a straight line) as they move to areas that had water or grass during previous drought periods (Lyons et al., 2013).

If we combine this animal tracking capability with regular remotely sensed images of vegetation quality, then we can start to answer questions about why certain animals move to certain areas. It also highlights the fact that it is very seldom good enough to put a fence around an area and call it preserved. Due to the changing nature of Africa's savannahs, animals need to be able to move to areas with water or forage, often outside reserves, when they become available. In the face of a changing climate and changing rainfall patterns, this ability to move long distance to reach vital resources remains one of the best adaptation responses that African ungulates may have to cope with climate change. However, fences, roads, and farmland may block these migrations across the landscape, putting those populations at risk of extinction (Section 5.1.1).

African conservation areas need to make provision for these animal movements, or risk conservation areas without sustainable animal populations. This presents a problem—how do we investigate options for large animals to move across a transformed rural landscape and minimise human-wildlife conflict while, on the other hand, still providing access to green grass or drinking water at a specific protected area? Both resources change in location and time of year, and only regular, detailed remote sensing of vegetation, combined with detailed animal movement studies, will provide the necessary picture in time and space.

The popularity and utility of these products have preceded and facilitated the expansion of other remote sensing applications in ecosystem monitoring. Among the most popular are radar products (Figure 10.4), which have become the standard method for obtaining elevation and other terrain data (NASA, 2009, 2013), as well as estimates of carbon stocks (Carreiras et al., 2012). Advances have also been made in using hyper-spectral imagery to monitor soil properties (Mashimbye et al., 2012) and even individual trees (Naidoo et al., 2012). LiDAR has enabled biologists to map three-dimensional vegetation, which can be used to explain animal movements (Loarie et al., 2013; Davies et al., 2016) and measure carbon stocks and forest loss (Burton et al., 2017).

Conservation biologists are often faced with the shifting baseline syndrome, where the reference points they use to measure their progress may be vastly different from earlier states.

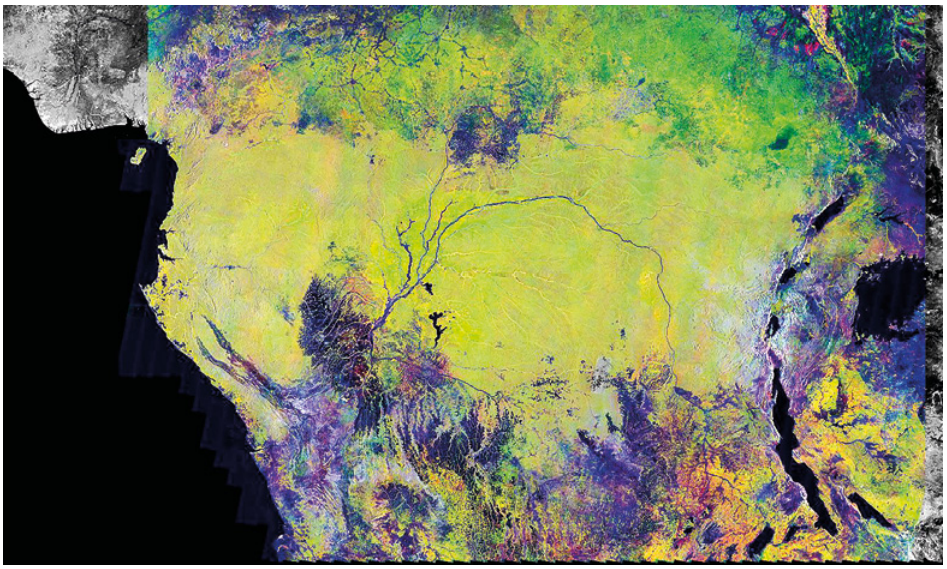


Figure 10.4 A radar composite of equatorial Africa obtained in 1996 by Japan's Earth Resources Satellite. The area shown covers about 7.4 million km2. Due to the filters applied to the image, yellow represents flooded forests, palm plantations, and urban areas; green represents forest, and black represents surface water. Savannahs may be black, blue, purple, or green, depending on the ecoregion type. Displays such as these allow researchers to map ecosystems, and ecosystem loss at a much larger scale than if only field observations were used. Image courtesy of NASA, https://images.nasa.gov/details-PIA01348.html, CC0.

Despite the opportunities presented by remote sensing, it is critical to remember that it is not a substitute for traditional field monitoring methods. Most importantly, remote sensing applications cannot be considered reliable without verification using field data (see Burton et al., 2017). Most remote sensing products are also relatively new, which does not allow for enough opportunities to compare across time. In the absence of historical remotely sensed data, geospatial analysts usually choose the best reference site currently available; this may expose those analysts to shifting baseline syndrome, because the chosen reference site may be vastly different from earlier states the scientists are actually interested in studying (Bunce et al., 2008; Papworth et al., 2009). Remote sensing is, therefore, not a cure-all for ecosystem monitoring challenges; it is simply a powerful tool to supplement traditional field-based monitoring.

This page titled 6.6: Ecosystem Monitoring is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by John W. Wilson & Richard B. Primack (Open Book Publishers) via source content that was edited to the style and standards of the LibreTexts platform.

- **10.1: Ecosystem Monitoring** by John W. Wilson & Richard B. Primack is licensed CC BY 4.0. Original source: https://doi.org/10.11647/OBP.0177.

SECTION OVERVIEW

6.7: Human impacts on terrestrial biogeochemistry

Learning Objectives

- Human activities have increased atmospheric carbon dioxide by about 40% over pre-industrial levels and more than doubled the amount of nitrogen available to ecosystems. Similar trends have been observed for phosphorus and other elements, and these changes have major consequences for biogeochemical cycles and climate change.
- In total, land in the United States absorbs and stores an amount of carbon equivalent to about 17% of annual U.S. fossil fuel emissions. U.S. forests and associated wood products account for most of this land sink. The effect of this carbon storage is to partially offset warming from emissions of CO2 and other greenhouse gases.
- Altered biogeochemical cycles together with climate change increase the vulnerability of biodiversity, food security, human health, and water quality to changing climate. However, natural and managed shifts in major biogeochemical cycles can help limit rates of climate change.

6.7.1: Introduction

6.7.2: Human-Induced Changes

6.7.3: Sinks and Cycles

6.7.4: Impacts and Options

6.7.5: References

Source:

Galloway, J. N., W. H. Schlesinger, C. M. Clark, N. B. Grimm, R. B. Jackson, B. E. Law, P. E. Thornton, A. R. Townsend, and R. Martin, 2014: Ch. 15: Biogeochemical Cycles. Climate Change Impacts in the United States: The Third National Climate Assessment, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 350-368. doi:10.7930/J0X63JT0. : http://nca2014.globalchange.gov/report/sectors/biogeochemical-cycles

6.7: Human impacts on terrestrial biogeochemistry is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

6.7.1: Introduction

Biogeochemical cycles involve the fluxes of chemical elements among different parts of the Earth: from living to non-living, from atmosphere to land to sea, and from soils to plants. They are called “cycles” because matter is always conserved and because elements move to and from major pools via a variety of two-way fluxes, although some elements are stored in locations or in forms that are differentially accessible to living things. Human activities have mobilized Earth elements and accelerated their cycles – for example, more than doubling the amount of reactive nitrogen that has been added to the biosphere since pre-industrial times.[1,2] Reactive nitrogen is any nitrogen compound that is biologically, chemically, or radiatively active, like nitrous oxide and ammonia, but not nitrogen gas (N_2). Global-scale alterations of biogeochemical cycles are occurring, from human activities both in the U.S. and elsewhere, with impacts and implications now and into the future. Global carbon dioxide emissions are the most significant driver of human-caused climate change. But human-accelerated cycles of other elements, especially nitrogen, phosphorus, and sulfur, also influence climate. These elements can affect climate directly or act as indirect factors that alter the carbon cycle, amplifying or reducing the impacts of climate change.

Climate change is having, and will continue to have, impacts on biogeochemical cycles, which will alter future impacts on climate and affect our capacity to cope with coupled changes in climate, biogeochemistry, and other factors.

Source:

Galloway, J. N., W. H. Schlesinger, C. M. Clark, N. B. Grimm, R. B. Jackson, B. E. Law, P. E. Thornton, A. R. Townsend, and R. Martin, 2014: Ch. 15: Biogeochemical Cycles. Climate Change Impacts in the United States: The Third National Climate Assessment, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 350-368. doi:10.7930/J0X63JT0. : http://nca2014.globalchange.gov/report/sectors/biogeochemical-cycles

6.7.1: Introduction is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

6.7.2: Human-Induced Changes

Human activities have increased atmospheric carbon dioxide by about 40% over pre-industrial levels and more than doubled the amount of nitrogen available to ecosystems. Similar trends have been observed for phosphorus and other elements, and these changes have major consequences for biogeochemical cycles and climate change.

The human mobilization of carbon, nitrogen, and phosphorus from the Earth's crust and atmosphere into the environment has increased 36, 9, and 13 times, respectively, compared to geological sources over pre-industrial times.[3] Fossil fuel burning, land-cover change, cement production, and the extraction and production of fertilizer to support agriculture are major causes of these increases.[4] Carbon dioxide (CO_2) is the most abundant of the heat-trapping greenhouse gases that are increasing due to human activities, and its production dominates atmospheric forcing of global climate change.5 However, methane (CH_4) and nitrous oxide (N_2O) have higher greenhouse-warming potential per molecule than CO_2, and both are also increasing in the atmosphere. In the U.S. and Europe, sulfur emissions have declined over the past three decades, especially since the mid-1990s, because of efforts to reduce air pollution.[6] Changes in biogeochemical cycles of carbon, nitrogen, phosphorus, and other elements – and the coupling of those cycles – can influence climate. In turn, this can change atmospheric composition in other ways that affect how the planet absorbs and reflects sunlight (for example, by creating small particles known as aerosols that can reflect sunlight).

State of the Carbon Cycle

The U.S. was the world's largest producer of human-caused CO_2 emissions from 1950 until 2007, when it was surpassed by China. U.S. emissions account for approximately 85% of North American emissions of CO_2 [7] and 18% of global emissions. [8,9] Ecosystems represent potential "sinks" for CO_2, which are places where carbon can be stored over the short or long term (see "Estimating the U.S. Carbon Sink"). At the continental scale, there has been a large and relatively consistent increase in forest carbon stocks over the last two decades,[10] due to recovery from past forest harvest, net increases in forest area, improved forest management regimes, and faster growth driven by climate or fertilization by CO_2 and nitrogen.[7,11] The largest rates of disturbance and "regrowth sinks" are in southeastern, south central, and Pacific northwestern regions.[11] However, emissions of CO_2 from human activities in the U.S. continue to increase and exceed ecosystem CO_2 uptake by more than three times. As a result, North America remains a net source of CO_2 into the atmosphere[7] by a substantial margin.

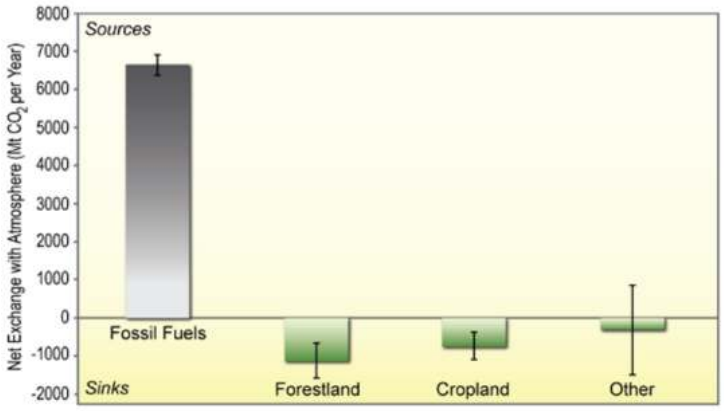


Figure 6.7.2.1:Major North American Carbon Dioxide Sources and Sinks. The release of carbon dioxide from fossil fuel burning in North America (shown here for 2010) vastly exceeds the amount that is taken up and temporarily stored in forests, crops, and other ecosystems (shown here is the annual average for 2000-2006). (Figure source: King et al. 20127). (Copyright; author via source)

Sources and Fates of Reactive Nitrogen

The nitrogen cycle has been dramatically altered by human activity, especially by the use of nitrogen fertilizers, which have increased agricultural production over the past half century.[1,2] Although fertilizer nitrogen inputs have begun to level off in the U.S. since 1980,[12] human-caused reactive nitrogen inputs are now at least five times greater than those from natural sources. [13,14,15,16] At least some of the added nitrogen is converted to nitrous oxide (N_2O), which adds to the greenhouse effect in Earth's atmosphere.

An important characteristic of reactive nitrogen is its legacy. Once created, it can, in sequence, travel throughout the environment (for example, from land to rivers to coasts, sometimes via the atmosphere), contributing to environmental problems such as the formation of coastal low-oxygen “dead zones” in marine ecosystems in summer. These problems persist until the reactive nitrogen is either captured and stored in a long-term pool, like the mineral layers of soil or deep ocean sediments, or converted back to nitrogen gas.[17,18] The nitrogen cycle affects atmospheric concentrations of the three most important human-caused greenhouse gases: carbon dioxide, methane, and nitrous oxide. Increased available nitrogen stimulates the uptake of carbon dioxide by plants, the release of methane from wetland soils, and the production of nitrous oxide by soil microbes.

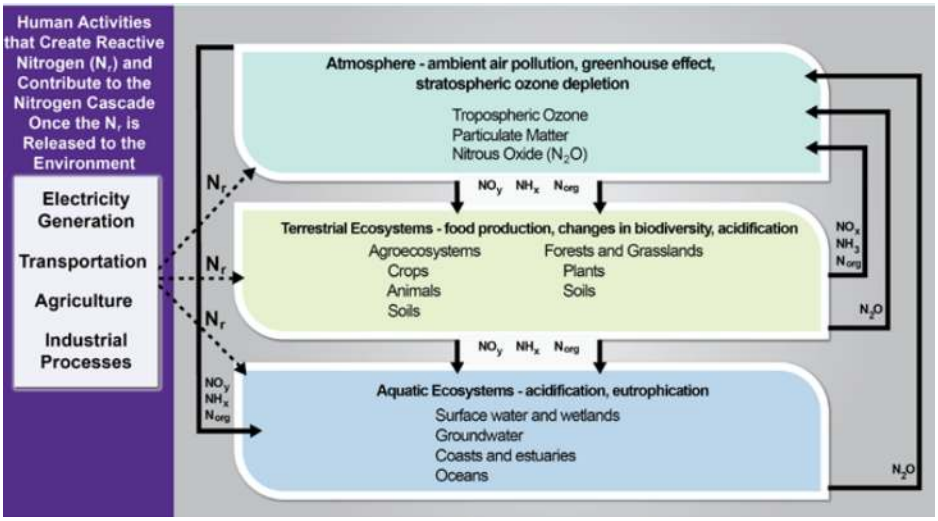


Figure 6.7.2.2: **Human Activities that Form Reactive Nitrogen and Resulting Consequences in Environmental Reservoirs**. Once created, a molecule of reactive nitrogen has a cascading impact on people and ecosystems as it contributes to a number of environmental issues. Molecular terms represent oxidized forms of nitrogen primarily from fossil fuel combustion (such as nitrogen oxides, NOx), reduced forms of nitrogen primarily from agriculture (such as ammonia, (NH_3), and organic forms of nitrogen (Norg) from various processes. NOy is all nitrogen-containing atmospheric gases that have both nitrogen and oxygen, other than nitrous oxide (N_2O). NHx is the sum of ammonia (NH_3) and ammonium (NH_4). (Figure source: adapted from EPA 2011;[13] Galloway et al. 2003;[17] with input from USDA. USDA contributors were Adam Chambers and Margaret Walsh). (Copyright; author via source)

Phosphorus and other elements The phosphorus cycle has been greatly transformed in the United States, [19] primarily from the use of phosphorus fertilizers in agriculture. Phosphorus has no direct effects on climate, but does have indirect effects, such as increasing carbon sinks by fertilizing plants. Emissions of sulfur, as sulfur dioxide, can reduce the growth of plants and stimulate the leaching of soil nutrients needed by plants.[20]

Source:

Galloway, J. N., W. H. Schlesinger, C. M. Clark, N. B. Grimm, R. B. Jackson, B. E. Law, P. E. Thornton, A. R. Townsend, and R. Martin, 2014: Ch. 15: Biogeochemical Cycles. Climate Change Impacts in the United States: The Third National Climate Assessment, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 350-368. doi:10.7930/J0X63JT0. : http://nca2014.globalchange.gov/report/sectors/biogeochemical-cycles

6.7.2: Human-Induced Changes is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

6.7.3: Sinks and Cycles

In total, land in the United States absorbs and stores an amount of carbon equivalent to about 17% of annual U.S. fossil fuel emissions. U.S. forests and associated wood products account for most of this land sink. The effect of this carbon storage is to partially offset warming from emissions of CO_2 and other greenhouse gases.

Considering the entire atmospheric CO_2 budget, the temporary net storage on land is small compared to the sources: more CO_2 is emitted than can be taken up (see “Estimating the U.S. Carbon Sink”).[7,21,22,23] Other elements and compounds affect that balance by direct and indirect means (for example, nitrogen stimulates carbon uptake [direct] and nitrogen decreases the soil methane sink [indirect]). The net effect on Earth’s energy balance from changes in major biogeochemical cycles (carbon, nitrogen, sulfur, and phosphorus) depends upon processes that directly affect how the planet absorbs or reflects sunlight, as well as those that indirectly affect concentrations of greenhouse gases in the atmosphere.

Carbon

In addition to the CO_2 effects described above, other carbon-containing compounds affect climate change, such as methane and volatile organic compounds (VOCs). As the most abundant non-CO_2 greenhouse gas, methane is 20 to 30 times more potent than CO_2 over a century timescale. It accounted for 9% of all human-caused greenhouse gas emissions in the United States in 2011,[8] and its atmospheric concentration today is more than twice that of pre-industrial times.[24,25] Methane has an atmospheric lifetime of about 10 years before it is oxidized to CO_2, but it has about 25 times the global warming potential of CO_2. An increase in methane concentration in the industrial era has contributed to warming in many ways.26 Methane also has direct and indirect effects on climate because of its influences on atmospheric chemistry. Increases in atmospheric methane and VOCs are expected to deplete concentrations of hydroxyl radicals, causing methane to persist in the atmosphere and exert its warming effect for longer periods.[25,27] The hydroxyl radical is the most important “cleaning agent” of the troposphere (the active weather layer extending up to about 5 to 10 miles above the ground), where it is formed by a complex series of reactions involving ozone and ultraviolet light.[3]

Nitrogen and Phosphorus

The climate effects of an altered nitrogen cycle are substantial and complex.[4,28,29,30,31] Carbon dioxide, methane, and nitrous oxide contribute most of the human-caused increase in climate forcing, and the nitrogen cycle affects atmospheric concentrations of all three gases. Nitrogen cycling processes regulate ozone (O3) concentrations in the troposphere and stratosphere, and produce atmospheric aerosols, all of which have additional direct effects on climate. Excess reactive nitrogen also has multiple indirect effects that simultaneously amplify and mitigate changes in climate. Changes in ozone and organic aerosols are short-lived, whereas changes in carbon dioxide and nitrous oxide have persistent impacts on the atmosphere.



Figure 6.7.3.1: Nitrogen Emissions. Figure shows how climate change will affect U.S. reactive nitrogen emissions, in Teragrams (Tg) CO_2 equivalent, on a 20-year (top) and 100-year (bottom) global temperature potential basis. Positive values on the vertical axis depict warming; negative values reflect cooling. The height of the bar denotes the range of uncertainty, and the white line denotes the best estimate. The relative contribution of combustion (dark brown) and agriculture (green) is denoted by the color shading. (Figure source: adapted from Pinder et al. 20122[8]). (Public Domain. US Global Change Research Program.)

The strongest direct effect of an altered nitrogen cycle is through emissions of nitrous oxide (N_2O), a long-lived and potent greenhouse gas that is increasing steadily in the atmosphere.[25,26] Globally, agriculture has accounted for most of the atmospheric rise in N_2O.[32,33] Roughly 60% of agricultural N_2O derives from elevated soil emissions resulting from the use of nitrogen fertilizer. Animal waste treatment accounts for about 30%, and the remaining 10% comes from crop-residue burning.[34] The U.S. reflects this global trend: around 75% to 80% of U.S. human-caused N_2O emissions are due to agricultural activities, with the majority being emissions from fertilized soil. The remaining 20% is derived from a variety of industrial and energy sectors.[35,36] While N_2O currently accounts for about 6% of human-caused warming,26 its long lifetime in the atmosphere and rising concentrations will increase N_2O-based climate forcing over a 100-year time scale.[33,37,38]

Excess reactive nitrogen indirectly exacerbates changes in climate by several mechanisms. Emissions of nitrogen oxides (NOx) increase the production of tropospheric ozone, which is a greenhouse gas.[39] Elevated tropospheric ozone may reduce CO_2 uptake by plants and thereby reduce the terrestrial CO_2 sink.[40] Nitrogen deposition to ecosystems can also stimulate the release of nitrous oxide and methane and decrease methane uptake by soil microbes.[41]

However, excess reactive nitrogen also mitigates changes in greenhouse gas concentrations and climate through several intersecting pathways. Over short time scales, NOx and ammonia emissions lead to the formation of atmospheric aerosols, which cool the climate by scattering or absorbing incoming radiation and by affecting cloud cover.[26,42] In addition, the presence of NOx in the lower atmosphere increases the formation of sulfate and organic aerosols.43 At longer time scales, NOx can increase rates of methane oxidation, thereby reducing the lifetime of this important greenhouse gas.

One of the dominant effects of reactive nitrogen on climate stems from how it interacts with ecosystem carbon capture and storage, and thus, the carbon sink. As mentioned previously, addition of reactive nitrogen to natural ecosystems can increase carbon storage as long as other factors are not limiting plant growth, such as water and nutrient availability.[44] Nitrogen deposition from human sources is estimated to contribute to a global net carbon sink in land ecosystems of 917 to 1,830 million metric tons (1,010 to 2,020

million tons) of CO_2 per year. These are model-based estimates, as comprehensive, observationally-based estimates at large spatial scales are hindered by the limited number of field experiments. This net land sink represents two components: 1) an increase in vegetation growth as nitrogen limitation is alleviated by human-caused nitrogen deposition, and 2) a contribution from the influence of increased reactive nitrogen availability on decomposition. While the former generally increases with increased reactive nitrogen, the net effect on decomposition in soils is not clear. The net effect on total ecosystem carbon storage was an average of 37 metric tons (41 tons) of carbon stored per metric ton of nitrogen added in forests in the U.S. and Europe.[45]

When all direct and indirect links between reactive nitrogen and climate in the U.S. are added up, a recent estimate suggests a modest reduction in the rate of warming in the near term (next several decades), but a progressive switch to greater net warming over a 100-year timescale.[28,29] That switch is due to a reduction in nitrogen oxide (NOx) emissions, which provide modest cooling effects, a reduction in the nitrogen-stimulated CO_2 storage in forests, and a rising importance of agricultural nitrous oxide emissions. Current policies tend to reinforce this switch. For example, policies that reduce nitrogen oxide and sulfur oxide emissions have large public health benefits, but also reduce the indirect climate mitigation co-benefits by reducing carbon storage and aerosol formation.

Changes in the phosphorus cycle have no direct effects on climate, but phosphorus availability constrains plant and microbial activity in a wide variety of land- and water-based ecosystems.[46,47] Changes in phosphorus availability due to human activity can therefore have indirect impacts on climate and the emissions of greenhouse gases in a variety of ways. For example, in land-based ecosystems, phosphorus availability can limit both CO_2 storage and decomposition46,48 as well as the rate of nitrogen accumulation.[49] In turn, higher nitrogen inputs can alter phosphorus cycling via changes in the production and activity of enzymes that release phosphorus from decaying organic matter, [50] creating another mechanism by which rising nitrogen inputs can stimulate carbon uptake.

Other Effects: Sulfate Aerosols

In addition to the aerosol effects from nitrogen mentioned above, there are both direct and indirect effects on climate from other aerosol sources. Components of the sulfur cycle exert a cooling effect through the formation of sulfate aerosols created from the oxidation of sulfur dioxide (SO_2) emissions.[26] In the United States, the dominant source of sulfur dioxide is coal combustion. Sulfur dioxide emissions rose until 1980, but have since decreased by more than 50% following a series of air-quality regulations and incentives focused on improving human health and the environment, as well as reductions in the delivered price of low-sulfur coal.[51] That decrease in emissions has had a marked effect on U.S. climate forcing: between 1970 and 1990, sulfate aerosols caused cooling, primarily over the eastern U.S., but since 1990, further reductions in sulfur dioxide emissions have reduced the cooling effect of sulfate aerosols by half or more.[42] Continued declines in sulfate aerosol cooling are projected for the future, [42] particularly if coal continues to be replaced by natural gas (which contains far fewer sulfur impurities) for electricity generation. Here, as with nitrogen oxide emissions, the environmental and socioeconomic tradeoffs are important to recognize: lower sulfur dioxide and nitrogen oxide emissions remove some climate cooling agents, but improve ecosystem health and save lives.[16,31,52]

Three low-concentration industrial gases are particularly potent for trapping heat: nitrogen trifluoride (NF_3), sulfur hexafluoride (SF_6), and trifluoromethyl sulfur pentafluoride (\(SF_{5}CF_{3}\)). None currently makes a major contribution to climate forcing, but since their emissions are increasing and their effects last for millennia, continued monitoring is important.). None currently makes a major contribution to climate forcing, but since their emissions are increasing and their effects last for millennia, continued monitoring is important.

Source:

Galloway, J. N., W. H. Schlesinger, C. M. Clark, N. B. Grimm, R. B. Jackson, B. E. Law, P. E. Thornton, A. R. Townsend, and R. Martin, 2014: Ch. 15: Biogeochemical Cycles. Climate Change Impacts in the United States: The Third National Climate Assessment, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 350-368. doi:10.7930/J0X63JT0. : http://nca2014.globalchange.gov/report/sectors/biogeochemical-cycles

6.7.3: Sinks and Cycles is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

6.7.4: Impacts and Options

Altered biogeochemical cycles together with climate change increase the vulnerability of biodiversity, food security, human health, and water quality to changing climate. However, natural and managed shifts in major biogeochemical cycles can help limit rates of climate change.

Climate change alters key aspects of biogeochemical cycling, creating the potential for feedbacks that alter both warming and cooling processes into the future. For example, as soils warm, the rate of decomposition will increase, adding more CO2 to the atmosphere. In addition, both climate and biogeochemistry interact strongly with environmental and ecological concerns, such as biodiversity loss, freshwater and marine eutrophication (unintended fertilization of aquatic ecosystems that leads to water quality problems), air pollution, human health, food security, and water resources. Many of the latter connections are addressed in other sections of this assessment, but we summarize some of them here because consideration of mitigation and adaptation options for changes in climate and biogeochemistry often requires this broader context.

Climate-Biogeochemistry Feedbacks

Both rising temperatures and changes in water availability can alter climate-relevant biogeochemical processes. For example, as summarized above, nitrogen deposition drives temperate forest carbon storage, both by increasing plant growth and by slowing organic-matter decomposition.[53] Higher temperatures will counteract soil carbon storage by increasing decomposition rates and subsequent emission of CO2 via microbial respiration. However, that same increase in decomposition accelerates the release of reactive nitrogen (and phosphorus) from organic matter, which in turn can fuel additional plant growth.[44] Temperature also has direct effects on net primary productivity (the total amount of CO2 stored by a plant through photosynthesis minus the amount released through respiration). The combined effects on ecosystem carbon storage will depend on the extent to which nutrients constrain both net primary productivity and decomposition, on the extent of warming, and on whether any simultaneous changes in water availability occur.[54] Similarly, natural methane sources are sensitive to variations in climate; ice core records show a strong correlation between methane concentrations and warmer, wetter conditions.[55] Thawing permafrost in polar regions is of particular concern because it stores large amounts of methane that could potentially be released to the atmosphere.

Biogeochemistry, Climate, and Interactions with Other Factors

Societal options for addressing links between climate and biogeochemical cycles must often be informed by connections to a broader context of global environmental changes. For example, both climate change and nitrogen deposition can reduce biodiversity in water- and land-based ecosystems. The greatest combined risks are expected to occur where critical loads are exceeded.[56,57] A critical load is defined as the input rate of a pollutant below which no detrimental ecological effects occur over the long-term according to present knowledge.[57] Although biodiversity is often shown to decline when nitrogen deposition is high due to fossil fuel combustion and agricultural emissions,[57,58] the compounding effects of multiple stressors are difficult to predict. Warming and changes in water availability have been shown to interact with nitrogen in additive or synergistic ways to exacerbate biodiversity loss.[59] Unfortunately, very few multi-factorial studies have been done to address this gap.

Human induced acceleration of the nitrogen and phosphorus cycles already causes widespread freshwater and marine eutrophication,[60,61] a problem that is expected to worsen under a warming climate.[61,62] Without efforts to reduce future climate change and to slow the acceleration of biogeochemical cycles, existing climate changes will combine with increasing inputs of nitrogen and phosphorus into freshwater and estuarine ecosystems. This combination of changes is projected to have substantial negative effects on water quality, human health, inland and coastal fisheries, and greenhouse gas emissions.[18,61]

Similar concerns – and opportunities for the simultaneous reduction of multiple environmental problems (known as “co-benefits”) – exist in the realms of air pollution, human health, and food security. For example, methane, volatile organic compounds, and nitrogen oxide emissions all contribute to the formation of tropospheric ozone, which is a greenhouse gas and has negative consequences for human health and crop and forest productivity.[37,63,64] Rates of ozone formation are accelerated by higher temperatures, creating a reinforcing cycle between rising temperatures and continued human alteration of the nitrogen and carbon cycles.[65] Rising temperatures also work against some of the benefits of air pollution control.[64] Some changes will trade gains in one arena for declines in others. For example, lowered NOx, NHx, and SOx emissions remove cooling agents from the atmosphere, but improve air quality.[16,31] Recent analyses suggest that targeting reductions in compounds like methane and black

carbon aerosols that have both climate and air-pollution consequences can achieve significant improvements in not only the rate of climate change, but also in human health.[31] Finally, reductions in excess nitrogen and phosphorus from agricultural and industrial activities can potentially reduce the rate and impacts of climate change, while simultaneously addressing concerns in biodiversity, water quality, food security, and human health.[66]

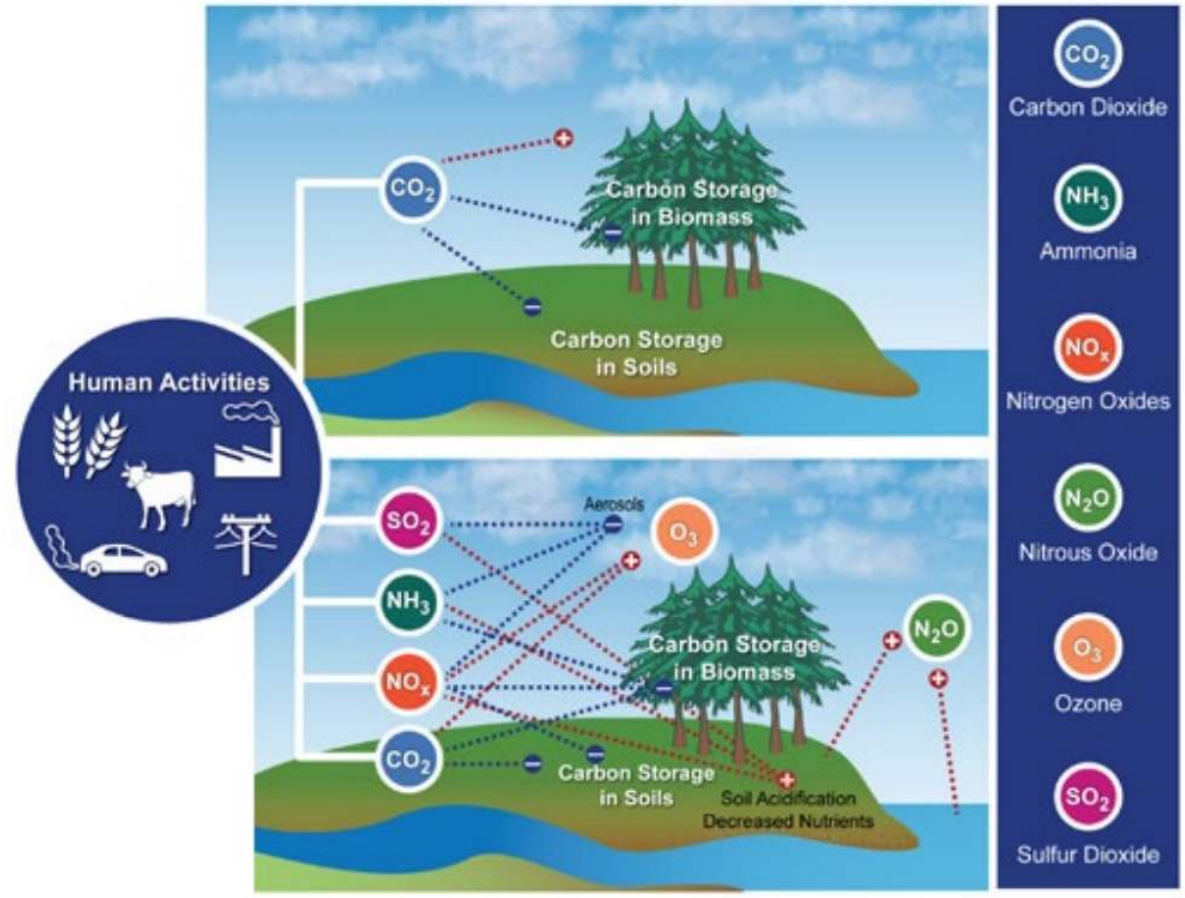


Figure 6.7.4.1: **Many Factors Combine to Affect Biogeochemical Cycles**. . Top panel shows the impact of the alteration of the carbon cycle alone on radiative forcing. The bottom panel shows the impacts of the alteration of carbon, nitrogen, and sulfur cycles on radiative forcing. SO2 and NH3 increase aerosols and decrease radiative forcing. NH3 is likely to increase plant biomass, and consequently decrease forcing. NOx is likely to increase the formation of tropospheric ozone (O3) and increase radiative forcing. Ozone has a negative effect on plant growth/biomass, which might increase radiative forcing. CO2 and NH3 act synergistically to increase plant growth, and therefore decrease radiative forcing. SO2 is likely to reduce plant growth, perhaps through the leaching of soil nutrients, and consequently increase radiative forcing. NOx is likely to reduce plant growth directly and through the leaching of soil nutrients, therefore increasing radiative forcing. However, it could act as a fertilizer that would have the opposite effect. (Public Domain. Galloway, et al. US Global Change Research Program)

Source:

Galloway, J. N., W. H. Schlesinger, C. M. Clark, N. B. Grimm, R. B. Jackson, B. E. Law, P. E. Thornton, A. R. Townsend, and R. Martin, 2014: Ch. 15: Biogeochemical Cycles. Climate Change Impacts in the United States: The Third National Climate Assessment, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 350-368. doi:10.7930/J0X63JT0. http://nca2014.globalchange.gov/repo...hemical-cycles

6.7.4: Impacts and Options is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

6.7.5: References

Source:

Galloway, J. N., W. H. Schlesinger, C. M. Clark, N. B. Grimm, R. B. Jackson, B. E. Law, P. E. Thornton, A. R. Townsend, and R. Martin, 2014: Ch. 15: Biogeochemical Cycles. Climate Change Impacts in the United States: The Third National Climate Assessment, J. M. Melillo, Terese (T.C.) Richmond, and G. W. Yohe, Eds., U.S. Global Change Research Program, 350-368. doi:10.7930/J0X63JT0. http://nca2014.globalchange.gov/repo...hemical-cycles

6.7.5: References is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

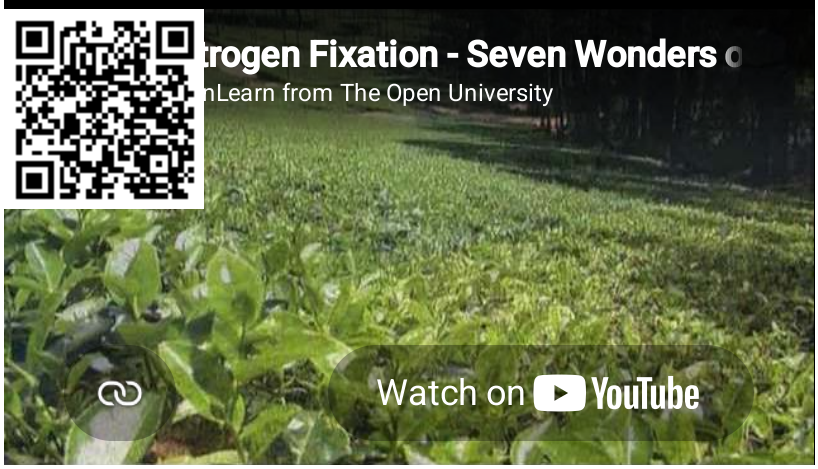
6.8: Summary



Maps take some time to build because we have to find or write matching materials. LibreTexts POV is that it is best to make available pages that we have finished rather than wait till the entire project is complete .This map is not completely finished, some pages are missing but we are workin' on it. . . If you know of some open materials which could be used here let us know info@libretexts.org (Public Domain ; Public Domain Pictures)







6.8: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

7: Wetland Ecosystems

7.1: Introduction
7.2: Types of wetlands
7.3: Productivity in wetland ecosystems
7.4: Organic matter storage in wetlands
7.5: Microbial metabolism in saturated sediments
7.6: Anaerobic and aerobic metabolic pathways
7.7: The redox ladder
7.8: Wetlands and water quality
7.9: Wetlands and global change
7.9.1: Introduction
7.9.2: Carbon Accumulation and GHG Emissions from Freshwater Wetlands (Including Permafrost) In a Changing Climate
7.9.3: Ecological Consequences for Freshwater Wetlands in a Changing Climate
7.9.4: Salt Marsh and Mangrove Response to a Changing Climate and Associated Sea Level Rise
7.10: Summary

7: Wetland Ecosystems is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.1: Introduction



Wetlands are lands that are wet for significant periods of time. They are common where water and land meet. Wetlands can be large flat areas or relatively small and steep areas.Wetlands are rich and unique ecosystems with many species that rely on both the land and the water for survival. Only specialized plants are able to grow in these conditions. Wetlands tend have a great deal of biological diversity. Wetland ecosystems can also be fragile systems that are sensitive to the amounts and quality of water present within them.

Types of Wetlands

There are a variety of different types of wetlands. **Marshes** are shallow wetlands around lakes, streams, or the ocean where grasses and reeds are common, but trees are not. Frogs, turtles, muskrats, and many varieties of birds are at home in marshes.

A **swamp** is a wetland with lush trees and vines found in a low-lying area beside slow-moving rivers. Like marshes, they are frequently or always inundated with water. Since the water in a swamp moves slowly, oxygen in the water is often scarce. Swamp plants and animals must be adapted for these low-oxygen conditions. Like marshes, swamps can be fresh water, salt water, or a mixture of both.

In an **estuary**, salt water from the sea mixes with fresh water from a stream or river. These semi-enclosed areas are home to plants and animals that can tolerate the sharp changes in salt content that the constant motion and mixing of waters creates. Estuaries contain brackish water, water that has more salt than fresh water but less than seawater. Because of the rapid changes in salt content, estuaries have many different habitats for plants and animals and extremely high biodiversity.



Marshes



Swamps

Ecological Role of Wetlands

 As mentioned above, wetlands are home to many different species of organisms. Although they make up only 5 percent of the area of the United States, wetlands contain more than 30 percent of the plant types. Many endangered species live in wetlands, so wetlands are protected from human use. Wetlands also offer protection from storm surges from tropical storm systems like hurricanes and provide safe, hidden protection for hatchlings.

Wetlands also play a key biological role by removing pollutants from water. For example, they can trap and use fertilizer that has washed off a farmer's field, and therefore they prevent that fertilizer from contaminating another body of water. Since wetlands naturally purify water, preserving wetlands also helps to maintain clean supplies of water.

CC licensed content, Shared previously

- Dynamic Earth: Introduction to Physical Geography. **Authored by**: R. Adam Dastrup. **Located at**: http://www.opengeography.org/physical-geography.html. **Project**: Open Geography Education. **License**: *CC BY-SA: Attribution-ShareAlike*
- Wetlands (Moscow, Russia). **Authored by**: WomEOS. **Provided by**: Lake sunset panorama. **Located at**: https://flic.kr/p/6cJRUV. **License**: *CC BY-SA: Attribution-ShareAlike*
- Fort Fisher State Recreation Area Salt Marsh. **Authored by**: Dincher. **Located at**: https://commons.wikimedia.org/wiki/File:Fort_Fisher_State_Recreation_Area_Salt_Marsh.JPG. **License**: *CC BY-SA: Attribution-ShareAlike*

Public domain content

- View of Bond Swamp National Wildlife Refuge, Georgia. **Authored by**: John and Karen Hollingsworth. **Provided by**: U.S. Fish and Wildlife Service. **Located at**: https://commons.wikimedia.org/wiki/File:View_of_Bond_Swamp_National_Wildlife_Refuge,_Georgia.jpg. **License**: *Public Domain: No Known Copyright*
- Dancing Tupelo Swamp. **Authored by**: Jill Rees and Ellen Starr. **Provided by**: USDA. **Located at**: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/il/programs/financial/eqip/?cid=nrcs141p2_030473. **License**: *Public Domain: No Known Copyright*

7.1: Introduction is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.2: Types of wetlands

Wetlands are lands that are wet for significant periods. They are common where water and land meet. Wetlands can be large flat areas or small and steep areas. Wetlands are vibrant and unique ecosystems with many species that rely on both the land and the water for survival. Only specialized plants can grow in these conditions. Wetlands tend to have a great deal of biological diversity. Wetland ecosystems can also be fragile systems sensitive to the amounts and quality of water present within them. Learn more about wetlands further from the U.S. Environmental Protection Agency.



"Sunrise at Viru Bog, Estonia" is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

Types of Wetlands

There are a variety of distinct types of wetlands. **Marshes** are shallow wetlands around lakes, streams, or the ocean where grasses and reeds are common, but trees are not. Frogs, turtles, muskrats, and many varieties of birds are at home in marshes.



"Bride Book Salt Marsh" is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

A **swamp** is a wetland with lush trees and vines found in a low-lying area beside slow-moving rivers. Like marshes, they are often or always inundated with water. Since the water in a swamp moves slowly, oxygen in the water is often scarce. Swamp plants and

animals must be adapted for these low-oxygen conditions. Like marshes, swamps can be freshwater, saltwater, or a mixture of both.



"National Wildlife Refuge, Georgia" by the United States Fish and Wildlife Service is licensed as Public Domain.

In an **estuary**, salt water from the sea mixes with fresh water from a stream or river. These semi-enclosed areas are home to plants and animals that can tolerate the sharp changes in salt content that the constant motion and mixing of waters create. Estuaries contain brackish water, water that has more salt than freshwater but less than seawater. Because of the rapid changes in salt content, estuaries have many different habitats for plants, animals, and extremely high biodiversity.



"Sierra Leone River Estuary" by the European Space Agency is licensed under the Creative Commons Attribution-ShareAlike 3.0 IGO license.

Ecological Role of Wetlands

As mentioned above, wetlands are home to many varied species of organisms. Although they make up only 5 percent of the United States area, wetlands contain more than 30 percent of the plant types. Many endangered species live in wetlands, so wetlands are protected from human use. Wetlands also offer protection from storm surges from tropical storm systems like hurricanes and supply safe, hidden protection for hatchlings.

Wetlands also play a critical biological role by removing pollutants from water. For example, they can trap and use fertilizer that has washed off a farmer's field, and therefore they prevent that fertilizer from contaminating another body of water. Since wetlands naturally purify water, preserving wetlands also help to maintain clean supplies of water.

Ponds and Lakes

Ponds and lakes are bordered by hills or low rises so that the water is blocked from flowing directly downhill. **Ponds** are small bodies of freshwater that usually have no outlet; ponds are often are fed by underground springs. **Lakes** are more substantial bodies of water. Lakes are usually freshwater, although the Great Salt Lake in Utah is just one exception. Water usually drains out of a lake through a river or a stream, and all lakes lose water to evaporation.

Large lakes have tidal systems and currents, and can even affect weather patterns. The Great Lakes in the United States contain 22 percent of the world's fresh surface water. The largest them, Lake Superior, has a tide that rises and falls several centimeters each day. The Great Lakes are large enough to alter the Northeastern United States' weather system by the "lake effect," which is an increase in snow downwind of the relatively warm lakes. The Great Lakes are home to countless species of fish and wildlife.

Lakes form in various ways: in depressions carved by glaciers, in calderas, and along tectonic faults, to name a few. Subglacial lakes are even found below an ice cap. As a result of geologic history and landmasses' arrangement, most lakes are in the Northern Hemisphere.

Limnology is the study of bodies of freshwater and the organisms that live there. The ecosystem of a lake is divided into three distinct sections:

- **Surface (littoral) zone** is the sloped area closest to the edge of the water.
- **Open-water zone** (also the photic or limnetic zone) has abundant sunlight.
- **Deep-water zone** (also the aphotic or profundal zone) has little or no sunlight.

There are several life zones found within a lake. In the littoral zone, sunlight promotes plant growth, which provides food and shelter to animals such as snails, insects, and fish. In the open-water zone, other plants and fish, such as bass and trout, live. The deep-water zone does not have photosynthesis since there is no sunlight. Most deep-water organisms are scavengers, such as crabs and catfish that feed on dead organisms that fall to the bottom of the lake. Fungi and bacteria aid in the decomposition in the deep zone. Though different creatures live in the oceans, ocean waters also have these same divisions based on sunlight with similar types of creatures that live in each of the zones.

Lakes are not permanent features of a landscape. Some come and go with the seasons, as water levels rise and fall. Over a longer time, lakes disappear when they fill with sediments, if the springs or streams that fill them diminish, or if their outlets grow because of erosion. When the climate of an area changes, lakes can either expand or shrink. Lakes may disappear if precipitation significantly diminishes.

This page titled 7.2: Types of wetlands is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Adam Dastrup via source content that was edited to the style and standards of the LibreTexts platform.

7.3: Productivity in wetland ecosystems

Wetlands are among the most productive ecosystems in the world, comparable to rain forests and coral reefs. They also are a **source of substantial biodiversity** in supporting numerous species from all of the major groups of organisms — from microbes to mammals. Physical and chemical features such as climate, topography (landscape shape), geology, nutrients, and hydrology (the quantity and movement of water) help to determine the plants and animals that inhabit various wetlands. Wetlands in Texas, North Carolina, and Alaska, for example, differ substantially from one another because of their varying physical and biotic nature.

Wetlands can be thought of as "biological supermarkets." They **produce great quantities of food** that attract many animal species. The complex, dynamic feeding relationships among the organisms inhabiting wetland environments are referred to as food webs. The combination of shallow water, high levels of inorganic nutrients, and high rates of primary productivity (the synthesis of new plant biomass through photosynthesis) in many wetlands is ideal for the **development of organisms that form the base of the food web** —for example, many species of insects, mollusks, and crustaceans. Some animals consume the above-ground live vegetation (herbivore-carnivore food web); others utilize the dead plant leaves and stems, which break down in the water to form small, nutrient-enriched particles of organic material called detritus.

As the plant material continues to break down into smaller and smaller particles, it becomes increasingly enriched (nutritious) due to bacterial, fungal and protozoan activity. This enriched proteinaceous material, including the various microbes that colonize it, feeds many small aquatic invertebrates and small fish. Many of these invertebrates and fish then serve as food for larger predatory amphibians, reptiles, fish, birds, and mammals. Numerous species of **birds and mammals rely on wetlands for food, water, and shelter, especially while migrating and breeding**.

Many animals need wetlands for part or all of their life-cycles. In late winter and early spring, for example, adult tiger salamanders migrate from uplands to vernal pools for **breeding and egg deposition**. The gilled larvae resulting from their fertilized eggs then develop further, eventually producing lungs. Therefore, they must leave the vernal pools for adjacent upland, generally forested, habitat as adults, where they are mainly subterranean. In this instance, a complex of wetlands within a forest matrix is important as the life-cycle requirements of the tiger salamanders change. Thus, for the existence of the tiger salamander, both wetlands and uplands are important and essential. This can similarly be said of other amphibians like the spotted salamander as well as many other animals.

The diversity of habitats in a watershed or larger landscape unit is also important for other ecological functions associated with wetlands. One such function, **biogeochemical cycling**, involves the biologic, physical, and chemical transformations of various nutrients within the biota, soils, water, and air. Wetlands are very important in this regard, particularly relating to nitrogen, sulfur, and phosphorous. A good example of this occurs in anaerobic (non-oxygenated) and chemically reduced wetland soils and the muddy sediments of aquatic habitats like estuaries, lakes, and streams, which support microbes that function in nitrogen and sulfur cycling. Upon death and decay, the nitrogen and sulfur in plant and animal biomass is released through mineralization. Much of this is eventually transformed into gaseous forms and released into the atmosphere, where it once again becomes available to certain plants and their associated nitrogen-fixing bacteria in the soil. This is literally a major defense for mud, since it is the anaerobic and chemically reducing conditions in the substrate, in conjunction with various microbes, that ensure the gaseous release of the nitrogen and sulfur. On the other hand, phosphorous does not have a gaseous form, but vascular plants in wetlands transform inorganic forms of phosphorus (that might otherwise be shunted into undesirable algal blooms) into organic forms in their biomass as they grow. Thus, **wetlands provide the conditions needed for the removal of both nitrogen and phosphorus from surface water.**

Scientists also point out that **atmospheric maintenance** is an additional wetland function. Wetlands **store carbon within their live and preserved (peat) plant biomass instead of releasing it to the atmosphere as carbon dioxide, a greenhouse gas** affecting global climates. Therefore, **wetlands world-wide help to moderate global climatic conditions**. On the other hand, filling, clearing and draining wetlands releases carbon dioxide.

Wetlands also play an important role in the **hydrologic cycle** -- a cycle we all experience quite readily, for example, with the precipitation from a thunderstorm and the evaporation of ponded water from a puddle or bird bath. **Wetlands can receive, store, and release water in various ways** -- physically through ground water and surface water, as well as biologically through transpiration by vegetation -- and therefore function in this very important global cycle.

Fish and wildlife use wetlands to varying degrees depending upon the species involved. Some live only in wetlands for their entire lives; others require wetland habitat for at least part of their life cycle; still others use wetlands much less frequently, generally for feeding. In other words, for many species wetlands are primary habitats, meaning that these species depend on them for survival; for others, wetlands provide important seasonal habitats, where food, water, and cover are plentiful.

For example, wetlands are essentially the permanent habitat of the beaver, muskrat, wood duck, clapper rail, mud minnow, wild rice, cattail, broadleaf arrowhead and swamp rose. For other species, such as largemouth bass, chain pickerel, woodcock, hooded warbler, otter, black bear, raccoon, and meadow vole, wetlands provide important food, water, shelter, or nesting habitat.

Numerous birds —including certain shorebirds, wading birds, and raptors, and many songbirds— feed, nest, and/or raise their young in wetlands. Migratory waterfowl, including ducks, geese, and swans, use coastal and inland wetlands as resting, feeding, breeding, or nesting grounds for at least part of the year. For example, in the Chesapeake Bay Region (a major wintering area for waterfowl), coastal wetlands supported an annual average of nearly 79,000 wintering black ducks over a 45-year period (1950-1994); over the same period, it supported an annual average of about 14,000 wintering pintails. Most of these ducks rely on the prairie potholes (depressional wetlands) in upper mid-western United States and adjacent Canada and interior wetlands in northeastern North America for nesting. Indeed, an international agreement to protect wetlands of international importance was developed because some species of migratory birds are completely dependent on certain wetlands and would become extinct if those wetlands were destroyed (read on for the economic values associated with these resources.)

Because they produce so much plant biomass and invertebrate life, **estuaries and their coastal marshes serve as important nursery areas** for the young of many game (recreational) and commercial fish and shellfish. Menhaden, flounder, sea trout, spot, croaker, and striped bass are among the more familiar fish that depend on coastal wetlands. Such areas are also critical nursery habitat for young commercial shrimp along the Southeast and Gulf Coasts. Freshwater fish, such as the chain pickerel and northern pike, use well-flooded or ponded wetlands as breeding and nursery areas. Some fish, like the brown bullhead and mud minnow, even subsist in wetlands that have natural low dissolved oxygen concentrations that unadapted species cannot endure. **In the Pacific Northwest, some wetlands release cooler water to salmon-bearing streams and rivers**; in places this is critical to the health of coldwater fish populations.

Excerpted from:

Wetland Functions and Values, Watershed Academy Web, US Environmental Protection Agency. Accessed on December 2023 https://cfpub.epa.gov/watertrain/moduleFrame.cfm?parent_object_id=262

7.3: Productivity in wetland ecosystems is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.4: Organic matter storage in wetlands

Organic Matter Decomposition

The organic matter generated by plants can be mineralized or accumulate in the soil matrix as sequestered organic carbon and other elements (Elsey-Quirk & Cornwell, 2022). Slow cycling of organic resources may indicate inherent resistance to enzymatic attack but may also be due to environmental factors, e.g., moisture, pH, temperature, particle size, and nutrient and oxygen availability (Li et al., 2022). Natural polymers (e.g., lignin, cellulose, hemicellulose, pectin) can bind to inorganic ions, clay, or other organic residues that protect these materials from degradation. Plant detritus can also be naturally enclosed to other molecules, e.g., cellulose is usually tightly associated with lignin, which limits cellulase access.

As oxygen-dependent peroxidases and dioxygenases primarily degrade lignin, lignified cellulose is not readily decomposed in anaerobic environments. Other oxidants, such as nitrate, metal oxides or sulfate (e.g., Equations 1 and 2, 4, 5, and 8, respectively) do not seem to promote anaerobic degradation of these resources, which leads to lignocellulose accumulation in water-saturated soils and sediments, as bogs (Kirk, 2004).

The storage of organic matter in soils and sediments constitutes an essential reservoir of carbon in the biosphere (Fenchel, et al., 2012). Storage is derived from balancing detritus input and mineralization (Six & Jastrow, 2002; Kirk, 2004). Element turnover is often quantified as mean residence time (MRT) or half-life (t½). The MRT of an element is defined as: (i) the average time the element resides in the medium at a steady state or (ii) the average time required to replenish the contents of the element at a steady state completely. The half-life of soil organic matter (Soil OM) is the time required for decomposing 50% of the currently existing stock.

The typical model used to describe the dynamic behavior or turnover of the Soil OM is the first order model, which assumes a constant (zero-order) input with a proportional (constant) mass loss per unit of time. For Soil OM, the MRT = 1⁄k; where k = coefficient of mass loss of organic matter (t-1); Six & Jastrow (2002). Depending on latitude, humidity, management, soil type and composition and method used for evaluation, the MRT of Soil OM ranges from tens to thousands of years (Stevenson, 1994; Six & Jastrow, 2002; Torn et al., 2009).

Evidence suggests that the persistence of organic carbon oxidation in the environment (soil and sediment) is determined by the interaction between substrates, microbial communities, and abiotic conditions. Therefore, organic matter turnover must be seen as dependent on microbial ecology and the state of a specific environment. Varying degrees of structural organization, microbial ability, and resource constraints within a given environment (soil aggregate, soil horizon) make it likely that identical organic compounds can be recycled at different rates due to variations in driving force (i.e., controlling environmental variables). Soil organic matter can be classified as a reservoir of reduced carbon in different states (Kleber, 2010). This categorization leads to observing the structural heterogeneity of the Soil OM and organic detritus. On a carbon basis, it is possible to distinguish that decomposing resources are formed by labile organic carbon (LC), soluble organic carbon (which will be released in dissolved form; DOC) and refractory particulate organic carbon (POCR; comprising basically of structural compounds); Bianchini Júnior & Cunha-Santino (2011)

Inputs and Exports of Organic Matter

Organic matter enters wetlands from various sources (i.e., point and diffuse) and in different ways (e.g., functions: impulse, step, linear, exponential, and periodic). The atmospheric sources (precipitation and dry deposition) that enter the air-water interface are partially dependent on the surface size of the aquatic environment and the amount of rainfall. In wetlands, detritus additions are naturally (and periodically) linked to: (i) the seasonality of the climate; (ii) GHG emissions (Bloom et al., 2012) and (iii) hydrological dynamics (Baker et al., 2009; Gilvear & Bradley, 2009; Grootjans & Van Diggelen, 2009).

Seasonal additions of detritus tend to be more evident in temperate than in tropical environments, where seasonal temperature variations are less intense (Gonçalves Junior et al., 2014). Hydrological dynamics (e.g., variations in aquifer height, fluviometric variations, runoff), which define the frequency and duration of floods, are an important source of allochthonous detritus (Barker & Maltby, 2009; Dise, 2009). Detritus enters a wetland from various sources and in several different ways. Sources of adduction related to the hydrological cycle, such as mass transported by flow or precipitation can often be characterized by periodic functions.

The general pattern of high spring/summer runoff, with the relatively low and constant flow for other seasons, is repeated in a very predictable way.

The origin of the detritus (autochthonous or allochthonous) can also interfere with cycling rates, with allochthonous resources being normally more inaccessible since potentially labile fractions have already been consumed, remaining detritus refractory fractions (Thurman, 1985; Gimenes et al., 2010). In wetlands, hydrology alters many variables related to detritus cycling, e.g., humidity depends on the flood regime, running water carries oxygen and nutrients, while in stagnant water, oxygen is quickly depleted, and nutrients are transformed into less available forms. Hydrological changes induced by climatic and anthropogenic disturbances can also define the rates and predominant composition of GHG emissions; for example, drainage lowers the water table and raises the oxygen content of the soil, increasing CO_2 emissions. In temperate wetlands, the highest emissions were found where the water level remained close to the soil surface, suggesting that mainly litter and not burial organic matter contributes to CH_4 emission (Wang et al., 2021). CH_4 emissions from drained wetland soils are generally negligible because soil carbon is preferentially oxidized to CO_2 (Hiraishi et al., 2014).

Three different reactions generate CH_4 under strictly reductive conditions (Boon, 2006). The first uses CO_2, acetate (HCOO-), or carbon monoxide (CO) to produce CH_4 (Equation 7.4.1). In the second reaction, CH_4 can be produced by the reduction of the methyl group of methyl compounds, such as methanol (Equation 7.4.2). In the third reaction, CH_4 is produced by the breakdown of acetate into methane and carbon dioxide (Equation 7.4.3).

$$\begin{gathered}\mathrm{CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O} \\ \mathrm{+ Energy\ (\Delta G_{00} = -31.3 kcal mol^{-1})}\end{gathered} \tag{7.4.1}$$

$$\begin{gathered}\mathrm{4CH_3OH \rightarrow 3CH_4 + CO_2 + 2H_2O} \\ \mathrm{+\ Energy\ \ (\Delta G_{00} = -76.2 kcal mol^{-1})}\end{gathered} \tag{7.4.2}$$

$$\begin{gathered}\mathrm{4CH_3COO^- + H_2O \rightarrow 3CH_4 + HCO_3^-} \\ \mathrm{+Energy\ \ (\Delta G_{00} = -7.4 kcal mol^{-1})}\end{gathered} \tag{7.4.3}$$

Anaerobic mineralization bioassays indicated that CO_2 is the main product, instead of CH_4 (Romeiro & Bianchini Júnior, 2006; Cunha-Santino & Bianchini Júnior, 2013; Bianchini Júnior & Cunha-Santino, 2016). The controlling factors of CH_4 production are: (i) availability of electron acceptors (Segers, 1998); (ii) quantity and quality of the organic matter supply (Bianchini Júnior et al., 2010); (iii) temperature (Romeiro & Bianchini Júnior, 2008); (iv) pH (Kiene, 1991; Cunha-Santino et al., 2006; Bloom et al., 2012); (v) coenzymes and prosthetic groups (Schlegel, 1997) and (vi) micronutrients (Banik et al., 1996; Basiliko & Yavitt, 2001). Interactions between these abiotic factors that influence metabolic pathways in generating of specific intermediate products can influence CH_4 production (Bergman et al., 1999). The intermediate compounds (e.g., methanol, propanol, formic acid, butyric acid) and acetate (Equation 34) is the main substrate for methanogenesis (Boone, 1991; Conrad, 1999).

The physical and biotic structure and resulting metabolism of a wetland ecosystem are tightly coupled to hydrological and chemical loads from the watershed (Wetzel, 2006). The importance of hydrology for the export of organic carbon is evident; in general, higher export rates are expected from wetlands that are open to water flow. Riparian wetlands provide large amounts of organic detritus to streams, including coarse detritus. It is evidence that watersheds that drain wetlands export more organic material but maintain more nutrients than watersheds that have no wetlands (Mitsch & Gosselink, 2015). In wetlands, the export of organic matter is predominantly associated with dissolved organic matter derived from relatively recalcitrant chemical compounds, often associated with the origin of lignin and cellulose structural tissues of higher plants and various products of bacterial degradation (Wetzel, 2006).

References

Baker, C., Thompson, J.R., & Simpson, M., 2009. Hydrological dynamics I: surface waters, flood and sediment dynamics. In: Barker, T., & Maltby, E., eds. The wetlands handbook. Chichester: Wiley-Blackwell, 120-168. http://dx.doi.org/10.1002/9781444315813.ch6.

Banik, S., Sen, M., & Sen, S.P., 1996. Effects of inorganic fertilizers and micronutrients on methane production from wetland rice (Oryza sativa L.). Biol. Fertil. Soils 21(4), 319-322. http://dx.doi.org/10.1007/BF00334910.

Barker, T., & Maltby, E., 2009. Introduction – The dynamics of wetlands. In: Barker, T., & Maltby, E., eds. The wetlands handbook. Chichester: Wiley-Blackwell, 113-119. http://dx.doi.org/10.1002/9781444315813.ch5.

Basiliko, N., & Yavitt, J.B., 2001. Influence of Ni, Co, Fe, and Na additions on methane production in Sphagnum-dominated northern American peatlands. Biogeochemistry 52(2), 133-153. http://dx.doi. org/10.1023/A:1006461803585

Bergman, I., Lundberg, P., & Nilsson, M., 1999. Microbial carbon mineralization in an acid surface peat: effects of environmental factors in laboratory incubations. Soil Biol. Biochem. 31(13), 1867-1877. http://dx.doi.org/10.1016/S0038-0717(99)00117-0.

Bianchini Júnior, I., Cunha-Santino, M.B., Romeiro, F., & Bitar, A.L., 2010. Emissions of methane and carbon dioxide during anaerobic decomposition of aquatic macrophytes from a tropical lagoon (São Paulo, Brazil). Acta Limnol. Bras. 22(2), 157-164. http://dx.doi.org/10.1590/S2179- 975X2010000200005.

Bianchini Júnior, I., & Cunha-Santino, M.B., 2011. Model parameterization for aerobic decomposition of plant resources drowned during man-made lakes formation. Ecol. Modell. 222(7), 1263-1271. http://dx.doi.org/10.1016/j.ecolmodel.2011.01.019.

Bianchini Júnior, I., & Cunha-Santino, M.B., 2016. CH_4 and CO_2 from decomposition of Salvinia auriculata Aublet, a macrophyte with high invasive potential. Wetlands 36(3), 557-564. http://dx.doi. org/10.1007/s13157-016-0765-4.

Bloom, A.A., Palmer, P.I., Fraser, A., & Reay, D.S., 2012. Seasonal variability of tropical wetland CH_4 emissions: the role of the methanogen-available carbon pool. Biogeosciences 9(8), 2821-2830. http:// dx.doi.org/10.5194/bg-9-2821-2012.

Boone, D.R. 1991. Ecology of methanogens. In: Rogers, J.E. & Whitman, W.B., eds. Microbial production and consumption of greenhouse gases: methane, nitrogen oxides, and halomethanes. Washington: American Society for Microbiology, 39-51.

Conrad, R., 1999. Contribution of hydrogen to methane production and control of hydrogen concentrations in methanogenic soils and sediment. FEMS Microbiol. Ecol. 28(3), 193-202. http://dx.doi. org/10.1111/j.1574-6941.1999.tb00575.x.

Cunha-Santino, M.B., Bianchini Júnior, I., Gianotti, E.P., & Silva, E., 2006. Degradação anaeróbia de macrófitas aquáticas da Lagoa do Infernão: metanogênese. In: Santos, J.E., Pires, J.S., & Moschini, L.E., eds. Estudos Integrados em Ecossistemas - Estação Ecológica de Jataí. São Carlos: EdUFSCar, 143-158.

Cunha-Santino, M.B., & Bianchini Júnior, I., 2013. Tropical macrophyte degradation dynamics in freshwater sediments: relationship to greenhouse gas production. J. Soils Sediments 13(8), 1461-1468. http://dx.doi.org/10.1007/s11368-013-0735-x

Dise, N.,2009. Biogeochemical dynamics III: the critical role of carbon in wetlands. In: Barker, T., & Maltby, E. eds. The wetlands handbook. Chichester: Wiley-Blackwell, 249-265. http://dx.doi. org/10.1002/9781444315813.ch11.

Elsey-Quirk, T., & Cornwell, J.C., 2022. Organic matter and nutrient cycling in coastal wetlands and submerged aquatic ecosystems in an age of rapid environmental change -The Anthropocene. J. Mar. Sci. Eng. 10(8), 1096. http://dx.doi.org/10.3390/ jmse10081096.

Fenchel, T., King, G.M., & Blackburn, T.H., 2012. Bacterial biogeochemistry: the ecophysiology of mineral cycling. Amsterdam: Academic Press, 312 p.

Gilvear, D.J., & Bradley, C., 2009. Hydrological dynamics II: groundwater and hydrological connectivity. In: Barker, T., & Maltby, E., eds. The wetlands handbook. Chichester: Wiley-Blackwell, 169-193. http://dx.doi.org/10.1002/9781444315813.ch7.

Gimenes, K.Z., Cunha-Santino, M.B., & Bianchini Júnior, I., 2010. Decomposição de matéria orgânica alóctone e autóctone em ecossistemas aquáticos. Oecol. Aust. 14(4), 1075-1112. http://dx.doi. org/10.4257/oeco.2010.1404.13.

Gonçalves Junior, J.F., de Souza Rezende, R., Gregório, R.S., & Valentin, G.C., 2014. Relationship between dynamics of litterfall and riparian plant species in a tropical stream. Limnologica 44, 40-48. http://dx.doi.org/10.1016/j.limno.2013.05.010.

Grootjans, A.P., & Van Diggelen, R., 2009. Hydrological dynamics III: hydro-ecology. In: Barker, T., & Maltby, E., eds. The wetlands handbook. Chichester: Wiley-Blackwell, 194-212. http://dx.doi. org/10.1002/9781444315813.ch8.

Hiraishi, T., Krug, T., Tanabe, K., Srivastava, N., Baasansuren, J., Fukuda, M. & Troxler, T.G., 2014. 2013 Supplement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Wetlands. Switzerland: IPCC.

Kiene, R.P., 1991. Production and consumption of methane in aquatic systems. In: Rogers, J.E., & Whitman, W.B., eds. Microbial production and consumption of greenhouse gases: methane, nitrogen oxides, and halomethanes. Washington: American Society for

Microbiology, 111-146.

Kirk, G., 2004. The biochemistry of submerged soils. Chichester: John Wiley & Sons, 291p. http://dx.doi. org/10.1002/047086303X

Kleber, M., 2010. What is recalcitrant soil organic matter? Environ. Chem. 7(4), 320-332. http://dx.doi. org/10.1071/EN10006.

Li, Y., Chen, Z., Chen, J., Castellano, M.J., Ye, C., Zhang, N., Miao, Y., Zheng, H., Li, J., & Ding, W., 2022. Oxygen availability regulates the quality of soil dissolved organic matter by mediating microbial metabolism and iron oxidation. Glob. Chang. Biol. 28(24), 7410-7427. PMid:36149390. http://dx.doi. org/10.1111/gcb.16445.

Mitsch, W.J., & Gosselink, J.G., 2015. Wetlands. Hoboken: John Wiley & Sons, 752 p

Romeiro, F., & Bianchini Júnior, I., 2008. Kinetic pathways for anaerobic decomposition of Ludwigia inclinata. Hydrobiologia 607(1), 103-111. http:// dx.doi.org/10.1007/s10750-008-9370-8.

Schlegel, H.G., 1997. Microbiología general. Barcelona: Omega, 672 p.

Segers, R., 1998. Methane production and methane consumption: a review of process underlying wetlands methane fluxes. Biogeochemistry 41(1), 23-51. http://dx.doi.org/10.1023/A:1005929032764.

Six, J., & Jastrow, J.D., 2002. Organic matter turnover. In: Lal, R., ed. Encyclopedia of soil science. New York: Marcel Dekker, 936-942.

Stevenson, F.J., 1994. Humus chemistry: genesis, composition, reactions. New York: John Willey, 512 p.

Thurman, E.M., 1985. Organic geochemistry of natural waters. Dordrecht: Martinus Nijhoff/Dr W. Junk, 497 p. http://dx.doi.org/10.1007/978-94-009- 5095-5.

Torn, M.S., Swanston, C.W., Castanha, C., & Trumbore, S.E., 2009. Storage and turnover of organic matter in soil. In: Senesi, N., Xing, B., & Huang, P.M., eds. Biophysico-chemical processes involving natural nonliving organic matter in environmental. Hoboken: John Wiley & Sons, 219-271. http:// dx.doi.org/10.1002/9780470494950.ch6.

Wang, H., Ho, M., Flanagan, N., & Richardson, C.J., 2021. The effects of hydrological management on methane emissions from Southeastern shrub bogs of the USA. Wetlands 41(7), 87. http://dx.doi. org/10.1007/s13157-021-01486-7.

Wetzel, R.G., 2006. Wetlands ecosystem processes. In: Batzer, D.P. & Sharitz, R.R. (Eds.) Ecology of freshwater and estuarine wetlands. Berkeley: University of California Press, 285-312.

Excerpted from

Cunha-Santino, M. B. D., & Bianchini Júnior, I. (2023). Reviewing the organic matter processing by wetlands. *Acta Limnologica Brasiliensia*, *35*, e19. Accessed December 2023 from https://www.scielo.br/j/alb/a/ypwb635W6PGrxXSZWPtgt4S CC-BY

7.4: Organic matter storage in wetlands is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Marcela Bianchessi da Cunha-Santino1 and Irineu Bianchini Júnior.

7.5: Microbial metabolism in saturated sediments

The requirements for organic matter reduction soil are: (i) the absence of oxygen; (ii) the presence of organic matter, and (iii) anaerobic bacterial activity (Sigee, 2005). Metabolic routes, reaction rates and the degree of organic matter reduction are influenced by the nature and content of organic matter, temperature, nature and content of electron acceptors, and pH (Wilson et al., 2011; Szafranek-Nakonieczna & Stepniewska, 2014; Boye et al., 2018; Li et al., 2021b). Growth rates (i.e., anabolism) are not invariant properties of organisms; they are markedly influenced by the prevailing environmental conditions, particularly by the complexity of the medium (substrate) and by the nature of the primary source of carbon and energy (Mandelstam et al., 1982).

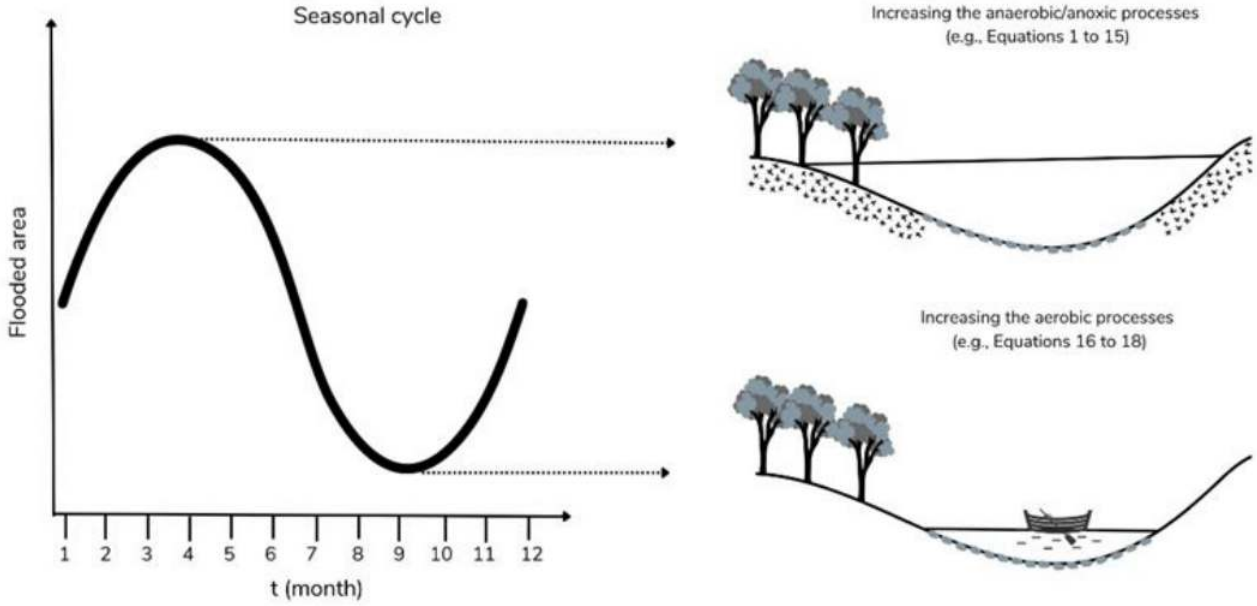


Figure 7.5.1: . Seasonal change in the predominant metabolic reactions in wetland soil as a function of the flood pulse (sensu Neiff, 1999). Hatched soil: predominantly anoxic or anaerobic processes. Equations for the aerobic and anaerobic reations are found on the next page

Usually, heterotrophic bacteria grow faster in a medium containing nutrients and complex organic compounds (e.g., amino acids, purines, pyrimidines, and vitamins). Facultative anaerobic bacteria growth in the presence of oxygen is generally faster than in its absence. Presumably, this reflects the greater efficiency with which organisms can generate ATP in an aerobic environment; comparisons of the energy yields of aerobic oxidations (e.g., Conn et al., 1987) with those of anaerobic respirations and fermentations illustrate these processes. Substances such as nitrate, whose assimilation requires extra energy demands on organisms, generally support growth at a lower rate than those such as ammonia, whose assimilation requires less energy input (Mandelstam et al., 1982).

Microorganisms can use a wide variety of organic compounds as a source of energy. The catabolism of these substances leads to the production of intermediate compounds that act as raw material suppliers for biosynthetic reactions and energy production (Williams & del Giogio, 2005). The CO_2 molecules produced in the decomposition (anaerobic: or aerobic decay: ; Figure 2) associate with water molecules and form carbonic acid; carbonic acid dissociations produce bicarbonates and carbonates; such dissociations generate H+ (Equation 7.5.1), thus acidify the medium (Langmuir, 1997).

In water-saturated soils, typical of wetlands, the degree of acidity partially depends on the amount of decomposed organic material and the degradation velocity. The degree of acidity is mainly proportional to the production of CO_2 . Other processes also contribute to the increase in the acidity are: (i) leaching (dissolutions) of detritus and subsequent dissociation of organic acids; (ii) nitrification (Equation 7.5.2); (iii) oxidations of sulfur compounds (Equations 7.5.3 to 7.5.5) and (iv) cation hydrolysis (Mihelcic, 1999; Cunha-Santino & Bianchini Júnior, 2002; Konhauser, 2007).

Increases in water electrical conductivity values due to soil submersion are related to releasing ions and compounds from leaching and mineralization processes. Changes in the color of water result from the formation and dissolution of inorganic colloids (e.g.,

iron and sulfur salts; Ciminelli et al., 2014; Li et al., 2021a; Equations 7.5.6 and 7.5.7), the release of pigments from plants detritus (Killops & Killops, 2013) and the formation of organic colloids (e.g., humic substances; Cunha-Santino et al., 2013).

$$\begin{aligned} &CO_2 + H_2O \rightarrow H_2CO_3 + H_2O \rightarrow H^+ \\ &+HCO_3{}^- + H_2O \rightarrow H^+ + CO_3{}^2 \end{aligned} \tag{7.5.1}$$

$$NH_4{}^+ + 2O_2 \rightarrow NO_3{}^- + H_2O + 2H^+ \tag{7.5.2}$$

$$H_2\ S + 2O_2 \rightarrow SO_4{}^{2-} + 2H^+ \tag{7.5.3}$$

$$S^0 + H_2O + 1\tfrac{1}{2}O_2 \rightarrow SO_4{}^{2-} + 2H^+ \tag{7.5.4}$$

$$S_2O_3{}^{2-} + H_2O + 2O_2 \rightarrow 2SO_4{}^{2-} + 2H^+ \tag{7.5.5}$$

$$CH_2O + 8H^+ + 4FeOOH \rightarrow CO_2 + 4Fe^{2+} + 7H_2O \tag{7.5.6}$$

$$3H_2\ S + 2FeOOH \rightarrow 2FeS + S^0 + 4H_2O \tag{7.5.7}$$

In anaerobic environments (e.g., sediments and saturated soils), the rates of detritus mineralization (which are derived mainly from the growth rates of microorganisms) usually are lower than those observed under aerobic conditions (Wetzel, 2001; Kirk, 2004; Mitsch & Gosselink, 2015). The tendency is for a greater stock of organic matter to occur in sediments, submerged soils, and hydrosols. The accumulation of plant residues in soils and sediments generally results from a long time of degradation of plant resources compared to the rates of input of allochthonous detritus and those from autochthonous primary production, i.e., chemosynthesis, oxygenic (Equation 7.5.8) or anoxygenic (Equations 7.5.9 to 7.5.10) (Conn et al., 1987). The lack of oxygen has drastic consequences for the decomposition of organic material. Obligate aerobic decomposers are limited to the more aerated patches of sediment. Fungi and bacteria that can break down refractory organic molecules mainly belong to this group. Although the half-life of such substances increases strongly in anaerobic soils, the more easily degradable organic matter will be transformed by anaerobes (facultative or obligate). These organisms use other electron acceptors to replace oxygen, which leads to interactions of the carbon cycle with those of nitrogen, manganese, iron, and sulfur (Verhoeven, 2009), as demonstrated by Equations 1 to 10.

$$\begin{aligned} &6CO_2 + 12H_2O + \text{Energy}\left(\Delta G_{00} = +686\text{kcalmol}^{-1}\right) \\ &\rightarrow C_6H_{12}O_6 + 6H_2O + 6O_2 \end{aligned} \tag{7.5.8}$$

Illegal pream-token (\)

Illegal pream-token (2)

Illegal pream-token (\)

References

Boye, K., Herrmann, A.M., Schaefer, M.V., Tfaily, M.M., & Fendorf, S., 2018. Discerning microbially mediated processes during redox transitions in flooded soils using carbon and energy balances. Front. Environ. Sci. 6, 15. http://dx.doi.org/10.3389/fenvs.2018.00015.

Conn, E.E., Stumpf, P.K., Bruening, G., & Doi, R.Y., 1987. Outlines of biochemistry. New York: John Wiley and Sons, 693 p.

Ciminelli, V.S.T., Barbosa, F.A.R., Tundisi, J.G., & Duarte, H.A., 2014. Recursos minerais, água e biodiversidade. Quím. Nova Esc. 8, 39-45.

Cunha-Santino, M.B., & Bianchini Júnior, I., 2002. Estequiometria da decomposição aeróbia de galhos, cascas serapilheira e folhas. In: Espíndola, E.L.G., Mauad, F.F., Schalch, V., Rocha, O., Felicidade, N., & Rietzler, A.C., eds. Recursos hidroenergéticos: usos, impactos e planejamento integrado. São Carlos: RiMa, 43-56.

Cunha-Santino, M.B., & Bianchini Júnior, I., 2013. Tropical macrophyte degradation dynamics in freshwater sediments: relationship to greenhouse gas production. J. Soils Sediments 13(8), 1461-1468. http://dx.doi.org/10.1007/s11368-013-0735-x.

Killops, S.D., & Killops, V.J., 2013. Introduction to organic geochemistry. Hoboken: Wiley-Blackwell, 408p.

Kirk, G., 2004. The biochemistry of submerged soils. Chichester: John Wiley & Sons, 291p. http://dx.doi.org/10.1002/047086303X.

Konhauser, K., 2007. Introduction to geomicrobiology. New York: Blackwell, 440 p.

Langmuir, D., 1997. Aqueous environmental geochemistry. Upper Saddle River: Prentice Hall, 600 p.

Li, B., Feng, M., Chen, X., Wang, Y., Shen, Y., & Wu, Q.L., 2021a. Abundant sediment organic matter potentially facilitates chemical iron reduction and surface water blackness in a Chinese deep lake. Environ. Pollut. 272, 116002. PMid:33246758. http://dx.doi.org/10.1016/j.envpol.2020.116002.

Li, Q., Leroy, F., Zocatelli, R., Gogo, S., Jacotot, A., Guimbaud, C., & Laggoun-Défarge, F., 2021b. Abiotic and biotic drivers of microbial respiration in peat and its sensitivity to temperature change. Soil Biol. Biochem. 153, 108077. http://dx.doi.org/10.1016/j.soilbio.2020.108077.

Mandelstam, J., McQuillen, K., & Dawes, I., 1982. Biochemistry of bacterial growth. Oxford: Blackwell, 449 p.

Mihelcic, J.R., 1999. Fundamentals of environmental engineering. New York. John Wiley, 352 p.

Mitsch, W.J., & Gosselink, J.G., 2015. Wetlands. Hoboken: John Wiley & Sons, 752 p.

Sigee, D.C., 2005. Freshwater microbiology. Chichester: John Wiley & Sons, 524 p.

Szafranek-Nakonieczna, A., & Stepniewska, Z., 2014. Aerobic and anaerobic respiration in profiles of Polesie Lubelskie peatlands. Int. Agrophys. 28(2), 219-229. http://dx.doi.org/10.2478/intag-2014-0011.

Verhoeven, J.T.A., 2009. Wetlands biochemical cycles and their interactions. In: Barker, T., & Maltby, E., eds. The wetlands handbook. Chichester: Wiley-Blackwell, 266-281. http://dx.doi.org/10.1002/9781444315813.ch12.

Wetzel, R.G., 2001. Limnology: lake and river ecosystems. San Diego: Academic Press, 1006 p.

Williams, P.J.B., & del Giogio, P.A., 2005. Respiration in aquatic ecosystems: history and background. In: Del Giorgio, P.A. & Williams, P.J.B., eds. Respiration in aquatic ecosystems. Oxford: Oxford University Press, 1-17. http://dx.doi.org/10.1093/acprof:oso/9780198527084.003.0001.

Wilson, J.S., Baldwin, D.S., Rees, G.N., & Wilson, B.P., 2011. The effects of short-term inundation on carbon dynamics, microbial community structure and microbial activity in floodplain soil. River Res. Appl. 27(2), 213-225. http://dx.doi.org/10.1002/rra.1352.

Excerpted from:

Cunha-Santino, M. B. D., & Bianchini Júnior, I. (2023). Reviewing the organic matter processing by wetlands. *Acta Limnologica Brasiliensia*, *35*, e19. Accessed December 2023 from https://www.scielo.br/j/alb/a/ypwb635W6PGrxXSZWPtgt4S CC-BY

7.5: Microbial metabolism in saturated sediments is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.6: Anaerobic and aerobic metabolic pathways

Microbial metabolism in anaerobic sediments, such as waterlogged wetlands, plays a crucial role in regulating several essential biogeochemical cycles, i.e., carbon, nitrogen, phosphorus, and sulfur (Inglett et al., 2005). Flooded soils support intense heterotrophic activity and many transformations are directly mediated by aerobic and anaerobic microorganisms (D'Angelo & Reddy, 1999). The reduction of submerged soil and sediment is a consequence of the anaerobic respiration of bacteria (Reddy & DeLaune, 2008). During anaerobic respiration, organic matter is oxidized, and the components of the medium are reduced (Boon, 2006; Madsen, 2015; Reddy & DeLaune, 2008; Hanson et al., 2013). Examples include: (i) denitrification (Equation 7.6.1); (ii) the (dissimilatory) reduction of nitrate and nitrite to ammonium (Equations 7.6.2 and 7.6.3) (iii) the reductions of manganese dioxide (Equations 7.6.4 and 7.6.5); (iv) the reduction of ferric hydroxide (Equations 7.6.6 and 7.6.7; (v) sulfate reduction (Equations 7.6.8 to 7.6.10) and (vi) carbon dioxide reduction (Equation 7.6.11):

$$\begin{gathered}\mathrm{C_6H_{12}O_6 + 6NO_3^- \rightarrow 2\ N_2 + 2NO_2^- + 4CO_2 + 2CO_3^{2-}} \\ \mathrm{+6H_2O + Energy\ (\Delta G_{00} = -590 kcal mol^{-1})}\end{gathered} \tag{7.6.1}$$

$$\begin{gathered}\mathrm{4CH_2O_2 + NO_3{}^- + 2H^+ \rightarrow 4CO_2 + NH_4{}^+ + 3H_2O} \\ \mathrm{+ Energy\ (\Delta G_{00} = -226 kcal mol^{-1})}\end{gathered} \tag{7.6.2}$$

$$\begin{gathered}\mathrm{3CH_2O_2 + NO_2^- + 2H^+ \rightarrow 3CO_2 + NH_4{}^+ + 2H_2O} \\ \mathrm{+ Energy\ (\Delta G_{00} = -166 kcal mol^{-1})}\end{gathered} \tag{7.6.3}$$

$$\begin{gathered}\mathrm{C_6H_{12}O_6 + 12MnO_2 + 24H^+ \rightarrow 6CO_2 + 12Mn^+ + 18H_2O} \\ \mathrm{+Energy\ (\Delta G_{00} = -484 kcal mol^{-1})}\end{gathered} \tag{7.6.4}$$

$$\begin{gathered}\mathrm{CH_3COO^- + 4MnO_2 + 9H^+ \rightarrow 2CO_2 + 4Mn^+ + 6H_2O} \\ \mathrm{+ Energy\ (\Delta G_{00} = -137 kcal mol^{-1})}\end{gathered} \tag{7.6.5}$$

$$\begin{gathered}\mathrm{C_6H_{12}O_6 + 24Fe(OH)_3 + 48H^+ \rightarrow 6CO_2 + 24Fe^{2+} + 66H_2O} \\ \mathrm{+ Energy\ (\Delta G_0' = -40 kcal mol^{-1})}\end{gathered} \tag{7.6.6}$$

$$\begin{gathered}\mathrm{CH_3COO^- +} \\ \mathrm{8Fe(OH)_3 + 17H^+ \rightarrow 2CO_2 + 8Fe^{2+} + 22H_2O} \\ \mathrm{+ Energy\ (\Delta G_0' = -11 kcal mol^{-1})}\end{gathered} \tag{7.6.7}$$

$$\begin{gathered}\mathrm{2\,(CH_2O) + H_2SO_4 \rightarrow 2CO_2 + 2H_2O + H_2\ S} \\ \mathrm{+\ Energy\ \ (\Delta G_{00} = -21 kcal mol^{-1})}\end{gathered} \tag{7.6.8}$$

$$\begin{gathered}\mathrm{CH_4 + SO_4{}^{2-} \rightarrow HCO_3{}^- + HS^- + H_2O} \\ \mathrm{+\ Energy\ \ (\Delta G_{00} = -20\ to\ -40 kcal mol^{-1})}\end{gathered} \tag{7.6.9}$$

$$\begin{gathered}\mathrm{CH_3COO^- + SO_4{}^{2-} + 3H^+ \rightarrow 2CO_2 + 2H_2O + H_2\ S} \\ \mathrm{+ Energy\ (\Delta G_0' = -13.7 kcal mol^{-1})}\end{gathered} \tag{7.6.10}$$

$$\begin{gathered}\mathrm{2CH_3CH_2OH + CO_2 + 17H^+ \rightarrow CH_4 + 2CH_3COO^- + 2H^+} \\ \mathrm{+Energy\ \ (\Delta G_{00} = -24 kcal mol^{-1})}\end{gathered} \tag{7.6.11}$$

In addition to anaerobic respiration, chemosynthesis also occurs; instead of organic matter, bacteria oxidize reduced inorganic compounds, such as ammonium (Equation 7.6.12 ; Madsen, 2015), sulfides, elemental sulfur (Equation (\PageIndex{12}\)) and other reduced sulfur compounds, such as thiosulphate (Wetzel, 2001).

$$\begin{gathered}\mathrm{NH_4{}^+ + NO_2^- \rightarrow N_2 + 2H_2O} \\ \mathrm{+ Energy\ (\Delta G_{00} = -86.2 kcal mol^{-1})}\end{gathered} \tag{7.6.12}$$

$$\begin{gathered}\mathrm{5\ S^o + 6NO_3{}^- + 2CO_3{}^{2-} \rightarrow 5SO_4{}^{2-} + 2CO_2 + H_2O + 3\ N_2} \\ \mathrm{+ Energy\ (\Delta G_0' = -179 kcal mol^{-1})}\end{gathered} \tag{7.6.13}$$

In anaerobic environments, such as water-saturated wetlands soils, fermentations occur when the electron acceptors are organic compounds (Schlegel, 1997). In this case, the oxidation may be incomplete, generating intermediate organic products (such as lactic acid and ethanol; Equations 7.6.14 and 7.6.15, respectively; Conn et al., 1987). Substrate catabolism may be incomplete (Denef et al., 2009), and the products (organic or inorganic) enter the anabolic routes of the decomposers and, consequently, are resynthesized and incorporated by these organisms (Fenchel, 2008). Other products are incorporated and/or converted into the class of non-cellular organic compounds, such as humic substances (Assunção et al., 2017).

$$\begin{aligned} &C_6H_{12}O_6 \rightarrow 2CH_3CHOHCOOH + 2H_2O \\ &+ \text{Energy} \ \left(\Delta G_0' = -32.4 \text{kcalmol}^{-1}\right) \end{aligned} \tag{7.6.14}$$

$$\begin{aligned} &C_6H_{12}O_6 \rightarrow 2CH_3CH_2OH + 2CO_2 \\ &+ \text{Energy} \left(\Delta G_{00} = -25.4 \text{kcalmol}^{-1}\right) \end{aligned} \tag{7.6.15}$$

The oxidation of organic matter (e.g., glucose, palmitic acid, and lactic acid; Equations 7.6.16, 7.6.17 and 7.6.18, respectively) is interrupted by soil submersion, and the aerobic microorganisms consume the oxygen in the soil and become quiescent or die. Facultative and obligate anaerobes then proliferate, using carbon compounds and oxidized components of the soil as substrate and dissimilation products of organic matter as electron acceptors in respiration (Sahrawat, 2004).

$$\begin{aligned} &C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O \\ &+ \text{Energy} \left(\Delta G_{00} = -686 \text{kcalmol}^{-1}\right) \end{aligned} \tag{7.6.16}$$

Illegal pream-token (\)

Illegal pream-token (\)

References

Assunção, A.W.A., Souza, B.P., Silva, W.T.L., Cunha-Santino, M.B., & Bianchini Júnior, I., 2017. Molecular changes of aquatic humic substances formed from four aquatic macrophyte under different oxygen conditions. Chem. Ecol. 33(10), 918-931. http://dx.doi.org/10.1080/02757540.2 017.1393532

Boon, P.I. 2006. Biogeochemistry and bacterial ecology of hydrologically dynamic wetlands. In: Batzer, D.P., & Sharitz, R.R., eds. Ecology of freshwater and estuarine wetlands. Berkeley: University of California Press, 115-176. https://doi.org/10.1525/california/9780520247772.003.0005.

Conn, E.E., Stumpf, P.K., Bruening, G., & Doi, R.Y., 1987. Outlines of biochemistry. New York: John Wiley and Sons, 693 p

D'Angelo, E.M., & Reddy, K.R., 1999. Regulators of heterotrophic microbial potentials in wetland soils. Soil Biol. Biochem. 31(6), 815-830. http://dx.doi. org/10.1016/S0038-0717(98)00181-3.

Denef, K., Plante, A.F., & Six, J., 2009. Characterization of soil organic matter. In: Kutsch, W.L., Bahn, M., & Heinemeyer, A., eds. Soil carbon dynamics: an integrated methodology. Cambridge: Cambridge University Press, 91-126.

Fenchel, T., 2008. The microbial loop - 25 years later. J. Exp. Mar. Biol. Ecol. 366(1-2), 99-103. http://dx.doi.org/10.1016/j.jembe.2008.07.013.

Hanson, T.E., Campbell, B.J., Kalis, K.M., Campbell, M.A., & Klotz, M.G., 2013. Nitrate ammonification by Nautilia profundicola AmH: experimental evidence consistent with a free hydroxylamine intermediate. Front. Microbiol. 4, 180. PMid:23847604. http://dx.doi.org/10.3389/fmicb.2013.00180.

Inglett, P.W., Reddy, K.R., & Corstanje, R., 2005. Anaerobic soils. In: Hillel, D., ed. Encyclopedia of soils in the environment. London: Academic Press, 72-78. http://dx.doi.org/10.1016/B0-12-348530-4/00178-8

Madsen, E.L., 2015. Environmental microbiology. From genomes to biogeochemistry. Malden: Blackwell, 592 p.

Reddy, K.R., & DeLaune, R.D., 2008. Biochemistry of wetlands: science and applications. Boca Raton: CRC Press, 800 p. http://dx.doi. org/10.1201/9780203491454.

Sahrawat, K.L., 2004. Organic matter accumulation in submerged soils. Adv. Agron. 81, 169-201. http://dx.doi.org/10.1016/S0065-2113(03)81004-0.

Schlegel, H.G., 1997. Microbiología general. Barcelona: Omega, 672 p.

Excerpted from

Cunha-Santino, M. B. D., & Bianchini Júnior, I. (2023). Reviewing the organic matter processing by wetlands. *Acta Limnologica Brasiliensia*, *35*, e19. Accessed December 2023 from https://www.scielo.br/j/alb/a/ypwb635W6PGrxXSZWPtgt4S CC-BY

7.6: Anaerobic and aerobic metabolic pathways is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Marcela Bianchessi da Cunha-Santino and Irineu Bianchini Júnior.

7.7: The redox ladder

When we develop or evaluate an analytical method, we often need to understand how the chemistry that takes place affects our results. Suppose we wish to isolate Ag^+ by precipitating it as AgCl. If we also need to control pH, then we must use a reagent that does not adversely affect the solubility of AgCl. It is a mistake to use NH_3 to adjust the pH, for example, because it increases the solubility of AgCl (see reaction 6.5.4). One of the primary sources of determinate errors in many analytical methods is failing to account for potential chemical interferences.

In this section we introduce the ladder diagram as a simple graphical tool for visualizing equilibrium chemistry. We will use ladder diagrams to determine what reactions occur when we combine several reagents, to estimate the approximate composition of a system at equilibrium, and to evaluate how a change to solution conditions might affect an analytical method.

> Although not specifically on the topic of ladder diagrams as developed in this section, the following papers provide appropriate background information: (a) Runo, J. R.; Peters, D. G. *J. Chem. Educ.* **1993**, *70*, 708–713; (b) Vale, J.; Fernández-Pereira, C.; Alcalde, M. *J. Chem. Educ.* **1993**, *70*, 790–795; (c) Fernández-Pereira, C.; Vale, J. *Chem. Educator* **1996**, *6*, 1–18; (d) Fernández- Pereira, C.; Vale, J.; Alcalde, M. *Chem. Educator* **2003**, *8*, 15–21; (e) Fernández-Pereira, C.; Alcalde, M.; Villegas, R.; Vale, J. *J. Chem. Educ.* **2007**, *84*, 520–525. Ladder diagrams are a great tool for helping you to think intuitively about analytical chemistry. We will make frequent use of them in the chapters to follow.

Ladder Diagrams for Acid–Base Equilibria

Let's use acetic acid, CH_3COOH, to illustrate the process we will use to draw and to interpret an acid–base ladder diagram. Before we draw the diagram, however, let's consider the equilibrium reaction in more detail. Acetic acid's acid dissociation reaction and equilibrium constant expression are

$$CH_3COOH(aq) + H_2O(l) \rightleftharpoons H_3O^+(aq) + CH_3COO^-(aq)$$

$$K_a = \frac{[CH_3COO^-]\,[H_3O^+]}{[CH_3COOH]} = 1.75 \times 10^{-5}$$

First, let's take the logarithm of each term in this equation and multiply through by –1

$$-\log K_a = 4.76 = -\log [H_3O^+] - \log \frac{[CH_3COO^-]}{[CH_3COOH]}$$

Now, let's replace $-\log[H_3O^+]$ with pH and rearrange the equation to obtain the result shown here.

$$\text{pH} = 4.76 + \log \frac{[CH_3COO^-]}{[CH_3COOH]} \tag{7.7.1}$$

Equation 7.7.1 tells us a great deal about the relationship between pH and the relative amounts of acetic acid and acetate at equilibrium. If the concentrations of CH_3COOH and CH_3COO^- are equal, then Equation 7.7.1 reduces to

$$\text{pH} = 4.76 + \log(1) = 4.76 + 0 = 4.76$$

If the concentration of CH_3COO^- is greater than that of CH_3COOH, then the log term in Equation 7.7.1 is positive and the pH is greater than 4.76. This is a reasonable result because we expect the concentration of the conjugate base, CH_3COO^-, to increase as the pH increases. Similar reasoning will convince you that the pH is less than 4.76 when the concentration of CH_3COOH exceeds that of CH_3COO^-.

Now we are ready to construct acetic acid's ladder diagram (Figure 6.6.1). First, we draw a vertical arrow that represents the solution's pH, with smaller (more acidic) pH levels at the bottom and larger (more basic) pH levels at the top. Second, we draw a horizontal line at a pH equal to acetic acid's pK_a value. This line, or step on the ladder, divides the pH axis into regions where either CH_3COOH or CH_3COO^- is the predominate species. This completes the ladder diagram.

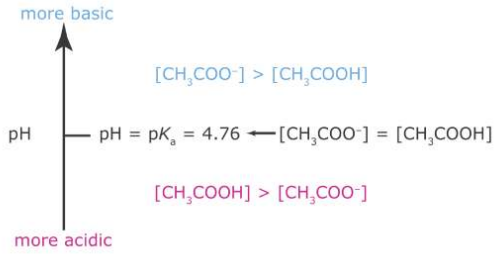


Figure 6.6.1 . Acid–base ladder diagram for acetic acid showing the relative concentrations of CH_3COOH and CH_3COO^-. A simpler version of this ladder diagram dispenses with the equalities and inequalities and shows only the predominate species in each region.

Using the ladder diagram, it is easy to identify the predominate form of acetic acid at any pH. At a pH of 3.5, for example, acetic acid exists primarily as CH_3COOH. If we add sufficient base to the solution such that the pH increases to 6.5, the predominate form of acetic acid is CH_3COO^-.

Example 6.6.1

Draw a ladder diagram for the weak base *p*-nitrophenolate and identify its predominate form at a pH of 6.00.

Solution

To draw a ladder diagram for a weak base, we simply draw the ladder diagram for its conjugate weak acid. From Appendix 11, the pK_a for *p*-nitrophenol is 7.15. The resulting ladder diagram is shown in Figure 6.6.2 . At a pH of 6.00, *p*-nitrophenolate is present primarily in its weak acid form.

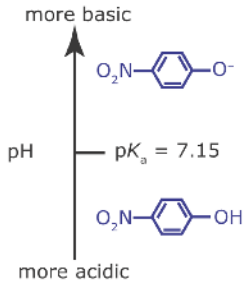


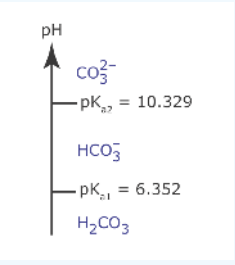
Figure 6.6.2 . Acid–base ladder diagram for *p*-nitrophenolate.

Exercise 6.6.1

Draw a ladder diagram for carbonic acid, H_2CO_3. Because H_2CO_3 is a diprotic weak acid, your ladder diagram will have two steps. What is the predominate form of carbonic acid when the pH is 7.00? Relevant equilibrium constants are in Appendix 11.

Answer

From Appendix 11, the pK_a values for H_2CO_3 are 6.352 and 10.329. The ladder diagram for H_2CO_3 is shown below. The predominate form at a pH of 7.00 is HCO_3^-.



A ladder diagram is particularly useful for evaluating the reactivity between a weak acid and a weak base. Figure 6.6.3 , for example, shows a single ladder diagram for acetic acid/acetate and for *p*-nitrophenol/*p*-nitrophenolate. An acid and a base can not co-exist if their respective areas of predominance do not overlap. If we mix together solutions of acetic acid and sodium *p*-nitrophenolate, the reaction

$$C_6H_4NO_2^-(aq) + CH_3COOH(aq) \rightleftharpoons CH_3COO^-(aq) + C_6H_4NO_2H(aq) \tag{7.7.2}$$

occurs because the areas of predominance for acetic acid and *p*-nitrophenolate do not overlap. The solution's final composition depends on which species is the limiting reagent. The following example shows how we can use the ladder diagram in Figure 6.6.3 to evaluate the result of mixing together solutions of acetic acid and *p*-nitrophenolate.

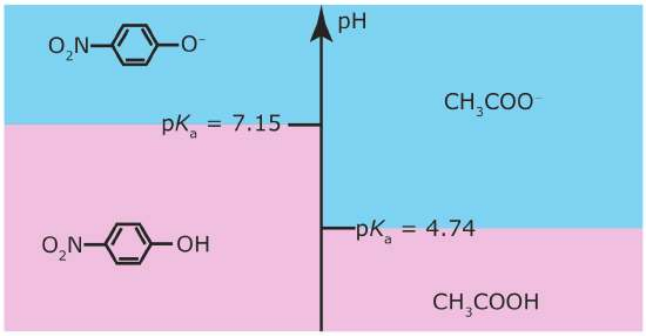


Figure 6.6.3 . Acid–base ladder diagram showing the areas of predominance for acetic acid/acetate and for *p*-nitrophenol/*p*-nitrophenolate. The areas shaded in blueshows the pH range where the weak bases are the predominate species; the weak acid forms are the predominate species in the areas shaded in pink.

Example 6.6.2

Predict the approximate pH and the final composition after mixing together 0.090 moles of acetic acid and 0.040 moles of *p*-nitrophenolate.

Solution

The ladder diagram in Figure 6.6.3 indicates that the reaction between acetic acid and *p*-nitrophenolate is favorable. Because acetic acid is in excess, we assume the reaction of *p*-nitrophenolate to *p*-nitrophenol is complete. At equilibrium essentially no *p*-nitrophenolate remains and there are 0.040 mol of *p*-nitrophenol. Converting *p*-nitrophenolate to *p*-nitrophenol consumes 0.040 moles of acetic acid; thus

$$\text{moles } CH_3COOH = 0.090 - 0.040 = 0.050 \text{ mol}$$
$$\text{moles } CH_3COO^- = 0.040 \text{ mol}$$

According to the ladder diagram, the pH is 4.74 when there are equal amounts of CH_3COOH and CH_3COO^-. Because we have slightly more CH_3COOH than CH_3COO^-, the pH is slightly less than 4.74.

Exercise 6.6.2

Using Figure 6.6.3 , predict the approximate pH and the composition of the solution formed by mixing together 0.090 moles of *p*-nitrophenolate and 0.040 moles of acetic acid.

Answer

The ladder diagram in Figure 6.6.3 indicates that the reaction between acetic acid and *p*-nitrophenolate is favorable. Because *p*-nitrophenolate is in excess, we assume the reaction of acetic acid to acetate is complete. At equilibrium essentially no acetic acid remains and there are 0.040 moles of acetate. Converting acetic acid to acetate consumes 0.040 moles of *p*-nitrophenolate; thus

$$\text{moles } p\text{-nitrophenolate } = 0.090 - 0.040 = 0.050 \text{ mol}$$

$$\text{moles } p\text{-nitrophenol } = 0.040 \text{ mol}$$

According to the ladder diagram for this system, the pH is 7.15 when there are equal concentrations of *p*-nitrophenol and *p*-nitrophenolate. Because we have slightly more *p*-nitrophenolate than we have *p*-nitrophenol, the pH is slightly greater than 7.15.

If the areas of predominance for an acid and a base overlap, then we do not expect that much of a reaction will occur. For example, if we mix together solutions of CH_3COO^- and *p*-nitrophenol, we do not expect a significant change in the moles of either reagent. Furthermore, the pH of the mixture must be between 4.76 and 7.15, with the exact pH depending upon the relative amounts of CH_3COO^- and *p*-nitrophenol.

We also can use an acid–base ladder diagram to evaluate the effect of pH on other equilibria. For example, the solubility of CaF_2

$$CaF_2(s) \rightleftharpoons Ca^{2+}(aq) + 2F^-(aq)$$

is affected by pH because F^- is a weak base. From Le Châtelier's principle, we know that converting F^- to HF will increase the solubility of CaF_2. To minimize the solubility of CaF_2 we need to maintain the solution's pH so that F^- is the predominate species. The ladder diagram for HF (Figure 6.6.4) shows us that maintaining a pH of more than 3.17 will minimize solubility losses.

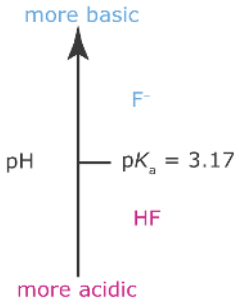


Figure 6.6.4 . Acid–base ladder diagram for HF. To minimize the solubility of CaF_2, we need to keep the pH above 3.17, with more basic pH levels leading to smaller solubility losses. See Chapter 8 for a more detailed discussion.

Ladder Diagrams for Complexation Equilibria

We can apply the same principles for constructing and interpreting an acid–base ladder diagram to equilibria that involve metal–ligand complexes. For a complexation reaction we define the ladder diagram's scale using the concentration of uncomplexed, or free ligand, pL. Using the formation of $Cd(NH_3)^{2+}$ as an example

$$Cd^{2+}(aq) + NH_3(aq) \rightleftharpoons Cd\,(NH_3)^{2+}\,(aq)$$

we can show that log K_1 is the dividing line between the areas of predominance for Cd^{2+} and for $Cd(NH_3)^{2+}$.

$$K_1 = 3.55 \times 10^2 = \frac{\left[Cd\,(NH_3)^{2+}\right]}{\left[Cd^{2+}\right]\left[NH_3\right]}$$

$$\log K_1 = \log\left(3.55 \times 10^2\right) = \log \frac{\left[Cd\,(NH_3)^{2+}\right]}{\left[Cd^{2+}\right]} - \log\left[NH_3\right]$$

$$\log K_1 = 2.55 = \log \frac{\left[Cd\,(NH_3)^{2+}\right]}{\left[Cd^{2+}\right]} + pNH_3$$

$$pNH_3 = \log K_1 + \log \frac{\left[Cd^{2+}\right]}{\left[Cd\,(NH_3)^{2+}\right]} = 2.55 + \log \frac{\left[Cd^{2+}\right]}{\left[Cd\,(NH_3)^{2+}\right]}$$

Thus, Cd^{2+} is the predominate species when pNH_3 is greater than 2.55 (a concentration of NH_3 smaller than 2.82×10^{-3} M) and for a pNH_3 value less than 2.55, $Cd(NH_3)^{2+}$ is the predominate species. Figure 6.6.5 shows a complete metal–ligand ladder diagram for Cd^{2+} and NH_3 that includes additional Cd–NH_3 complexes.

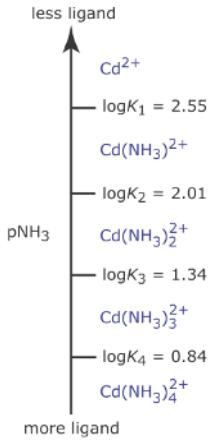


Figure 6.6.5 . Metal–ligand ladder diagram for Cd^{2+}–NH_3 complexation reactions. Note that higher-order complexes form when pNH_3 is smaller (which corresponds to larger concentrations of NH_3).

Example 6.6.3

Draw a single ladder diagram for the $Ca(EDTA)^{2-}$ and the $Mg(EDTA)^{2-}$ metal–ligand complexes. Use your ladder diagram to predict the result of adding 0.080 moles of Ca^{2+} to 0.060 moles of $Mg(EDTA)^{2-}$. EDTA is an abbreviation for the ligand ethylenediaminetetraacetic acid.

Solution

Figure 6.6.6 shows the ladder diagram for this system of metal–ligand complexes. Because the predominance regions for Ca^{2+} and $Mg(EDTA)^{2-}$ do not overlap, the reaction

$$Ca^{2+}(aq) + Mg(EDTA)^{2-}(aq) \rightleftharpoons Ca(EDTA)^{2-}(aq) + Mg^{2+}(aq)$$

proceeds essentially to completion. Because Ca^{2+} is the excess reagent, the composition of the final solution is approximately

$$\text{moles } Ca^{2+} = 0.080 - 0.060 = 0.020 \text{ mol}$$

$$\text{moles } Ca(EDTA)^{2-} = 0.060 \text{ mol}$$

$$\text{moles } Mg^{2+} = 0.060 \text{ mol}$$

$$\text{moles } Mg(EDTA)^{2-} = 0 \text{ mol}$$

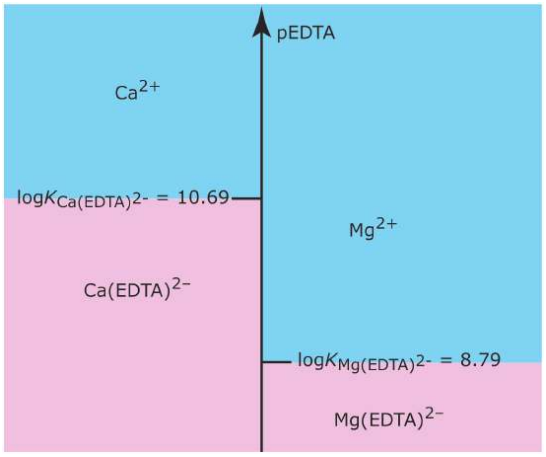


Figure 6.6.6 . Metal–ligand ladder diagram for $Ca(EDTA)^{2-}$ and for $Mg(EDTA)^{2-}$. The areas shaded in blueshow the pEDTA range where the free metal ions are the predominate species; the metal–ligand complexes are the predominate species in the areas shaded in pink.

The metal–ligand ladder diagram in Figure 6.6.5 uses stepwise formation constants. We also can construct a ladder diagram using cumulative formation constants. For example, the first three stepwise formation constants for the reaction of Zn^{2+} with NH_3

$$Zn^{2+}(aq) + NH_3(aq) \rightleftharpoons Zn(NH_3)^{2+}(aq) \quad K_1 = 1.6 \times 10^2$$

$$Zn(NH_3)^{2+}(aq) + NH_3(aq) \rightleftharpoons Zn(NH_3)_2^{2+}(aq) \quad K_2 = 1.95 \times 10^2$$

$$Zn(NH_3)_2^{2+}(aq) + NH_3(aq) = Zn(NH_3)_3^{2+}(aq) \quad K_3 = 2.3 \times 10^2$$

suggests that the formation of $Zn(NH_3)_3^{2+}$ is more favorable than the formation of $Zn(NH_3)^{2+}$ or $Zn(NH_3)_2^{2+}$. For this reason, the equilibrium is best represented by the cumulative formation reaction shown here.

$$Zn^{2+}(aq) + 3NH_3(aq) \rightleftharpoons Zn(NH_3)_3^{2+}(aq) \quad \beta_3 = 7.2 \times 10^6$$

> Because K_3 is greater than K_2, which is greater than K_1, the formation of the metal-ligand complex $Zn(NH_3)_3^{2+}$ is more favorable than the formation of the other metal ligand complexes. For this reason, at lower values of pNH_3 the concentration of $Zn(NH_3)_3^{2+}$ is larger than the concentrations of $Zn(NH_3)^{2+}$ or $Zn(NH_3)_2^{2+}$. The value of β_3 is
>
> $$\beta_3 = K_1 \times K_2 \times K_3$$

To see how we incorporate this cumulative formation constant into a ladder diagram, we begin with the reaction's equilibrium constant expression.

$$\beta_3 = \frac{\left[Zn(NH_3)_3^{2+}\right]}{\left[Zn^{2+}\right]\left[NH_3\right]^3}$$

Taking the log of each side

$$\log \beta_3 = \log \frac{\left[Zn(NH_3)_3^{2+}\right]}{\left[Zn^{2+}\right]} - 3 \log \left[NH_3\right]$$

and rearranging gives

$$pNH_3 = \frac{1}{3}\log \beta_3 + \frac{1}{3}\log \frac{\left[Zn^{2+}\right]}{\left[Zn(NH_3)_3^{2+}\right]}$$

When the concentrations of Zn and $Zn(NH_3)_3^{2+}$ are equal, then

$$pNH_3 = \frac{1}{3}\log \beta_3 = 2.29$$

In general, for the metal–ligand complex ML_n, the step for a cumulative formation constant is

$$pL = \frac{1}{n}\log \beta_n$$

Figure 6.6.7 shows the complete ladder diagram for the Zn^{2+}–NH_3 system.

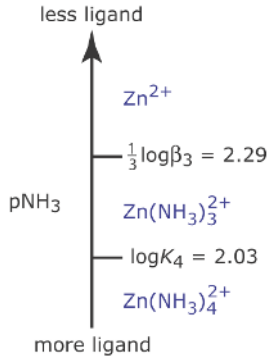


Figure 6.6.7 . Ladder diagram for Zn^{2+}–NH_3 metal–ligand complexation reactions showing both a step based on a cumulative formation constant and a step based on a stepwise formation constant.

Ladder Diagrams for Oxidation/Reduction Equilibria

We also can construct ladder diagrams to help us evaluate redox equilibria. Figure 6.6.8 shows a typical ladder diagram for two half-reactions in which the scale is the potential, *E*.

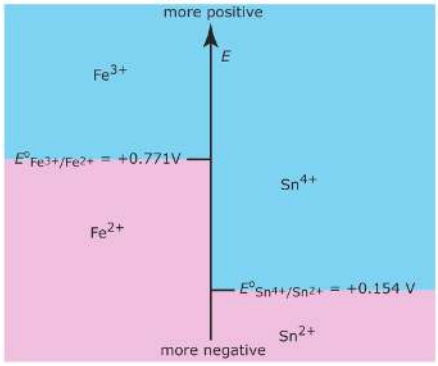


Figure 6.6.8 . Redox ladder diagram for Fe^{3+}/Fe^{2+} and for Sn^{4+}/Sn^{2+}. The areas shaded in blueshow the potential range where the oxidized forms are the predominate species; the reduced forms are the predominate species in the areas shaded in pink. Note that a more positive potential favors the oxidized form.

The Nernst equation defines the areas of predominance. Using the Fe^{3+}/Fe^{2+} half-reaction as an example, we write

$$E = E^\circ - \frac{RT}{nF}\ln\frac{[Fe^{2+}]}{[Fe^{3+}]} = 0.771 - 0.05916\log\frac{[Fe^{2+}]}{[Fe^{3+}]}$$

At a potential more positive than the standard state potential, the predominate species is Fe^{3+}, whereas Fe^{2+} predominates at potentials more negative than E^o. When coupled with the step for the Sn^{4+}/Sn^{2+} half-reaction we see that Sn^{2+} is a useful reducing agent for Fe^{3+}. If Sn^{2+} is in excess, the potential of the resulting solution is near +0.154 V.

Because the steps on a redox ladder diagram are standard state potentials, a complication arises if solutes other than the oxidizing agent and reducing agent are present at non-standard state concentrations. For example, the potential for the half-reaction

$$UO_2^{2+}(aq) + 4H_3O^+(aq) + 2e^- \rightleftharpoons U^{4+}(aq) + 6H_2O(l)$$

depends on the solution's pH. To define areas of predominance in this case we begin with the Nernst equation

$$E = +0.327 - \frac{0.05916}{2}\log\frac{[U^{4+}]}{[UO_2^{2+}]\,[H_3O^+]^4}$$

and factor out the concentration of H_3O^+.

$$E = +0.327 + \frac{0.05916}{2}\log[H_3O^+]^4 - \frac{0.05916}{2}\log\frac{[U^{4+}]}{[UO_2^{2+}]}$$

From this equation we see that the area of predominance for UO_2^{2+} and U^{4+} is defined by a step at a potential where $[U^{4+}] = [UO_2^{2+}]$.

$$E = +0.327 + \frac{0.05916}{2}\log\left[H_3O^+\right]^4 = +0.327 - 0.1183\text{pH}$$

Figure 6.6.9 shows how pH affects the step for the UO_2^{2+} /U^{4+} half-reaction.

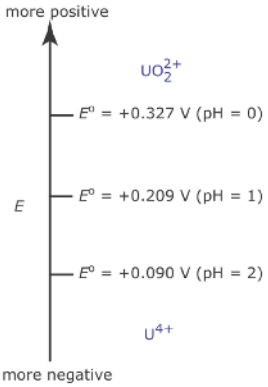


Figure 6.6.9 . Redox ladder diagram for the UO_2^{2+}/U^{4+} half-reaction showing the effect of pH on the step defined by the standard state's potential.

7.7: The redox ladder is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **6.6: Ladder Diagrams** by David Harvey is licensed CC BY-NC-SA 4.0.

7.8: Wetlands and water quality

Wetlands are valuable to us because they greatly influence the flow and quality of water. They help **improve water quality, including that of drinking water, by intercepting surface runoff and removing or retaining inorganic nutrients, processing organic wastes, and reducing suspended sediments** before they reach open water. For example, as the runoff water passes through wetlands, they retain or process excess nitrogen and phosphorus, decompose organic pollutants, and trap suspended sediments that would otherwise clog waterways and affect fish and amphibian egg development.

In performing this filtering function, wetlands save us a great deal of money. A 1990 study showed that, the Congaree Bottomland Hardwood Swamp in South Carolina, removes a quantity of pollutants that would be equivalent to that removed annually by a $5 million waste water treatment plant. Another study at a 2,500 acre wetland in Georgia, indicated that it saves $1 million in water pollution abatement costs annually. **Wetlands also reduce environmental problems, such as algal blooms, dead zones, and fish kills, that are generally associated with excess nutrient loadings**. However, the capacity of wetlands to function this way is not unlimited, and too much surface runoff carrying sediments, nutrients, and other pollutants can degrade wetlands and thus the societal services they provide.

In addition to improving water quality through filtering, **some wetlands maintain stream flow during dry periods; others replenish groundwater**. Many Americans, of course, depend on groundwater for drinking. The Floridian aquifer system, for instance, is one of the more productive ground water sources in the United States. It occurs across the entire state of Florida, and into southern Georgia, and portions of South Carolina and Alabama. This huge subsurface reservoir produces some of the cleanest water in the nation. Its primary source is rainwater that filters through hundreds of feet of sand and rock. One calculation for 5-acre Florida cypress swamp recharging groundwater was that, if 80 percent of swamp was drained, available ground water would be reduced by an estimated 45 percent.

Flood Protection

Because of their low topographic position relative to uplands (e.g., isolated depressions, floodplains), **wetlands store and slowly release surface water, rain, snowmelt, groundwater and flood waters.** Trees and other **wetland vegetation also impede the movement of flood waters and distribute them more slowly over floodplains**. This combined water storage and slowing action **lowers flood heights and reduces erosion downstream and on adjacent lands**. It also helps **reduce floods and prevents waterlogging of agricultural lands**. Wetlands within and downstream of urban areas are particularly valuable in this regard, counteracting the greatly increased rate and volume of surface-water runoff from pavement and buildings.

Preserving and restoring wetlands, together with other water retention, can often provide the level of flood protection otherwise provided by expensive dredging operations and levees. The preservation of wetlands also results in many other benefits to society, such as the protection of ecologically significant fish and wildlife habitat. A good example of this is the Mississippi River's bottomland hardwood-riparian wetlands, which once stored at least 60 days of floodwater and represented significant fish and wildlife habitat. They now store only 12 days of floodwater because most have been filled, leveed, or drained, with substantial loss of fish and wildlife habitat. Another good example is Minnesota, where the cost of replacing the natural flood control function of 5000 acres of drained wetlands was found to be $1.5 million annually.

To quote Henry Wessman, the mayor of Grand Forks, ND: "The total cost of flood damage is born by taxpayers again and again as the flood waters come. I offer as a suggestion to compensate farmers within the area to actually retain natural wetlands. If you look at the costs of compensating farmers for such activities as opposed to the almost annual cost of flood protection and flood fighting within a city such as Grand Forks, you would realize that over the long haul, you are doing yourself a much greater service by retaining that water rather than by continually paying for flood damage."

Therefore, **in addition to their fish and wildlife values, wetlands reduce the likelihood of flood damage to homes, businesses, and crops in agricultural areas**. They also help **control increases in the rate and volume of runoff in urban areas**. This protection results in less monetary flood damage (and related insurance costs), as well as protection of human health, safety, and welfare.

LibreTexts

Excerpted from:

Wetland Functions and Values, Watershed Academy Web, US Environmental Protection Agency. Accessed on December 2023 https://cfpub.epa.gov/watertrain/moduleFrame.cfm?parent_object_id=262

7.8: Wetlands and water quality is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

SECTION OVERVIEW

7.9: Wetlands and global change

7.9.1: Introduction

7.9.2: Carbon Accumulation and GHG Emissions from Freshwater Wetlands (Including Permafrost) In a Changing Climate

7.9.3: Ecological Consequences for Freshwater Wetlands in a Changing Climate

7.9.4: Salt Marsh and Mangrove Response to a Changing Climate and Associated Sea Level Rise

---

7.9: Wetlands and global change is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by William R. Moomaw, G. L. Chmura & N. Roulet, Gillian T. Davies, C. M. Finlayson, B. A. Middleton, Susan M. Natali, J. E. Perry and Ariana E. Sutton-Grier.

7.9.1: Introduction

Wetlands In a Changing Climate: The Science

The United Nations Framework Convention on Climate Change calls for the “stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system” (UNFCCC 1992). The Paris Climate Agreement in 2015 (UNFCCC 2017) established a goal of keeping global average temperature increase substantially less than 2°C above the preindustrial value, and making every attempt to keep it below 1.5^{o}C.

In order to have a two in three probability of keeping global average temperature from rising by more than 2°C, it is essential to have “negative emissions" of GHGs; in other words, meeting the goal of the Paris Climate Agreement requires active removal and sequestration of atmospheric C (Sanderson et al. 2016). Sequestration is used here to refer to the photosynthetic removal of CO_2 from the atmosphere and its conversion into cellulose and other carbon compounds in plants, and its conversion from decaying plants into soil organic matter. Ricke and Caldeira (2014) have shown that peak warming occurs within about one decade after a pulse of CO_2 is added to the atmosphere. Hence the benefits of avoided CO_2 emissions will be manifested within the lifetimes of people who acted to avoid those emissions. Solomon et al. (2009) have shown that after peak warming is reached, effects will persist for 1000 years. IPCC estimates that depending upon the scenario, “about 15 to 40% of CO_2 emitted by 2100 will remain in the atmosphere longer than 1000 years” (Ciais et al. 2013) affecting 40 generations. Hence avoiding emissions of GHGs to the atmosphere is recommended to be a prime consideration that benefits both present and future generations.

For most types of wetlands, the bulk of sequestered carbon is in the soils rather than in the plant communities. Draining these wetlands to convert them to agriculture as has been done in many countries and regions including Indonesia, Malaysia, Russia, New Zealand, Florida Everglades and in Northern Europe, allows soil organic matter to be oxidized and release CO_2 into the atmosphere. When mangroves are removed for coastal development and for aquaculture, or forested wetlands are harvested, additional carbon is released from soils and harvest residues. In the Southeast United States, a major wood pellet fuel industry has developed where the carbon in the wood is released as CO_2 immediately upon combustion. The use of wood pellets to replace coal for electricity, on the mistaken assumption that it is carbon neutral, is expected to grow substantially by 2050 (IEA 2017), further degrading forested wetlands while adding large amounts of CO_2 to the atmosphere.

CO_2, added to the atmosphere by human activity, is the primary GHG responsible for climate change, followed by CH_4 and N_2O (Myhre et al 2014). These gases move among the natural reservoirs of terrestrial and marine plants, soils, oceans and the atmosphere. Human activity has reduced the size and capacity of these reservoirs while increasing GHG emissions (Ciais et al. 2013). Altering albedo (solar reflectivity from the earth’s surfaces) from land use change can increase or decrease global warming. Climate forcing (heat trapping) from black C (particulate matter from fossil fuel and biofuels combustion (Bond et al. 2013) is a significant contributor to global warming.

The average annual anthropogenic CO_2 emissions for the period 2006-2015 are estimated to be 10.3 $PgCy^{-1}$ (Petagrams C per year or 10^{15} grams C per year) with 9.3±0.5 $PgCy^{-1}$ from fossil fuels and industrial processes and 1.0±0.5 $PgCy^{-1}$ from land use change (Fig. 7.9.1.1, Le Quéré et al. 2016). The total CO_2 emissions from fossil fuels and industrial processes between 1750 and 2011 are estimated to be 375±30 PgC, and the total amount from land use change is estimated to be 180±80 PgC. Therefore, nearly one-third of CO_2 added to the atmosphere from human activity has come from deforestation and oxidation of disturbed soil organic matter (Ciais et al. 2013). By November 2017, CO_2 in the atmosphere had increased to 865 PgC or 406 ppm (NOAA 2018).

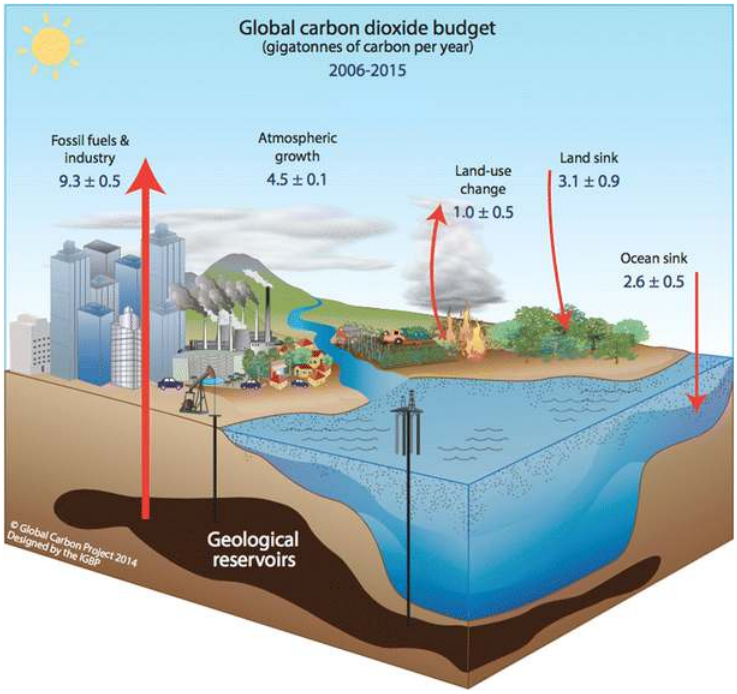


Figure 7.9.1.1: (Le Quéré et al. 2016). Note that approximately 10% of annual emissions are from land use change, and that the land sink removes an amount equal to about 33% of annual emissions. This value is calculated by difference, and wetlands are not counted separately from the total land sink. 1 Gigatonne (10^9 tonnes) equals 1 Petagram (10^{15}) grams. Le Quéré et al. 2016

The net annual increase of CO_2 in the atmosphere each year is 4.5±0.5 $PgCy^{-1}$ or slightly less than half of annual emissions, and concentrations have increased by over 40% above preindustrial levels. The biosphere has been the major means for removing and sequestering atmospheric CO_2 for over 300 million years, but its potential to be a major resource for addressing climate change has been underappreciated in current policy discussions. Each year, 2.6±0.5 PgC equal to about 25% of annual emissions is removed by the ocean's phytoplankton or is dissolved in the ocean's waters. The difference between total emissions to the atmosphere and net removals by the oceans requires that an additional amount of CO_2 equivalent to 3.1±0.9 $PgCy^{-1}$ would need to be removed by terrestrial ecosystems to balance the carbon flows. This is nearly 30% of annual anthropogenic emissions from all sources. This analysis only reports estimates of the aggregate removal of CO_2 by the terrestrial biosphere (all plants and soils), and does not explicitly consider the specific role of wetlands as either a source or a sink (Fig. 7.9.1.1, Le Quéré et al. 2016).

CH_4 has a 100-year Global Warming Potential more than 28 times that of CO_2 (Myhre et al. 2013). It is estimated that between 1750 and 2011 human activity has increased atmospheric CH_4 by a factor of 2.5 from 1984 to 4954 Tg CH_4 y^{-1} (722 ppb to1803 ppb) (1 Teragram CH_4 is 10^{12} grams CH_4) (Ciais et al. 2013). Currently the major sources of emissions arise from fossil fuel usage (85-105 Tg CH_4 y^{-1}), ruminant livestock (87-94 Tg CH_4 y^{-1}), landfills and waste (67-90 Tg CH_4 y^{-1}), and rice production (33-40 Tg CH_4 y^{-1}). Average annual anthropogenic emissions of CH_4 from all these sources between 2000 and 2009 total between 272-329 Tg CH_4 y^{-1} CH_4 is removed from the atmosphere at a rate of 492-785 Tg CH_4 y^{-1} mostly by atmospheric chemistry with small contributions from soil oxidation (Fig. 7.9.1.2) (Ciais et al. 2013). CH_4 emissions from wetlands are between 177 and 284 Tg CH_4 y^{-1}, with an additional 8-73 Tg CH_4 y^{-1} emitted from freshwater sources.

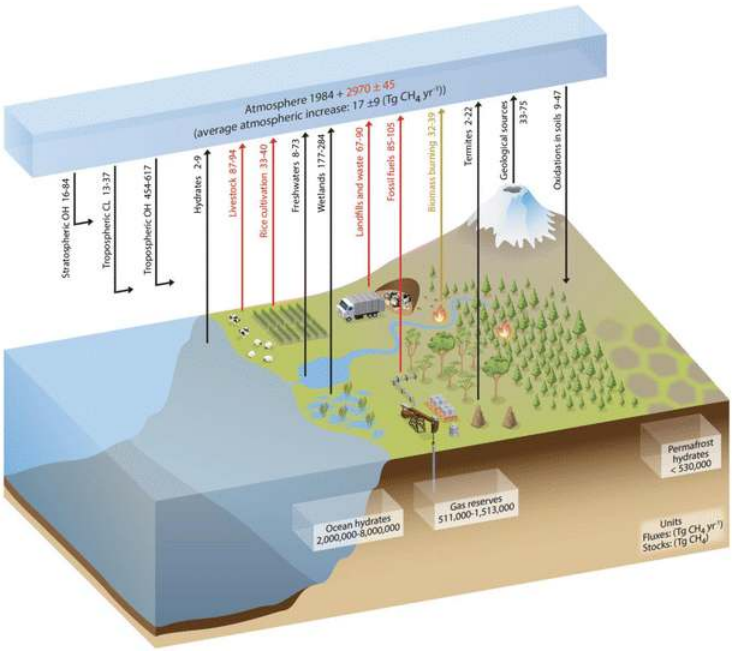


Figure 7.9.1.2: **Global methane (CH_4) budget.** Average annual anthropogenic emissions of methane from fossil fuels, ruminants, landfills and waste and rice cultivation between 2000 and 2009 total between 272-329 Tgy^{-1}. For comparison, natural wetlands and freshwaters are estimated to release between 185-357 Tgy^{-1}. CH_4 is removed at a rate of 492-785 Tgy^{-1} mostly by atmospheric chemistry with small contributions by soil oxidation. Black arrows represent natural emissions, red arrows represent anthropogenic emissions since 1750, and the brown arrow represents emissions from both natural fires and anthropogenic biomass burning. 1 Teragram equals10^{12} grams. Ciais et al. 2013

Nitrous oxide (N_2O) has a radiative forcing ~300 times that of CO_2. It is a byproduct of both nitrification (under aerobic conditions) and denitrification (under anaerobic conditions), and thus can be produced in wetland soils (Megonigal et al. 2004). However, freshwater and saltwater wetland soils are a source of N_2O only if they receive excessive levels of reactive nitrogen – otherwise they may be a sink for this potent GHG (e.g. Auget et al 2014, Chmura et al. 2016).

While natural solutions have focused on the role of forests to remove and sequester CO_2, there is substantially more C sequestered in soils than in vegetation. The range of estimates for carbon sequestered in vegetation is 450-650 PgC, while the estimate for C stored in soils is 1500-2400 PgC with an additional 1700 PgC estimated to be in permafrost (Ciais et al. 2013). The large amount of carbon sequestered in wetlands is discussed in subsequent sections. As soils warm, and as permafrost thaws, these soils release their stored C as CO_2 or CH_4 resulting from microbial decomposition of soil organic carbon (SOC). These feedback emissions trap additional heat and warm the planet further. A first priority is to avoid disturbing wetlands and keep temperatures from rising as much as possible. As the subsequent sections illustrate, wetlands can play a significant role in addressing climate change by sequestering C, and by providing climate resiliency and adaptation while providing additional ecosystem services.

To limit excessive warming, it is necessary to stabilize CO_2, CH_4, N_2O and other GHG concentrations in the atmosphere at an appropriate level, by decreasing emission rates and increasing removal rates. There are three basic strategies for accomplishing this goal.

1. Reduce the addition of GHGs into the atmosphere from fossil fuels, biofuels, industry, agriculture and other sources to near zero.
2. Prevent the climate and land-use mediated release of additional GHGs (CO_2, CH_4, N_2O) from wetlands, including wetlands underlain by permafrost, from deforestation and forest degradation, and from all soils including degraded grassland and agricultural soils.
3. Increase the capacity of natural systems including wetlands to actively remove CO_2 from the atmosphere and sequester the C for the long-term.

References

Bond TC, Doherty SJ, Fahey DW, Forster PM, Berntsen T, DeAngelo BJ, Flanner MG, Ghan S, Kärcher B, Koch D, Kinne S, Kondo Y, Quinn PK, Sarofim MC, Schultz MG, Schulz M, Venkataraman C, Zhang H, Zhang S, Bellouin N, Guttikunda SK, Hopke PK, Jacobson MZ, Kaiser JW, Klimont Z, Lohmann U, Schwarz JP, Shindell D, Storelvmo T, Warren SG, Zender CS (2013) Bounding the role of black carbon in the climate system: A scientific assessment. Journal of Geophysical Research – Atmospheres 118:2169–8996. https://doi.org/10.1002/jgrd.50171

Chmura GL, Kellman L, van Ardenne L, Guntenspergen GR (2016) Greenhouse gas fluxes from salt marshes exposed to chronic nutrient enrichment. PLOS One 11(2):e0149937. https://doi.org/10.1371/journal.pone.0149937

Ciais P, Sabine C, Bala G, Bopp L, Brovkin V, Canadell J, Chhabra A, DeFries R, Galloway J, Heimann M, Jones C, Le Quéré C, Myneni RB, Piao S, Thornton P (2013) Carbon and other biogeochemical cycles, Chapter 6. In: Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: the physical science basis. Contribution of Working Group I to the fifth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

IEA (2017) Technology Roadmap:Delivering Siustainable Bioenergy. International Energy Agency https://www.iea.org/.../IEA_Bioenergy_Roadmap_FoldOut_2ndEdition_WEB.pdf

Le Quéré C, Andrew RM, Canadell JG, Sitch S, Korsbakken JI, Peters GP, Manning AC, Boden TA, Tans PP, Houghton RA, Keeling RF, Alin S, Andrews OD, Anthoni P, Barbero L, Bopp L, Chevallier F, Chini LP, Ciais P, Currie K, Delire C, Doney SC, Friedlingstein P, Gkritzalis T, Harris I, Hauck J, Haverd V, Hoppema M, Klein Goldewijk K, Jain AK, Kato E, Körtzinger A, Landschützer P, Lefévre N, Lenton A, Lienert S, Lombardozzi D, Melton JR, Metzl N, Millero F, Monteiro PMS, Munro DR, Nabel JEMS, Nakaoka S, O'Brien K, Olsen A, Omar AM, Ono T, Pierrot D, Poulter B, Rödenbeck C, Salisbury J, Schuster U, Schwinger J, Séférian R, Skjelvan I, Stocker BD, Sutton AJ, Takahashi T, Tian H, Tilbrook B, van der Laan-Luijkx IT, van der Werf GR, Viovy N, Walker AP, Wiltshire A, Zaehle S (2016) Global carbon budget 2016. Earth System Science Data 8:605–649. https://doi.org/10.5194/essd-8-605-2016 http://www.earth-syst-sci-data.net/8/605/2016/essd-8-605-2016.pdf. Accessed 1 Sept 2017

Megonigal JP, Hines ME, Visscher PT (2004) Anaerobic metabolism: Linkages to trace gases and aerobic processes. In: Schlesinger WH (ed) Biogeochemistry. Elsevier-Pergamon, Oxford, pp 317–424

Myhre G, Shindell D, Breon FM, Collins W, Fugelestvedt J, Huang J, Lamarque J-F, Lee D, Mendoza B, Nakajima T, Robock A (2013) Anthropogenic and natural radiative forcing. In: Stocker TF, Qin D, Plattner G-K et al (eds) Climate change 2013: The physical science basis. Contributions of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University, Cambridge

NOAA 2018 National Oceanic and Atmospheric Administration Earth Monitoring and Research Laboratory Global Monitoring Division Trends in atmospheric carbon dioxide Recent global monthly mean CO_2 https://www.esrl.noaa.gov/gmd/ccgg/trends/global.html Accessed 26 Feb 2018

Ricke KL, Caldeira K (2014) Maximum warming occurs about one decade after a carbon dioxide emission. Environmental Research Letters 9:124002. https://doi.org/10.1088/1748-9326/9/12/124002

Sanderson BM, O'Neill BC, Tebaldi C (2016) What would it take to achieve the Paris temperature targets? Geophysical Research Letters 43:7133–7142. https://doi.org/10.1002/2016GL069563Return to ref 2016 in article

Solomon S, Plattnerb G-K, Knuttic R, Friedlingstein P (2009) Irreversible climate change due to carbon dioxide emissions. PNAS 106(6):61704–61709

UNFCCC (1992) United Nations Framework Convention on Climate Change. United Nations. 1992. FCCC/INFORMAL/84. GE.05-62220 (E) 200705 http://unfccc.int/essential_background/convention/items/6036.php Accessed 15 Oct 2017

UNFCCC (2017) United Nations Framework Convention on Climate Change, Paris Agreement-Status of Ratification http://unfccc.int/paris_agreement/items/9444.php Accessed 15 Oct 2017

LibreTexts

Excerpted from:

Moomaw, W.R., Chmura, G.L., Davies, G.T. *et al.* Wetlands In a Changing Climate: Science, Policy and Management. *Wetlands* **38**, 183–205 (2018). Accessed December 2023 https://doi.org/10.1007/s13157-018-1023-8 CC-BY

7.9.1: Introduction is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

7.9.2: Carbon Accumulation and GHG Emissions from Freshwater Wetlands (Including Permafrost) In a Changing Climate

Wetland conservation has important implications for atmospheric C cycles, since a substantial portion of the soil C pool is stored in wetlands. Northern high latitude and tropical peatlands store more than 600 PgC (Gorham 1991; Hugelius et al. 2014), which is among the largest reserves in the world (Köchy et al. 2015). This amount is more than two-thirds as much as is stored in the atmosphere and comparable to the amount stored in global forest biomass (Pan et al. 2011). Wetland conditions are critical for C accumulation and storage since decomposition in these systems is limited by a lack of oxygen due to water saturation (Brinson et al. 1981). Therefore, when plant productivity exceeds decomposition there is a net accumulation of soil C. This process eventually develops deep peat deposits, which may accumulate for thousands of years. In high latitudes of the Northern Hemisphere, the accumulation process is further intensified by the presence of permafrost, which can have contrasting effects on hydrology, leading to either wetland formation or loss (Sannel and Kuhry 2008). The negative climate feedback (i.e. net cooling effect) that results from increased plant productivity and the long-term C accumulation and storage by wetlands is, in part, offset by CH_4 emissions from freshwater wetlands (Turetsky et al. 2014). Freshwater wetlands represent the largest natural source of CH_4, releasing approximately 180 – 220 Pg CH_4 yr^{-1} (Mikaloff Fletcher et al. 2004, Kirschke et al. 2013). However, wetlands that accumulate peat account for less than a quarter of all wetland CH_4 emissions (Turetsky et al. 2014 and references therein).

The influence of future climate on wetland soils C will depend upon the same factors that facilitated C accumulation in these systems: water saturated soils and minimal modification of wetlands through land-use change, and in the case of high latitude peat lands, low temperature. Globally, temperature, low oxygen (due to soil saturation), and the chemical and physical form of the organic matter, are the primary factors limiting decomposition in wetlands. Changes in precipitation and evapotranspiration patterns, which alter the water balance of wetland ecosystems, will substantially influence wetland C cycling. However, the magnitude, directionality, and seasonality of projected hydrologic changes are regionally variable (Collins et al. 2013), and therefore, the fate of soil C stored in wetlands will depend on local conditions. In contrast, changes in the global energy balance, usually manifested by an increase in temperature, are most likely to accelerate the decomposition rate of wetland organic C stored at the soil surface. Deeper C pools may be unaffected unless there are associated changes in hydrology (van Groenigen et al 2016). These potential losses of below ground C may also be partially offset by increased primary productivity.

The greenhouse gas dynamics of permafrost regions differ in important ways from liquid water wetlands. The microbial metabolism of soil carbon is greatly reduced when the soils are frozen for long periods. Thawing changes the availability of oxygen and liquid water, and activates bacterial metabolism, which leads to a relatively abrupt increase in emissions of either or both carbon dioxide and methane. In addition, the low solubility of methane in water causes the accumulation of this gas in bubbles under the permafrost layer. Thawing releases these bubbles, which substantially contributes to this abrupt emission increase. In permafrost regions, increased temperature will have both direct and indirect effects on wetland C storage; permafrost thaw can dramatically affect hydrology in the Arctic, but the C consequences of that change are dependent upon landscape conditions (Olefeldt et al. 2016). Permafrost thaw can lead to wetland drainage because permafrost restricts vertical water flow. As the permafrost thaws to deeper soil layers or is completely thawed, the perched water table may be lowered, resulting in drier surface soils. Permafrost-mediated wetland drainage can lead to substantial C losses because of higher rates of aerobic bacterial metabolism. However, permafrost thaw can also result in ground collapse that can cause wetland formation and substantially increase CH_4 emissions from permafrost ecosystems (Christensen et al. 2004; Natali et al. 2015; Schuur et al. 2015).

The effects of climate changes on wetland C storage will be determined largely by the extent to which the wetlands have been modified through land-use change (Petrescu. et al 2015). Altering wetlands can increase the vulnerability of the organic C pool by weakening the self-regulating feedbacks that exist in many peatland systems (Frolking et al. 2010). Land use change that affects wetland hydrology has had substantial impacts on wetland structure and function. Draining wetlands decreases CO_2 uptake and increases rates of microbial decomposition and CO_2 release (Mietten et al 2017). Soil C is also lost by peat extraction, drainage and other disturbance (Laıne et al. 2014; Evans et al. 2015; Strain, et al 2017). The hydrologic changes can be so large that they result in massive losses of C to the atmosphere, such as occurred during the fires in tropical peatlands in Southeast Asia (Page et al. 2002).

While the drainage of natural wetlands for conversion to agricultural land results in net losses of soil organic C, radiative forcing from wetland conversion depends on relative changes in the direction and magnitude of two major GHGs: CO_2 and CH_4 (Petrescu

et al. 2015). Despite a decline in CH_4 emissions following wetland drainage, wetland conversion to cropland results in a significant net increase in atmospheric radiative forcing (heat trapping) (Petrescu et al. 2015). On the other hand, land use changes that cause flooding and creation of wetlands can alter C pools through the saturation and burial of organic C (Knoll et al. 2014). Despite the potential for C sequestration, reservoir formation leads to increased GHG emissions, primarily because of CH_4 emissions from ponded water and highly fluctuating water levels in reservoirs compared to natural lakes (Deemer et al. 2016; Hayes et al. 2017).

Increased atmospheric CO_2 is projected to almost double current freshwater wetland CH_4 emissions, primarily due to warmer temperatures as well as enhanced precipitation (Shindell et al. 2004). The increase in CH_4 emissions under high CO_2 concentrations will primarily result from increased emission rates from tropical wetlands and from wetland expansion in northern high latitudes (Shindell et al. 2004; van Groenigen et al. 2011). The response of wetlands to future climate scenarios will also vary across wetland systems. For example, Wu and Roulet (2014) suggest that ombrotrophic (rain-fed) peatlands will maintain structure and function, but fen-like systems that rely on terrestrial water inputs are much more vulnerable to climate change. Land use and climate-mediated changes in CH_4 emissions from freshwater wetlands can produce a large increase in radiative forcing (heat trapping) in decades to several centuries, but in the long-term (century-millennia), C sequestration by wetlands represents, at present, a net cooling effect (Frolking and Roulet 2007; Neubauer and Megonigal 2015). However, land use, land use change, and fire can cause abrupt changes in soil C storage in wetlands, switching these long-term C sinks to sources of C to the atmosphere (Joosten et al. 2016)

References

Brinson MM, Lugo AE, Brown S (1981) Primary productivity, decomposition and consumer activity in freshwater wetlands. Annual Review of Ecology and Systematics 12:123–161

Christensen TR, Johansson T, Åkerman J, Mastepanov M, Malmer N, Friborg T, Crill P, Svensson BH (2004) Thawing sub-arctic permafrost: Effects on vegetation and methane emissions. Geophysical Research Letters 31:L04501. https://doi.org/10.1029/2003GL018680

Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichefet T, Friedlingstein P, Gao X, Gutowski WJ, Johns T, Krinner G, Shongwe M, Tebaldi C, Weaver AJ, Wehner M (2013) Long-term climate change: Projections, commitments and irreversibility. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Doschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp 1029–1136. https://doi.org/10.1017/CBO9781107415324.024

Deemer BR, Harrison JA, Li S, Beaulieu JJ, DeLSontro T, Barros N, Bezerra-Neto JF, Powers SM, Dos Santos MA, Vonk JA (2016) Greenhouse gas emissions from reservoir water surfaces: A new global synthesis. BioScience 66:949–964. https://doi.org/10.1093/biosci/biw117

Evans CD, Renou-Wilson F, Strack M (2015) Inclusion of waterborne carbon in IPCC greenhouse gas emissions accounting for drained and re-wetted peatlands: Development of the methodology and future research needs. Aquatic Sci. https://doi.org/10.1007/s00027-015-0447-y

Frolking S, Roulet NT (2007) Holocene radiative forcing impact of northern peatland carbon accumulation and methane emissions. Global Change Biology 13(5):1079–1088

Gorham E (1991) Northern peatlands: Role in the carbon cycle and probable responses to climatic warming. Ecological Applications 1(2):182–195

Hayes NM, Deemer BR, Corman JR, Razavi NR, Strock KE (2017) Key differences between lakes and reservoirs modify climate signals: A case for a new conceptual model. Limnology and Oceanography Letters 2(2):47–62

Hugelius G, Strauss J, Zubrzycki S, Harden JW, Schuur EAG, Ping CL, Schirrmeister L, Grosse G, Michaelson GJ, Koven CD, O'Donnell JA, Elberling B, Mishra U, Camill P, Yu Z, Palmtag J Kuhry P (2014) Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. Biogeosciences 11:6573–6593. https://doi.org/10.5194/bg-11-6573-2014

Joosten H, Couwenberg J, von Unger M (2016) International carbon policies as a new driver for peatland restoration. In: Bonn A, Allot T, Evans M, Joosten H, Stoneman R (eds) Peatland restoration and ecosystem services: Science, policy and practice.

Cambridge University Press/British Ecological Society, Cambridge, pp 291–313

Kirschke S, Bousquet P, Ciais P, Saunois M, Canadell JG, Dlugokencky EJ, Bergamaschi P, Bergmann D, Blake DR, Bruhwiler L, Cameron-Smith P, Castaldi S, Chevallier F, Feng L, Fraser A, Heimann M, Hodson EL, Houweling S, Josse B, Fraser PJ, Krummel PB, Lamarque J, Langenfelds RL, Le Quéré C, Naik V, O'Doherty S, Palmer PI, Pison I, Plummer D, Poulter B, Prinn RG, Rigby M, Ringeval B, Santini M, Schmidt M, Shindell DT, Simpson IJ, Spahni R, Steele LP, Strode SA, Sudo K, Szopa S, van der Werf GR, Voulgarakis A, van Weele M, Weiss RF, Williams JE, Zeng G (2013) Three decades of global methane sources and sinks. Nature Geoscience. 6:813–823. https://doi.org/10.1038/ngeo1955

Knoll LB, Vanni MJ, Renwick WJ, Kollie S (2014) Burial rates and stoichiometry of sedimentary carbon, nitrogen and phosphorus in Midwestern US reservoirs. Freshwater Biology 58(11):2342–2353. https://doi.org/10.1111/fwb.12438

Köchy M, Hiederer R, Freibauer A (2015) Global distribution of soil organic carbon – Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world. SOIL 1:351–365. https://doi.org/10.5194/soil-1-351-2015

Laıne A, Nakamura H, Nishii K, Miyasaka T (2014) A diagnostic study of future evaporation changes projected in CMIP5 climate models. Clim Dyn 42:2745–2761. https://doi.org/10.1007/s00382-014-2087-7

Mietten et al 2017

Mikaloff Fletcher SE, Tans PP, Bruhwiler LM, Miller JB, Heimann M (2004) CH_4 sources estimated from atmospheric observations of CH_4 and its $^{13}C/^{12}C$ isotopic ratios: 2. Inverse modeling of CH_4 fluxes from geographical regions. Global Biogeochemical Cycles 18, GB4005. https://doi.org/10.1029/2004GB002224

Natali SM, Schuur EAG, Mauritz M, Schade JD, Celis G, Crummer KG, Johnston C, Krapek J, Pegoraro E, Salmon VG, Webb EE (2015) Permafrost thaw and soil moisture driving CO_2 and CH_4 release from upland tundra. J Geophys Res Biogeosci 120:525–537. https://doi.org/10.1002/2014JG002872

Neubauer SC, Megonigal JP (2015) Moving beyond global warming potentials to quantify the climatic role of ecosystems. Ecosystems 18:1000–1013. https://doi.org/10.1007/s10021-015-9879-4

Pan Y, Birdsey RA, Fang J, Houghton R, Kauppi PE, Kurz WA, Phillips OL, Shvidenko A, Lewis SL, Canadell JG, Ciais P, Jackson RB, Pacala SW, McGuire AD, Piao S, Rautiainen A, Sitch S, Hayes D (2011) A large and persistent carbon sink in the world's forests. Science 333(6045):988–993

Page SE, Siegert F, Rieley JO, Boehm H-DV, Jaya A, Limin S (2002) The amount of carbon released from peat and forest fires in Indonesia during 1997. Nature 420:61–65. https://doi.org/10.1038/nature01131

Petrescu AL, Tuovinen J-P, Baldocchi DD, Desai AR, Roulet NT, Vesala T, Dolman AJ, Oechel WC, Marcolla B, Friborg T, Rinne J, Matthes JH, Merbold L, Meijide A, Kiely G, Sottocornola M, Sachs T, Zona D, Varlagin A, Lai DY-F, Veenendaal E, Parmentier F-JW, Skiba U, Lund M, Hensen A, van Huissteden J, Flanagan LB, Shurpali NJ, Grünwald N, Humphreys ER, Jackowicz-Korczyński M, Aurela MA, Laurila T, Grüning C, Corradi CAR, Schrier-Uijl AP, Christensen TR, Tamstorf MP, Mastepanov M, Martikainen PJ, Verma S, Bernhofer BC, Cescatti A (2015) The uncertain climate footprint of wetlands under human pressure. PNAS 112(15):4594–4599. https://doi.org/10.1073/pnas.1416267112

Sannel ABK, Kuhry P (2008) Radiocarbon dating of peat profiles from the subarctic ecoclimatic region, west-central Canada. PANGAEA. https://doi.org/10.1594/PANGEA.786457

Schuur EAG, McGuire AD, Grosse G, Harden J, Hayes DJ, Hugelius H, Koven CD, Kuhry P, Lawrence DM, Natali SM, Olefeldt D, Romanovsky VE, Schädel C, Turetsky M, Treat C, Vonk J (2015) Climate change and the permafrost carbon feedback. Nature 520:171–179. https://doi.org/10.1038/nature1433

Shindell DT, Walter BP, Faluvegi G (2004) Impacts of climate change on methane emissions from wetlands. Geophysical Research Letters 31(21):L21202

Strain MA, van Belzen J, Comandini P, Wong J, Bouma TJ, Airoldi L (2017) The role of changing climate in driving the shift from perennial grasses to annual succulents in a Mediterranean saltmarsh. Journal of Ecology 105:13741385

Turetsky MR, Kotowska A, Bubier J, Dise NB, Crill P, Hornibrook ERC, Minkkinen K, Moore TR, Myers-Smith IH, Nykänen H, Olefeldt D, Rinne J, Saarnio S, Shurpali N, Tuittila E-S, Waddington JM, White JR, Wickland KP, Wilmking M (2014) A synthesis

of methane emissions from 71 northern, temperate, and subtropical wetlands. Global Change Biology 20:2183–2197. https://doi.org/10.1111/gcb.12580

van Groenigen KJ, Osenberg CW, Hungate BA (2011) Increased soil emissions of potent greenhouse gases under increased atmospheric CO_2. Nature 475:214–216

Wu J, Roulet NT (2014) Climate change reduces the capacity of northern peatlands to absorb the atmospheric carbon dioxide: The different responses of bogs and fens. Global Biogeochem Cycles 28:1005–1024. https://doi.org/10.1002/2014GB004845

Excerpted from:

Moomaw, W.R., Chmura, G.L., Davies, G.T. *et al.* Wetlands In a Changing Climate: Science, Policy and Management. *Wetlands* **38**, 183–205 (2018). Accessed December 2023 https://doi.org/10.1007/s13157-018-1023-8 CC-BY

7.9.2: Carbon Accumulation and GHG Emissions from Freshwater Wetlands (Including Permafrost) In a Changing Climate is shared under a CC BY license and was authored, remixed, and/or curated by LibreTexts.

7.9.3: Ecological Consequences for Freshwater Wetlands in a Changing Climate

Freshwater wetlands may be altered by climate change in all geographic regions of the world (Junk et al. 2013). A changed climate will alter hydrology, and functionality may be impaired by increased temperatures, drought or flooding events, CO_2 increases, and/or salinity intrusion (Junk et al. 2013). These changes will affect critical functions and ecosystem services such as carbon storage, biodiversity support, wildlife habitat and water quality (Junk et al. 2013). Negative impacts related to climate change will be compounded by synergies with other stressors, such as invasive species and land use change, thereby potentially increasing both the difficulties in managing and restoring wetlands, and the risk of endemic species extinctions (Erwin 2009).

Despite these challenges, some freshwater wetlands may be relatively resilient to climate change (Baron et al. 2002; Middleton and Souter 2016) within certain boundaries of temperature, precipitation, water level, salinity intrusion, and storm activity (Poff et al. 2002; Bernstein et al. 2007). At the same time, salinity intrusion poses specific threats to coastal freshwater wetlands because many species in these ecosystems are intolerant of salinity (Keddy 2010). Also, these species often have lower levels of production if salinity levels become too high (Middleton 1999; Sutter et al. 2014; Middleton and Souter 2016). A recent review synthesizes the state of our knowledge on how salinization associated with climate change will impact these wetlands (Herbert et al. 2015).

Climate change poses threats to non-coastal freshwater wetlands as well; hydrology is shifting as many local water regimes have become wetter or drier in recent decades (Fig. 7.9.3.1) (Mallakpour and Villarini 2015). In particular, megadroughts predicted by climate models (Cook et al. 2015) may dry Midwestern and Southwestern wetlands in North America with severe consequences for both wetlands and society. Severe droughts could impair the ability of these wetlands to maintain services including water quality, water supply, flood control, storm protection, and direct harvests of fish, animals, and plants, ultimately with severe negative impacts on ecosystem function and biodiversity (Baron et al. 2002; Middleton and Souter 2016). In addition, reduced winter snowpack and earlier snowmelt are impacting northern freshwater wetlands by altering the timing and magnitude of stream flows (Lawler 2009). In northern areas with permafrost, vegetation structure completely changes after permafrost melts (Malhotra and Roulet 2015). In fact, climate change is already changing community composition, species distribution, phenology, physiology and invasive species presence (Lawler 2009).

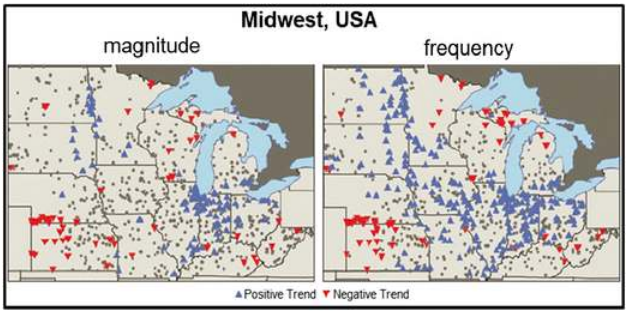


Figure 7.9.3.1: **The magnitude and *frequency* of flood events in the Midwestern United States from 1962−2011.** Triangles show trends of flooding at U.S.G.S. gage stations with trends (positive, negative, neutral; blue triangle, red triangle, and gray circle, respectively; from Mallakpour and Villarini 2015)

Unfortunately, many of the world's freshwater wetlands are already stressed by increased land-use pressure, so that additional hydrological alteration can contribute to an overall decrease in resilience to climate change (Baron 2002; Middleton and Souter 2016). Human alteration is commonplace throughout river corridors, challenging management as the impacts of upstream alterations accumulate along the waterway (See Fig. 7.9.3.2). (DuBowy 2013; Tockner and Stanford 2002). As demands for river resources increase, such problems are expected to worsen (Baron et al. 2002). Flowing water is compromised by river re-engineering practices, even though moving water generally improves oxygenation and plant health (Middleton 1999). Also, upriver freshwater extraction in tidal freshwater wetlands coupled with sea level rise can cause the salinification of surface and ground water, with accompanying stress and even the collapse of tidal vegetation in the freshwater reaches of estuaries (Perry and Atkinson 2009; Middleton and Souter 2016).

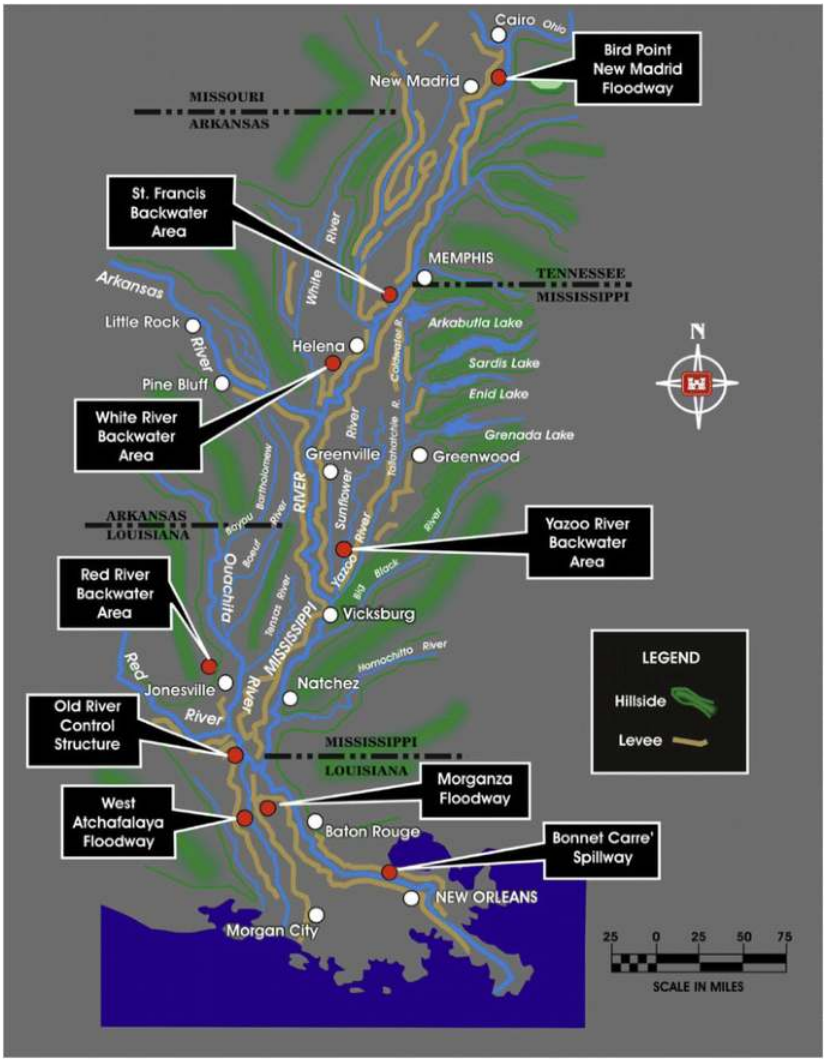


Figure 7.9.3.2: The hydrologic changes in the Mississippi River and tributaries for navigation and development include straightening, deepening, levee construction and damming. These engineering practices influence ecosystem processes across the floodplain and channel of this big river system (DuBowy 2013)

Fortunately, emerging research suggests that vegetation collapse sometimes can be avoided by hydrologic remediation (Souter et al. 2010). Freshwater remediation can reduce salinity and revive freshwater forests stressed by salinity intrusion, if the vegetation is not fatally damaged (Middleton et al. 2015; Middleton and Souter 2016). Such techniques could become critical for maintaining future ecosystem health and services (Baron et al. 2002; Middleton and Souter 2016). To date, there is no report of long-term monitoring of the survival of vegetation following remediation, so any long-term benefits are untested (Middleton and Souter 2016). Managers may need to carefully monitor the effects of traditional techniques and adjust the timing and/or intensity of management actions accordingly (Jackson and Hobbs 2009; Middleton et al. 2017).

One harbinger of ecosystem change is that the early life history stages of foundation species (species with a strong role in structuring communities) are increasingly unsuccessful at the hot or dry edges of their ranges, noting that juveniles are more sensitive to environmental extremes than adult plants (Jackson and Hobbs 2009). Without regeneration, vegetation enters a relict state (Williams et al. 1999). Worldwide examples of relict foundation species are growing, and such vegetation may be poised for abrupt decline if disturbance removes adult vegetation (Middleton et al. 2017). There are several indicators that some freshwater wetlands are poised for collapse at the edges of their ranges, and the loss of all but relict species is a key indicator of that problem (Middleton et al 2017). Thus, freshwater wetlands face a myriad of challenges in the face of climate change.

References

Baron JS, Poff NL, Angermeier PL, Dahm CN, Gleick PH, Hairston NG Jr, Jackson RB, Johnston CA, Richter BD, Steinman AD (2002) Meeting ecological and societal needs for freshwater. Ecological Applications 12:1247–1260

Bernstein L, Bosch P, Canziani O, Chen Z, Christ R, Davidson O, Hare W, Huq S, Karoly D, Kattsov V, Kundzewicz Z, Liu J, Lohmann U, Manning M, Matsuno T, Menne B, Metz B, Mirza M, Nicholls N, Nurse L, Pachauri R, Palutikof J, Parry M, Qin D, Ravindranath N, Reisinger A, Ren J, Riahi K, Rosenzweig C, Rusticucci M, Schneider S, Sokona Y, Solomon S, Stott P, Stouffer R, Sugiyama T, Swart R, Tirpak D, Vogel C, Yoh G. (2007) Intergovernmental Panel on Climate Change. Fourth Assessment Report, Climate Change 2007: synthesis report. IPCC Secretariat, World Meteorological Organization, Geneva

Cook BI, Ault TR, Smerdon JE (2015) Unprecedented 21st century drought risk in the American Southwest and Central Plains. Science Advances 1:e1400082

DuBowy PJ (2013) Mississippi River ecohydrology: Past, present and future. Ecohydrology and Hydrobiology 13:73–83

Erwin KL (2009) Wetlands and global climate change: the role of wetland restoration in a changing world. Wetlands Ecology and Management 17:71–84. https://doi.org/10.1007/s11273-008-9119-1

Herbert ER, Boon P, Burgin AJ, Neubauer SC, Franklin RB, Ardon M, Hopfensperger KN, Lamers LPM, Gell P (2015) A global perspective on wetland salinization: Ecological consequences of a growing threat to freshwater wetlands. Ecosphere 6(10):1–43. https://doi.org/10.1890/ES14-00534.1

Jackson ST, Hobbs RJ (2009) Ecological restoration in the light of ecological history. Science 325:567–569. https://doi.org/10.1126/science.1172977

Junk WJ, An S, Finlayson CM, Gopal B, Kvet J, Mitchell A, Mitsch WJ, Robarts RD (2013) Current state of knowledge regarding the world's wetlands and their future under global climate change: a synthesis. Aquatic Sciences 75:151–167. https://doi.org/10.1007/s00027-012-0278-z

Keddy PA (2010) Wetland ecology principles and conservation, 2nd edn. Cambridge University Press, NY, p 497

Lawler JJ (2009) Climate change adaptation strategies for resource management and conservation planning. The Year in Ecology and Conservation Biology, 2009. Annals of the New York Academy of Sciences 1162:79–98. https://doi.org/10.1111/j.1749-6632.2009.04147.x

Malhotra A, Roulet NT (2015) Environmental correlates of peatland carbon fluxes in a thawing landscape: Do transitional thaw stages matter? Biogeosciences 12(10):3119–3130

Mallakpour I, Villarini G (2015) The changing nature of flooding across the central United States. Nature Climate Change 5:250–254

Middleton BA (1999) Wetland restoration, flood pulsing and disturbance dynamics. John Wiley and Sons, New York

Middleton BA, Johnson D, Roberts B (2015) Hydrologic remediation for the Deepwater Horizon Incident drove ancillary primary production increase in coastal swamps. Ecohydrology 8:838–850

Middleton BA, Souter N (2016) Functional integrity of wetlands, hydrologic alteration and freshwater availability. Ecosystem Health and. Sustainability 2(1):e01200. https://doi.org/10.1002/ehs2.1200

Middleton BA, Boudell J, Fisichelli N (2017) Using management to address vegetation stress related to land-use and climate change. Restoration Ecology 26:1–4

Perry JE, Atkinson RB (2009) York River tidal marshes. Journal of Coastal Research 57:43–52

Poff NL, Brinson MM, Day JW Jr (2002) Aquatic ecosystems and global climate change. Potential impacts on inland freshwater and coastal wetland ecosystems in the United States. Pew Charitable Trust, Philadelphia http://www.pewtrusts.org/~/media/legacy/uploadedfiles/wwwpewtrustsorg/reports/protecting_ocean_life/envclimateaquaticecosystemspdf.pdf

Souter NJ, Cunningham S, Little S, Wallace T, McCarthy B, Henderson M (2010) Evaluation of a visual assessment method for tree condition of eucalypt floodplain forests. Ecological Management and Restoration 11:210–214

Sutter LS, Perry JE, Chambers RM (2014) Tidal freshwater marsh plant responses to low level salinity increases. Wetlands 34:167–175

Tockner K, Stanford JA (2002) Riverine flood plains: Present state and future trends. Environmental Conservation 29:308–330

Williams K, Ewel KC, Stumpf RP, Putz FE, Workman TW (1999) Sea−level rise and coastal forest retreat on the west coast of Florida, USA. Ecology 80:2045–2063

Excerpted from:

Moomaw, W.R., Chmura, G.L., Davies, G.T. *et al.* Wetlands In a Changing Climate: Science, Policy and Management. *Wetlands* **38**, 183–205 (2018). Accessed December 2023 https://doi.org/10.1007/s13157-018-1023-8 CC-BY

7.9.3: Ecological Consequences for Freshwater Wetlands in a Changing Climate is shared under a CC BY license and was authored, remixed, and/or curated by LibreTexts.

7.9.4: Salt Marsh and Mangrove Response to a Changing Climate and Associated Sea Level Rise

Salt Marsh and Mangrove Response to a Changing Climate and Associated Sea Level Rise

Saltwater coastal wetlands are generally found in sheltered waters and include mangrove forests, seagrass meadows, and tidal salt marshes. These wetlands host incredibly productive plant communities, which take up substantial amounts of C via photosynthesis, and store a significant fraction of that C in their wet, anaerobic soils (Chmura et al. 2003; Donato et al. 2011; Fourqurean et al. 2012). This C has been termed “coastal wetland blue carbon.” These vegetated saltwater coastal ecosystems represent an estimated 0.2% of the area of the ocean, but have C stocks equivalent to 50% of the C buried in ocean sediments (Duarte et al. 2013). As such, saltwater coastal blue C wetlands are some of the most C rich ecosystems on the planet (See Fig. 3 in McLeod et al. 2011). Thus, there are growing efforts to include saltwater wetlands in international climate protection activities and policy frameworks (Wylie et al. 2016; Howard et al. 2017).

Salt marshes and mangrove swamps have accumulated C-rich soil for centuries to millennia as sea levels have slowly risen increasing levels of plant production (See Fig. 3 in McLeod et al. 2011). These wetland soils accumulate vertically through three synergistic processes (See Fig. 7 in Fitzgerald et al. 2008). The belowground growth adds volume to the soil and the aboveground portion helps trap inorganic sediment carried in tidal waters that regularly flood the soil. Extended saturation of the soil reduces the rate of decomposition of soil organic matter, thereby enabling the persistence of the effective blue C sink. Increasing soil volume results in raised surface elevation of the wetland, so that on decadal scales its elevation roughly tracks sea level rise (e.g. Chmura et al. 2001; Ellison 2008). This increase in elevation is accompanied by lateral expansion of the marsh or mangrove swamp over tidal flats in the lower intertidal zone and inland over adjacent terrestrial ecosystems. The vegetation that occupies intertidal niches has evolved a suite of mechanisms to tolerate flooding by saline water, but at a greater expenditure of energy (e.g. Mendelssohn et al. 1982). There is a limit to this tolerance.

Saltwater wetlands provide significant ecosystem services. Mangroves and salt marshes help to slow and attenuate waves and storm surge, reducing the flooding and erosion of ocean coastal communities (Shepard et al. 2011; Arkema et al. 2013). One study suggests that U.S. marine saltwater wetlands provide $23.2 billion dollars of storm protection every year (Costanza et al. 2008) while another study estimates that every hectare of salt marsh provides US $8,234 dollars, or US $3,334 per acre, in storm protection, on average, per year (Barbier et al. 2011). Since a warmer climate contributes to increased storm intensity (Trenberth et al. 2015), enhancing these protective measures is seen as a cost-effective way to protect coastal communities and infrastructure. The storm protection qualities of wetlands are leading many policy and decision makers to consider more investments in protecting or restoring coastal wetlands and other ecosystems to provide the climate adaptation benefits of natural storm and erosion reduction (Barbier 2014; Sutton-Grier et al. 2015).

The impact of climate warming, its associated sea level rise and changes in precipitation patterns will vary considerably within and among tidal marshes. Few studies have looked at combined effects of sea level rise and other aspects of climate change. Feher et al. (2017) reviewed the literature on the influence of changing temperature and precipitation regimes on tidal saline wetlands. They found that for several ecosystem properties and many regions there was still insufficient evidence to make generalized predictions.

Research, however, has demonstrated differences due to climate zones and vegetation. For instance, where growing seasons are limited by cold temperatures, such as the coast of the northern Northwest Atlantic, studies have shown that a warmer climate would marginally increase decomposition, but will increase plant production and soil carbon storage (Charles and Dukes 2009; Gedan and Bertness 2010; Kirwan et al. 2014), although the effect of a rise of sea level was not addressed. On the Mediterranean coast, experimentally increased temperature, decreased precipitation and increased inundation period caused vegetation to shift from a perennial grass to an annual succulent (Strain et al. 2017).

There are two major ways that climate change is expected to impact all saltwater wetlands. Climate warming is expected to increase rates of sea level rise, resulting in loss of wetland area through “coastal squeeze,” particularly in areas surrounded by urbanized uplands (e.g. Torio and Chmura 2013). This has been identified as the largest climate change threat for mangroves (Gilman et al. 2008). Secondly, warmer temperatures will allow poleward shifts in flora and fauna that can result in significant changes in the saltwater tidal habitat, thereby altering its ecosystem services, including ability to store blue C, and in some cases causing the release of CO_2 from the blue C sink as described below.

Modification of estuarine hydrology or increased rates of sea level rise can increase the hydroperiod (duration of flooding) beyond the thresholds tolerated by intertidal vegetation. Climate warming will increase rates of sea level rise primarily from continued melting of the world's ice sheets and glaciers and the thermal expansion of a warming ocean (Church et al. 2013). As the magnitude and rate of ice sheet melting is difficult to model, predicted rates of sea level rise vary, but it is accepted that increasing rates of sea level rise and its impact will be felt on all coastlines, most severely on those already subject to subsidence (sinking). One modest projection, a 0.6 m (2 ft) rise in global (eustatic) sea level by 2100, would translate to an increase of 0.61 m (2.3 ft) at New York City and 1.07 m (3.5 ft) in Galveston Texas. The greatest uncertainty is the rate of melting of ice sheets covering Antarctica and Greenland. There is nothing magical about the year 2100, and it is certain that sea levels will continue to rise for centuries under all current scenarios. A recent report considers six possible outcomes for global mean sea level rise by 2100 ranging from 0.3 meters with a 100% probability to an intermediate projection of 1.0 m with a 17% probability. If recent estimates for Antarctica ice melt are included there is a 0.1% probability that the rise could reach 2.5 m (NOAA 2017). See Fig. 7.9.4.1.

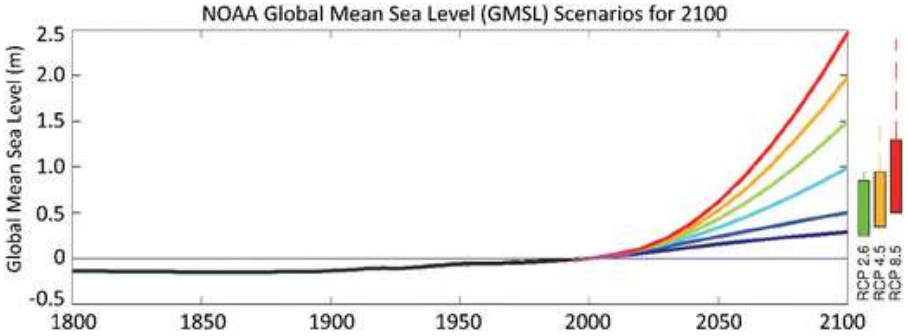


Figure 7.9.4.1: **Past and projected changes in global mean sea level rise.** (NOAA 2017a). Mean sea level rise and projections to 2100 under alternative IPCC scenarios. The lowest rise is 0.30 meters (100% probability), the intermediate is 1.0 meters (17% probability) and the highest is to 2.5 meters (0.1% probability)

Within tidal wetlands the effects of increased rates of sea level rise will be most strongly felt at the lower elevations where vegetation will most rapidly succumb and soil accretion will cease (e.g. Kirwan et al. 2010). Without living vegetation, the submerged wetland soil and its C stock can be exposed to erosion and possibly to oxidation of the organic matter, returning centuries of stored CO_2 back to the atmosphere. The fate of soil organic matter eroded from wetlands is an increasingly important science question that is not yet resolved (e.g. DeLaune and White 2012). If the upland adjacent to the tidal wetland is not developed and slopes are gentle, then the wetland can migrate inland, limiting the loss of area (but not necessarily blue carbon stocks). However, if this land is developed or if natural topography is steep, the structures or grade will prevent migration, putting the marsh or mangrove in a coastal squeeze (Torio and Chmura 2013). The potential for coastal squeeze is high on many of the world's coastlines, particularly on the highly urbanized bays and estuaries of the U.S., such as San Francisco Bay in California and the shore of New York City on Jamaica Bay (Hopper and Meixler 2016). The loss of wetland area due to coastal squeeze means loss of all its ecosystem services including essential habitat for fish and wildlife, loss of the ability of the system to store additional C and loss of its capacity to buffer inland development from the impacts of storms. One opportunity to decrease the amount of salt marsh loss that is likely to occur with sea level rise is to actively plan for future inland marsh migration now. There have been a few innovative studies considering how to plan for marsh migration including one that examined which wetlands along the Gulf coast of the U.S. are most threatened by projected future urban development. This information can be used to identify migration corridors for these wetlands and set priorities for current protection to prevent future coastal squeeze (Enwright et al. 2016). Another study examined two conditions to determine which marshes along the U.S. Northeast and Mid-Atlantic coast are likely to be resilient to sea level rise by examining the current health of the marsh as well as its potential to migrate inland (Anderson and Barnett 2017).

Climate warming has a direct impact on salt marshes and mangrove swamps by increasing poleward migration of their flora and fauna. Such changes are most observable where species' populations occur near the edge of their biogeographic ranges. In fact, globally, mangroves are expanding their range from tropical and subtropical climes, to invade salt marshes on adjacent warm temperate coasts (e.g. Godoy and DeLacerda 2015). Studies are finding that climate-changed-induced movement of mangroves into saltmarsh with warming temperatures is resulting in increases in the carbon stored in biomass and soils in marine and estuarine mangroves. This is because mangrove forests have some of the highest average C storage per land area in unmanaged terrestrial ecosystems (Doughty et al. 2015, Kelleway et al. 2015). As mangroves replace salt marsh vegetation, soil C may increase (Bianchi et al. 2013). However, such invasions significantly change habitat structure and we know little about impacts on biotic interactions,

potential lags for co-evolved species to shift, or challenges to mosquito control management (Dale et al. 2013). While SLR is expected to enable mangroves to migrate inland where other obstacles do not occur, the example of mangrove dieback in northern Australia (Duke et al. 2017) shows that the impact of climate is more complex, with changes in the regional climate patterns resulting in lower rainfall and tidal depression during the hot part of the year being suggested as the cause of the dieback.

Several studies have documented that increasing salinity in upstream reaches of an estuary will decrease biomass accumulation of foundation freshwater plant species (Sutter et al. 2014, 2015; Neubauer et al. 2005). In microcosm studies, Sutter et al. (2015) found that even smooth cordgrass (*Spartina alterniflora*), a salt marsh foundation species on the western Atlantic, had reduced growth when exposed to increased salinity and grown with the invasive strain of tall common reed (*Phragmites australis*).

An example of range extension of benthic fauna is found in the herbaceous salt marsh fiddler crab (*Uca pugnax*) that burrows in marsh soil. Historically, the range of the fiddler crab has been limited along the northwestern Atlantic coast to waters south of Cape Cod, Massachusetts. Its range recently has expanded northward where it has been observed on the coast of New Hampshire (Johnson 2014). The effect of fiddler crabs on C storage has been studied in Virginia salt marshes where Thomas and Blum (2010) found that 74% more root material was decomposed in marshes with fiddler crab burrows. Unless potential predators and competitors accompany crab migration, this range extension could lead to significant release of CO_2 to the atmosphere from northern salt marsh C sinks.

Saltwater wetlands are effective natural C sinks until they are disturbed, degraded, or destroyed by draining them for urban development, agriculture, aquaculture or by other means. Rising sea levels will also degrade these ecosystems. Disturbing wetland hydrology can enable oxygen to oxidize stored soil organic matter. Drying wetland soils increases microbial decomposition of stored organic C causing these natural sinks to become sources of CO_2 emissions (Pendleton et al. 2012). Preventing loss of these ecosystems is a priority to avoid additional GHG emissions. Restoring degraded or lost saltwater wetlands can regenerate their ability to remove and sequester CO_2 from the atmosphere.

References

Anderson MG, Barnett A (2017) Resilient coastal sites for conservation in the Northeast and Mid-Atlantic US. The Nature Conservancy, Eastern Conservation Science

Arkema KK, Guannel G, Verutes G, Wood SA, Guerry A, Ruckelshaus M, Kareiva P, Lacayo M, Silver JM (2013) Coastal habitats shield people and property from sea-level rise and storms. Nature Climate Change 3:913–918j. https://doi.org/10.1038/NCLIMATE1944

Barbier EB, Hacker SD, Kennedy C, Koch EW, Stier AC, Silliman BR (2011) The value of estuarine and coastal ecosystem services. Ecological Monographs 81(2):169–193

Bianchi TS, Allison MA, Zhao J, Li X, Comeaux RS, Feagin RA, Kulawardhana RW (2013) Historical reconstruction of mangrove expansion in the Gulf of Mexico: Linking climate change with carbon sequestration in coastal wetlands. Estuarine, Coastal and Shelf Science 119:7–16. https://doi.org/10.1016/j.ecss.2012.12.007

Charles H, Dukes JS (2009) Effects of warming and altered precipitation on plant and nutrient dynamics of a New England salt marsh. Ecological Applications 19:1758–1773

Chmura GL, Anisfeld S, Cahoon D, Lynch J (2003) Global carbon sequestration in tidal, saline wetland soils. Global Biogeochemical Cycles 17(4):22.1–22.12. https://doi.org/10.1029/2002GB001917

Church JA, Monselesan D, Gregory JM (2013) Marzeion B (2013) Evaluating the ability of process based models to project sea-level change. Environmental Research Letters 8:014051 (8pp). https://doi.org/10.1088/1748-9326/8/1/014051

Costanza R, Perez-Maqueo O, Martinez ML, Sutton P, Anderson SJ, Mulder K (2008) The value of coastal wetlands for hurricane protection. Ambio 37(4):241–248

Dale P, Eslami-Andargoli L, Knight J (2013) The impact of encroachment of mangroves into saltmarshes on saltwater mosquito habitats. Journal of Vector Ecology 38(2):330–338

Delaune RD, White JR (2012) Will coastal wetlands continue to sequester carbon in response to an increase in global sea level? A case study of the rapidly subsiding Mississippi river deltaic plain. Climatic Change 110:297. https://doi.org/10.1007/s10584-011-

0089-6

Donato DC, Kauffman JB, Murdiyarso D, Kurnianto S, Stidham M, Kanninen M (2011) Mangroves among the most carbon-rich forests in the tropics. Nature Geoscience 4:293–297. https://doi.org/10.1038/ngeo1123

Doughty CL, Langley JA, Walker WS, Feller IC, Schaub R, Chapman SK (2015) Mangrove expansion rapidly increases coastal wetland storage. Estuaries and Coasts. https://doi.org/10.1007/s12237-015-9993-8

Duarte CM, Losada IJ, Hendriks IE, Mazarrasa I, Marbà N (2013) The role of coastal plant communities for climate change mitigation and adaptation. Nature Climate Change 3:961–968. https://doi.org/10.1038/nclimate1970

Duke NC, Kovacs JM, Griffiths AD, Preece L, Hill DJE, van Oosterzee P, Mackenzie J, Morning HS, Burrows D (2017) Large-scale dieback of mangroves in Australia's Gulf of Carpentaria: A severe ecosystem response, coincidental with an unusually extreme weather event. Marine and Freshwater Research. https://doi.org/10.1071/MF16322

Ellison JC (2008) Long-term retrospection on mangrove development using sediment cores and pollen analysis: A review. Aquatic Botany 89:93–104

Enwright NM, Griffith KT, Osland MJ (2016) Barriers to and opportunities for landward migration of coastal wetlands with sea level rise. Frontiers in Ecology and the Environment 14(6):307–316

Feher LC, Osland MJ, Griffith KT, Grace JB, Howard RJ, Stagg CL, Enwright NM, Krauss KW, Gabler CA, Day RH, Rogers K (2017) Linear and nonlinear effects of temperature and precipitation on ecosystem properties in tidal saline wetlands. Ecosphere 8(10):e01956. https://doi.org/10.1002/ecs2.1956

FitzGerald DM, Fenster MS, Argow BA, Buynevich IV (2008) Coastal impacts due to sea-level rise. Annual Review of Earth and Planetary Sciences 36:601–647

Fourqurean JW, Duarte CM, Kennedy H, Marbà N, Holmer M, Mateo MA, Apostolaki ET, Kendrick GA, Krause-Jensen D, McGlathery KJ, Serrano O (2012) Seagrass ecosystems as a globally significant carbon stock. Nature Geoscience 5:505–509. https://doi.org/10.1038/ngeo1477

Gedan KB, Bertness MD (2010) How will warming affect the salt marsh foundation species Spartina patens and its ecological role? Oecologia 164:479–487

Gilman EL, Ellison JC, Duke NC, Field C (2008) Threats to mangroves from climate change and adaptation options: A review. Aquatic Botany 89:237–250

Godoy MDP, DeLacerda LD (2015) Mangroves response to climate change: A review of recent findings on mangrove extension and distribution. Anais da Academia Brasileira de Ciências (Annals of the Brazilian Academy of Sciences) https://doi.org/10.1590/0001-3765201520150055 Accessed 15 Oct 2017

Hopper T, Meixler MS (2016) Modeling coastal vulnerability through space and time. PLoS ONE 11(10):e0163495. https://doi.org/10.1371/journal

Howard J, Sutton-Grier A, Herr D, Kleypas J, Landis E, Mcleod E, Pidgeon E, Simpson S (2017) Clarifying the role of coastal and marine systems in climate mitigation. Frontiers in Ecology and Environment 15:42–50

Johnson DS (2014) Fiddler on the roof: A northern range extension for the marsh fiddler crab *Uca pugnax*. Journal of Crustacean Biology 34(5):671–673

Kelleway JJ, Saintilan N, Macreadie PI, Skilbeck CG, Zawadzki A, Ralph PJ (2015) Seventy years of continuous encroachment substantially increases "blue carbon" capacity as mangroves replace intertidal salt marshes. Global Change Biology 22:1097–1109. https://doi.org/10.1111/gcb.13158

Kirwan ML, Guntenspergen GR, Langley JA (2014) Temperature sensitivity of organic-matter decay in tidal marshes. Biogeosciences 11:4801–4808

McLeod E, Chmura GL, Bouillon S, Salm R, Björk M, Duarte CM, Lovelock CE, Schlesinger WH, Silliman BR (2011) A blueprint for blue carbon: Toward an improved understanding of the role of vegetated coastal habitats in sequestering CO_2. Frontiers in Ecology Environment 9(10):552–560 Figure 3. https://doi.org/10.1890/110004

Mendelssohn IA, McKee KL, Postek MT (1982) Sublethal stresses controlling *Spartina alterniflo*ra productivity. In: Gopal B, Turner RE, Wetzel RG, Whigham DF (eds) Wetlands: Ecology and management. National Institute of Ecology, Jaipur, pp 223–242

Neubauer SC, Givler K, Valentine SK, Megonigal JP (2005) Seasonal patterns and plant - mediated controls of subsurface wetland biogeochemistry. Ecology 86(12):3334–3344

NOAA 2017 National Oceanic and Atmospheric Administration U.S. Department of Commerce, U.S. Geological Survey, U.S. Environmental Protection Agency and Rutgers University, Global and regional sea level rise scenarios for the United States NOAA Technical Report NOS CO-OPS 083 techrpt83_Global_and_Regional_SLR_Scenarios_for_the_US_final.pdf. Accessed 19 Apr 2017

Pendleton L, Donato DC, Murray BC, Crooks S, Jenkins WA, Sifleet S, Craft C, Fourqurean JW, Kauffman JB, Marba N, Megonigal JP, Pidgeon E, Herr D, Gordon D, Baldera A (2012) Estimating global "blue carbon" emissions from conversion and degradation of vegetated coastal ecosystems. PLoS ONE 7(9):e43542. https://doi.org/10.1371/journal.pone.0043542

Shepard CC, Crain CM, Beck MW (2011) The protective role of coastal marshes: A systematic review and meta-analysis. PLoS ONE 6(11):e27374. https://doi.org/10.1371/journal.pone.0027374

Strain MA, van Belzen J, Comandini P, Wong J, Bouma TJ, Airoldi L (2017) The role of changing climate in driving the shift from perennial grasses to annual succulents in a Mediterranean saltmarsh. Journal of Ecology 105:13741385

Sutter LS, Perry JE, Chambers RM (2014) Tidal freshwater marsh plant responses to low level salinity increases. Wetlands 34:167–175

Sutter LA, Chambers RM, Perry JE (2015) Seawater intrusion mediates species transition in low salinity, tidal marsh vegetation. Aquatic Botany 122:32–39

Sutton-Grier AE, Bamford H, Wowk K (2015) Future of our coasts: the potential for natural and hybrid infrastructure to enhance the resilience of our coastal communities, economies and ecosystem. Environmental Science & Policy 51:137–148

Thomas CR, Blum LK (2010) Importance of the fiddler crab *Uca pugnax* to salt marsh soil organic matter accumulation. Marine Ecology Progress Series 414:167–177. https://doi.org/10.3354/meps08708

Torio D, Chmura GL (2013) Assessing coastal squeeze of tidal wetlands. Journal of Coastal Research 29(5):1049–1061. https://doi.org/10.2112/JCOASTRES-D-12-00162.1

Trenberth KE, Fasullo JT, Shepherd TG (2015) Attribution of climate extreme events. Nature Climate Change 5:725–730

Wylie L, Sutton-Grier AE, Moore A (2016) Keys to successful blue carbon projects: Lessons learned from global case studies. Marine Policy 65:76–84

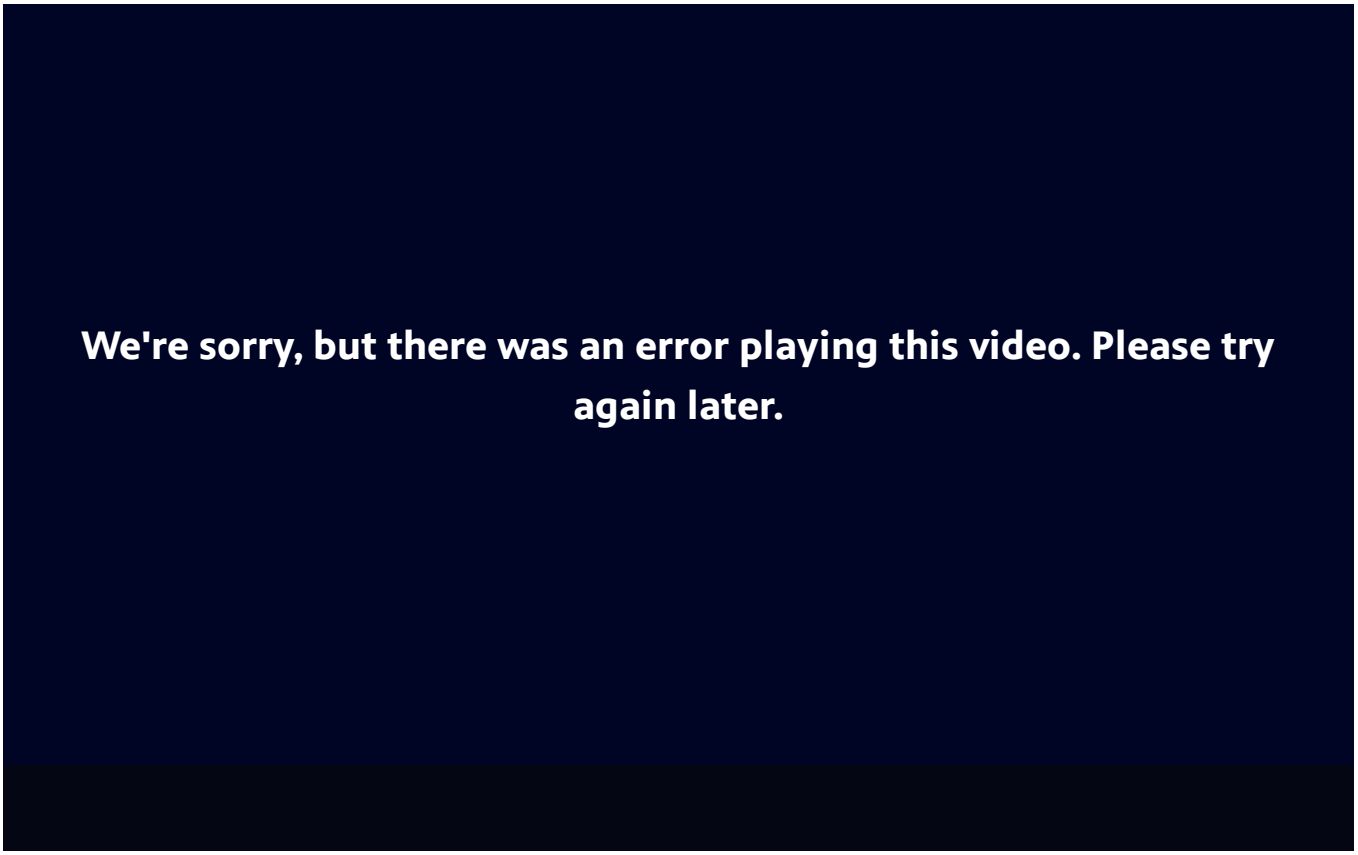
Excerpted From:

Moomaw, W.R., Chmura, G.L., Davies, G.T. *et al.* Wetlands In a Changing Climate: Science, Policy and Management. *Wetlands* **38**, 183–205 (2018). Accessed December 2023 https://doi.org/10.1007/s13157-018-1023-8 CC-BY

7.9.4: Salt Marsh and Mangrove Response to a Changing Climate and Associated Sea Level Rise is shared under a CC BY license and was authored, remixed, and/or curated by LibreTexts.

7.10: Summary

A short video from Public Television about wetlands



7.10: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

8: Inland waters

8.1: Introduction
8.2: The structure of aquatic ecosystems
8.3: The special properties of water
8.4: Inland waters and their watersheds
8.4.1: The Physical Template
8.4.2: Watershed Structure
8.4.3: Watershed Functions
8.4.4: Summary
8.5: Lakes
8.5.1: Freshwater Lakes
8.5.2: Saline Lakes
8.6: Streams and Rivers
8.7: Estuaries- Where the turf meets the surf
8.8: Waves and Beaches
8.9: Human impacts on inland waters
8.9.1: Fresh Water Supply and the Water Cycle
8.9.2: Water Usage
8.9.3: Water Scarcity and Solutions
8.9.4: Data Dive- Aqueducts of Rome
8.9.5: Review
8.10: Summary

8: Inland waters is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.1: Introduction





8.1: Introduction is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.2: The structure of aquatic ecosystems

Abiotic factors in Aquatic Biomes

Like terrestrial biomes, the **aquatic biome** is influenced by abiotic factors. In the case of this biome, the abiotic factors include light, temperature, flow, and dissolved nutrients.The aquatic medium—water— has different physical and chemical properties than air. Even if the water in a pond or other body of water is perfectly clear (there are no suspended particles), water, on its own, absorbs light. As one descends deep enough into a body of water, eventually there will be a depth at which the sunlight cannot reach. While there are some abiotic and biotic factors in a terrestrial ecosystem that shade light (like fog, dust, or insect swarms), these are not usually permanent features of the environment. The importance of light in aquatic biomes is central to the communities of organisms found in both freshwater and marine systems because it controls productivity through photosynthesis.

Definition: Aquatic Biome

The **aquatic biome** is the largest of all the biomes, covering about 75 percent of Earth's surface. This biome is usually divided into two categories: freshwater and marine. Typically, freshwater habitats are less than 1 percent salt. Marine life, however, has to be adapted to living in a habitat with a high concentration of salt. Freshwater habitats include ponds, lakes, rivers, and streams, while marine habitats include the ocean and salty seas.

In addition to light, solar radiation warms bodies of water and many exhibit distinct layers of water at differing temperatures. The water temperature affects the organisms' rates of growth and the amount of dissolved oxygen available for respiration.

The movement of water (flow) is also important in many aquatic biomes. In rivers, the organisms must obviously be adapted to the constant movement of the water around them, but even in larger bodies of water such as the oceans, regular currents and tides impact availability of nutrients, food resources, and the presence of the water itself.

Finally, all natural water contains dissolved salts. Fresh water contains low levels of such dissolved substances because the water is rapidly recycled through evaporation and precipitation. The oceans have a relatively constant high salt content. Aquatic habitats at the interface of marine and freshwater ecosystems have complex and variable salt environments that range between freshwater and marine levels. These are known as brackish water environments. Lakes located in closed drainage basins concentrate salt in their waters and can have extremely high salt content that only a few and highly specialized species are able to inhabit.

Dissolved nutrients such as nitrogen, phosphorus, and silica, occur in low amounts but are important for microscopic aquatic plants known as phytoplankton. Phytoplankton serve as the basis of the food chain in most marine systems. Phytoplankton growth is often limited by the availability of these, and other nutrients, and thus systems with high nutrients support rapid growth of phytoplankton who in turn support diverse communities of aquatic organisms. Nutrients availability in freshwater systems is often high. In marine systems, nutrient availability is highly variable and is influenced by wind, ocean depth, proximity to shore, and circulation of deep ocean currents.

Freshwater Systems

The aquatic (freshwater, or partly freshwater) biome include lakes, ponds, and wetlands (standing water) as well as rivers and streams (flowing water). Freshwater biomes are found in terrestrial landscapes and are therefore connected with abiotic and biotic factors influencing these terrestrial biomes.

Lakes and Ponds

Lakes and **ponds** can range in area from a few square meters to thousands of square kilometers. A lake is an area filled with water, localized in a basin, surrounded by land, and distinct from any river or other outlet that serves to feed or drain the lake. Lakes are typically larger and deeper than ponds, which also lie on land, though there are no official or scientific definitions. Most lakes are fed and drained by rivers and streams.

Temperature is an important abiotic factor affecting living things found in lakes and ponds. During the summer in temperate regions, **thermal stratification** of deep lakes occurs when the upper layer of water is warmed by the Sun and does not mix with deeper, cooler water (Figure 8.2.1). The process produces a sharp transition between the warm water above and cold water beneath

and has a large influence on the animal and plant life inhabiting the lake. The two layers do not mix until cooling temperatures and winds break down the stratification and the water in the lake mixes from top to bottom.

During the period of stratification, most of the productivity occurs in the warm, well-illuminated, upper layer, while dead organisms slowly rain down into the cold, dark layer below where decomposing bacteria and cold-adapted species such as lake trout exist. Like the ocean, lakes and ponds have a photic layer in which photosynthesis can occur. Phytoplankton (algae and cyanobacteria) are found here and provide the base of the food web of lakes and ponds. Zooplankton, such as rotifers and small crustaceans, consume these phytoplankton. At the bottom of lakes and ponds, bacteria in the aphotic zone break down dead organisms that sink to the bottom.

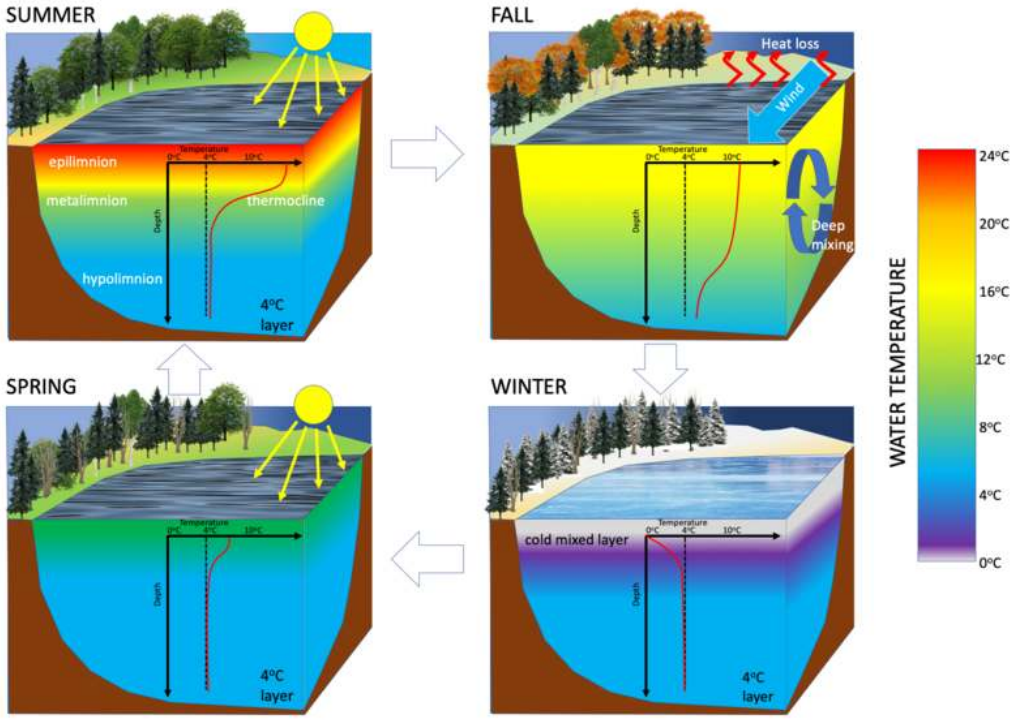


Figure 8.2.1: The effect of seasons on mixing and stratification of large freshwater lakes. (CC-BY-SA 4.0: Dimictic lake by wikimedia Commons)

Rivers and Streams

Rivers and the narrower **streams** that feed into the rivers are continuously moving bodies of water that carry water from the source or headwater to the mouth at a lake or ocean. The largest rivers include the Nile River in Africa, the Amazon River in South America, and the Mississippi River in North America (Figure 8.2.2).



(a) (b)

Figure 8.2.2: Rivers range from (a) narrow and shallow to (b) wide and slow moving. (modification of work by Cory Zanker; credit b: modification of work by David DeHetre)

Abiotic features of rivers and streams vary along the length of the river or stream. Streams begin at a point of origin referred to as source water. The source water is usually cold, low in nutrients, and clear. The channel (the width of the river or stream) is narrower here than at any other place along the length of the river or stream. Headwater streams are of necessity at a higher elevation than the mouth of the river and often originate in regions with steep grades leading to higher flow rates than lower elevation stretches of the river.

Headwater streams are often clear because they are closer to their origin and fast moving, resulting in minimal siltation. Photosynthesis here is mostly attributed to algae that are growing on rocks; the swift current inhibits the growth of phytoplankton. While photosynthesis in the water may be partially blocked by overhead tree branches and leaves, those same leaves add energy to the stream. When the leaves decompose, the organic material and nutrients in the leaves are returned to the water. The leaves also support a food chain of invertebrates that eat them and are in turn eaten by predatory invertebrates and fish. Plants and animals have adapted to this fast-moving water. For instance, some species of mayfly (phylum Arthropoda) have flattened bodies and legs with modified claws to help them cling to the underside of submerged rocks. This body form reduces drag and allows these species to benefit from the high oxygen concentrations in fast-moving currents without being dislodged. Freshwater trout species are an important predator in these fast-moving rivers and streams.

As the river or stream flows away from the source, the width of the channel gradually widens, the current slows, and the temperature characteristically increases. The increasing width results from the increased volume of water from more and more tributaries. Gradients are typically lower farther along the river, which accounts for the slowing flow. With increasing volume can come increased silt, and as the flow rate slows, the silt may settle, thus increasing the deposition of sediment. Phytoplankton can also be suspended in slow-moving water. Therefore, the water will not be as clear as it is near the source. The water is also warmer as a result of longer exposure to sunlight and the absence of tree cover over wider expanses between banks. Worms and insects can be found burrowing into the mud. Predatory vertebrates include waterfowl, frogs, and fishes. In heavily silt-laden rivers, these predators must find food in the murky waters, and, unlike the trout in the clear waters at the source, these vertebrates cannot use vision as their primary sense to find food. Instead, they are more likely to use taste or chemical cues to find prey.

When a river reaches the ocean or a large lake, the water typically slows dramatically and any silt in the river water will settle. Rivers with high silt content discharging into oceans with minimal currents and wave action will build deltas, low-elevation areas of sand and mud, as the silt settles onto the ocean bottom. Rivers with low silt content or in areas where ocean currents or wave action are high create estuarine areas where the freshwater and saltwater mix.

Wetlands

Wetlands are environments in which the soil is either permanently or periodically saturated with water. Wetlands are different from lakes and ponds because they are shallower. Because of this, they exhibit a near continuous cover of emergent vegetation. Emergent vegetation consists of wetland plants that are rooted in the soil but have portions of leaves, stems, and flowers extending above the water's surface. There are several types of wetlands including marshes, swamps, bogs, mudflats, and salt marshes (Figure 8.2.3).



Figure 8.2.3: Located in southern Florida, Everglades National Park is vast array of wetland environments, including sawgrass marshes, cypress swamps, and estuarine mangrove forests. Here, a great egret walks among cypress trees. (credit: NPS)

Freshwater **marshes** and **swamps** are characterized by slow and steady water flow. **Bogs** develop in depressions where water flow is low or nonexistent. Bogs usually occur in areas where there is a clay bottom with poor percolation. Percolation is the movement of water through the pores in the soil or rocks. The water found in a bog is stagnant and oxygen depleted because the oxygen that is used during the decomposition of organic matter is not replaced. As the oxygen in the water is depleted, decomposition slows. This leads to organic acids and other acids building up and lowering the pH of the water. At a lower pH, nitrogen becomes unavailable to plants. This creates a challenge for plants because nitrogen is an important limiting resource. Some types of bog plants (such as sundews, pitcher plants, and Venus flytraps) capture insects and extract the nitrogen from their bodies. Bogs have low net primary productivity because the water found in bogs has low levels of nitrogen and oxygen.

Estuaries

Estuaries are biomes that occur where a river, a source of freshwater, meets the ocean. Therefore, both freshwater and saltwater are found in the same vicinity; mixing results in a diluted (brackish) salt water. Estuaries form protected areas where many of the offspring of crustaceans, mollusks, and fish begin their lives. Salinity is an important factor that influences the organisms and the adaptations of the organisms found in estuaries. The salinity of estuaries varies and is based on the rate of flow of its freshwater sources. Once or twice a day, high tides bring salt water into the estuary. Low tides occurring at the same frequency reverse the current of salt water (Figure 8.2.4).



Figure 8.2.4: As estuary is where fresh water and salt water meet, such as the mouth of the Klamath River in California. (credit: U.S. Army Corps of Engineers)

The daily mixing of fresh water and salt water is a physiological challenge for the plants and animals that inhabit estuaries. Many estuarine plant species are halophytes, plants that can tolerate salty conditions. Halophytic plants are adapted to deal with salt water spray and salt water on their roots. In some halophytes, filters in the roots remove the salt from the water that the plant absorbs. Animals, such as mussels and clams (phylum Mollusca), have developed behavioral adaptations that expend a lot of energy to function in this rapidly changing environment. When these animals are exposed to low salinity, they stop feeding, close their shells, and switch from aerobic respiration (in which they use gills) to anaerobic respiration (a process that does not require oxygen). When high tide returns to the estuary, the salinity and oxygen content of the water increases, and these animals open their shells, begin feeding, and return to aerobic respiration.

Marine Systems

The ocean is a continuous body of saltwater that is relatively uniform in chemical composition. It is a weak solution of mineral salts and decayed biological matter. The ocean is categorized by several zones (Figure 8.2.5). All of the ocean's open water is referred to as the **pelagic realm** (or zone). The **benthic realm** (or zone) extends along the ocean bottom from the shoreline to the deepest parts of the ocean floor. From the surface to the bottom or the limit to which photosynthesis occurs is the photic zone (approximately 200 m or 650 ft). At depths greater than 200 m, light cannot penetrate; thus, this is referred to as the aphotic zone. The ocean is, on average, 4267 m or 14,000 ft deep. Thus, the majority of the ocean is aphotic and lacks sufficient light for photosynthesis.

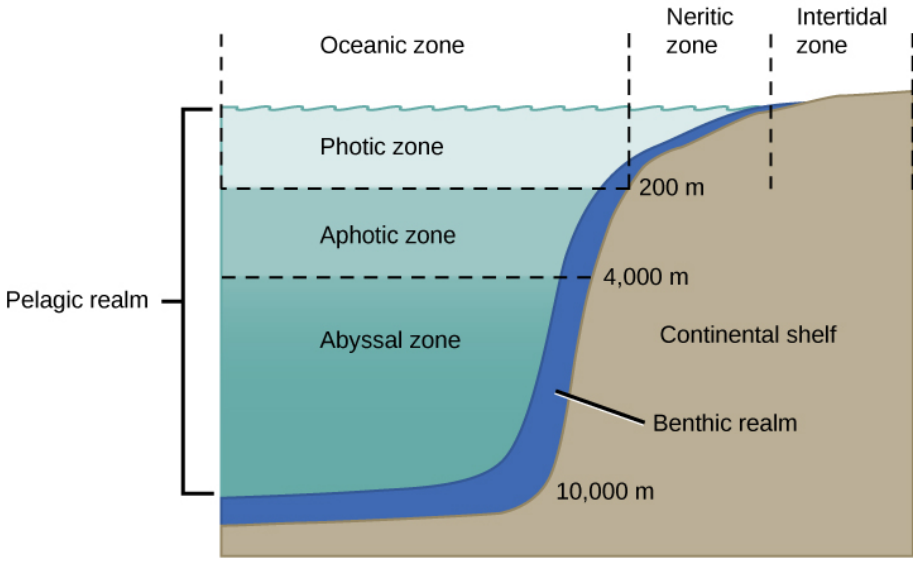


Figure 8.2.5: The ocean is divided into different zones based on water depth, distance from the shoreline, and light penetration. (CC BY; Open Stax)

Intertidal Zone

The **intertidal zone** (Figure 8.2.5) is the oceanic region that is closest to land. With each tidal cycle, the intertidal zone alternates between being inundated with water and left high and dry. Generally, most people think of this portion of the ocean as a sandy beach. In some cases, the intertidal zone is indeed a sandy beach, but it can also be rocky, muddy, or dense with tangled roots in mangrove forests.

The intertidal zone is an extremely variable environment because of tides. Organisms may be exposed to air at low tide and are underwater during high tide. Therefore, living things that thrive in the intertidal zone are often adapted to being dry for long periods of time. The shore of the intertidal zone is also repeatedly struck by waves and the organisms found there are adapted to withstand damage from the pounding action of the waves (Figure 8.2.6). The exoskeletons of shoreline crustaceans (such as the shore crab, *Carcinus maenas*) are tough and protect them from desiccation (drying out) and wave damage. Another consequence of the pounding waves is that few algae and plants establish themselves in constantly moving sand or mud. Thus in sandy or mud bottom intertidal communities many organisms burrow into the substrate.


Figure 8.2.6: Sea stars, sea urchins, and mussel shells are often found in the intertidal zone, shown here in Kachemak Bay, Alaska. (credit: NOAA)

Neritic Zone

The **neritic zone** is the relatively shallow part of the ocean above the drop-off of the continental shelf, approximately 200 meters (660 ft) in depth. From the point of view of marine biology it forms a relatively stable and well-illuminated environment for marine life, from plankton up to large fish and corals, while physical oceanography sees it as where the oceanic system interacts with the coast. When the water is relatively clear, photosynthesis can occur in the neritic zone. The water contains silt and is well-

oxygenated, low in pressure, and stable in temperature. These factors all contribute to the neritic zone having the highest productivity and biodiversity of the ocean. Phytoplankton, including photosynthetic bacteria and larger species of algae, are responsible for the bulk of this primary productivity. Zooplankton, protists, small fishes, and shrimp feed on the producers and are the primary food source for most of the world's fisheries. The majority of these fisheries exist within the neritic zone.

Coral Reefs

Coral reefs are ocean ridges formed by marine invertebrates living in warm shallow waters within the photic zone of the ocean. They are found within 30˚ north and south of the equator. Corals are adapted to warm clear waters. The Great Barrier Reef is a well-known, and largest reef system in the world, located several miles off the northeastern coast of Australia. Other coral reefs are fringing islands, which are directly adjacent to land, or atolls, which are circular reefs surrounding a former island that is now underwater.

The coral-forming colonies of animals (related to jelly fish and sea anemones) secrete a calcium carbonate skeleton. These calcium-rich skeletons slowly accumulate, thus forming the underwater reef (Figure 8.2.7. Corals found in shallower waters (at a depth of approximately 60 m or about 200 ft) have a mutualistic relationship with photosynthetic algae. The relationship provides corals with the majority of the nutrition and the energy they require. The waters in which these corals live are nutritionally poor and, without this mutualism, it would not be possible for large corals to grow because there are few planktonic organisms for them to feed on. Some corals living in deeper and colder water do not have a mutualistic relationship with algae; these corals must obtain their energy exclusively by feeding on plankton using stinging cells on their tentacles.

Coral reefs are one of the most diverse biomes. It is estimated that more than 4000 fish species inhabit coral reefs. These fishes can feed on coral, the cryptofauna (invertebrates found within the calcium carbonate structures of the coral reefs), or the seaweed and algae that are associated with the coral. These species include predators, herbivores, or planktivores. Predators are animal species that hunt and are carnivores or "flesh eaters." Herbivores eat plant material, and planktivores eat plankton.

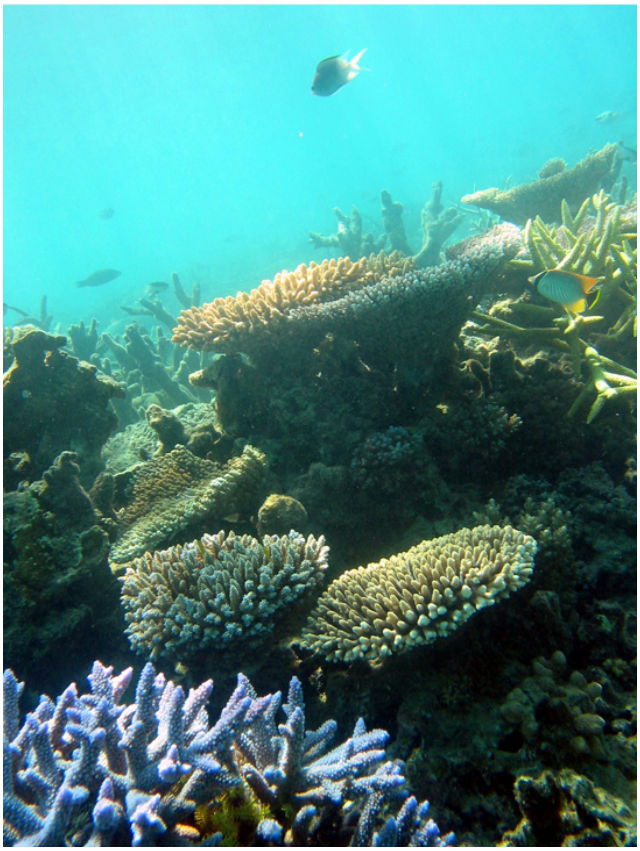


Figure 8.2.7: Coral reefs are formed by the calcium carbonate skeletons of coral organisms, which are marine invertebrates in the phylum Cnidaria. (credit: Terry Hughes)

Kelp Forests

Kelp forests are underwater areas with a high density of kelp, which covers a large part of the world's coastlines. They are recognized as one of the most productive and dynamic ecosystems on Earth (Figure 8.2.8). Although algal kelp forest combined with coral reefs only cover 0.1% of Earth's total surface, they account for 0.9% of global primary productivity. Kelp forests occur worldwide throughout temperate and polar coastal oceans.



Figure 8.2.8: The kelp (a type of brown algae) of a kelp forest creates 3-dimensional structure and habitat for a diversity of fish, invertebrates, and mammals. (public domatin; Sara Kappus taken on Santa Catalina Island)

Kelp forests are mostly found in cool, shallow, nutrient-rich water near coasts. Kelp is a brown alga, which requires access to light in order to photosynthesize - this is the reason for their abundance in shallow coastal water, Kelp forests provide a unique habitat for marine organisms and are a source for understanding many ecological processes. Over the last century, they have been the focus of extensive research, particularly in trophic ecology, and continue to provoke important ideas that are relevant beyond this unique ecosystem. For example, kelp forests can influence coastal oceanographic patterns and provide many ecosystem services

Kelp forests often draw comparisons to terrestrial forests, however the two ecosystems are distinct in that kelp is an algae - not a member of the Plant Kingdom. The structure of all kelp bears three universal morphological characteristics: holdfasts, stipes, and fronds. The holdfasts, which serve as an anchor for kelp, must be attached to hard substrates such as rocks or sand. Although they resemble roots on terrestrial plants, holdfasts do not transport water or nutrients through their stipe. The stipe, analogous to the stem, gives structural support to the algae. Fronds are blades extending from the stipe. The blades serve as the photosynthetic and nutrient uptake center for the organism.

Kelp forests are an important foundation species, which provide large areas of habitat within the ocean. Kelp forests have a wide variety of inhabitants within their boundaries: invertebrates frequently graze on the blades, fish find shelter in the fronds, and invertebrates - such as brittle stars, sea stars, anemones, sponges and tunicates - live in the holdfast. Kelp is also the single largest source of fixed carbon within their ecosystems. Pieces of decomposing kelp sink deeper into the ocean, or wash up on beaches, providing nutrients organisms in other ecosystems. Thus, kelp derived particulate organic matter (POM) is an important food source for fauna both within and outside the kelp forest.

Since kelp is capable of attenuating waves and dampening their energy, many animals are able to seek refuge in the forests. Various species of kelp tolerate ocean storms differently. In fact, along the California coast giant kelp has been known to dominate in years with less turbulent sea conditions, while bull kelp is more dominant in open waters and years of turbulent conditions.

Oceanic zone

Photic Zone

Beyond the neritic zone is the open ocean area known as the oceanic zone, which is further divided into 3 layers based on sunlight availability (the photic zone, aphotic zone and abyssal zone) (Figure 8.2.5). Within the Photic zone there is sufficient light for net photosynthesis to occur. Abundant phytoplankton and zooplankton support populations of fish, sea birds, and marine mammals. Nutrients are scarce and this is a relatively less productive part of the marine biome when compared to the Neritic zone. When photosynthetic organisms and the organisms that feed on them die, their bodies fall to the bottom of the ocean and provide food for deep ocean scavengers.

Aphotic Zone

The depth at which less than one percent of sunlight reaches begins the aphotic zone. While most of the ocean's biomass lives in the photic zone, the majority of the ocean's water lies in the aphotic zone. Bioluminescence is more abundant than sunlight in this zone.

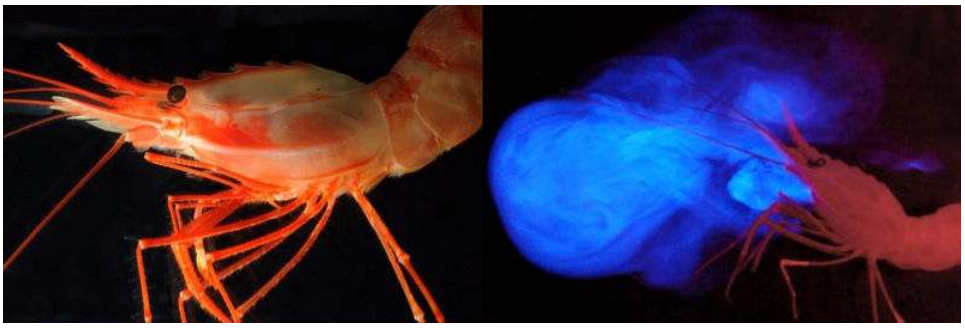


Figure 8.2.9: The deep-sea pandalid shrimp Heterocarpus ensifer and a photo of the same animal 'vomiting' light from glands located near its mouth.Photographed during the Bioluminescence and Vision 2015 expedition, as scientists sought to gain a better understanding of how and why organisms create their own light. (CC BY-SA; NOAA Ocean Explorer via Wikipedia Commons)

Abyssal zone

The deepest part of the ocean is the **abyssal zone**, which is at depths of 4000 m or greater. The abyssal zone (Figure 8.2.5) is very cold and has very high pressure, very low or no oxygen content, and high nutrient content as the dead and decomposing material that drifts down from the layers above. There are a variety of invertebrates and fishes found in this zone, but the abyssal zone does not have photosynthetic organisms. Chemosynthetic bacteria use the hydrogen sulfide and other minerals emitted from deep hydrothermal vents. These chemosynthetic bacteria use the hydrogen sulfide as an energy source and serve as the base of the food chain found around the vents.

Benthic Realm

Beneath the oceanic zone is the benthic realm, the deepwater region beyond the continental shelf (Figure 8.2.5). The bottom of the benthic realm is composed of sand, silt, and dead organisms. Temperature decreases as water depth increases. This is a nutrient-rich portion of the ocean because of the dead organisms that fall from the upper layers of the ocean. Because of this high level of nutrients, a diversity of fungi, sponges, sea anemones, marine worms, sea stars, fishes, and bacteria exists.

Attribution:

This page is a modified derivative of:

- 20.4 Aquatic and Marine Biomes by OpenStax CC BY
- Lake by Wikipedia, license CC-BY-SA 3.0
- Neritic Zone by Wikipedia, license CC-BY-SA 3.0
- Aphotic Zone Wikipedia, license CC-BY-SA 3.0
- Kelp Forests by Wikipedia, license CC-BY-SA 3.0
- Kelp Forest Ecosystems by LibraText, license CC BY-NC-SA 4.0

8.2: The structure of aquatic ecosystems is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.3: The special properties of water

A water molecule is composed of one oxygen and two hydrogen atoms that are joined together by **polar covalent** bonds. Covalent mean that the atoms share electrons, instead of completely giving up electrons to one another. **Polar** means that the electrons are not shared equally. These polar covalent bonds (Figure 8.2.1), along with the molecular shape, cause the water molecule to have a slightly positive charge on the hydrogen end and a slightly negative charge on the oxygen side. Water's charges are generated because oxygen is more electronegative than hydrogen, making it more likely that a shared electron would be found near the oxygen nucleus than the hydrogen nucleus, thus generating the partial negative charge near the oxygen. This gives water molecules their properties of attraction.

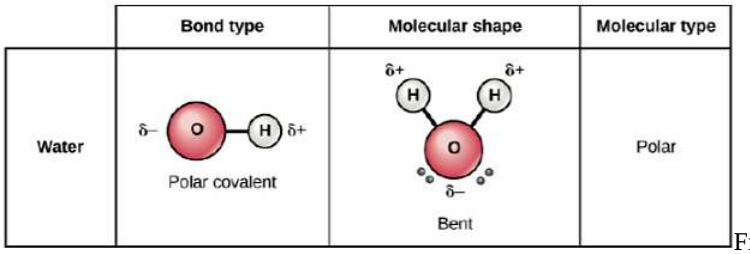


Figure 8.3.1: Polarity of the water molecule due to the uneven distribution of electrons in its covalent bond. C (From OpenStax Concepts of Biology text)

Hydrogen Bonds

Due to water's polarity, each water molecule attracts other water molecules as oppositely charged ends of the molecules attract each other. When this happens, a weak interaction occurs between the *positive* hydrogen end from one molecule and the *negative oxygen* end of another molecule. This interaction is called a **hydrogen bond**. This hydrogen bonding contributes to the following water's unique properties.

1. Water is the universal solvent

2. Exists in nature as a solid, liquid, and gas

3. The density of ice is less than liquid water

4. Water has a high surface tension

5. Water has a high heat capacity

6. Water exists as a liquid at room temperature It is important to note here that even we are only focusing on water in this text book, hydrogen bonding also occurs in other substances that have polar molecules.

Physical State of Water on Earth

Water on Earth can naturally exist as either solid, liquid or gas depending on the prevailing temperature and pressure conditions. The formation of hydrogen bonds (described above) is an important quality of liquid water that is crucial to life as we know it on Earth. As water molecules make hydrogen bonds with each other, liquid water takes on some unique physical and chemical characteristics when compared to other liquids. In liquid water, hydrogen bonds are constantly being formed and broken as the water molecules slide past each other. The energy of the moving water molecules (kinetic energy) is responsible for breaking the bonds. When heat is added to water (increasing temperature), the kinetic energy of the molecules goes up and more bonds are broken. As more heat is added to boiling water, the higher kinetic energy of the water molecules causes the hydrogen bonds to break completely and allow them to escape into the air as water vapor. On the other hand, when the temperature of water is reduced and water freezes, the water molecules form a crystalline structure maintained by hydrogen bonding (since there isn't enough energy to break the hydrogen bonds). The crystalline structure, ice, has a more open structure than the liquid form of water. The open structure of ice (Figure 8.3.2) makes ice less dense than liquid water, a phenomenon not seen in the solidification of other liquids.

The lower density of ice, illustrated in Figure 8.3.2, causes it to float at the surface of liquid water, such as an iceberg in the ocean or ice cubes in a glass of ice water. In lakes and ponds, ice will form on the surface of water creating an insulating barrier that

protects animals and plant life that live in the water from freezing. Without this layer of insulating ice, plants and animals living in the water would freeze in the solid block of ice and not survive. The ice crystals that form upon freezing would rupture the delicate membranes essential for the function of living cells, irreversibly damaging them.

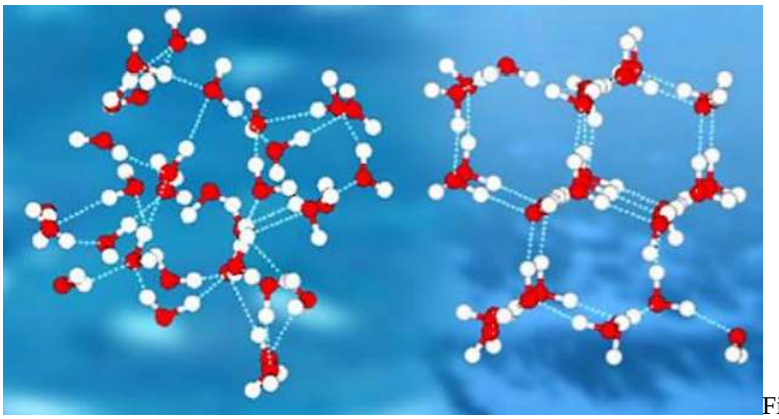


Figure 8.3.2: Hydrogen bonding makes ice less dense than liquid water. The lattice structure water is more condensed (left structure) than that of ice (right structure). The lattice structure of ice makes it less dense than freely flowing molecules of liquid water, enabling ice to float on liquid water. (Image credit: Lynn Yarris, http://www2.lbl.gov/Science-Articles...ter-solid.html)

High Heat Capacity

Water has the highest **specific heat capacity** of any liquid. Water's high heat capacity is a property caused by hydrogen bonding among the water molecules. **Specific heat** is defined as the amount of heat one gram of a substance must absorb or lose to change its temperature by one degree Celsius. For water, this amount is one **calorie**. It takes water a long time to heat up and a long time to cool down. In fact, the specific heat capacity of water is about five times more than that of sand. This explains why land cools faster than the sea. Due to its high heat capacity, warm-blooded animals use water to disperse heat more evenly and maintain temperature in their bodies: it acts in a similar manner to a car's cooling system, transporting heat from warm places to cool places, causing the body to maintain a more even temperature.

Heat of Vaporization

Water also has a high **heat of vaporization**, the amount of energy required to change one gram of a liquid substance to a gas. A considerable amount of heat energy (586 calories) is required to accomplish this change in water. This process occurs on the surface of water. As liquid water heats up, hydrogen bonding makes it difficult to separate the liquid water molecules from each other, which is required for it to enter the gas phase (steam). Thus, water acts as a heat sink and requires much more heat to boil than liquids such as ethanol, whose hydrogen bonds are weaker. Eventually, as water reaches its boiling point of 100° Celsius (212° Fahrenheit), the heat can break the hydrogen bonds between the water molecules, and the kinetic energy between the water molecules allows them to escape from the liquid as a gas. Even when below its boiling point, water's individual molecules acquire enough energy from other water molecules such that some surface water molecules can escape and vaporize: this process is known as **evaporation**.

Since hydrogen bonds need to be broken for water to evaporate means that a substantial amount of energy is used in the evaporation process. As the water evaporates, energy is taken up by the process, cooling the environment where the evaporation is taking place. In many living organisms, including in humans, the evaporation of sweat, which is 90 percent water, allows the organism to cool so that homeostasis of body temperature can be maintained.

Water is a Solvent

Since water is a polar molecule with slightly positive and slightly negative charges, ions and polar molecules can readily dissolve in it. Water is, therefore, referred to as a **solvent**, because it is capable of dissolving more substances (polar substances) than any other liquid. The charges associated with these molecules will form hydrogen bonds with water, surrounding the particle with water molecules. This is very important as it enables water to dissolve various chemicals and distribute them within living organisms, including taking toxic substances out of living things, and in the environment.

Water's Cohesive and Adhesive Properties

Have you ever filled a glass of water to the very top and then slowly added a few more drops? Before it overflows, the water forms a dome-like shape above the rim of the glass (Figure 8.3.3)



Figure 8.3.3: Water in a glass form a dome shape above the glass due to cohesive forces of attraction among water molecules. Photo Credit: Sam Mutiti

This water can stay above the glass because of its **cohesive** properties. In cohesion, water molecules are attracted to each other (because of hydrogen bonding), keeping the molecules together at the liquid-gas (water-air) interface, although there is no more room in the glass. Cohesion allows for the development of **surface tension**, the capacity of a substance to resist rupture when placed under tension or stress. This is also why water forms droplets when placed on a dry surface rather than being flattened out by gravity (Figure 8.3.3)



Figure 8.3.4: Beading up of water due strong cohesive forces between water molecules (Water USGS, right hand photo credit: J Schmidt; National Park Service).

When a steel needle is placed carefully on water it does not sink even though steel is denser (heavier) than the water. Cohesion and surface tension keep the hydrogen bonds of water molecules intact and support the item floating on the top. It is even possible for an insect to "float" on water if it sits gently without breaking the surface tension, as shown in Figure 8.3.5.



Figure 8.3.5: The weights of the needle and water strider are pulling the surface downward; at the same time, the surface tension is pulling it up, suspending them on the surface of the water and keeping them from sinking. (Credit: Cory Zanker (left) and Tim Vickers (right)

Another important property of water is **adhesion**, or the attraction between water molecules and other molecules. This attraction is sometimes stronger than water's cohesive forces, especially when water is exposed to charged surfaces such as on the inside of thin glass tubes known as capillary tubes. Adhesion is observed when water "climbs" up the tube placed in a glass of water: notice that the water appears to be higher on the sides of the tube than in the middle. This is because the water molecules are attracted to the charged glass walls of the capillary tube more than they are to each other and, therefore, adhere to it. This type of adhesion is called **capillary action**, and is illustrated in Figure 8.3.6. This process is also involved in the movement of water and nutrients from the soil around the root systems to other parts of plants above the ground.

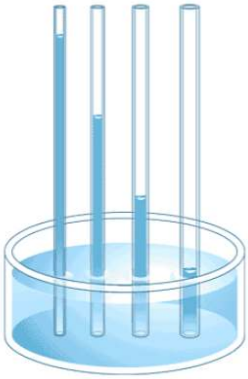


Figure 8.3.6: Capillary action in a glass tube is caused by the adhesive forces exerted by the internal surface of the glass exceeding the cohesive forces between the water molecules themselves. (Credit: moodle.clsd.k12.pa.us/distric...s/0-13-115540- 7/ch23/ch23_s5_1.html)

8.3: The special properties of water is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **8.2: Properties of Water** by Caralyn Zehnder, Kalina Manoylov, Samuel Mutiti, Christine Mutiti, Allison VandeVoort, & Donna Bennett is licensed CC BY-NC-SA 4.0. Original source: https://oer.galileo.usg.edu/cgi/viewcontent.cgi?article=1003&context=biology-textbooks.

SECTION OVERVIEW

8.4: Inland waters and their watersheds

8.4.1: The Physical Template

8.4.2: Watershed Structure

8.4.3: Watershed Functions

8.4.4: Summary

8.4: Inland waters and their watersheds is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.4.1: The Physical Template

The Physical Template

Within the watershed, various forms of matter, including water, are in constant cyclic flow. Through these processes, an abiotic (non-living) template of air, water, and soil is formed, upon which life can exist. The physical template of watershed structure is ultimately determined by varying combinations of climatic, geomorphic, and hydrologic processes.

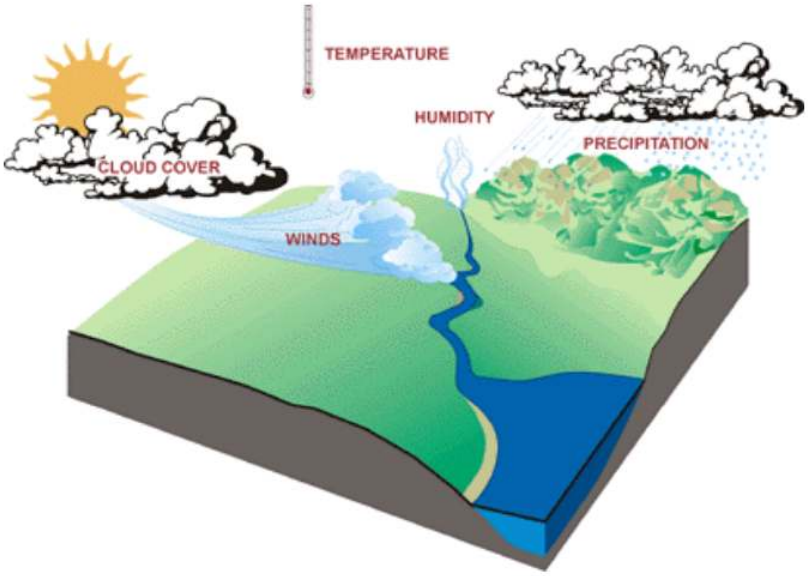


Figure 8.4.18.4.1 Climatic factors. (USEPA Introduction to Watershed Ecology)

Climatology, the science of climate and its causes, becomes important in understanding regional issues in watershed science (Figure 8.4.18.4.1). Though sometimes used synonymously with weather, climate is actually a distinct term with important ecological ramifications. Climate refers to an aggregate of both average and extreme conditions of temperature, humidity, and precipitation (including type and amount), winds, and cloud cover, measured over an extended period of time. Weather refers to present day environmental conditions; current temperatures and meteorological events make up weather, not climate. Long-term weather trends establish averages which become climatic regimes. Climate heavily influences watershed vegetation communities, stream flow magnitude and timing, water temperature, and many other key watershed characteristics.

Geology is defined as the science centered around the study of various earth structures, processes, compositions, characteristics, and histories. Geomorphology, however, refers specifically to the study of the landforms on the earth and the processes that change them over time (Figure 5). Fluvial geomorphology, referring to structure and dynamics of stream and river corridors, is especially important to understanding the formation and alteration of the stream or river channel as well as the flood plain and associated upland transitional zone; this is a critical discipline for effective, long-term watershed management. Figure 3. The physical template determines watershed structure Figure 5. Geomorphology helps explain river and watershed form. Figure 4. Climatic factors

One of the life-sustaining cycles we are most familiar with is the hydrologic cycle (Figure 8.4.28.4.2). This cycle is a natural, solar-driven process of evaporation, condensation, precipitation, and runoff.

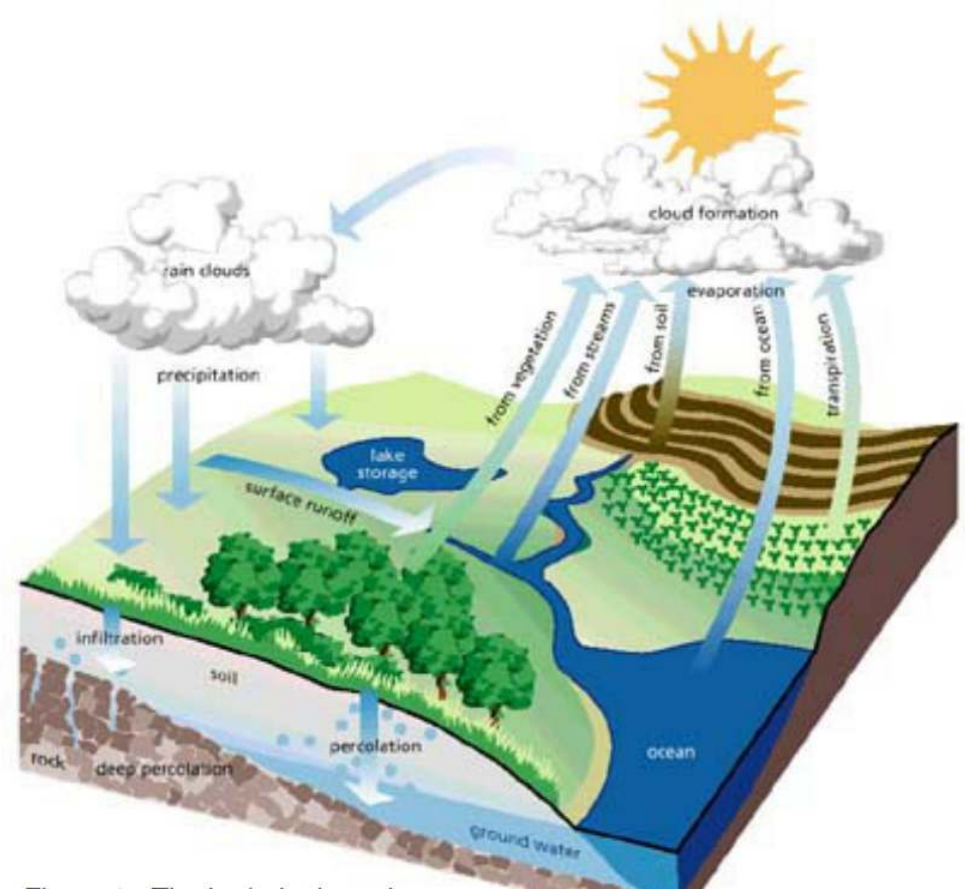


Figure 8.4.28.4.2 The hydrologic cycle. (USEPA Introduction to Watershed Ecology)

Hydrology is the science of water, as it relates to the hydrologic cycle. More specifically, it is the science of water in all its forms (liquid, gas, and solid) on, in and over the land areas of the earth, including its distribution, circulation and behavior, its chemical and physical properties, together with the reaction of the environment (including all living things) on water itself. The global water budget (Table 8.4.18.4.1) adds further insight into the water resources of our planet.

Table 8.4.18.4.1: The global water budget

Saltwater	Percent of global water		Freshwater	Percent of global water
Oceans	97.542		Glaciers/Ice	1.806
Salt lakes/seas	0.004		Ground/Soil: root zone 0-2500 feet 2500-12500 feet	 0.001 0.303 0.331
			Fresh Lakes Reservoirs	0.01 <0.001
			Atmosphere	0.001
			Rivers	<0.001
Total Salt	97.546%		Total Fresh	2.454%

The three elements of the physical template and other factors also interact significantly in determining the structure and composition of a watershed and its biotic communities. In this diagram (Figure 8), the strength of influence on the watershed is represented in two ways: the relative size of the arrows, and the level of control indicated at the right side of the diagram.

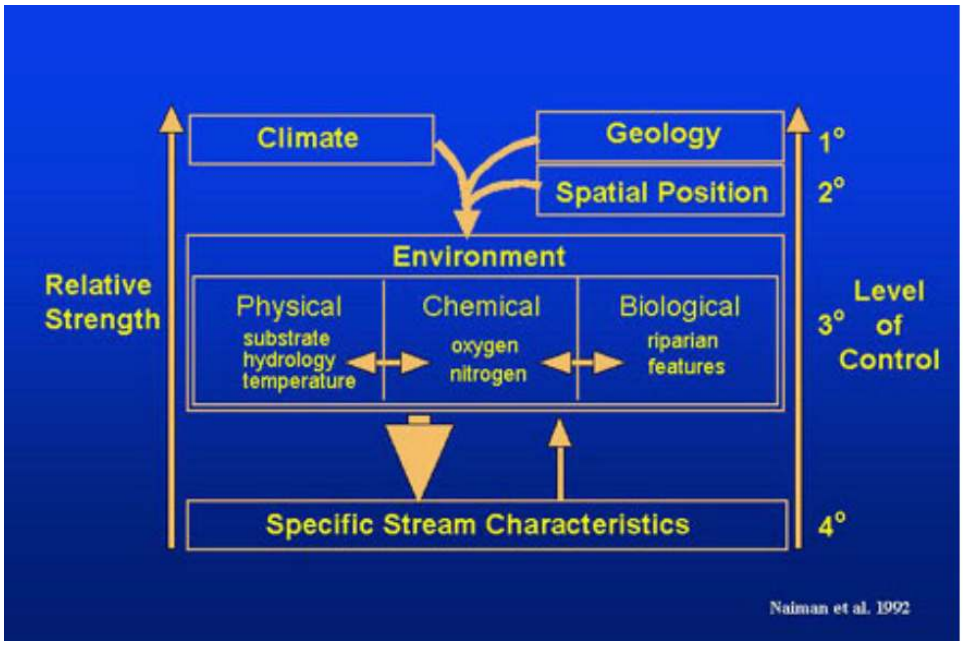


Figure 8.4.38.4.3 How elements of the physical template interact. (USEPA Introduction to Watershed Ecology)

As a result of different combinations of these formative processes, different types of watersheds are created (Figure 9). Here are some examples that show how different from one another watersheds of different origin and physical template conditions can be.

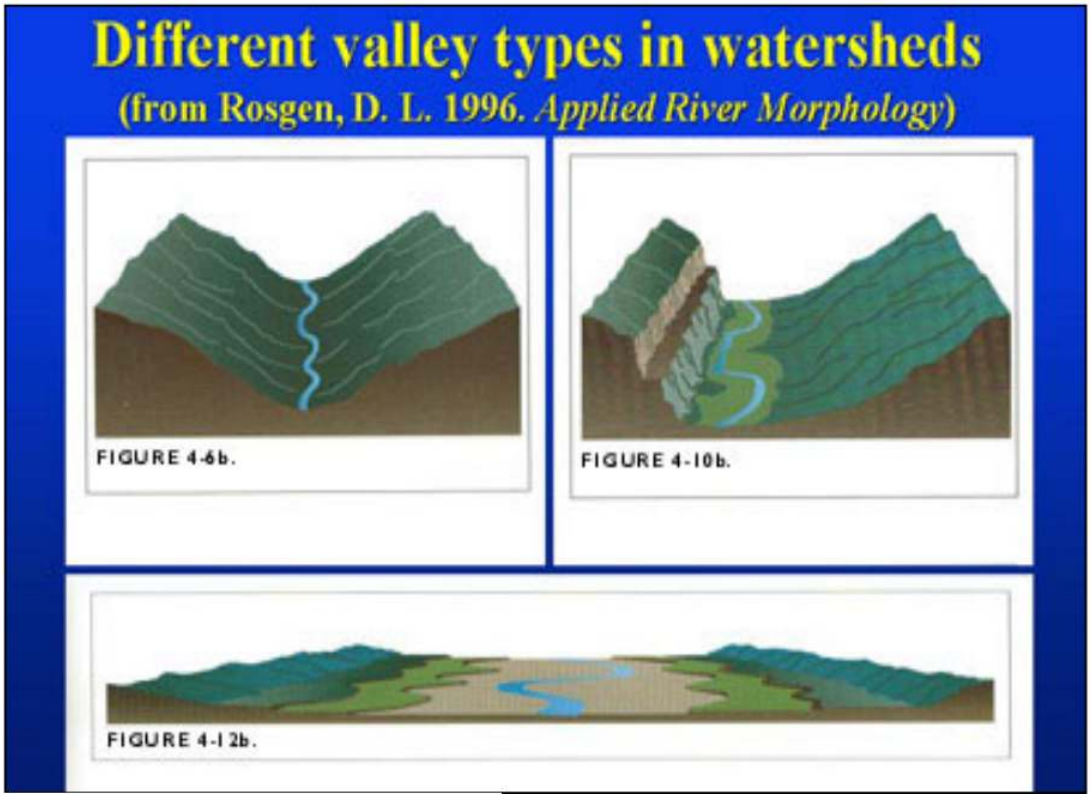


Figure 8.4.48.4.4 6b. Valley Type II, moderately steep, gentle sloping side slopes often in colluvial valleys. 10b. Valley Type VI, moderately steep, fault controlled valleys. 12b. Valley Type VIII, wide, gentle valley slope with well developed floodplain adjacent to river terraces. (USEPA Introduction to Watershed Ecology)

Thomas C. O'Keefe, Scott R. Elliott and Robert J. Naiman**,** Introduction to Watershed Ecology, USEPA Watershed Academy Web, Accessed on December 2023, https://cfpub.epa.gov/watertrain/module

8.4.1: The Physical Template is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

8.4.2: Watershed Structure

Watershed structure includes structure of flowing waters (mainly rivers and streams with associated riverine wetlands and riparian zones), still waters (lakes and associated basin-type wetlands and shorelands), and upland areas of watersheds.

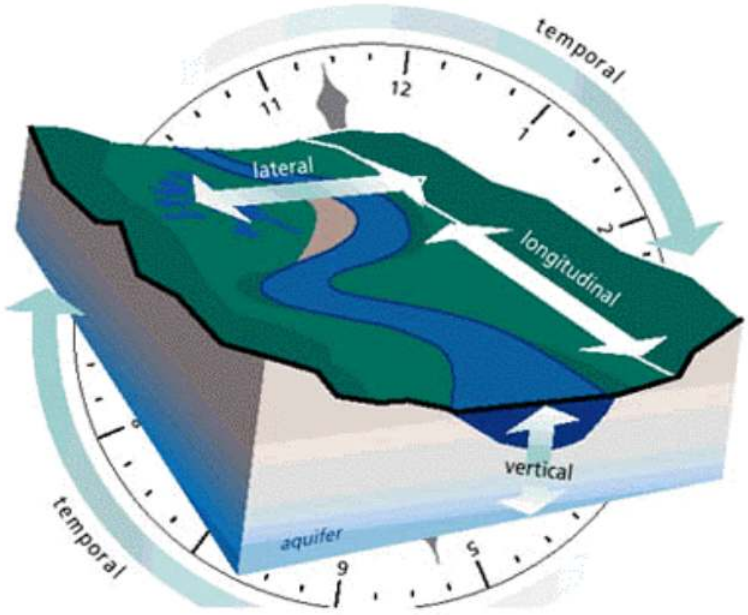


Figure 8.4.2.1: . A four-dimensional concept of watershed structure (after Ward, 1989). (USEPA Introduction to Watershed Ecology, public domain)

Flowing (Lotic) Systems

The US has more than 3.5 million miles of flowing water systems, which include springs and seeps, rivers, streams, creeks, brooks and side channels.

The Four-Dimensional Concept (Figure 8.4.2.1) recognizes that lotic systems' structure exists in a four-dimensional framework, as below:

Longitudinal (in an upstream and downstream direction) - Flowing water systems commonly go through structural changes en route from their source to mouth. Three zones are usually recognized – headwaters, where flow is usually lowest of any where along the system, slope is often steepest, and erosion is greater than sediment deposition; transfer zone, the middle range of the stream where slope usually flattens somewhat, more flow appears, and deposition and erosion are both significant processes; and the downstream end's depositional zone, where flow is highest but slope is minimal and deposition of sediment significantly exceeds erosion most of the time.

Lateral (across the channel, floodplains and hillslopes) - Again, significant variation occurs among stream types, but a common pattern includes the channel, the deepest part of which is called the thalweg; low floodplains that a re flooded frequently, and higher floodplains (e.g., the 100-year or 500-year) that are rarely inundated; terraces, which are former floodplains. A four-dimensional concept of watershed structure (after Ward, 1989) downcutting stream no longer floods; and hillslopes or other upland areas extending up-gradient to the watershed boundary.

Vertical (surface waters, ground water and their interactions) - It is always important to recognize that water bodies are not purely surface features; rivers and streams constantly interact with groundwater aquifers and exchange water, chemicals, and even organisms. Over its entire length, a stream often varies between influent reaches where surface water leaks downward into the aquifer, and effluent reaches where the stream receives additional water from the aquifer.

Temporal (through time, from temporary response to evolutionary change) - The dimension of time is important because rivers and streams are perpetually changing. Structure as described in the other three dimensions above should never be considered

permanent, and watershed managers should always think of structure not just as what is there now, but in terms of the structural changes in progress and their rates of occurrence.

Recognition of different types of streams and rivers is mostly reliant on channel form and function. For more on stream and river categories see the Stream Channel Classification box on the following page.

Structure in Upland Areas of Watersheds



Figure 8.4.2.2: . Aerial views reveal the complexity of landscape pattern, which implies many traits of a watershed's ecological functions. (USEPA Introduction to Watershed Ecology, public domain)

The physical form of the uplands in watersheds can vary greatly, in ways beyond the scope of this discussion. Here we focus only on the distribution of and variations in vegetation and land use, which together create the element of watershed structure called landscape pattern (Figure 8.4.2.2). Vegetation and land use patterns in watersheds are known to have many significant influences on the condition of the water bodies they drain into; this topic is explored in greater detail in the Academy 2000 module on Watershed Change (http://www.epa.gov/watertrain/a gents).

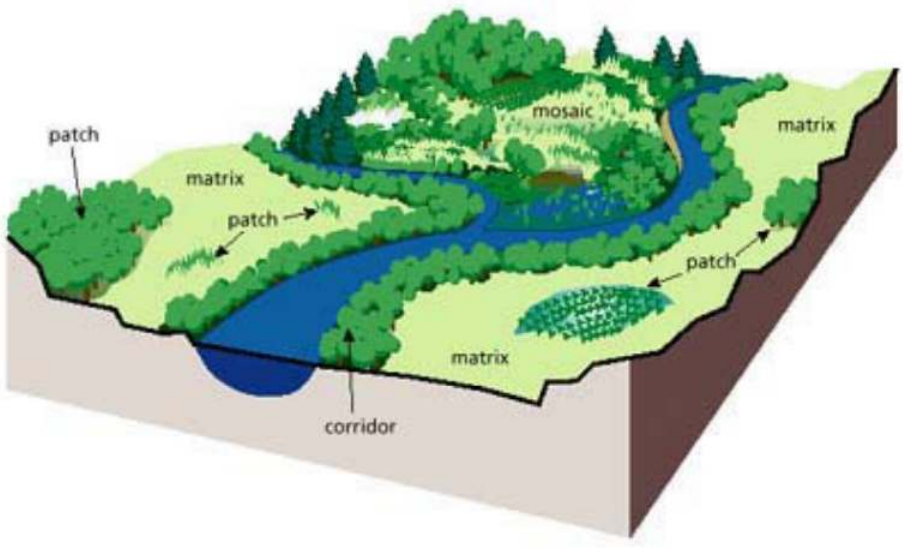


Figure 8.4.2.3: Useful units for describing landscape structure and pattern. (USEPA Introduction to Watershed Ecology, public domain)

Landscape patterns. Landscape ecology offers a simple set of concepts and terms for identifying basic landscape patterns: matrix, patch, and mosaic (Figure 8.4.2.3). The ecological term matrix refers to the dominant (> 60 percent) land cover, while a patch is a nonlinear area that is less abundant and different from the matrix. A mosaic is a collection of different patches comprising an area where there is no dominant matrix. Various patch types have been described in the Examples of Different Patch Types box below. Basically, the most obvious landscape patterns are formed by combinations of native vegetation communities, unvegetated areas, and land use patterns.

Landscape pattern change. The individual patches in a landscape can change, and so can the entire landscape change in pattern and/or composition. Disturbances and various landscape processes maintain a constant dynamic, referred to as a shifting mosaic. Some landscapes remain in a "dynamic equilibrium" and, although changing steadily from place to place, retain an important quality called mosaic stability. A well-managed forestry operation, for example, would exhibit over the long term a constantly shifting set of locations where mature forest occurred, but at the same time sustains the relative proportions of forested and nonforested land in the area. Or, a landscape may evolve toward a new type of pattern and composition (e.g., via timber clearcutting, suburban sprawl, abandonment and succession of agricultural lands back to forest, or landscape change due to disease, fire, or global warming). It is always important, when analyzing landscape pattern and landscape change, to remain aware that the spatial resolution of your information (how small a landscape feature you can detect) may or may not be sufficient to detect all the landscape changes of possible significance that may be occurring.

Vegetational patterns. Upland vegetation structure varies spatially, following various biogeographical patterns based on climate, physiography, soils, disturbance regimes, and their interactions (Figure 8.4.2.4). Vegetation communities are areas where a few species of plants dominate and establish a characteristic form or structure, within which a potentially large number of less abundant organisms also exist. Nationwide, there are hundreds of vegetation community types; the Society of American Foresters recognizes over 80 forest types alone (SAF, 1980). As a first step in analyzing vegetational patterns, it is easier to recognize a few generalized upland vegetation types based on their growth form, including:

- Forests (deciduous, evergreen and mixed)
- Shrublands
- Grasslands
- Forbs (broad-leaved herbs)

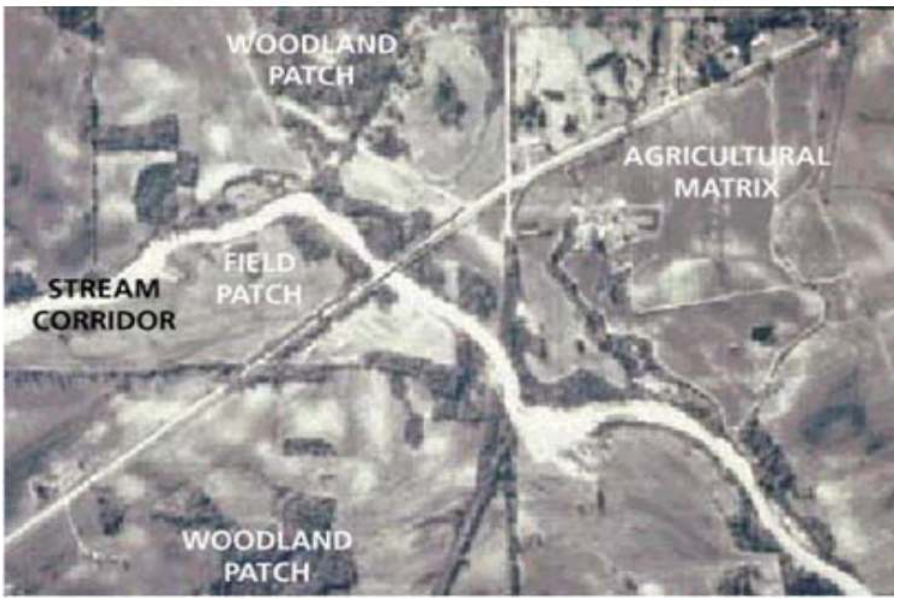


Figure 8.4.2.4:Examples of matrix, patch, and corridor. (USEPA Introduction to Watershed Ecology, public domain)

These categories are commonly found on land cover maps likely to be available in the GIS data for most watersheds, and can be consulted to give a general sense of vegetation patterns in the watershed. Human activity has carved up and fragmented many of the natural vegetation patterns that formerly covered our watersheds. Without human influence, however, vegetation patterns would not be uniform due to different vegetation communities arising from different environmental conditions (e.g. variations in moisture

and temperature due to slope and aspect) and events (e.g., fire, pest outbreak). In the West, the “rain shadow” is a common, basic example of how vegetation varies with physical position.

Example: Southern Sierra As moisture-laden air is pushed east, it begins to slowly rise as it reaches the graduallysloping foothills of the southern Sierra. As the air rises, pressure increases, and moisture in the air begins to condense. As the air continues eastward, it continues upward to the crest of the range, and moisture within the air is seemingly wrung-out, much like a sponge. The eastern slope of the Sierra is quite dramatic, topographically. Indeed, Mt. Whitney, the highest point in the coterminous United States, is just a matter of miles from North America’s lowest point (Death Valley, CA). Much of the eastern slope is quite dry, since the air is now deprived of much of its moisture.

Land-use patterns. Increasingly, the landscape structure and pattern we see is the result of widespread human activity. In all fields of environmental management including watershed management, analysis of land use types, patterns, and trends is commonplace. Because multiple uses occur in many locations and some land uses are not in themselves a visible landscape feature, mappers often use term land cover to describe the delineation of landscape structure and pattern formed by the dominant land uses and remaining vegetation communities. Some common land cover categories (indicating land uses within the areas) include:

- urban land (residential, commercial, industrial, mixed)
- agriculture (row crops, field crops, pasture)
- transportation (roads, railroads, airports)
- rangelands
- silviculture
- mining/extractive areas

Like vegetation patterns, the land use patterns in a watershed can be studied through GIS data or maps. Human-dominated landscapes, just as natural landscapes, are shifting mosaics that often progress through a series of changes in what is dominant.

Example: Rural Upstate New York Pre-settlement landscapes in upstate New York were probably almost entirely forested. By the mid to late 19th century, clear-cutting had removed the forest from most areas. In the first part of the 20th century, cropland and pasture became widespread and dominated 70% or more of the landscape, with remnant scattered forest patches, wetlands and lakes. By the second half of this century, a trend in migration into cities and towns led to an assorted pattern of landscapes, including urban/suburban-dominated areas with forest and agricultural patches, heavily agricultural landscapes, and transitional rural lands with decreasing agriculture and increasing forest regrowth.

Excerpted from:

Thomas C. O'Keefe, Scott R. Elliott and Robert J. Naiman**,** Introduction to Watershed Ecology, USEPA Watershed Academy Web, Accessed on December 2023, https://cfpub.epa.gov/watertrain/module

---

8.4.2: Watershed Structure is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

8.4.3: Watershed Functions

Among the most essential functions that occur in healthy watersheds are:

- Transport and storage
- Cycling and transformation
- Ecological succession

Transport and storage (of water, energy, organisms, sediments, and other materials)

Because a watershed is an area that drains to a common body of water, one of its main functions is to temporarily store and transport water from the land surface to the water body and ultimately (for most watersheds) onward to the ocean. But, in addition to moving the water, watersheds and their water bodies also transport sediment and other materials (including pollutants), energy, and many types of organisms. It is important when recognizing the transport function to also recognize temporary retention or storage at different locations in the watershed



Figure 8.4.3.1: Sediment transport and deposition. USEPA, Introduction to Watershed Ecology, Public Domain

Transport and Storage. As matter physically moves through the watershed, there are a number of terms which arise relative to various stages of cycling. Availability refers not just to the presence of an element in a system, but also speaks to the usability of a given agent. For instance, nitrogen gas may be plentiful in and around dam spillways, but N2 is not a usable form for most aquatic organisms, and thus the availability of nitrogen is compromised. Detachment refers to the release of matter from an anchoring point, and its subsequent movement. Transport, a process most evident in stream channels, involves the movement of a material through a system. Deposition refers to a given endpoint within a cycle. Integration refers to the assimilation of matter into a site or organism following depositional processes (see Naiman and Bilby 1998). An example using these terms is included below.

Example: Seston Seston (suspended particulate matter) is an important food source for a few key aquatic organisms. In the context of the above continuum, seston's availability within a given stream system may refer to the source. In a hypothetical example, nitrogen and phosphorus, both limiting in many freshwater ecosystems, are made available to a stream during a period of elevated discharge. Attached algae (periphyton) begins to grow on rocks and submerged surfaces (availability). During growth and maturation, algal cells begin to slough off from the rocks and logs (detachment) and the current carries these cells, rich in chlorophyll, downstream (transport). Hydropsychid caddisfly larvae, which feed on seston, spin elaborate nets (much like a spider's web), anchoring these nets on cobble substrates or other submerged features. These nets catch the suspended particulates as they drift with the current (deposition). The caddisfly larvae then feed off particles captured in the net (assimilation); algal cells represent high quality food for these organisms and they may, as a population, exert an influence over the relative abundance of various particle-types.

Transport and storage of water. One can view a watershed as an enormous precipitation collecting and routing device, but transportation and storage of water actually involves a complicated mix of many smaller processes (which are bolded in the following text). Even before precipitation reaches the ground (Figure 28), it interacts with vegetation. Trees and other vegetation are responsible for interception and detention of some of the rainfall, leading to some evaporation and also slowing the amount reaching the ground via throughfall and giving it time for better infiltration to groundwater (one form of storage). Saturation of soils, occurring when precipitation exceeds infiltration, leads to overland flow and, over longer time frames, drainage network development (Figure 8.4.3.2). The consistent flow of water in channels affects and shapes channel development and morphology in ways that seek dynamic equilibrium with the job to be done (moving water downstream). Recall also this module's earlier discussion of the longitudinal profile development of rivers and streams, and how upper, middle and lower zones of streams generally have very different forms to handle very different sets of functions, many related to transport and storage of water.

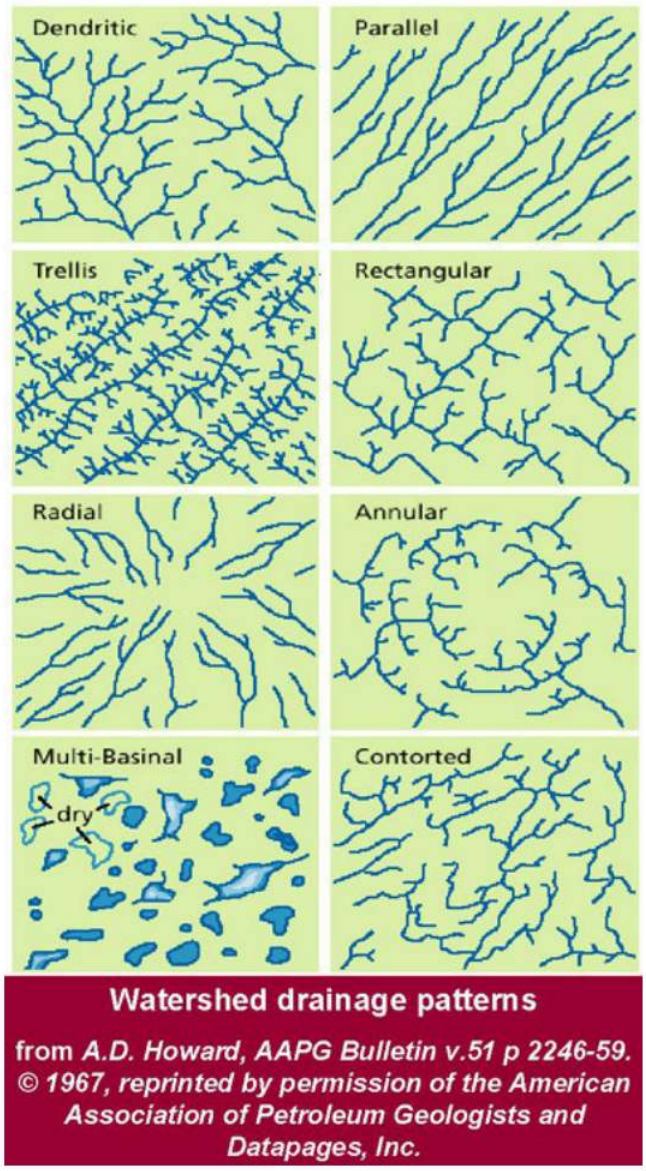


Figure 8.4.3.2: Common drainage patterns USEPA, Introduction to Watershed Ecology, Public Domain

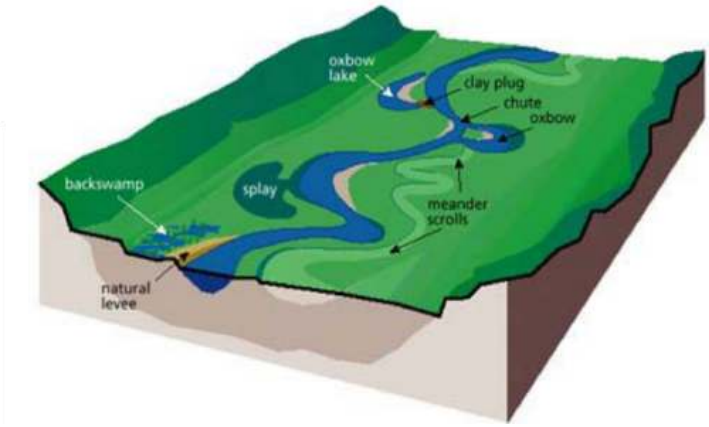


Figure 8.4.3.3: Dynamics of precipitation. USEPA, Introduction to Watershed Ecology, Public Domain

Transport and storage of sediments. Watersheds also collect and transport sediments as a major function. Sediment transport and storage is a complex network of smaller watershed processes, like the water processes described above, and actually is inseparable from water transport and storage. Sediment related processes mostly involve erosion and deposition, but sediment transport and storage also play a longer-term role in soil development.

The drainage network development and channel development (Figure 8.4.3.2 and 8.4.3.3) discussed above appears to be dominated by erosion at first glance, but the redeposition of sediments on floodplains is an important function that rejuvenates soils and influences the productivity and diversity of stream corridor ecosystems.

Cycling and Transformation.

Cycling and transformation are another broad class of natural functions in watersheds. Various elements and materials (including water) are in constant cycle through watersheds, and their interactions drive countless other watershed functions. Figure 31, for example (overleaf), illustrates interactions of the carbon and nitrogen cycles with stream biota and the resulting influence on dissolved oxygen. Elements like carbon, nitrogen, and phosphorus comprise the watershed's most important biogeochemical cycles. Cycling involves an element of interest's transport and storage, change in form, chemical transformation and adsorption.

Nutrient Spiraling. The flow of energy and nutrients in ecosystems are cyclic, but open-ended. True systems, in both an environmental and energetic context, are either "open" (meaning that there is some external input and/or output to the cyclic loop) or "closed" (meaning that the system is selfcontained). In watersheds, streams and rivers represent an open-system situation where energy and matter cycles, but due to the unidirectional flow, the matter does not return to the spot from whence it came. Also, nutrients "spiral" back and forth among the water column, the bodies of terrestrial and aquatic organisms, and the soil in the stream corridor en route downstream. Hence, the concept of nutrient "spiraling" implies both movement downstream and multiple exchanges between terrestrial and aquatic environment, as well as between biotic and abiotic components of the watershed.

The Cycling of Carbon and Energy. In food webs, carbon and the subsequent synthesized energy is cycled through trophic (food web) levels. Energy transfer is considered inefficient, with less than 1 percent of the usable solar radiation reaching a green plant being typically synthesized by consumers, and a mere 10 percent of energy being typically converted from trophic level to trophic level by consumers.

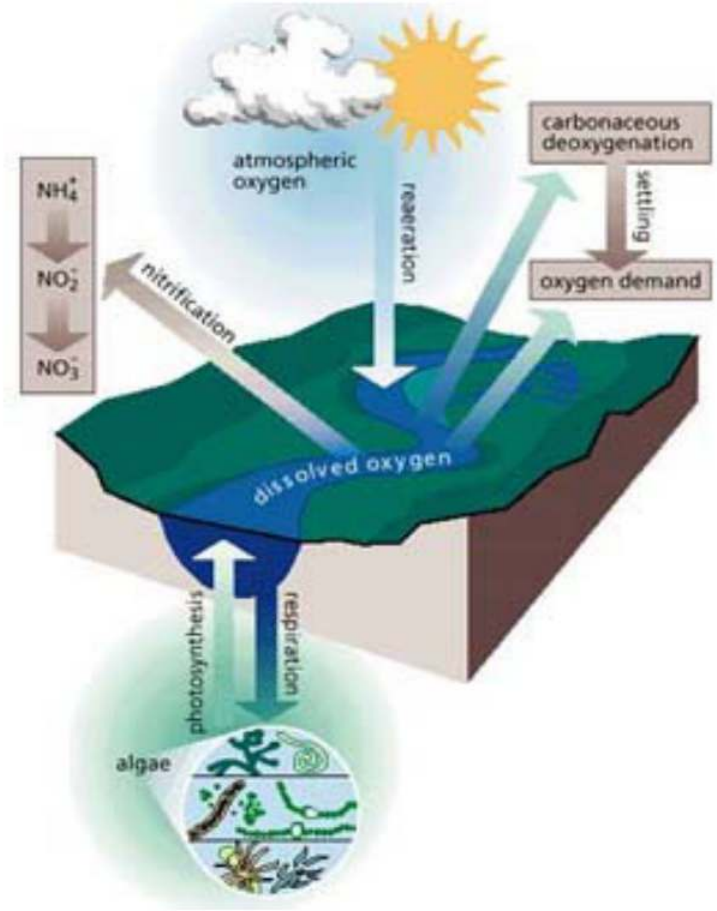


Figure 8.4.3.4: Carbon and nitrogen cycles affect stream biota. USEPA, Introduction to Watershed Ecology, Public Domain

Nitrogen and Phosphorus limitation. Most watershed systems (both the aquatic and terrestrial realms) are either N or P limited, in that these are the required elements which are at the lowest availability. As a general rule, the N:P ratio should be 15:1. A lower ratio would indicate that N is limiting, a higher ratio places P in that role. Commonly P is the limiting factor. Often, the slightest increase in P can trigger growth, as in algal blooms in an aquatic setting. In N and P limited systems, an input of either element above and beyond normal, "natural" levels may lead to eutrophication. The stream corridor is often a mediator of upland-terrestrial nutrient exchanges. As N and P move down through subsurface flow, riparian root systems often filter and utilize N and P, leaving less to reach the stream. This has a positive influence on those already nutrientoverloaded bodies of water, but would not necessarily be a positive influence on organisms struggling to find food in very clean, nutrient-limited headwaters streams. Microbes also denitrify significant amounts of N to the atmosphere . Still, N-fixers, like alder, may serve as sources of N for the stream channel, and groundwater pathways between the stream and the streamside forest may provide significant quantities of nitrogen.

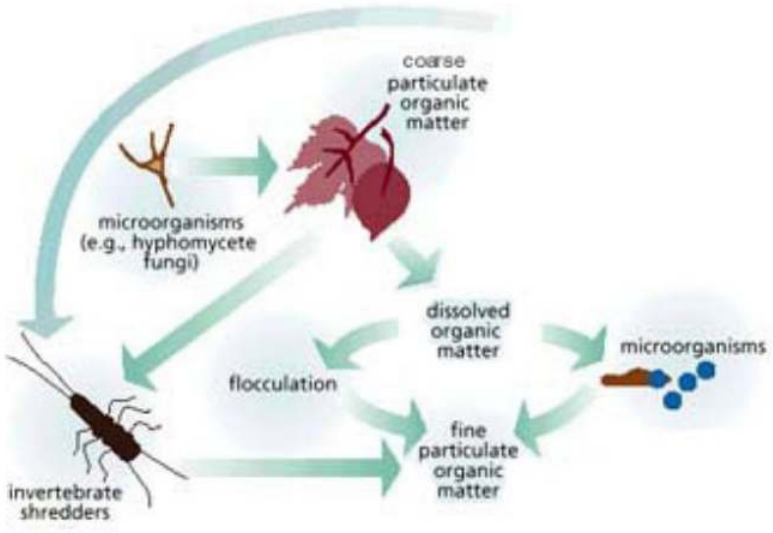


Figure 8.4.3.4: A simplified decomposition process. USEPA, Introduction to Watershed Ecology, Public Domain

Decomposition (Figure 8.4.3.4). Decomposition involves the reduction of energy-rich organic matter (detritus), mostly by microorganisms (fungi, bacteria, and protozoa) to CO2, H2O and inorganic nutrients. Through this process they both release nutrients available for other organisms and transform organic material into energy usable by other organisms. In lakes, much of the decomposition occurs in the waters prior to sedimentation. In the headwater reaches of streams, external sources of carbon from upland forests are a particularly important source of organic material for organisms and decomposition of microscopic particles occurs very rapidly. The bacteria and fungi modify the organic material through decomposition and make it an important food source for invertebrate and vertebrate detritivores, thereby reinserting these nutrients and materials into the watershed's aquatic and terrestrial food webs.

Decomposition is influenced by moisture, temperature, exposure, type of microbial substrate, vegetation, etc. Specifically, temperature and moisture affect the metabolic activity on the decomposing substrate. Nutritional value (as well as palatability) of the decomposing structure will also affect the time involved in complete breakdown and mineralization. Decomposition involves the following processes:

- The leaching of soluble compounds from dead organic matter
- Fragmentation
- Bacterial and fungal breakdown
- Consumption of bacterial and fungal organisms by animals
- Excretion of organic and inorganic compounds by animals
- Clustering of colloidal organic matter into larger particles

The process of death and consumption, along with the leaching of soluble nutrients from the decomposing substrate, release minerals contained in the microbial and detrital biomass. This process is known as mineralization.

Excerpted from:

Thomas C. O'Keefe, Scott R. Elliott and Robert J. Naiman**,** Introduction to Watershed Ecology, USEPA Watershed Academy Web, Accessed on December 2023, https://cfpub.epa.gov/watertrain/module

---

8.4.3: Watershed Functions is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

8.4.4: Summary

This section on watershed has introduced the most basic features of an exceedingly complex type of natural system -- the physical template, and watershed structure and function. In closing, examine the graphic of watershed dynamics below and think particularly about all the interactions occurring even in this simplified example. It is truly important to appreciate the natural processes at work, and how they are beneficial to our communities as well as our ecosystems. Even more, it is crucial to recognize how change affects watersheds and can jeopardize these benefits in very costly ways, when a normal change becomes great enough to be a change of concern.

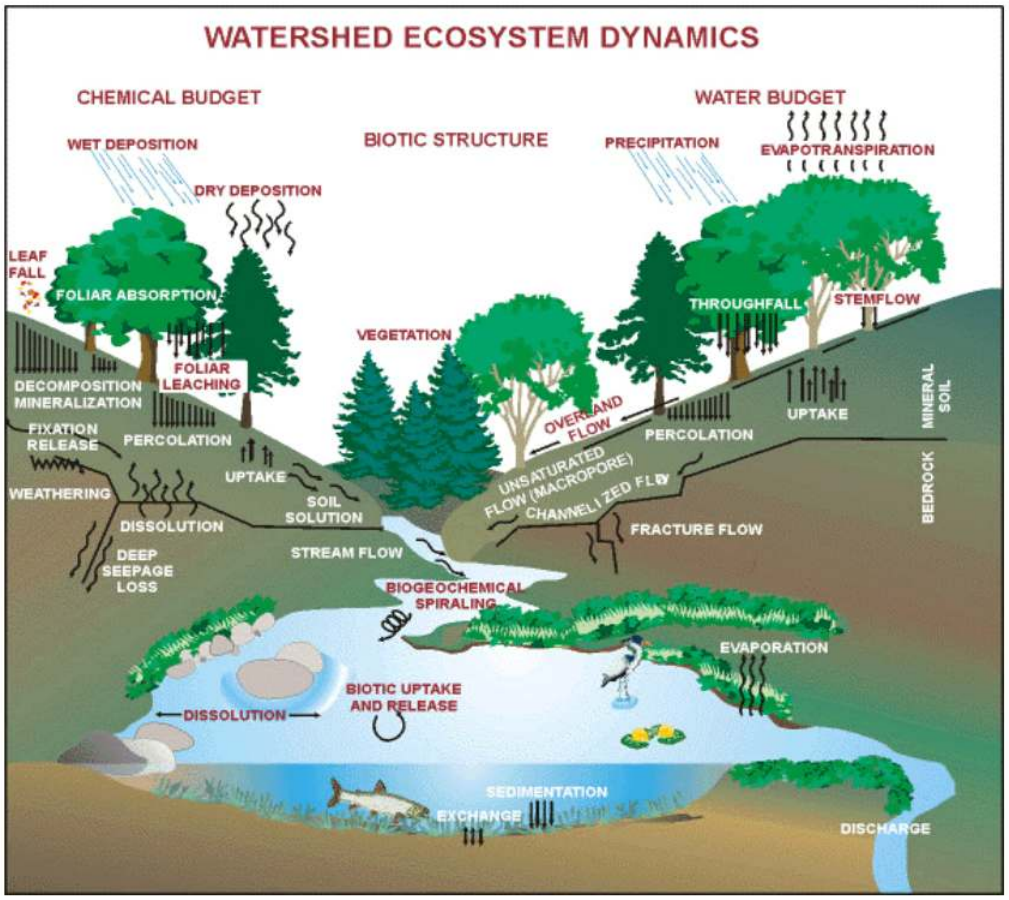


Figure 8.4.4.1: Watershed ecosystem dynamics summary (redrafted from Johnson and Van Hook, 1989. Analysis of biogeochemical cycling processes in Walker Branch Watershed.) US EPA Introduction to Watershed Ecology Public Domain

Excerpted from:

Thomas C. O'Keefe, Scott R. Elliott and Robert J. Naiman**,** Introduction to Watershed Ecology, USEPA Watershed Academy Web, Accessed on December 2023, https://cfpub.epa.gov/watertrain/module

8.4.4: Summary is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

8.5: Lakes

8.5: Lakes is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.5.1: Freshwater Lakes

Introduction

Freshwater lakes are inland bodies of water, typically formed from depressions in the earth's surface. These basins usually form from continental extension, strike slip faulting, or as sag basins. They are bounded by sills, and usually fed by streams and rivers from the surrounding environment. If there are channels that output water and sediment, the lake is hydrologically open. Hydrologically closed lakes (only water and sediment input) may also be freshwater if their dissolved ion concentrations are low enough. Freshwater has a solute concentration below approximately 5 grams per liter.

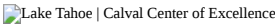


Figure 1: Lake Tahoe. https://calval.cr.usgs.gov/apps/lake-tahoe

Water is the primary media of sediment transport into and out of lakes, but wind may also deposit or erode sediment. Liquid water flow from channels is turbulent and depends on channel width, depth, and flow speed. Flow is largely constrained to above 10 to 20 meters lake depth; this upper oxygenated portion is called the epilimnion. Wind driven surface currents may be up to 30 centimeters per second in speed. They produce waves which may transport sediment around the lake, but wind usually only erodes sediment above the water level. Waves also mix the water and facilitate diffusion of oxygen in the upper layers of the lake.

Sediment Transport Processes

Sediment is usually carried to and from lakes by water from channels, but wind may also transport fine sediment. Flows from channels that move sediment are turbulent, with coarser grains under upper planar flow and fine grains being carried in suspension. Slower flow speeds produce dunes and current ripples, which are asymmetrical. Waves are also responsible for shaping lake shorelines, by both eroding and depositing sediment to make beaches. Waves may form from both air and water flows over the lake. They tend to produce wave ripples, which are symmetrical.

Grains move via in several ways within channels and lakes, depending upon the lift force produced by the transporting fluid:

- **Rolling** occurs when a grain rolls across the bed's surface (yes, really). They roll due to friction between the substrate and the fluid flow.
- **Saltation** occurs when a grain is lifted off the bed surface and falls back down in a series of jumps. The lift force isn't sufficient to carry the grain indefinitely.
- **Suspension** is when a grain is lifted into a flow and carried continuously. This is common for clay sized grains, but larger grains may be suspended in more turbulent flows.



Figure 2: Colorado River Draining into Lake Mead. https://www.nps.gov/lake/learn/understanding-the-lakes.htm

add a description of what the reader should pay attention to, e.g. why is part of it red, what is the light band above the water level

Figure 2 shows the Colorado River draining into Lake Mead. Note the plumes of fine rust colored sediment being transported in suspension. The lighter bands of rock above the shoreline are previous shorelines. Fine to coarse grained sediment may be carried into a lake. Fine sediment is carried in the epilimnion for away from the lake margins, while coarse sediment immediately drops out of suspension into the hypolimnion and forms turbidites. Grading of beds is typically fining upwards, especially near the mouths of rivers transporting sediment.

Lake levels are determined by inputs of water and sediment. Many freshwater lakes rely on sediment and fluid input from rivers, so I suggest visiting the chapter on them here for more information.

Waves are also really important, helping make beaches and erode shorelines. I recommend adding something about them, especially since you mention wave ripples as one of the most common structures in lake deposits.

Characteristics of Deposited Sediment

Sand and mud sized grains are most common, but carbonates, evaporites, and organic material may also be present. Sand is most common on the shoreline, but may also form parts of the lake floor in shallower areas. Mud may be in deep lake or overbank floodplain deposits. Cobbles and gravel occur in turbidites, mostly at their bases.

Structures in lake deposits resemble those in other low energy environments, but with their own unique features. Wave ripples and parallel laminations are the most common structures. Wave ripples only appear in shallow or nearshore lake deposits. This is because they form from wind currents, which produce elliptical motion at the lake's surface. This is gradually translated vertically into standing waves at the bed surface, leading to symmetrical ripples.

Other structures include root casts, mud cracks, and signs of bioturbation. Root casts and mud cracks form when the lake's shoreline recedes, exposing finer silt and clay to air. Plants colonize this exposed sediment and leave root casts. Mud cracks form due to the desiccation of the sediment, but because clay is cohesive, the bed doesn't fall apart. Bioturbation is when organisms dig into and churn sediment to make living burrows, escape from predators, or forage for food.



Figure 3: Feeding traces from an unknown animal; Death Valley, CA. http://www.pitt.edu/~cejones/GeoImages/5SedimentaryRocks/SedStructures/Bioturbation1.html

Be sure to describe where the root cases and mud cracks form relative to the lake shoreline. They require changes in lake level to form, typically, so adding something about what determines lake level would be helpful (especially given your image of Lake Mead above)

The most common organisms found in lake sediments are gastropods, bivalves, and ostracods. Shells and calcareous algae also form the source of much calcareous mudstone called , which may be deposited at the margins or deeper depths of lakes. Vertebrates are also known. Charophytes, a type of green algae, are also indicative of freshwater lake environments due to their intolerance of saline water.



Figure 4: Devonian charophytes from Rockford, IL. https://www.eeob.iastate.edu/faculty/DrewesC/htdocs/fossil-buttons.htm

Typical Vertical Sequence of Freshwater Lake Facies

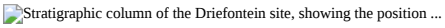


Figure 5: A stratigraphic column showing fluviolacustrine deposits. https://www.researchgate.net/publication/221818423_The_Earliest_Post-Paleozoic_Freshwater_Bivalves_Preserved_in_Coprolites_from_the_Karoo_Basin_South_Africa/figures?lo=1

It would be very helpful to connect the movement of the shoreline and fining/coarsening upward cycles to lake level changes

Flow typically transitions from more to less energetic over time. This is reflected in bedding layers as fining upward, where gravel transitions to sand, then silt and mud. Figure 5 above shows fluviolacustrine deposits from a Triassic locality in South Africa, specifically those in the Burgersdorp formation. You can see both abrupt and continuous changes in grain size and sedimentary structures. The sinuous lines represent erosional surfaces, while the linear ones represent gradual change. The coarsening upward beds, changing from mudstone to fine sand, show an increase in flow speed and turbulence, suggesting a regression of the shoreline. Conversely, the fining upward beds suggest a transgression due to the lack of turbulent flow allowing mud and silt to settle from suspension. Lake shorelines fluctuate frequently over time, which we can see through the presence of erosional surfaces, changes in grain size, and sedimentary structures. These differences in grain size and structures reflect lake margin and deep lake facies. Lake margin facies are typically coarse grained with sand or gravel, and their most common structures are wave ripples and current ripples. Deep lake facies typically consist of fine mud and frequently have planar laminations.

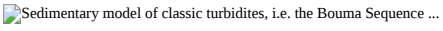


Figure 6: The Bouma Sequence. https://www.researchgate.net/publication/335685532_Verification_of_gravity-flow_models_case_study_from_the_Lower_Eocene_sediments_Tylmanowa_site_SE_Poland

Deep lake facies consist of clay and silt sized sediments, with their main structures being parallel laminations. The Bouma sequence above depicts a turbidite deposit, with gravel and coarse sand transitioning to very fine sand and then mud. These deposits form when debris flows drop sediments down a sloping surface, frequently near the mouths of rivers depositing sediment. The gravel and coarse sand drops from suspension first, then the fine sand, with the mud eventually settling from suspension when the flow loses enough energy. Turbidites are explained in more detail here. Variations in this sequence also appear in alluvial deposits, explained in more detail here. Overbank deposits in lakes may also form fine laminated mud, but unlike deep lake facies, the mud frequently has root casts.

Case Study: The Yixian Formation

This is where things get interesting!

For more than the past 30 years, many exceptional fossils have been found in China. One of their most famous and productive formations is the Yixian Formation located in the Liaoning province of northeastern China. It's early Cretaceous in age, between approximately 139-122 Ma. It is divided into at least three members: the lowermost Lujiatun Member, composed of fluvial and volcaniclastic rocks, the Xiatulaigou Member, composed of basaltic lava flows and associated breccias, and the Jianshangou Member (Wang et al, 2016).

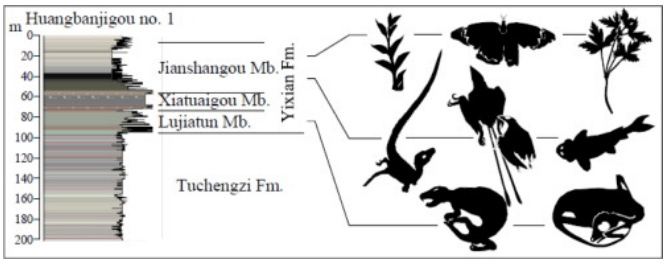
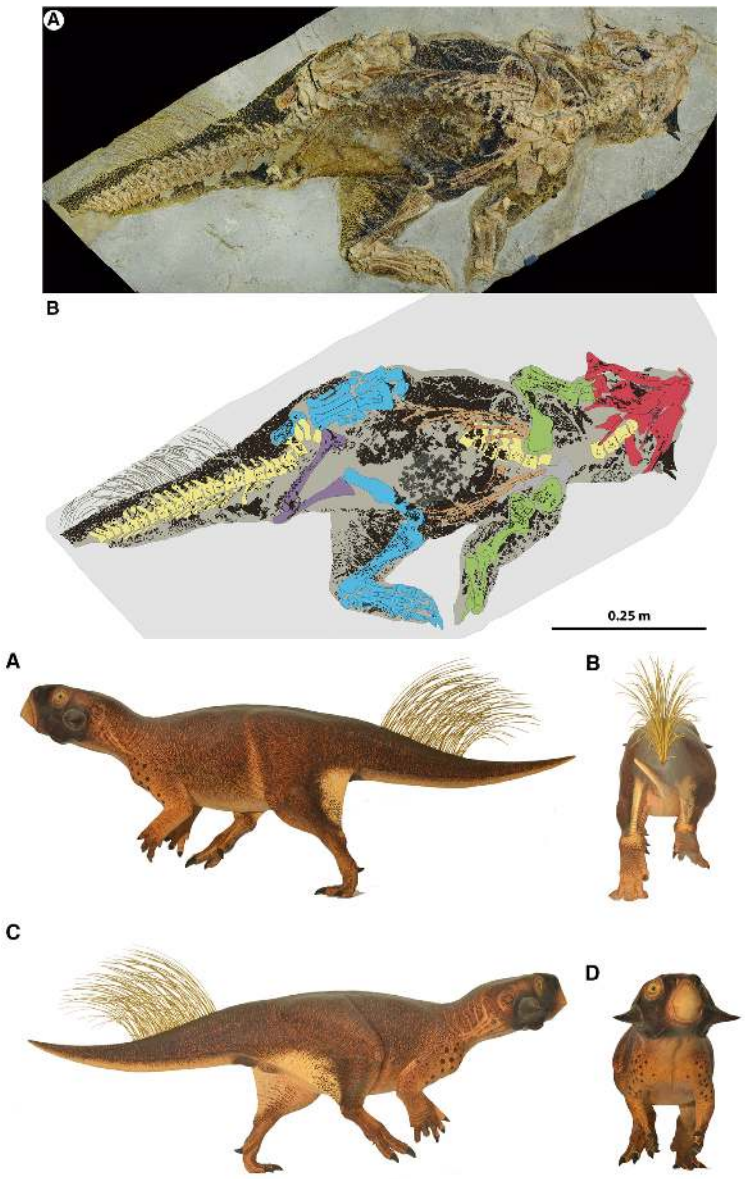


Figure 7: Diagram showing the Yixian and underlying Tuchengzi formations in a stratigraphic column. Taken from Wang et al. 2016.

The **Jianshangou Member** is a lacustrine component of the Yixian formation, and is highly fossiliferous. It has 4 subunits, or beds, with distinctive characteristics (Wang et al, 2016). The basal most is the Dajianshanzi bed, which is largely composed of volcaniclastic sediments and has freshwater bivalves. Then there is the Anjiagou Bed, composed of fine laminated mud interbedded with fine volcanic ash, which produced the first known feathered dinosaur *Sinosauropteryx* and birds such as *Confuciusornis*. Further up section is the Hengdaozi Bed, consisting of light mudstones and sandstones interbedded with ash. It preserves large amounts of insect and angiosperm fossils. Finally, the Huangbanjigou Bed consists of mudstones with varying degrees of lamination along with ash. The Jianshangou Member generally preserves a remarkably complete fossil record of the ecosystem, including gymnosperms, angiosperms, insects, molluscs, non-avian dinosaurs, birds, and mammals. Collectively, they're referred to as the **Jehol Biota**. It is the dinosaurs from this and similar beds that revealed the prevalence of feathers among non-avian dinosaurs, and they contain some of the earliest known flowering plants as well, such as *Archaefructus*.

One of the dinosaurs preserved in the Yixian beds is *Psittacosaurus*, a small dinosaur related to *Triceratops*. Several species are known. The exceptional preservation occurs due to the very fine mud and volcanic sediments, along with a very low energy depositional environment. This individual likely died and stayed in place, and was buried by successive layers of mud and ash settling from suspension. This allowed not only the bones, but imprints of the scaly skin and other soft tissues to be fossilized. The skin is preserved as residues of calcium phosphate, which the scales contained in life (Vinther et al, 2016). The preservation is so fine that impressions of the melanosomes, the tiny organelles that produce pigment, are visible under an electron microscope (Vinther et al, 2016)! Additionally, the scales show conspicuously different textures on different areas of the body.



Figures 8 and 9: *Psittacosaurus sp*. SMF R 4970 under cross polarized light and a model reconstruction. The darker portions of the specimen in Figure 7 are melanin residues, showing pigmentation variations on different areas of the body. The dinosaur was buried on its back, and the body was slightly crushed by the overburden. Both figures taken from Vinther et al. 2016.

Summary

Freshwater lakes consist of inland bodies of water supplied by rivers and streams. Waves are common physical phenomena that shape the deposition and sedimentary structures of beds. If we're incredibly lucky, as we are with the Yixian formation, lakes can provide us with a magnificent view of prehistoric life almost literally in the flesh.

References

Bohacs, K. M. (1999). Lake-basin type, source potential, and hydrocarbon character; an integrated sequence-stratigraphic-geochemical framework. *AAPG Bulletin*, *83*(11), 1878. doi:10.1306/E4FD42AB-1732-11D7-8645000102C1865D

Li, Q., Gao, K.-Q., Meng, Q., Clarke, J. A., Shawkey, M. D., D'Alba, L., . . . Vinther, J. (2012a). Reconstruction of Microraptor and the evolution of iridescent plumage. *Science (New York, N.Y.)*, *335*(6073), 1215. doi:10.1126/science.1213780

Li, Q., Gao, K.-Q., Meng, Q., Clarke, J. A., Shawkey, M. D., D'Alba, L., . . . Vinther, J. (2012b). Reconstruction of Microraptor and the evolution of iridescent plumage. *Science (New York, N.Y.), 335*(6073), 1215. doi:10.1126/science.1213780

Nichols, G. (2009). Sedimentology and stratigraphy. In (2nd ed. ed.). Chichester, UK ; Hoboken, NJ: Wiley-Blackwell.

Pan, Y., Sha, J., Zhou, Z., & Fürsich, F. T. (2013). The Jehol Biota: Definition and distribution of exceptionally preserved relicts of a continental Early Cretaceous ecosystem. *Cretaceous Research, 44*(C), 30-38. doi:10.1016/j.cretres.2013.03.007

Vinther, J., Nicholls, R., Lautenschlager, S., Pittman, M., Kaye, Thomas g., Rayfield, E., . . . Cuthill, Innes c. (2016). 3D Camouflage in an Ornithischian Dinosaur. *Current Biology, 26*(18), 2456-2462. doi:10.1016/j.cub.2016.06.065

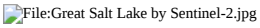
Wang, Y., Olsen, P. E., Sha, J., Yao, X., Liao, H., Pan, Y., . . . Rao, X. (2016). Stratigraphy, correlation, depositional environments, and cyclicity of the Early Cretaceous Yixian and ?Jurassic-Cretaceous Tuchengzi formations in the Sihetun area (NE China) based on three continuous cores. *Palaeogeography, Palaeoclimatology, Palaeoecology, 464*, 110-133. doi:10.1016/j.palaeo.2016.06.043

8.5.1: Freshwater Lakes is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

8.5.2: Saline Lakes

Introduction

Saline lakes occur in the hydrographically lowest areas of arid environments within closed-drainage basins (Hardie et al, 1978). They can be categorized into several categories: ephemeral, perennial, acidic, and alkaline (Hardie et al, 1978). Moreover, they have several depositional sub-environments: alluvial fans, sandflats, dry mudflats, dune fields, perennial and ephemeral stream floodplains, perennial and ephemeral lakes, springs – travertine and tufa flats and mounds, and shoreline features (Hardie et al, 1978).



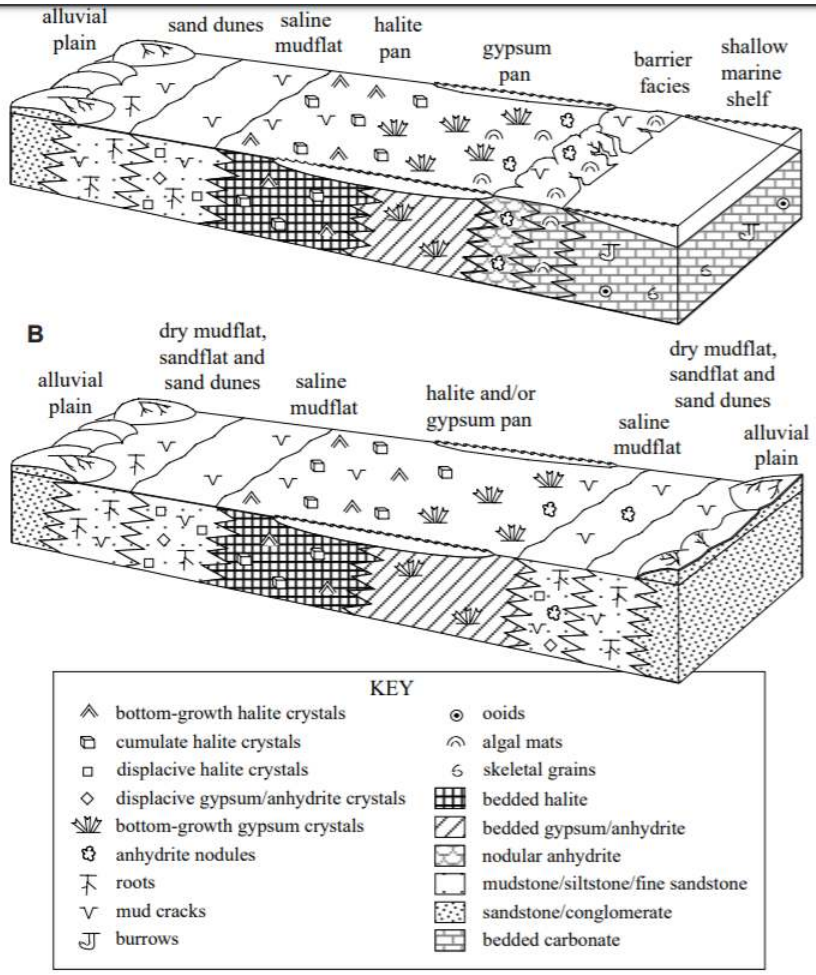
(Above, there is a picture of the Great Salt Lake in Utah, USA. It is one of the largest lakes in the United States aside from the Great Lakes and is largely considered a saline lake. Various portions of the lake are too salty for anything other than a few highly-specialized species.)

Fluids

A lake is called a saline lake if the water in it is made up of more than 5000 ppm of dissolved solute (Hardie et al, 1978). The amount of precipitation/influx of water into saline lakes is less than the amount that is evaporating and water flow into and out of the saline lake is non-existent or at least restricted (Hardie et al, 1978). Considering these circumstances, saline lakes are very low energy depositional environments. Also, the influx into this environment may come from perennial or ephemeral streams, springs, groundwater, or un-channeled sheet flow from storms (Hardie et al, 1978).

Sediment Transport Processes

Sediment transport into saline lakes is typically driven by fluid flow from storm flooding (Hardie et al, 1978). And so fluid flow is high in the alluvial fan sub-environment and slows as it passes through sandflats, then mudflats, and then into the saline lakes. Once storm floodwaters reach the mudflat environment, they have moderate velocities, are shallow, and un-channeled (Hardie et al, 1978). Evaporative pumping is another major way sediment (in the form of salts) is transported within saline lakes, particularly in the mudflat facies/sub-environment where this process creates surface efflorescence (Hardie et al, 1978). The dune field facies is associated with the bottom of the alluvial fan, sandflat, and both of the inflow stream floodplain sub-environments of saline lakes in that when any of these areas lack efflorescent salt crusts, they can be eroded and have sediment deposited on them by wind (Hardie et al, 1978).



(Above, is a figure from Benison and Goldstein 2001. It illustrates a cross-section of a salt lake depositional environment [this is the bottom cross-section] and a cross-section of a coastal marina [the cross-section on top]. Notice that the pan above occupies the area where the lake would normally be after it is recharged with water.)

Blue-green algae are associated with the precipitation of tufas and travertines in both spring orifices and inflow streams; however, the cause of this association may be due to the metabolisms of the algae or inorganic precipitation of carbonates that trap the algae (Hardie et al, 1978).

Characteristics of Deposited Sediment

Saline lakes are subdivided into several sub-environments that can be used as facies: alluvial fans, sandflats, dry mudflats, dune fields, perennial and ephemeral stream floodplains, perennial and ephemeral lakes, springs – travertine and tufa flats and mounds, and shoreline features (Hardie et al, 1978).

Alluvial Fans and Sandflats

Saline lakes are surrounded by alluvial fans that are typically created by stream point flows and have gravelly wedges that then fan out with very coarse sediments (Hardie et al, 1978). Alluvial fans are typically the products of storm-flooding and so this type of flow creates four kinds of deposits: shallow braid channel deposits, fills inside deeper incised channels, sieve deposits, and debris flow deposits (Hardie et al, 1978). On the surface of the alluvial fan, the braid channel deposits appear as coarse gravel bars that are isolated by braided shallow stream channels (Hardie et al, 1978). In vertical sections, the braid stream deposits in alluvial fans appear as cross-cutting lenses that are made up of framework-supported boulders and pebbles that may be imbricated and may have gaps in the framework filled in by sand (Hardie et al, 1978). Incised channel fill deposits can be several meters thick and have sediment that fines upward in grain size (Hardie et al, 1978). Evaporative pumping of groundwater beneath an alluvial fan's surface

can also create calcite pore cement, caliche crusts, coatings, root moulds, and nodules within the fan's sediment (Hardie et al, 1978). Within ephemeral and perennial saline lakes, the water has very little velocity (Hardie et al, 1978).

The deposits of the alluvial fans then lead into sandy aprons, which are referred to as the sandflats facies. The sediment in sandflats are finer than those in the alluvial fans and have pore and vug fillings that are likely to be gypsum, which can dehydrate into anhydrite, or high-Mg calcite (Hardie et al, 1978). The sandflats can also have wind ripples and longitudinal trains (Hardie et al, 1978).

Dry Mudflats

The dry mudflat facies is commonly associated with ephemeral saline lakes and are in general differentiated from the ephemeral saline lake facies in that they preserve depositional structures. At the surface, the mudflats will usually have polygonal mudcracks which are desiccation cracks and a thin layer of saline crusts which are surface efflorescences (Hardie et al 1978). Two types of saline crusts can form: thin, brittle, and dense crust made up of micritic alkaline earth carbonates; and puffy, porous, and crystalline crusts that are made up of soluble minerals like halite, trona, thermonatrite, or thenardite (Hardie et al, 1978). In addition, the soluble minerals in the second type of saline crust are usually dissolved by storm flooding and so their presence is recorded via the disruption and destruction of the top part of each depositional unit (Hardie et al, 1978). However, the sedimentary features of dry mudflats are not well known, but it is hypothesized that they could have small, millimeter-scale laminae and crystals of gypsum, glauberite, etc. (Hardie et al, 1978). The flow velocity, shallowing storm flood water, and traction deposition in dry mudflats could also produce sedimentary structures that are fine grained but also upper flow regime bedforms (Hardie et al, 1978).

Ephemeral Lakes

Ephemeral lakes are a type of saline lake that is shallow and every few years dries up leaving behind a layer of salt(s) precipitated by the evaporating brine (Hardie et al, 1978). These lakes are cyclically recharged via storm water runoff and then gradually recede as the water evaporates, which leads to the formation of two sub-facies – a salt pan facies and a saline mudflat facies (Hardie et al, 1978). The salt pan facies occurs in the lowest part of the lake bed and it is characterized by layered salts, while the saline mudflat facies occurs around the salt pan and is characterized by muddy clastic sediment that contains salt mineral crystals (Hardie et al, 1978). Also, the saline mudflat sub-facies may have the following features: root casts, When the turbulence from the storm flooding ceases, silt and clay-sized grains settle from suspension they create a thin lamina that extends across the recharged saline lake (Hardie et al, 1978). Wind-induced waves may then cause the formation of wave ripples on the mud lamina or rework the mud lamina into a silt clay lenticular lamination (Hardie et al, 1978). In the salt pan facies, this thin mud lamina can become the site of bacterial reduction of sulfate and creates iron sulfides and sulfuric acid, which give the sediment a black coloring (Hardie et al, 1978). Additionally, the salt pan facies is characterized by couplets of layers following a storm flooding of the lake – the thin mud lamina that is black in color and has intergrown salt crystals and an overlying, thick layer of crystalized salts (Hardie et al, 1978).

Perennial Lakes

Perennial saline lakes are a type of saline lake that remains deeply water-filled to some extent for most years. The strata of perennial lakes reflect the fact that they are constantly depositing saline material onto the lake bottom and so a major structure within these strata are crystalized salt laminae (Hardie et al, 1978). Clastic mud laminae can also be deposited at irregular points throughout the strata of a perennial lake as a result of different storm flood events (Hardie et al, 1978). The salt crystal laminae can also be altered diagenetically as a result of bacterial reduction into a different material, such as sulfide, or through recrystallization which would erase the original depositional structure (Hardie et al, 1978).

Dune Fields

Erosion from wind action on the bottom of the alluvial fan, sandflat, and both of the inflow stream floodplains sub-environments creates the dune fields, deflation flats, and hollows that are characteristic of the dune fields facies (Hardie et al, 1978). In some cases where the sediments of a dry saline lake bed are eroded, the sand making up the dunes in this facies can be made up of salt mineral grains like gypsum (Hardie et al, 1978).

Shoreline

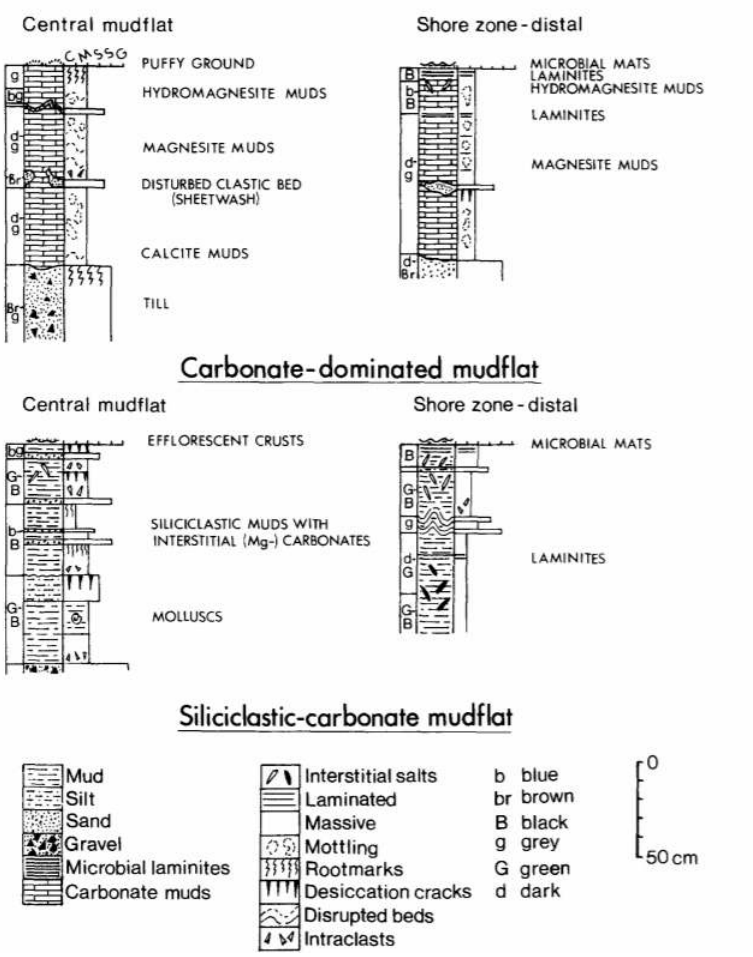
The shoreline facies is partly influenced by groundwater discharge from springs that may be located in them (Renaut and Long 1989). As a result, travertine deposits may be found on the shoreline along with spring mounds (Renaut and Long 1989). Typically,

shoreline deposits vary based on the saline lake's water level and this leads to the deposition of sparingly-soluble salts during high stands and deposition of more readily soluble salts during low stands (Renaut and Long 1989). In some parts of the shoreline, wind-action can cause certain areas of the lakeshore to have more salts deposited on them - resulting in a thicker salt layer. Deposits on the shoreline may consist of carbonate muds and their deposits are typically laminated and/or display mottling (Renaut and Long 1989).

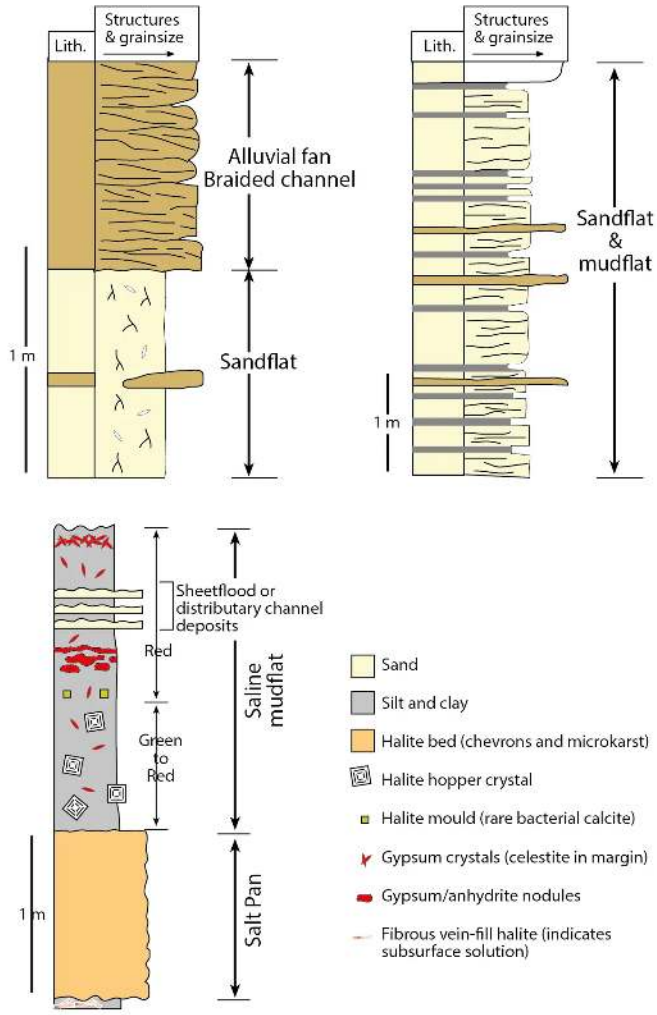
Biological Deposition

In the saline lake sub-environments, bioturbation may come in the form of halophyte plant roots, invertebrate burrows, bird feeding pits, or terrestrial animal burrows (Hardie et al, 1978). The alluvial fan, sandflats, dunes, and mudflats facies tend to have these features, which interrupt or destroy depositional layers (Hardie et al, 1978). Blue-green algal mats can accumulate on the bottom of saline lakes and it is diagenetically possible for them to be converted into kerogen, which has been associated with oil shales (Hardie et al, 1978). Additionally, the shoreline facies is commonly occupied by small cyanobacterial mats (Renaut and Long 1989).

Vertical Sequence of Facies



(The above was taken from Renaut and Long 1989. It depicts stratigraphic columns of Carbonate-dominated and Siliclastic-carbonate mudflat facies within the saline lakes of the Cariboo Plateau in British Columbia, Canada. The mudflat facies and shoreline facies are interlinked in that depending on the water level of the saline lake the shoreline can move further up a mudflat or down to its lowest extent.)



(Source. The above illustrates a stratigraphic section of Bristol Dry Lake, USA. Notice that alluvial fan facies starts off with the coarsest grains and that the sediments closer to the saline lake have finer sediment. Also, note that the bottom saline mudflat section illustrates a periodic increase in grain size due to faster water flow from storm flooding.)

Saline lake deposits can be differentiated from marine deposits in that they are halite dominated, have many soil features, lack marine fossils, and the scarcity of carbonates (Benison and Goldstein, 2001). Additionally, the strata associated with saline lake deposits have a combination of aeolian, flood, desiccation, and evaporation -related features (Benison and Goldstein, 2001).

Summary

Saline lakes are a unique depositional environment whose deposits are influenced by largely periodic water flow from storm flooding, but they are also shaped by aeolian processes and the surfacing of groundwater through springs. Along the cross-section of this environment towards the lake-itself, water flow is expected to decrease into a near or complete standstill. The deposits as a result decrease in grain size towards the lake - largely forming mud layers at the perimeter of and within the saline lake. As the lake level changes, the type of salts deposited change and the level change defines the saline mudflat subfacies and dry mudflat facies as the border between these is the shoreline. Sedimentary structures are highly deformed within the saline lake due to disturbance from salt crystal growth and many of the facies along the shoreline and within the lake are deposited upon by microorganisms including cyanobacteria. Lastly, the habitat surround saline lakes is in danger of disappearing putting a variety of organisms that rely on it in danger (Further Reading).

References

Benison, K.C., and Goldstein R.H. “Evaporites and Siliciclastics of the Permian Nippewalla Group of Kansas, USA: a Case for Non-Marine Deposition in Saline Lakes and Saline Pans.” *Sedimentology*, vol. 48, no. 1, 2001, pp. 165–188., doi:10.1046/j.1365-3091.2001.00362.x.

Hardie, L.A., Smoot, J.P., Eugster, H.P. “Evaporites and Siliciclastics of the Permian Nippewalla Group of Kansas, USA: a Case for Non-Marine Deposition in Saline Lakes and Saline Pans.” *Modern and Ancient Lake Sediments*. Special Publications of the International Association of Sedimentologists, no. 2, 1978, pp. 7-41.

Renaut, R.W., and Long, P.R. "Sedimentology of the saline lakes of the Cariboo Plateau, Interior British Columbia, Canada." *Sedimentary Geology*, vol. 64, 1989, pp. 239-264.

8.5.2: Saline Lakes is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

8.6: Streams and Rivers

Freshwater in streams, ponds, and lakes is an essential part of the hydrologic cycle if only because of its importance to living creatures. Along with wetlands, these freshwater regions contain a tremendous variety of organisms. **Streams** are bodies of water that consist of constant motion, called a **current**. Geologists recognize many categories of streams depending on their size, depth, speed, and location. Creeks, brooks, tributaries, bayous, and rivers might all be lumped together as streams. In streams, water always flows downhill, but the form that downhill movement takes varies with rock type, topography, and many other factors.



"Stream Landscape in British Columbia" is licensed under Public Domain.

Streams are the most important agents of erosion and transportation of sediments on Earth's surface. They are responsible for the creation of much of the topography that we see around us. They are also places of great beauty and tranquility, and of course, they supply much of the water that is essential to our existence. However, streams are not always peaceful and soothing. During large storms and rapid snowmelts, they can become raging torrents capable of moving cars and houses and destroying roads and bridges. When they spill over their banks, they can flood vast areas, devastating populations, and infrastructure. (11 Water – An Introduction to Geology, n.d.)

Parts of a Stream

There are a variety of distinct types of streams. A stream originates at its **source**s, such as high mountains where snows collect in winter and melt in summer, or a source might be a **spring**. A stream may have more than one source, and when two streams come together, it is called a **confluence**. The smaller of the two streams is a **tributary** of the larger stream. A stream may create a pool where the water slows and becomes more profound.

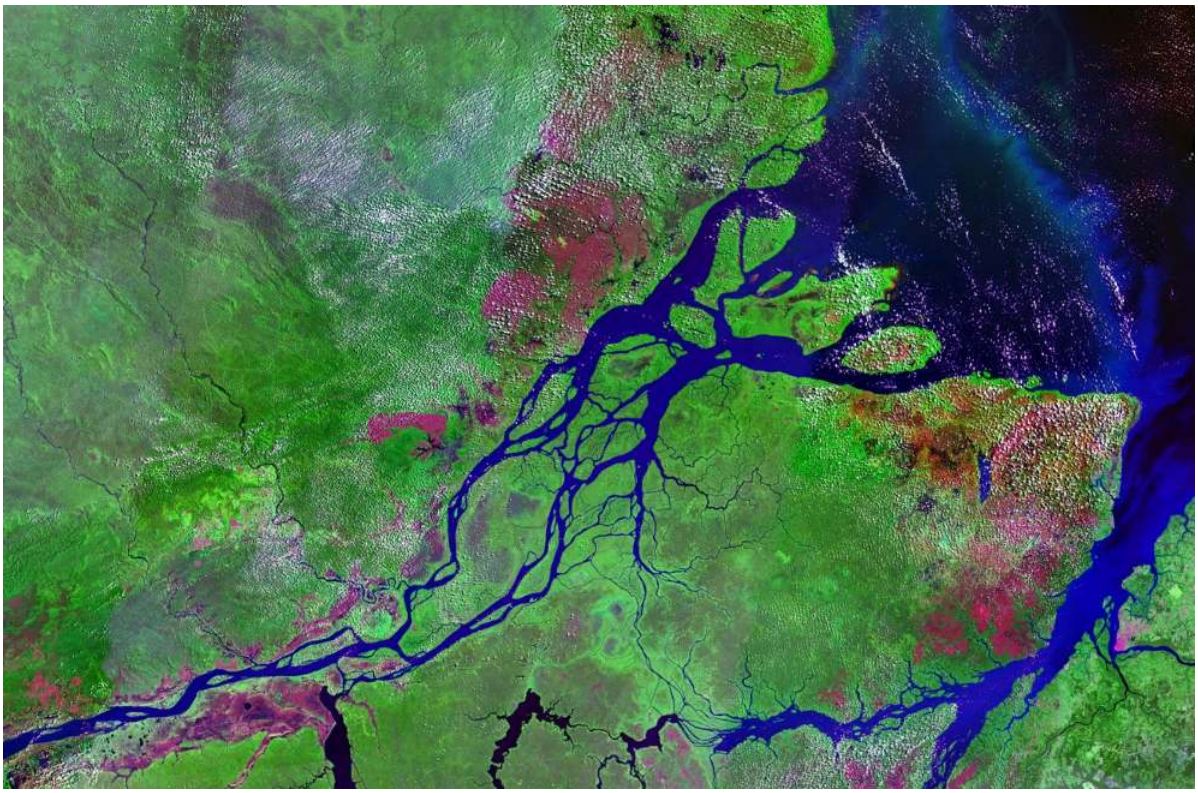


"River Source" by Tony Ferrie is licensed under the Creative Commons Attribution-ShareAlike 2.0 Generic.



"Flooding at the junction of the Mississippi and Ohio River" by NASA Landsat Science is licensed under Public Domain.

The point at which a stream comes into a large body of water, like an ocean or a lake, is called the **mouth**. Where the stream meets the ocean or lake, it is called an **estuary**. The mix of fresh and saltwater, where a river runs into the ocean, creates a diversity of environments where many different organisms create unique ecosystems.

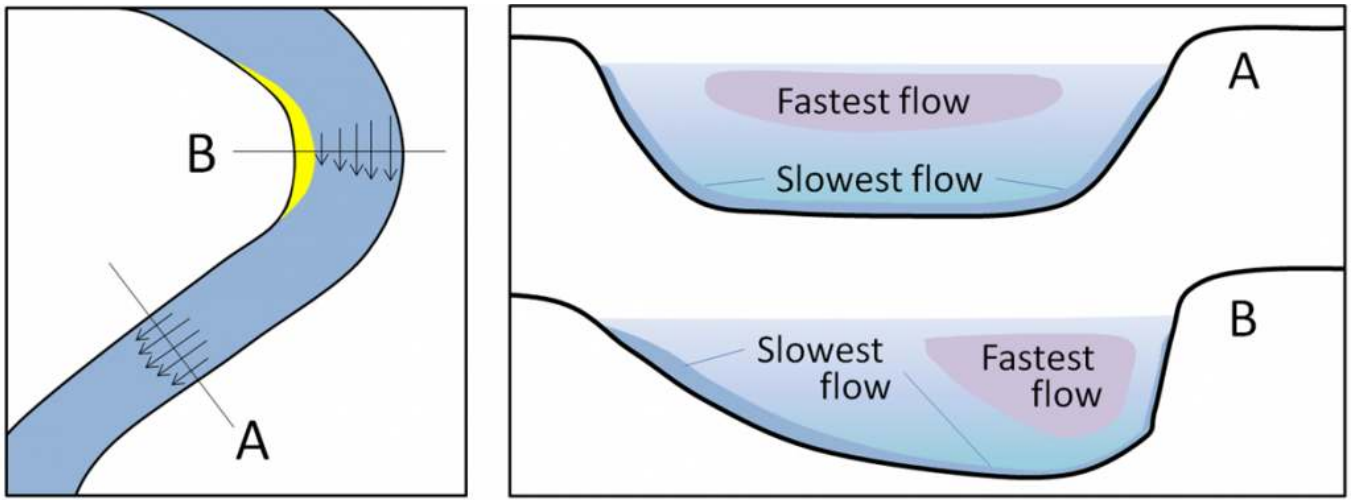


"Mouths of the Amazon River" by NASA is licensed under Public Domain.

Stream Erosion and Deposition

Flowing water is a fundamental mechanism for both erosion and deposition. Water flow in a stream is primarily related to the stream's gradient, but the stream channel's geometry also controls it. The water flow velocity is decreased by friction along the stream bed, so it is slowest at the bottom and edges and fastest near the surface and in the middle. The velocity below the surface is typically slightly higher than right at the surface because of friction between the water and the air. On a curved section of a stream, flow is fastest on the outside and slowest on the inside.

Other factors that affect stream-water velocity are the size of sediments on the stream bed – because large particles tend to slow the flow more than small ones – and the discharge, or volume of water passing a point in a unit of time. During a flood, the water level always rises, so there is a more cross-sectional area for the water to flow in. However, as long as a river remains confined to its channel, the velocity of the water flow also increases.

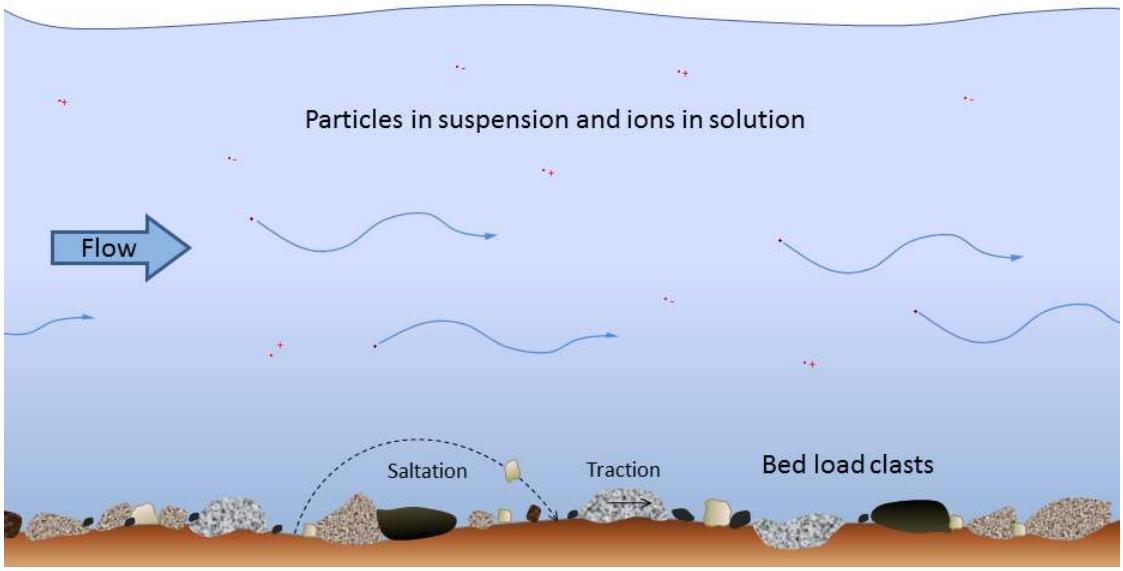


"Relative Velocity of Stream Flow" by Steven Earle is licensed under the Creative Commons Attribution 4.0 International license.

Large particles rest on the bottom, bedload, and may only be moved during rapid flows under flood conditions. They can be moved by **saltation** (bouncing) and by **traction** (being pushed along by the force of the flow). Smaller particles may rest on the bottom some of the time, where they can be moved by saltation and traction, but they can also be held in suspension in the flowing water, especially at higher velocities. Streams that flow fast tend to be turbulent (flow paths are chaotic, and the water surface appears rough), and the water may be muddy, while those that flow more slowly tend to have laminar flow (straight-line flow and a smooth water surface) and clear water. Turbulent flow is more effective than laminar flow at keeping sediments in suspension. Stream water also has a **dissolved load**, representing roughly 15 percent of the mass of material transported, and includes ions such as calcium and chloride in solution. The solubility of these ions is not affected by flow velocity.

The faster the water is flowing, the larger the particles that can be kept in suspension and transported within the flowing water. However, as Swedish geographer Filip Hjulström dis-covered in the 1940s, the relationship between grain size and the likelihood of a grain being eroded, transported, or deposited is not as simple as one might imagine. Consider, for example, a 1 mm grain of sand. If it is resting on the bottom, it will remain there until the velocity is high enough to erode it. However, once it is in suspension, that same 1 mm particle will remain in suspension as long as the velocity does not drop below 10 centimeters per second (cm/s.) For a 10 mm gravel grain, the velocity is 105 cm/s to be eroded from the bed but only 80 cm/s to remain in suspension.

On the other hand, a 0.01 mm silt particle only needs a velocity of 0.1 cm/s to remain in suspension but requires 60 cm/s to be eroded. A tiny silt grain requires a higher velocity to be eroded than a grain of sand that is 100 times larger. For clay-sized particles, the discrepancy is even more significant. In a stream, the most easily eroded particles are small sand grains between 0.2 mm and 0.5 mm. Anything smaller or larger requires a higher water velocity to be eroded and entrained in the flow. The main reason for this is that those small particles, especially the tiny grains of clay, have a strong tendency to stick together, so they are challenging to erode from the stream bed.



"Particles in Suspension and Ions in Solution" by Steven Earle is licensed under the Creative Commons Attribution 4.0 International license.

It is essential to be aware that a stream can erode and deposit sediments simultaneously. At 100 cm/s, for example, silt, sand, and medium gravel will be eroded from the stream bed and transported in suspension, coarse gravel will be held in suspension, pebbles will be both transported and deposited, and cobbles and boulders will remain stationary on the stream bed.

A stream typically reaches its highest velocity when it is close to flooding over its banks, known as the **bank-fullstage**. As soon as the flooding stream overtops its banks and occupies the broad area of its flood plain, the water has a much larger area to flow through, and the velocity drops significantly. At this point, sediment that was being carried by the high-velocity water is deposited near the edge of the channel, forming a natural bank or levée. (11 Water – An Introduction to Geology, n.d.)

Stream Types

Stream channels can be straight or curved, deep and slow, or rapid and choked with coarse sediments. The cycle of erosion has some influence on the nature of a stream, but several other factors are essential. Youthful streams that are actively downcutting their channels tend to be relatively straight and are typically ungraded (meaning that rapids and falls are frequent). Youthful streams commonly have a step-pool morphology, meaning that the stream consists of a series of pools connected by rapids and waterfalls. They also have steep gradients and steep and narrow **V-shaped valleys**, in some cases steep enough to be called **canyons**.



"Black Canyon and Gunnison River" is a V-shaped river taken by Terry Foote and is licensed under the Creative Commons Attribution 4.0 International license.

In mountainous terrains, steep youthful streams typically flow into broad and low-gradient **U-shaped glaciated valleys**. The youthful streams have high sediment loads, and when they flow into the lower-gradient glacial valleys where the velocity is not high enough to carry all of the sediment, braided patterns develop, characterized by a series of narrow channels separated by gravel bars.



"U-Shaped Valley in Norway" by Petr Smerkl is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

Braided streams can develop anywhere there is more sediment than a stream can transport. One such environment is in volcanic regions, where explosive eruptions produce enormous amounts of unconsolidated material that gets washed into streams. The Coldwater River next to Mt. St. Helens in Washington State is an excellent example.

A stream that occupies a vast, flat flood plain with a low gradient typically carries only sand-sized and finer sediments and develops a sinuous flow pattern. When a stream flows around a corner, the water on the outside has farther to go and tends to flow faster. This leads to erosion of the banks on the outside of the curve, deposition on the inside, and formation of a point bar. Over time, the sinuosity of the stream becomes increasingly exaggerated, and the channel migrates around within its flood plain, forming a meandering pattern.



"Braided Stream" by the National Oceanic and Atmospheric Administration is licensed under Public Domain.

The meander in the photo below has reached the point where the thin neck of land between two parts of the channel is about to be eroded. When this happens, **an oxbow lake** will form. Finally, at the point where a stream enters a still body of water, a lake, or the ocean, sediment is deposited and a delta forms.

Rivers

Rivers are the largest types of stream, moving substantial amounts of water from higher to lower elevations. The Amazon River, the planet's most enormous river, has a flow rate of nearly 220,000 cubic meters per second! People have used rivers since the beginning of civilization as a source of water, food, transportation, defense, power, recreation, and waste disposal.



"Red River Meanders" by the US Department of Agriculture is licensed under Public Domain.

Divides

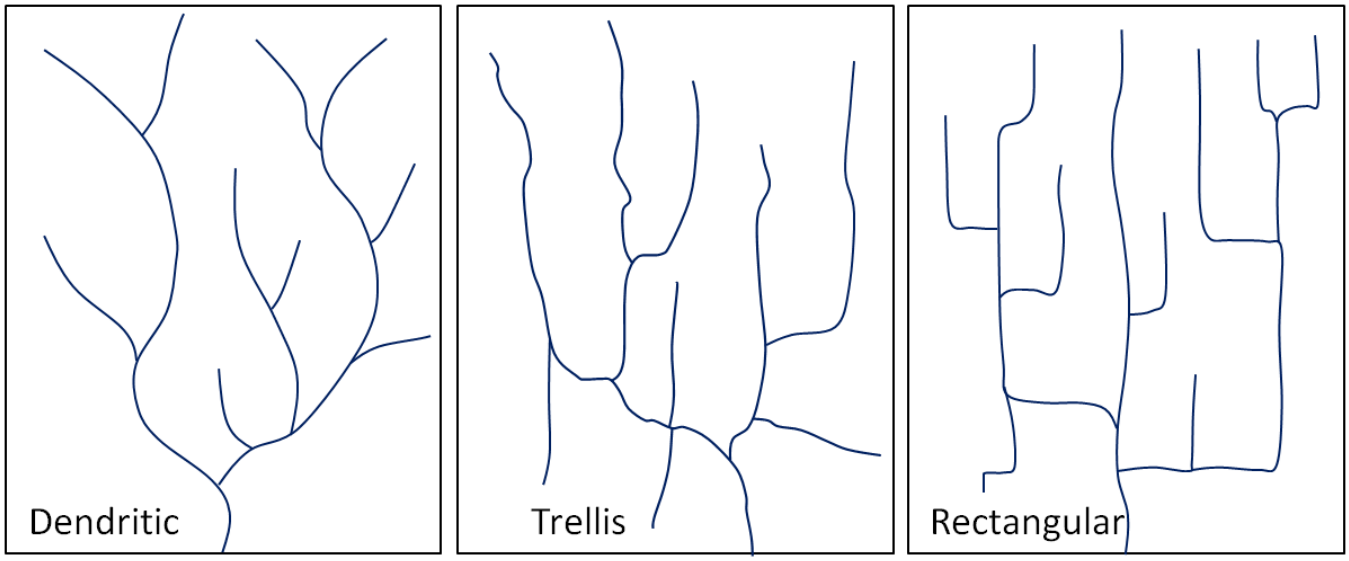
A **divide** is a topographically high area that separates a landscape into different water basins. The rain that falls on the north side of a ridge flows into the northern drainage basin, and rain that falls on the south side flows into the southern drainage basin. On a

much grander scale, entire continents have divides, known as **continental divides**.



"North America Water Divides" is licensed under the Creative Commons Attribution-ShareAlike 3.0 Unported license.

The pattern of tributaries within a drainage basin depends mainly on the type of rock beneath, and on structures within that rock (folds, fractures, faults, etc.). **Dendritic patterns**, which are by far the most common, develop in areas where the rock (or unconsolidated material) beneath the stream has no particular fabric or structure and can be eroded equally easily in all directions. Examples would be granite, gneiss, volcanic rock, and sedimentary rock that has not been folded. **Trellis drainage** patterns typically develop where sedimentary rocks have been folded or tilted and then eroded to varying degrees depending on their strength. The Rocky Mountains of B.C. and Alberta are an excellent example of this, and many of the drainage systems within the Rockies have trellis patterns. **Rectangularpatterns** develop in areas with little topography and a system of bedding planes, fractures, or faults that form a rectangular network. The fourth type of drainage pattern, which is not specific to a drainage basin, is known as **radial**. Radial patterns form around isolated mountains (such as volcanoes) or hills, and the individual streams typically have dendritic drainage patterns. (11 Water – An Introduction to Geology, n.d.)



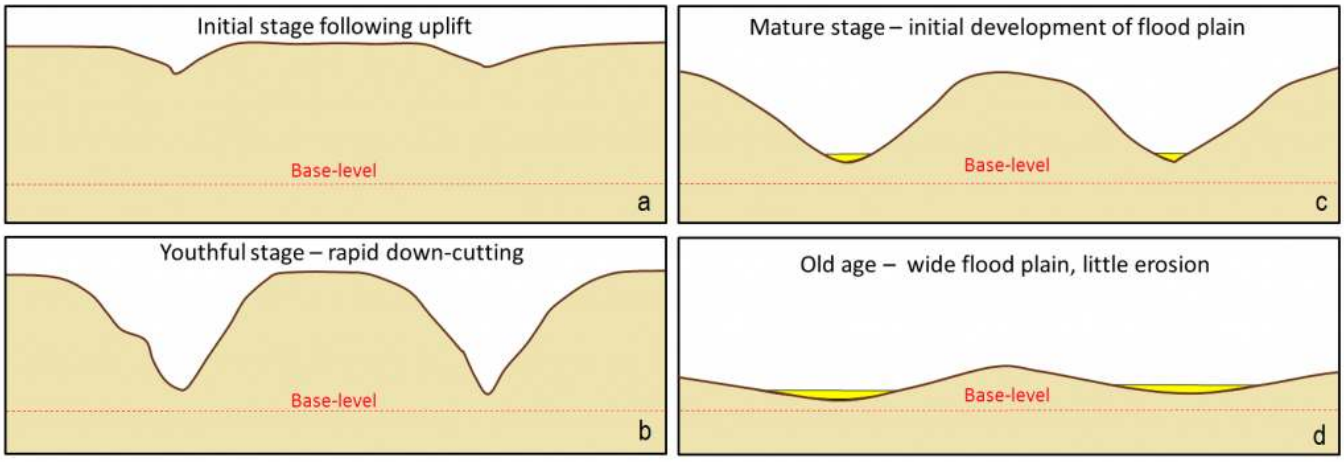
"Stream Patterns" by Steven Earle is licensed under the Creative Commons Attribution 4.0 International license.

Over geological time, a stream will erode its drainage basin into a smooth profile. If we compare this with an ungraded stream like Cawston Creek, we can see that graded streams are steepest in their headwaters, and their gradient gradually decreases toward their mouths. Ungraded streams have steep sections at various points, and typically have rapids and waterfalls at numerous locations along their lengths.

The ocean is the ultimate base level, but lakes and other rivers act as base levels for many smaller streams. Engineers can create an artificial base level on a stream by constructing a dam. Sediments accumulate within the flood plain of a stream, and then, if the base level changes, or if there is less sediment to deposit, the stream may cut down through those existing sediments to form terraces.

In the late 19th century, American geologist William Davis proposed that streams and the surrounding terrain develop in a cycle of erosion. Following tectonic uplift, streams erode quickly, developing deep V-shaped valleys that follow relatively straight paths. Gradients are high, and profiles are ungraded. Rapids and waterfalls are common. During the mature stage, streams erode more broad valleys and start to deposit thick sediment layers. Gradients are slowly reduced, and grading increases. In old age, streams are surrounded by rolling hills, and they occupy broad sediment-filled valleys. Meandering patterns are common.

Davis's work was done long before the idea of plate tectonics, and he was not familiar with the impacts of glacial erosion on streams and their environments. While some parts of his theory are out of date, it is still a useful way to understand streams and their evolution.



"David's Cycle of Erosion" by Steven Earle is licensed under the Creative Commons Attribution 4.0 International license.

This page titled 8.6: Streams and Rivers is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Adam Dastrup via source content that was edited to the style and standards of the LibreTexts platform.

8.7: Estuaries- Where the turf meets the surf

Estuaries are partially enclosed bodies of water where the salt water is diluted by fresh water input from land, creating **brackish** water with a salinity somewhere between fresh water and normal seawater. Estuaries include many bays, inlets, and sounds, and are often subject to large temperature and salinity variations due to their enclosed nature and smaller size compared to the open ocean.

Estuaries can be classified geologically into four basic categories based on their method of origin. In all cases they are a result of rising sea level over the last 18,000 years, beginning with the end of the last ice age; a period that has seen a rise of about 130 m. The rise in sea level has flooded coastal areas that were previously above water, and prevented the estuaries from being filled in by all of the sediments that have been emptied into them.

The first type is a **coastal plain estuary**, or **drowned river valley**. These estuaries are formed as sea level rises and floods an existing river valley, mixing salt and fresh water to create the brackish conditions where the river meets the sea. These types of estuaries are common along the east coast of the United States, including major bodies such as the Chesapeake Bay, Delaware Bay, and Narragansett Bay (Figure 8.7.1). Coastal plain estuaries are usually shallow, and since there is a lot of sediment input from the rivers, there are often a number of depositional features associated with them such as spits and barrier islands.

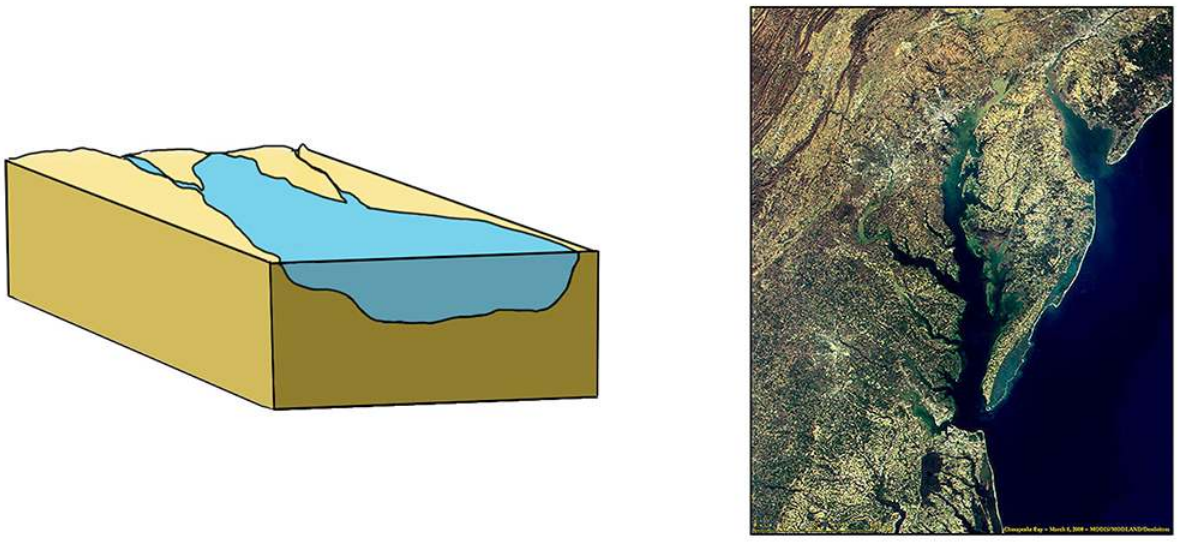


Figure 8.7.1 A coastal plain estuary. Sea level has risen and flooded what was once a river valley. The satellite image shows Chesapeake Bay and Delaware Bay, two coastal plain estuaries (left: JR, right: NASA, Public Domain via Wikimedia Commons).

The presence of sand bars, spits, and barrier islands can lead to **bar-built estuaries**, where a barrier is created between the mainland and the ocean. The water that remains inside the sand bar is cut off from complete mixing with the ocean, and receives freshwater input from the mainland, creating estuarine conditions (Figure 8.7.2).

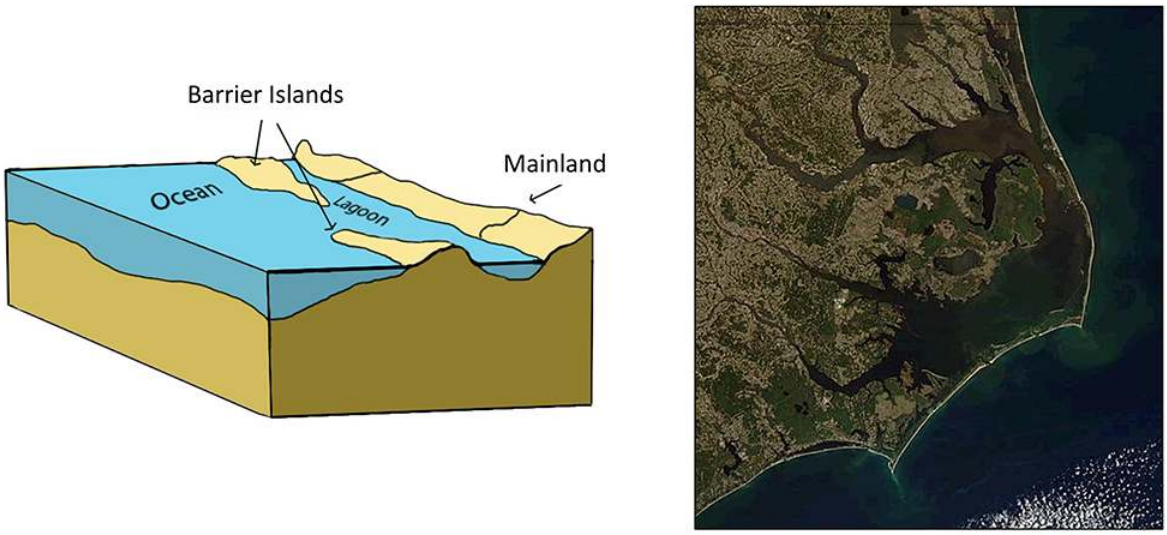


Figure 8.7.2 A bar-built estuary. Sand bars and barrier islands have partially isolated a lagoon from the rest of the ocean. Freshwater input into the lagoon from the mainland creates brackish conditions in the estuary. At right is a satellite image of Pamlico Sound, North Carolina, a bar-built estuary surrounded by spits and barrier islands (left: JR, right: NASA, Public Domain via Wikimedia Commons).

Fjords are estuaries formed in deep, U-shaped basins that were carved out by advancing glaciers. When the glaciers melted and retreated, sea level rose and filled these troughs, creating deep, steep-walled fjords (Figure 8.7.3). Fjords are common in Norway, Alaska, Canada, and New Zealand, where there are mountainous coastlines once covered by glaciers.

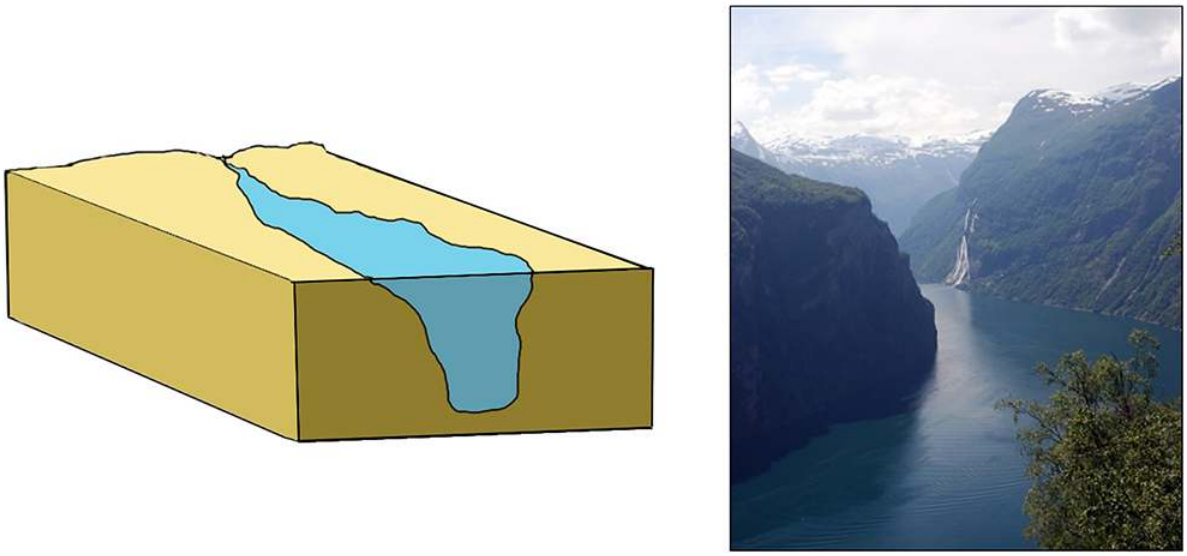


Figure 8.7.3 A fjord is a deep estuary that was carved out by glacial movements. At right is Geirangerfjord, Norway (left: JR, right: Fgmedia, [CC-BY-SA-3.0], via Wikimedia Commons).

Tectonic estuaries are the result of tectonic movements, where faulting causes some sections of the crust to subside, and those lower elevation sections then get flooded with seawater. San Francisco Bay is an example of a tectonic estuary (Figure 8.7.4).

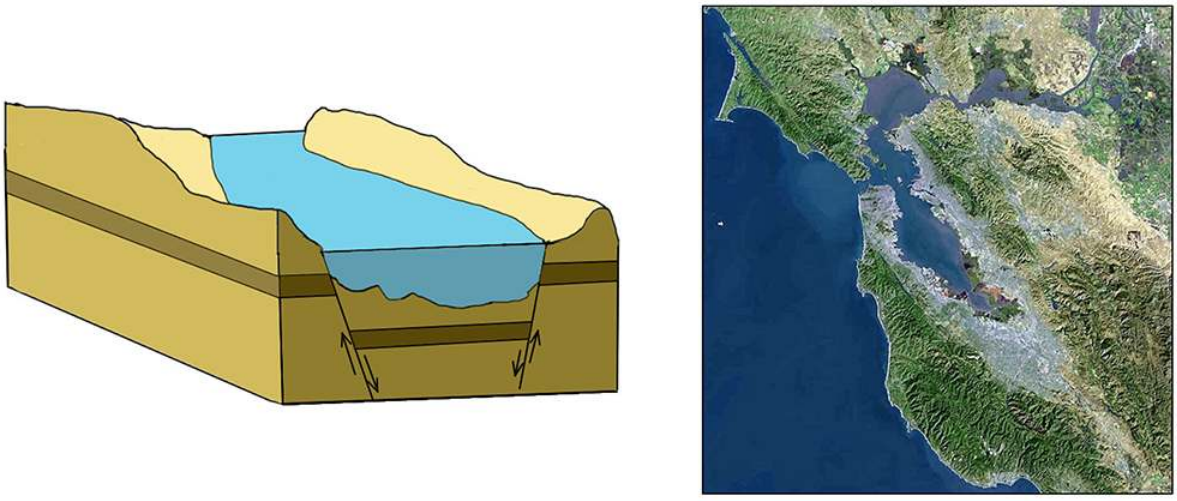


Figure 8.7.4 A tectonic estuary, formed from the subsidence of crust along fault lines, and the subsequent filling by seawater. San Francisco Bay is a tectonic estuary, shown at right (left: JR, right: USGS, Public Domain, via Wikimedia Commons).

Estuaries are also classified based on their salinity and mixing patterns. The amount of mixing of fresh and salt water in an estuary depends on the rate at which fresh water enters the head of the estuary from river input, and the amount of seawater that enters the estuary mouth as a result of tidal movements. The input of fresh water is reflected in the **flushing time** of the estuary. This refers to the time it would take for the in-flowing fresh water to completely replace all the fresh water currently in the estuary. Seawater input is measured by the **tidal volume**, or **tidalprism**, which is the average volume of sea water entering and leaving the estuary during each tidal cycle. In other words, it is the volume difference between high and low tides. The interaction between the flushing time, tidal volume, and the shape of the estuary will determine the extent and type of water mixing within the estuary.

In a **vertically mixed**, or **well-mixed** estuary there is complete mixing of fresh and salt water from the surface to the bottom. In a particular location the salinity is constant at all depths, but across the estuary the salinity is lowest at the head where the fresh water enters, and is highest at the mouth, where the seawater comes in. This type of salinity profile usually occurs in shallower estuaries, where the shallow depths allow complete mixing from the surface to the bottom.

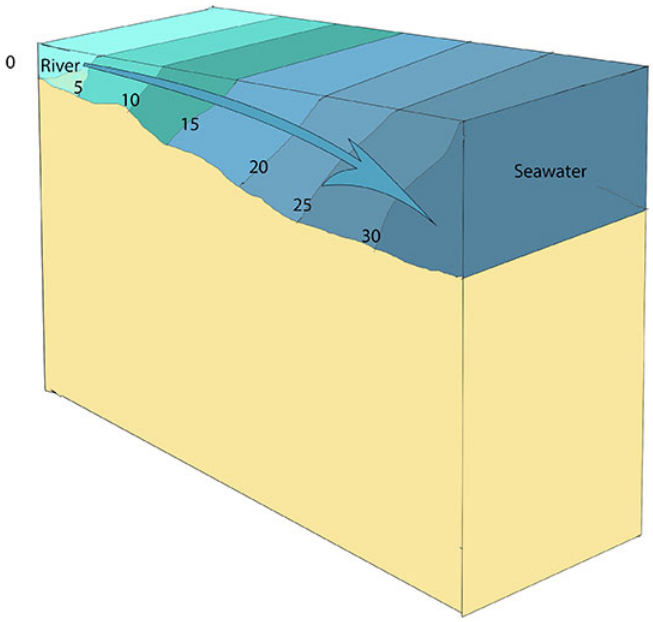


Figure 8.7.5 A well-mixed estuary. The shallow basin allows nearly complete mixing or fresh and seawater from top to bottom. Salinities are in ppt (JR).

Slightly stratified or **partially mixed** estuaries have similar salinity profiles to vertically mixed estuaries, where salinity increases from the head to the mouth, but there is also a slight increase in salinity with depth at any point. This usually occurs in deeper estuaries than those that are well-mixed, where waves and currents mix the surface water, but the mixing may not extend all the way to the bottom.

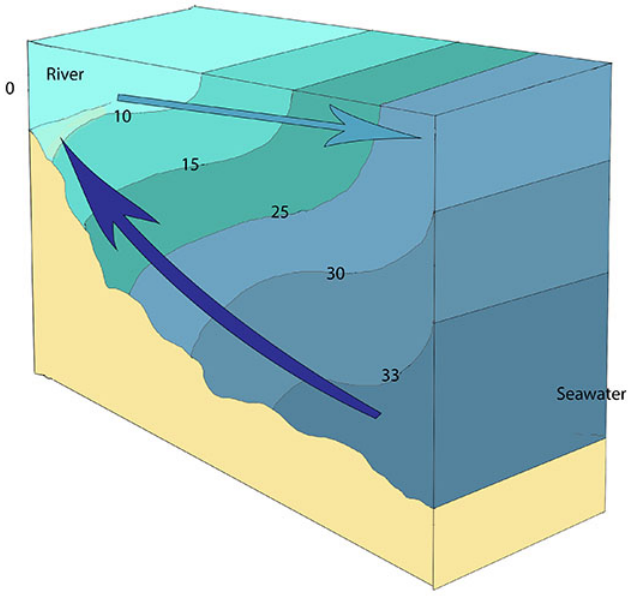


Figure 8.7.6 A slightly-stratified estuary. Generally deeper than a well-mixed estuary, the inflow of seawater (dark blue arrow) and fresh water (light blue arrow) create an estuary where salinity increases with depth, and at the surface when moving from the head to the mouth of the estuary. Salinities are in ppt (JR).

A **salt wedge** estuary occurs where the outflow of fresh water is strong enough to prevent the denser ocean water to enter through the surface, and where the estuary is deep enough that surface waves and turbulence have little mixing effect on the deeper water. Fresh water flows out along surface, salt water flows in at depth, creating a wedge shaped lens of seawater moving along the bottom. The surface water may remain mostly fresh throughout the estuary if there is no mixing, or it can become brackish depending on the level of mixing that occurs.

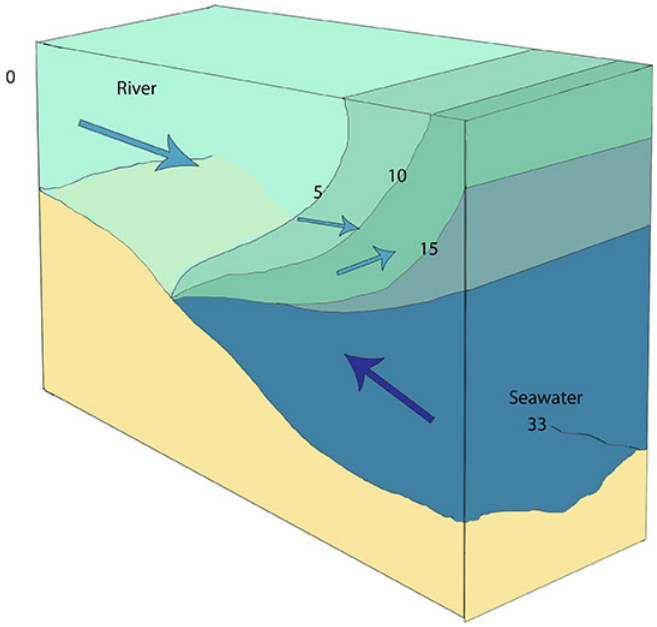


Figure 8.7.7 A salt wedge estuary. Strong river outflow (light blue arrows) creates a layer of mostly fresh water that sits on top of a wedge of encroaching seawater along the bottom (dark blue arrow). Salinities are in ppt (JR).

Highly stratified profiles are found in very deep estuaries, such as in fjords. Because of the depth, mixing of fresh and salt water only occurs near the surface, so in the upper layers salinity increases from the head to the mouth, but the deeper water is of standard ocean salinity.

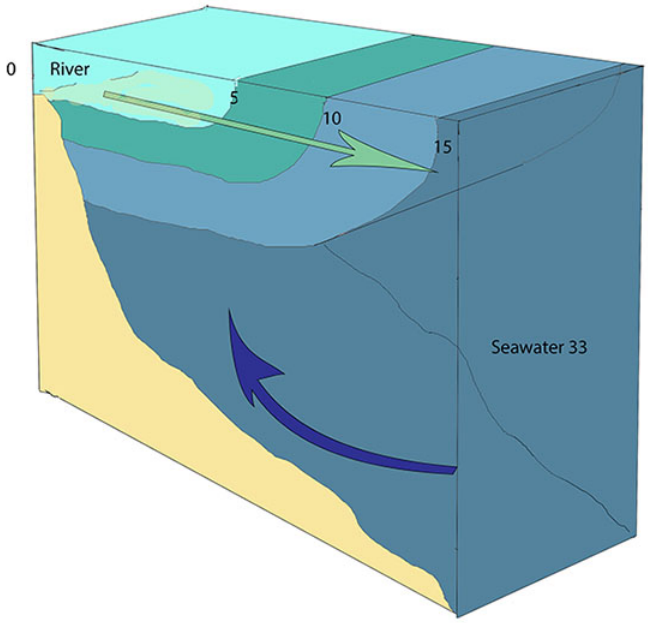
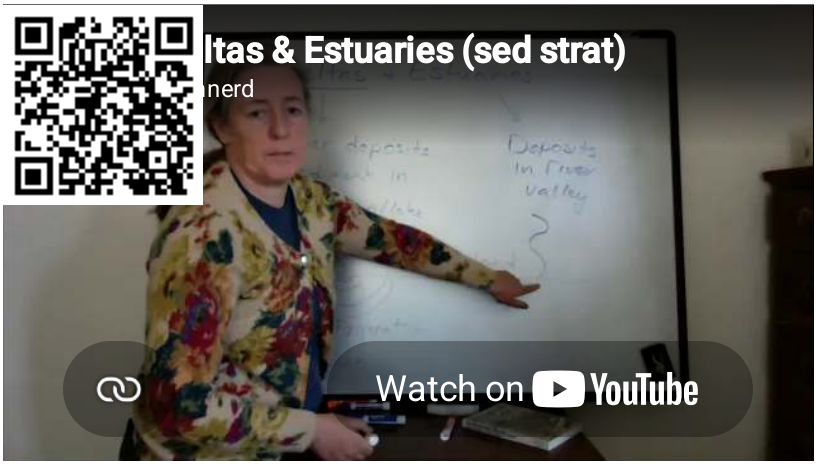


Figure 8.7.8 A highly stratified estuary. Strong river outflow and a deep basin prevent mixing between surface and bottom water, creating an estuary that is vertically stratified. Salinities are represented in ppt (JR).

Estuaries are very important commercially, as they are home to the majority of the world's metropolitan areas, they serve as ports for industrial activity, and a large percentage of the world's population lives near estuaries. Estuaries are also very important biologically, especially in their role as the breeding grounds for many species of fish, birds, and invertebrates.



8.7: Estuaries- Where the turf meets the surf is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **13.6: Estuaries** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

Waves and Beaches

Introduction:

Beaches are a type of landform that typically occur along coastlines and are present alongside bodies of water. They usually experience waves or current flow behavior that deposits and reworks sedimentary grains as well as their geologic structure over time. The fluids at work in shaping this environment include water and air (laminar, turbulent, and transitional flow types).

Sediment transport processes:

The key transport processes that dominate in shaping this type of environment include waves and turbulent flows via storms, lower to higher energy, respectively. Flow speeds present in this environment are almost always too fast (experiencing higher Reynolds' Number) to allow for mud-sized grains to accumulate. Rather, small, medium, and larger sedimentary grains tend to be transported with a very poorly sorted accumulation at the depositional site (1).



There are many different flow speed and turbulence variations present in waves. The ebb and flow nature of the tides at varying depths modulate the maximal and minimal flow speeds experienced among the waves (Reynolds' Number). These factors all lead to a multitude of sedimentary variations along beaches.

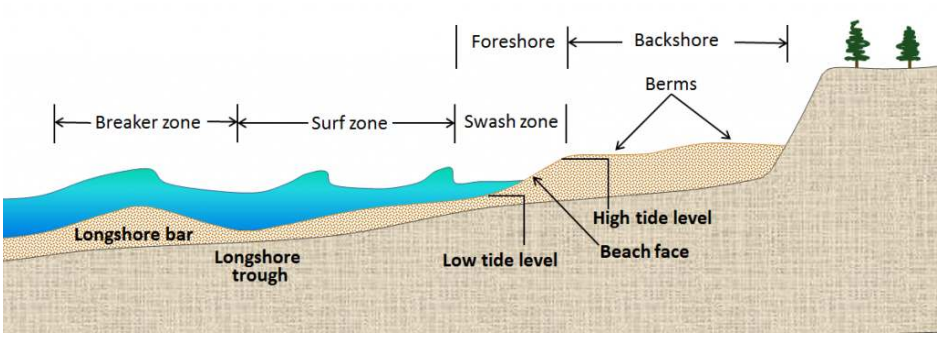
At the breaker zone, wave turbulence of varying flow speeds is experienced. As the waves move in from offshore (greater depth) carrying sedimentary grains of varying sizes and shapes closer to the swash zone (shallower depths), the flow speed, and therefore turbulence, increases rapidly. The suspended grains transported by these waves promptly deposit medium-to-larger/coarser sized sedimentary grains at a varying gradient along the breaker zone (2). The accumulation of sedimentary grains at this site tends to be poorly sorted and sub-rounded to angular, forming planar laminae dipping towards the ocean at the contact site. As the waves move back offshore from the breaker zone (high turbulence and flow speeds) they tend to carry and transport some medium-to-heavy grains back as the tide recedes from the swash zone, others are left deposited at the base.

In contrast to the breaker zone, turbulence and flow speeds of waves located in the offshore zone tend to be lower in energy and more consistent. Flow energy (Reynolds' Number) and turbulence decreases with increasing depths, which is due to the difference in aeolian influence present on the surface of the water versus the lack thereof at greater depths. Larger and coarser grains tend to be carried and transported closer to the surface in this zone with smaller grains experiencing transportation and deposition closer at the base (1).

During periods of high tides, such as storms, the behavior and location of grain transportation and deposition shift greatly. As the ebb and flow of high tides act along the shoreline, the tides will usually reach up and over the berm (maximum distance the waves can reach under normal non-stormy conditions) and deposit sedimentary grains further in along the shoreline. The presence of high winds under these conditions create aeolian dunes along the drier portions of the beach face.

Characteristics of deposited sediment:

The common sedimentary structures that are characteristic of this environment are produced by the processes of wave ripples, planar lamination, cross-bedding, and bioturbation (3). Sedimentary grains of all varying sizes and shapes are deposited along the coastline - depending on the turbulence experienced by the waves' ebb and flow (Reynolds' number at the time of sediment deposition). The principle of original horizontality, as is the case with sediment and stratigraphic units, is very much at play in this environment.



Grains deposited along and past the berm zone of the beach tend to be smoother in texture and less dense due to ease of transportation. Under non-stormy conditions, this zone experiences aeolian behavior and produces wind ripples, cross-bedding, and dunes as sedimentary structures. Grains tend to be well sorted with varying grain shapes present (1).

Grains deposited at the breaker zone tend to be poorly sorted, medium to coarser/larger in size and range from sub-rounded to angular in shape. This is due to the gradual decrease in depth as the wave approaches the breaker zone and thus increasing in turbulence. The sedimentary structures present at this zone tend to be wave ripples and planar laminae with a dip towards the ocean (6).

In the offshore zone, the transportation and deposition of sedimentary grains vary in size, sorting, and shape depending on the depths of the current. This is due to the circular motion of waves caused by the behavior of wind blowing across the surface. This aeolian influence across the ocean surface decreases gradually with increasing depths until being eliminated entirely just prior to reaching the ocean floor. The sedimentary structures present within this zone include cross-bedding lamination, planar lamination, and bioturbation via marine organism behavior.

Typical vertical sequence of facies:

Walther's Law of Facies states that the vertical depositional succession of stratigraphic facies reflects the lateral changes in sedimentary layers witnessed within the environment. It also states that when a depositional environment shifts laterally, stratigraphic layers will be deposited on top of one another (6). Walther's Law is very much at play in a beach environment. As the overall sea level rises and falls, the present stratigraphic layers will be altered in accordance with Walther's Law. Going from the beach, to the breaker zone where shallow marine rocks reside, to the offshore zone where deep marine rocks reside, all these layers will shift laterally reflecting the behavior experienced by the tide. Parts of the aforementioned lateral layers will also be deposited on top of another in accordance with the ebb and flow of the waves.

*Quick note regarding Walther's Law. If erosion, non-deposition, or a break in deposition have taken place in the stratigraphy - Walther's Law will not work in these cases (4).

Walther's Law - before/after

Stratigraphic Column (7)

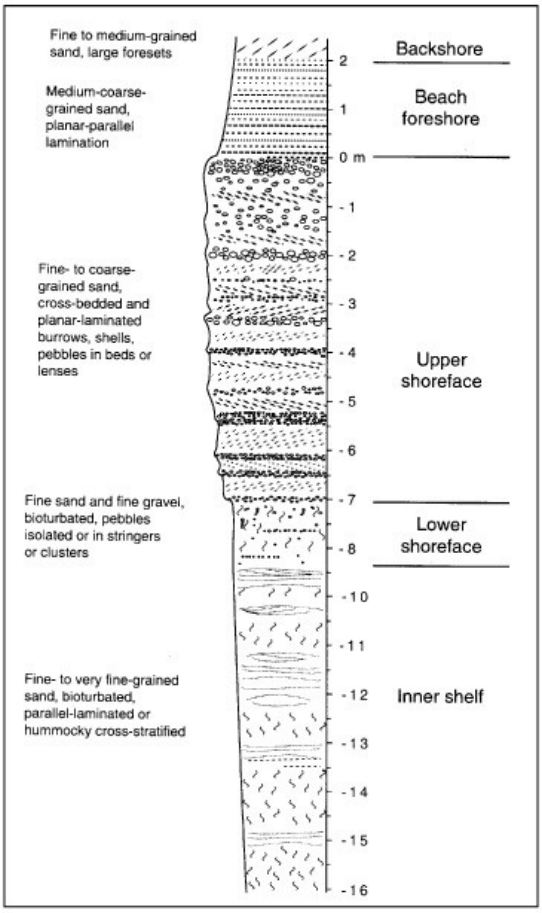
Summary:

Beaches are a type of landform that typically occur along coastlines and are present alongside bodies of water. They usually experience waves or current flow behavior that deposits and reworks sedimentary grains as well as their geologic structure over time. The vertical depositional succession of stratigraphic facies reflects the lateral changes in sedimentary layers witnessed within this environment. Common sedimentary structures that are characteristic of this environment are produced by the processes of wave ripples, planar lamination, cross-bedding, and bioturbation. The key transport processes that dominate in shaping this type of environment include waves and turbulent flows via storms, lower to higher energy (Reynolds' Number), respectively.

Keywords:

Laminar flow, Turbulent flow, Depositional environments, Grain sizes, Beaches, Paleo-beach, Principle of Original Horizontality, Grain transport, Coastline, bioturbation, Wave ripples, Current ripples, Dunes, Planar lamination, Stratigraphic columns, Trace/body fossils, Tides, Translational flow.

References:



1) https://geo.libretexts.org/Courses/University_of_California_Davis/UCD_GEL_109%3A_Sediments_and_Strata_(Sumner), Dawn Sumner UCD

2) https://www.youtube.com/watch?v=U9wtyfwlhCc&feature=youtu.be, "*Underwater in the breaker zone - small waves in the Bahamas*", Dawn Sumner UCD

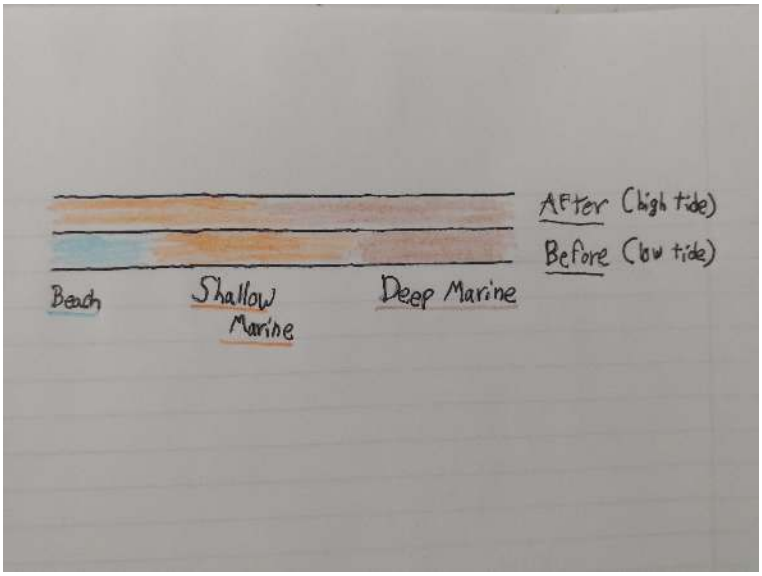
3) https://www.youtube.com/watch?v=CHZtSPZcdz0, "*Beaches and Waves*", Dawn Sumner UCD

4) https://www.youtube.com/watch?v=UG1WuSUoCcw, "*Walther's Law*", Dawn Sumner UCD

5) https://journals.lib.unb.ca/index.ph...icle/view/3154

6) Nichols Sedimentology and Stratigraphy

7) Clifton H.E. (2005) Coastal Sedimentary Facies. In: Schwartz M.L. (eds) Encyclopedia of Coastal Science. Encyclopedia of Earth Science Series. Springer, Dordrecht



8.8: Waves and Beaches is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

SECTION OVERVIEW

8.9: Human impacts on inland waters

Chapter Hook

The aqueducts of Rome are a huge feat of ancient achievement and ingenuity. Not only did they provide water for drinking and bathing but contributed to agriculture, hygiene (through sewers and plumbing), culture, and industry. These stone waterways took about 500 years to complete (312 BC to 226 AD) and contributed significantly to the success of Rome. Water is essential for survival, so other ancient civilizations found ways to move water as well; however, none even came close to being as successful or magnificent.



Figure 8.9.*a*: Roman aqueduct. Image by Pixabay (Public Domain)

https://geo.libretexts.org/@go/page/24621

8.9.1: Fresh Water Supply and the Water Cycle

8.9.2: Water Usage

8.9.3: Water Scarcity and Solutions

8.9.4: Data Dive- Aqueducts of Rome

8.9.5: Review

Attribution

Modified by Rachel Schleiger (CC-BY-NC).

This page titled 8.9: Human impacts on inland waters is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **13: Water Resources** by Melissa Ha and Rachel Schleiger is licensed CC BY-NC 4.0.

8.9.1: Fresh Water Supply and the Water Cycle

Water, air, and food are the most important natural resources to people. Humans can live only a few minutes without oxygen, less than a week without water, and about a month without food. Water also is essential for our oxygen and food supply. Plants breakdown water and use it to create oxygen during the process of photosynthesis.

Human babies are approximately 75% water, and adults are 60% water. Our brain is about 85% water, blood and kidneys are 83% water, muscles are 76% water, and even bones are 22% water. We constantly lose water by perspiration. In temperate climates we should drink about two quarts of water per day, and people in hot desert climates should drink up to 10 quarts of water per day. Loss of 15% of body-water usually causes death.

Water Reservoirs

Water is the only common substance that occurs naturally on Earth in three forms: solid, liquid and gas. The **hydrosphere** is the area of Earth where water movement and water storage occurs. **Water reservoirs** are the locations where water is stored. (Note that this term can also refer to artificial lakes created by dams.) Water is found as a liquid on the surface (rivers, lakes, oceans) and beneath the surface (groundwater), as ice (polar ice caps and glaciers), and as water vapor in the atmosphere. Figure 8.9.1.a illustrates the average time that an individual water molecule may spend in the Earth's major water reservoirs. **Residence time** is a measure of the average time an individual water molecule stays in a particular reservoir.

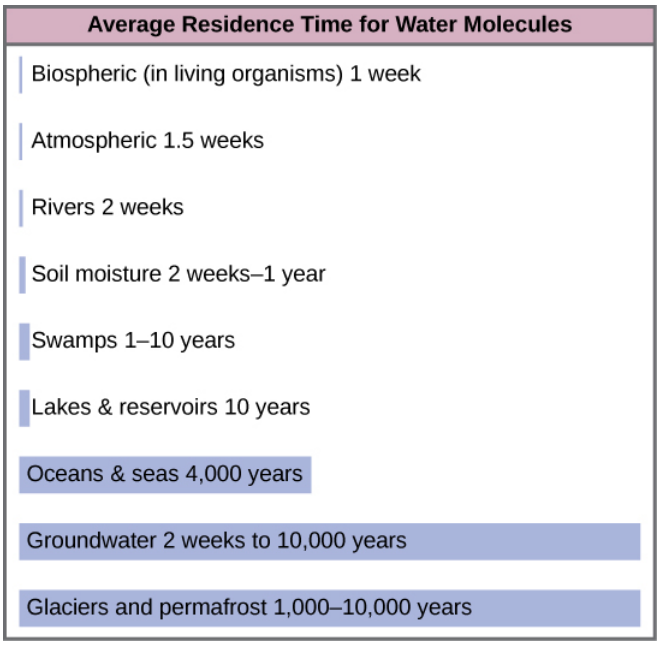
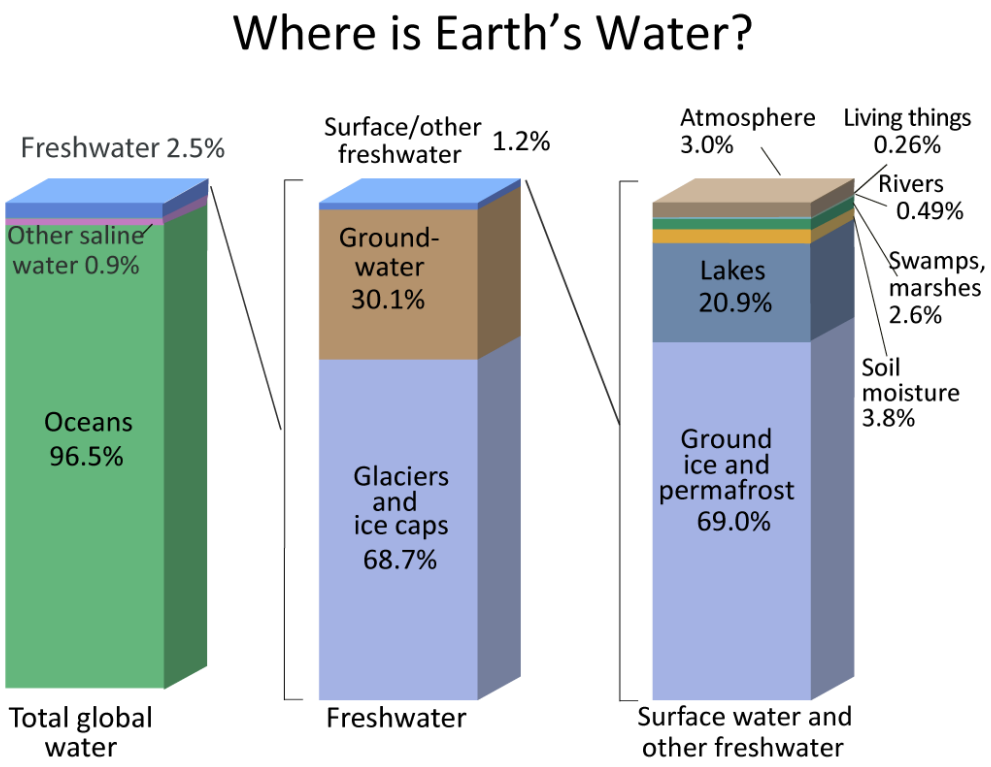


Figure 8.9.1.a: Average residence time that water remains in each reservoir. Water remains in organisms for about one week, in the atmosphere for 1.5 weeks, in rivers for two weeks, as soil moisture from two weeks to a year, in swamps for 1-10 years, in lakes for 10 years, in oceans and seas for 4,000 years, as groundwater for 2 weeks to 10,000 years, and in glaciers or as permafrost for 1,000-10,000 years. Image from OpenStax (CC-BY).

Earth is truly the Water Planet. The abundance of liquid water on Earth's surface distinguishes us from other bodies in the solar system. About 71% of Earth's surface is covered by water, and approximately half of Earth's surface is obscured by clouds (also made of water) at any time. There is a very large volume of water on our planet, about 1.4 billion cubic kilometers (km^3) = 330 million cubic miles, or about 53 billion gallons per person on Earth. All of Earth's water could cover the United States to a depth of 145 km (90 mi).

Despite the vast volumes of water on Earth, only 2.5% is fresh water (figure 8.9.1.b), and only 0.01% is available for humans to use. If all of world's water was shrunk to the size of 1 gallon, then the total amount of fresh water would be about 1/3 cup, and the amount of readily usable fresh water would be 2-3 tablespoons. Much of Earth's fresh water is trapped in glaciers and polar ice caps (figure 8.9.1.c), and this water is inconveniently located, mostly in Antarctica and Greenland. Shallow **groundwater** (water located below the Earth's surface) is the largest reservoir of usable fresh water. Many organisms are dependent on surface waters, such as

lakes and rivers, which comprise a small fraction of fresh water on Earth. A lack of these surface waters can have negative effects on ecosystems.



Source: Igor Shiklomanov's chapter "World fresh water resources" in Peter H. Gleick (editor), 1993, Water in Crisis: A Guide to the World's Fresh Water Resources. (Numbers are rounded).

Figure 8.9.1.*b*: Bar chart of the distribution of Earth's water, including total global water, fresh water, and surface water. Oceans are the largest of Earth's water reservoirs, and only 2.5% of Earth's water is fresh water. Most of this fresh water (68.7%) is in glaciers and ice caps, 30.1% is groundwater, and 1.2% is surface and other fresh water. Of this 1.2%, 3.0% is found in the atmosphere, 0.26% in living things, 0.49% in rivers, 2.6% in swamps and marshes, 3.8% as soil moisture, 20.9% in lakes, and 69.0% as ground ice and permafrost.



Image from USGS/Igor Shiklomanov (public domain).

Figure 8.9.1.*c*: Mountain Glacier in Argentina. Glaciers are the largest reservoir of fresh water, but they are not used much as a water resource directly by society because of their distance from most people. Source: Luca Galuzzi – www.galuzzi.it

The Water Cycle

The **water cycle** (hydrologic cycle) shows the movement of water through different reservoirs, which include oceans, atmosphere, glaciers, groundwater, lakes, rivers, and organisms (figure 8.9.1.*d*). Solar energy, which warms the oceans and other surface waters, and gravity drive the motion of water in the water cycle. This leads to **evaporation** (liquid water to water vapor) of liquid surface water, **sublimation** (ice to water vapor) of frozen water, and **transpiration** (the loss of water from plants to the atmosphere). When water in the soil is taken up by plant roots, it moves through tubes in the plant (the vascular system), evaporates within the space of the leaf, and transpires through the stomata (small microscope openings) of the leaves. Ecologists combine transpiration and evaporation into a single term that describes water returned to the atmosphere: **evapotranspiration**. Thus, large amounts of water move into the atmosphere as water vapor.

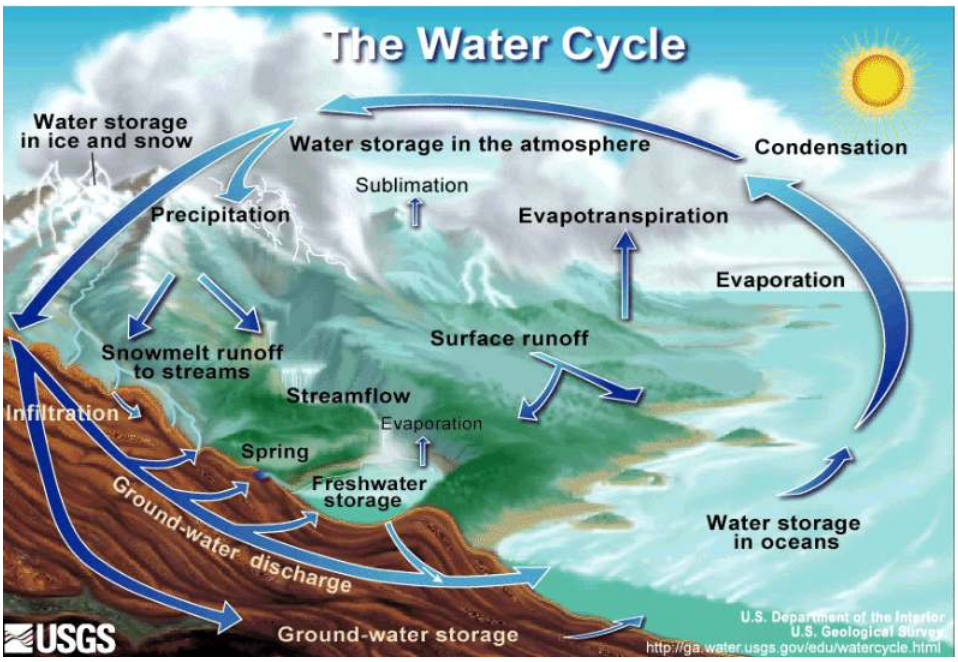


Figure 8.9.1.*d*: The water cycle. Arrows depict movement of water to different reservoirs located above, at, and below Earth's surface. Evapotranspiration and sublimination add water to the atmosphere, where it condenses to form clouds. Water is thus stored in the atmosphere. Through precipitation, water returns to the Earth's surface. When precipitation occurs at high elevation, water is stored as snow or ice. When the snow melts, the resultant runoff can add to streams. Water from precipitation can also infiltrate to replenish groundwater storage or form surface runoff. Groundwater discharge, water from streams, and surface runoff flows into bodies of fresh water and ocean. Source: United States Geological Survey

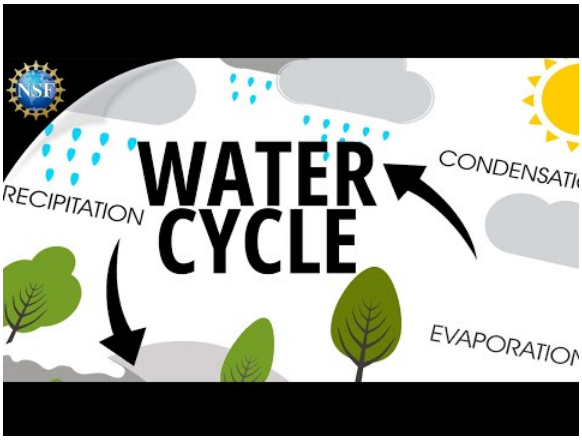
Water vapor in the atmosphere can migrate long-distances from ocean to over land by way of prevailing winds. Over the ocean or land, the air can cool and cause the water to **condense** back into liquid water. This usually happens in the form of very small water droplets that form around a microscopic piece of dust or salt called condensation nuclei. These small water droplets are visible as a cloud. Clouds build and once the water droplets are big enough, they fall to earth as **precipitation** (rain, snow, hail, or sleet), which returns water to Earth's surface.

Precipitation that reaches land can immediately return to the atmosphere, add to groundwater, or form surface runoff. In most natural terrestrial environments, rain encounters vegetation before it reaches the soil surface. A significant percentage of water evaporates immediately from the surfaces of plants or directly from the soil surface. Groundwater is replenished when water **infiltrates** into the soil and finally fills the pore spaces between particles in dirt, sand, and gravel or in the fissures in rocks. Groundwater slowly moves through rock and unconsolidated materials and some of it eventually reaches the surface again, where it discharges as springs and into streams, lakes, and the ocean. Many streams flow not because they are replenished from rainwater directly but because they receive a constant inflow from the groundwater below. Also, surface water in streams and lakes can infiltrate again to recharge groundwater. Therefore, the surface water and groundwater systems are connected. Groundwater can also ultimately flow into the ocean through subsurface groundwater flow, but some groundwater is found very deep in the bedrock and can persist there for millennia. **Surface runoff** is the flow of fresh water over land either from rain or melting ice. Runoff can

make its way through streams and lakes to the oceans. Surface runoff will occur only if the soil becomes saturated with water in a heavy rainfall.

The steps of the water cycle are explained in the video below.





Salinity and the Water Cycle

An important part of the water cycle is how water varies in salinity, which is the abundance of dissolved ions in water. The salt water in the oceans is highly saline, with about 35,000 mg of dissolved ions per liter of seawater. Evaporation is a distillation process that produces nearly pure water with almost no dissolved ions. As water vaporizes, it leaves the dissolved ions in the original liquid phase. Eventually, condensation forms clouds and sometimes precipitation. After rainwater falls onto land, it dissolves minerals in rock and soil, which increases its salinity. Rain and surface runoff are major ways in which minerals, including phosphorus and sulfur, are cycled from land to water. The environmental effects of runoff were discussed in Biogeochemical Cycles. Fresh water (such as lakes, rivers, and near-surface groundwater) has a relatively low salinity.

Human Interactions with The Water Cycle

Humans alter the water cycle by extracting large amounts of fresh water from surface waters as well as groundwater (see Water Usage). Additionally, changes in land use, such as deforestation, agriculture, and urbanization, reduce vegetation cover, which reduces infiltration and increases surface runoff. (Vegetation naturally traps precipitation as it falls, slows the flow surface runoff, and increases the infiltration rate.) This intensifies flooding and exacerbates erosion, lowering soil quality and causing sediment pollution in water. Additionally, humans redirect the flow of water by building dams and aqueducts (figure $8.9.1.e$). So much water

is removed or redirected from the Colorado River in the western United States that, despite its considerable size, in some years it is dry before reaching the sea in Mexico. In an extreme example, the Aral Sea in Central Asia has decreased to only 10% of its initial size after water was diverted for agriculture (see this case study for more details).



Figure 8.9.1.*e*: The California Aqueduct carries water needed for agriculture from Northern California to Southern California. Image by USGS (public domain).

Water Resources

Freshwater resources are ultimately replenished by precipitation. This water can then be obtained from surface water, such as rivers and lakes, and from aquifers, which store groundwater.

Primary Fresh Water Resources: Precipitation

Precipitation levels are unevenly distributed around the globe, affecting fresh water availability (figure 8.9.1.*f*). In general, due the uneven heating of the Earth and global air circulation cells resulting from the Earth's rotation, air rises near the equator and near 60° north and south latitude and sinks at the poles and 30° north and south latitude. As discussed in Climate Effects on Biomes, the intense sunlight at the equator heats air, causing it to rise and cool, which decreases the ability of the air mass to hold water vapor and results in frequent rainstorms. Around 30 degrees north and south latitude, descending air conditions produce warmer air, which increases its ability to hold water vapor and results in dry conditions. Both the dry air conditions and the warm temperatures of these latitude belts favor evaporation.

The size of continents, mountains, prevailing winds, ocean circulation patterns, and even the proximity of bodies of water can affect local climate patterns. For example, when cold winds blow across the relatively warm Great Salt Lake, the air warms, which causes it to pick up moisture. This local increase in the moisture content of the air may eventually fall as snow or rain on nearby mountains, a phenomenon known as “lake-effect precipitation”.

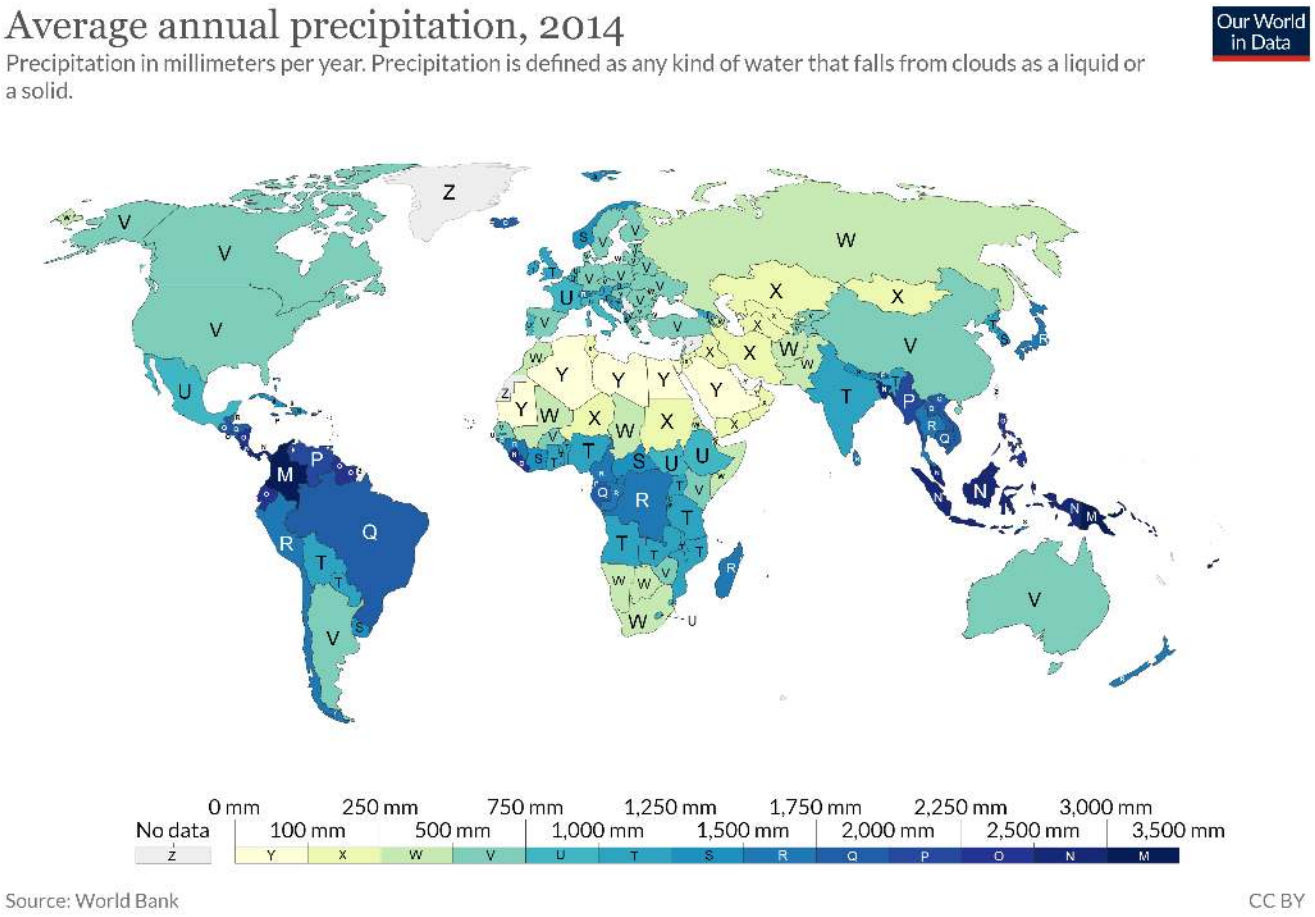


Figure 8.9.1.*f*: Annual average precipitation by country in 2014. Countries with the highest rainfall (3000-3500 mm) are Nicaragua, Costa Rica, Panama, Columbia, Sierra Leone (in West Africa), Bangladesh, and many of the islands is Southeast Asia. Countries with the least rainfall (0-100 mm) are Saudi Arabia and several North African countries. The United States has intermediate rainfall (500-750 mm). The world's rainforests grow in areas with high rainfall while areas with very little rainfall are deserts. Image modified from Our World in Data/World Bank (CC-BY).

In the United States, the 100th Meridian roughly marks the boundary between the humid and arid parts of the country (figure 8.9.1.*g*). Irrigation is required to grow crops west of the 100th Meridian. In the West, surface water is stored in reservoirs (artificial lakes) and mountain snowpacks and strategically released through a system of canals during times of high use.

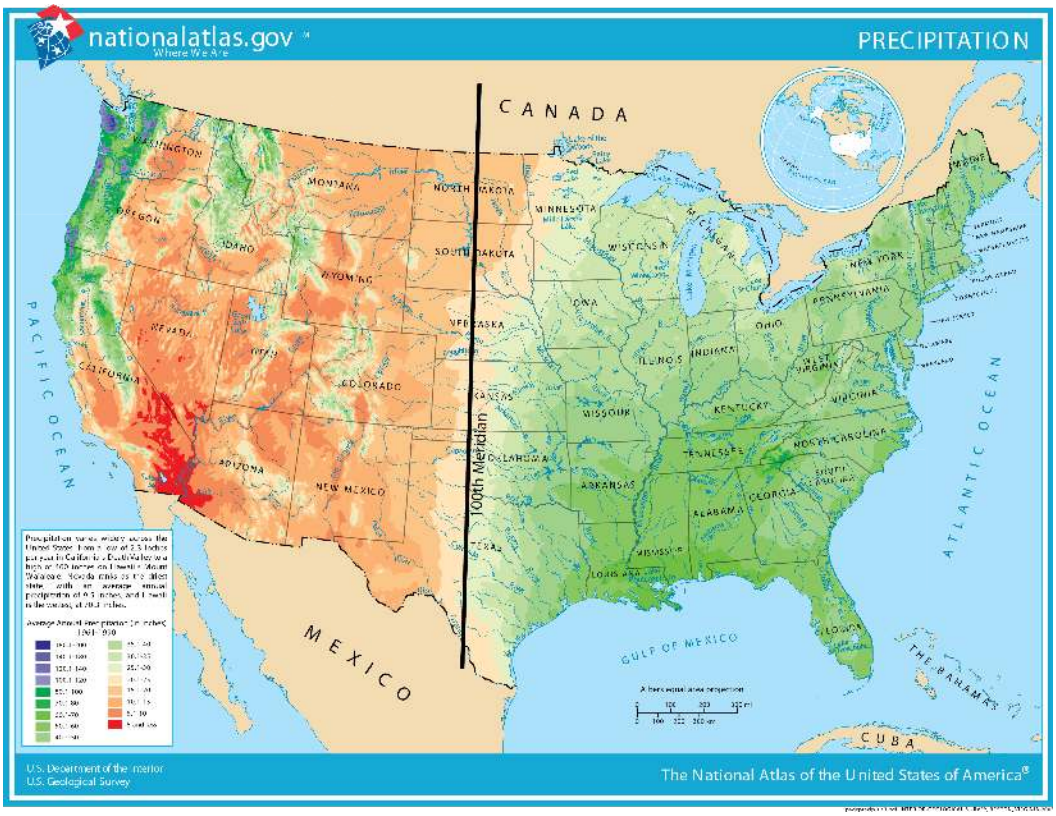


Figure 8.9.1.*g*: Distribution of precipitation in the United States. The 100th Meridian is approximately where the average precipitation transitions from relatively wet to dry. Eastern states are mostly green, indicating greater precipitation than most western states, which are mainly orange or red (indicating low precipitation). (Source: U.S. Geological Survey)

Surface Water: Rivers and Lakes

Rivers are an important water resource for irrigation of cropland and drinking water for many cities around the world. Flowing water from rain and melted snow on land enters river channels by surface runoff (figure 8.9.1.*h*) and seepage from surrounding land. The geographic area drained by a river and its tributaries is called a **watershed**. The Mississippi River watershed includes approximately 40% of the U.S., a measure that includes the smaller watersheds, such as the Ohio River and Missouri River that help to comprise it. Rivers that have had international disputes over water supply include the Colorado (Mexico, southwest U.S.), Nile (Egypt, Ethiopia, Sudan), Euphrates (Iraq, Syria, Turkey), Ganges (Bangladesh, India), and Jordan (Israel, Jordan, Syria).



Figure 8.9.1.*h*: Surface runoff, part of overland flow in the water cycle Source: James M. Pease at Wikimedia Commons

In addition to rivers, lakes can also be an excellent source of fresh water for human use. They usually receive water from surface runoff and groundwater. By building dams, people create artificial lakes (**reservoirs**).

Groundwater Resources

Although most people in the world use surface water, groundwater is a much larger reservoir of usable fresh water, containing more than 30 times more water than rivers and lakes combined. A large area of subsurface, porous rock unit or sediment that holds extractable groundwater is an **aquifer**. The **saturated zone** of an aquifer is where groundwater completely fills pore spaces in earth materials. The **water table** is the uppermost level at which the pores are fully saturated with water (figure 8.9.1.*i*).

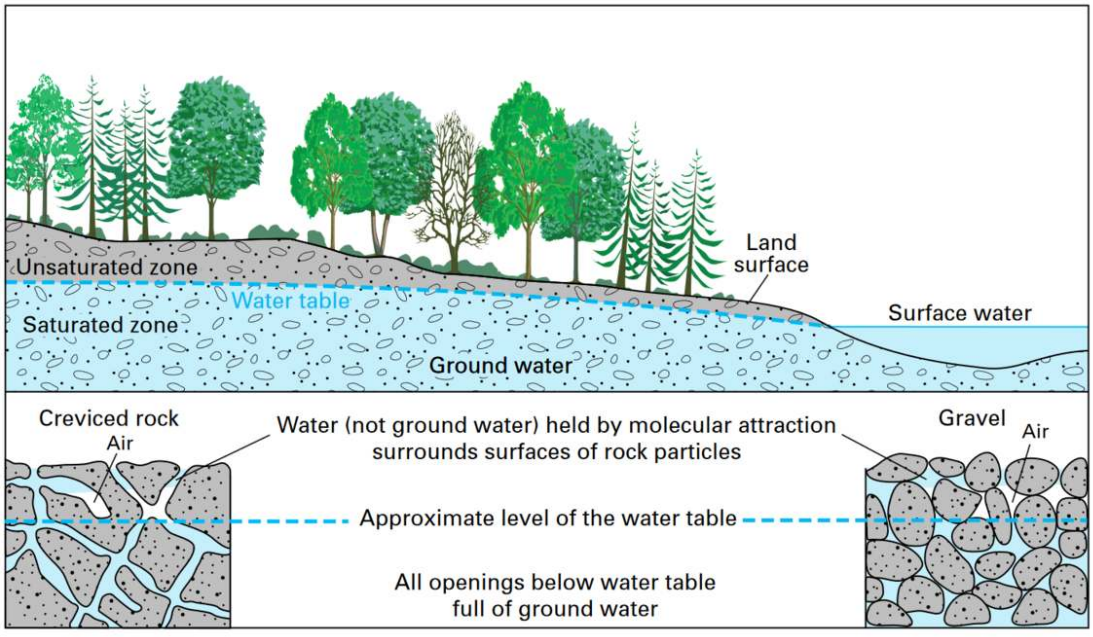


Figure 8.9.1.*i*: The unsaturated and saturated zones of an aquifer, which stores groundwater. The water table is the upper level of the saturated zone. The land surface is covered in vegetation, and the top of the surface water is continuous with the water table. The bottom of the image shows creviced rock (left) and gravel (right), both of which contain spaces that store air and water. Above the water table,

water (not groundwater) is held by molecular attraction and surrounds the surfaces of rock particles. All openings below the water table are full of water table. Image by USGS (public domain)

A combination of a place to put water (porosity) and the ability to move water (permeability) makes a good aquifer. **Porosity** is a measure of the open space in rocks –expressed as the percentage of open space that makes up the total volume of the rock or sediment material. **Permeability** is a measure of the interconnectedness of pores in a rock or sediment. The connections between pores allows for that material to transmit water. Porosity and permeability are functions of soil particle composition. For example, clays generally have very high porosity, but the pores are poorly connected, thereby causing low permeability.

Aquifers are commonly drilled, and wells installed, to provide water for agriculture and personal use. In many cases aquifers are being depleted faster than they are being replenished by water infiltrating down from above. Groundwater is a particularly important resource in arid climates, where surface water may be scarce. In addition, groundwater is the primary water source for rural homeowners, providing 98% of that water demand in the U.S.

As groundwater is pumped from water wells, there usually is a localized drop in the water table around the well called a **cone of depression** (figure 8.9.1.*j*). When there are a large number of wells that have been pumping water for a long time, the regional water table can drop significantly. This is called **groundwater mining**, which can force the drilling of deeper, more expensive wells that commonly encounter more saline groundwater. Rivers, lakes, and artificial lakes (reservoirs) can also be depleted due to overuse. Some large rivers, such as the Colorado in the U.S. and Yellow in China, run dry in some years.

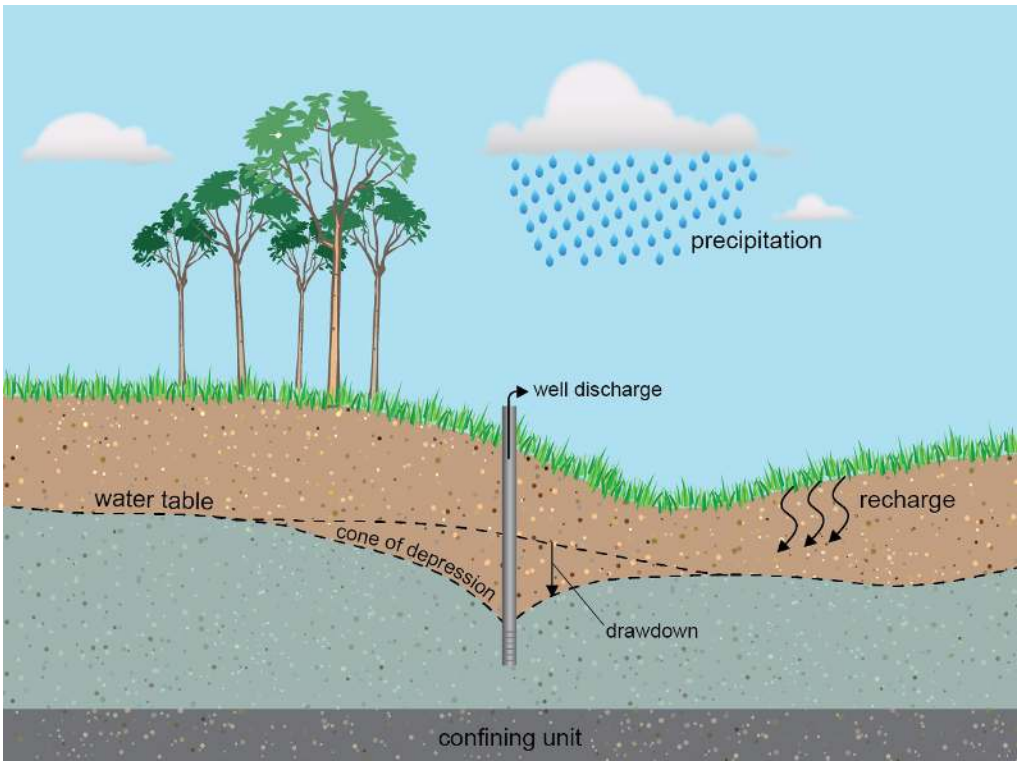


Figure 8.9.1.*j*: As groundwater is extracted from the well, the water table of the aquifer dips, forming a cone of depression. Drawdown is between the water table and the cone of depression. Precipitation is represented by water droplets falling from a cloud. It soaks into the ground, recharging the aquifer. At the bottom of the aquifer is the confining unit, which is impermeable to water. Image by Tara Gross/USGS (public domain).

Another water resource problem associated with groundwater mining is **saltwater intrusion**, where overpumping of freshwater aquifers near ocean coastlines causes salt water to enter freshwater zones. The drop of the water table around a cone of depression in an aquifer can change the direction of regional groundwater flow, which could send nearby pollution toward the pumping well instead of away from it. Finally, problems of **subsidence** (gradual sinking of the land surface over a large area) and **sinkholes** (rapid sinking of the land surface over a small area) can develop due to a drop in the water table. Because the pores in the aquifer collapse as subsidence occurs, this permanently reduces the capacity of the aquifer to hold water in the future.

Groundwater is replenished through infiltration, seepage from surface water (lakes, rivers, reservoirs, and swamps), surface water deliberately pumped into the ground, irrigation, and underground wastewater treatment systems (septic tanks). **Recharge areas** are locations where surface water infiltrates into the ground rather than running into rivers or evaporating (figure 8.9.1.k). Recharge areas are generally the topographically highest location of an aquifer. They are characterized by streams that lie below the water table and sediment or rock that allows infiltration into the subsurface. Wetlands, for example, are excellent recharge areas. Recharge areas mark the beginning of groundwater flow paths.

Recharge can be induced through the aquifer management practice of aquifer storage and recovery. Injection wells allow for humans to increase the rate of recharge into an aquifer system by pumping water into an aquifer (figure 8.9.1.k). Injection wells are regulated by state and federal governments to ensure that the injected water is not negatively impacting the quality or supply of the existing groundwater in the aquifer. Some aquifers are capable of storing significant quantities of water, allowing water managers to use the aquifer system like a surface reservoir. Water is stored in the aquifer during periods of low water demand and high water supply and later extracted during times of high water demand and low water supply.

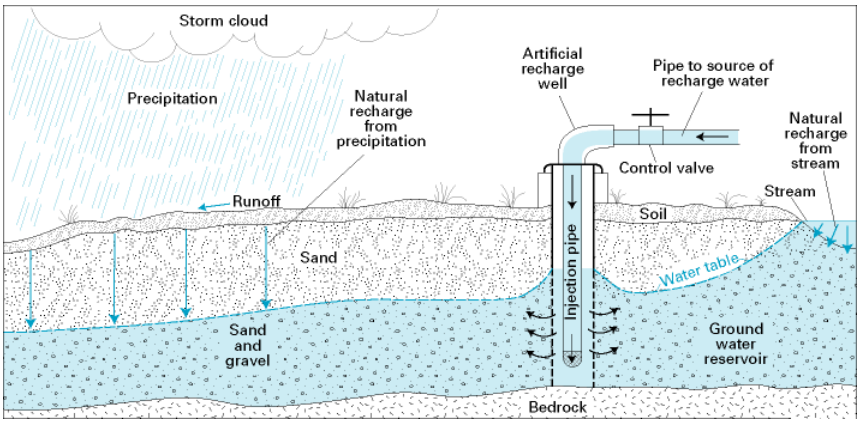


Figure 8.9.1.k: Groundwater can be replenished through natural recharge areas as well as artificial injection wells. In this diagram, precipitation emerges from a storm cloud. Some of the water washes away as runoff, and some water infiltrates through the sand (natural recharge from precipitation). The water fills the sand and gravel in the aquifer (labeled ground water reservoir). Below the aquifer is bedrock, which is impermeable to water. Another natural source of recharge is from streams. Through an artificial recharge well, water flows from a pipe through a control valve into an injection pipe. This also adds to the groundwater. Image by USGS (public domain).

Attributions

- Water Cycle and Fresh Water Supply, Water Supply Problems and Solutions, and Biogeochemical Cycles from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- The Water Cycle, Water Use and Distribution, and Groundwater from *An Introduction to Geology* by Chris Johnson et al. (licensed under CC-BY-NC-SA)
- Biogeochemical Cycles and the Flow of Energy in the Earth System from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (licensed under CC-BY). Download for free at CNX.
- Biogeochemical Cycles from *Biology 2e* by OpenStax (CC-BY). Access for free at openstax.org.

This page titled 8.9.1: Fresh Water Supply and the Water Cycle is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **13.1: Fresh Water Supply and the Water Cycle** by Melissa Ha and Rachel Schleiger is licensed CC BY-NC-SA 4.0.

8.9.2: Water Usage

Freshwater supply is one of the most important ecosystem services. In 2014, global water consumption was 3999 km^3per year (over 1,000 trillion gallons!). The greatest use of this water is for irrigation in agriculture, but significant quantities of water are also extracted for public and municipal use, as well as industrial applications and power generation (figure 8.9.2.a).

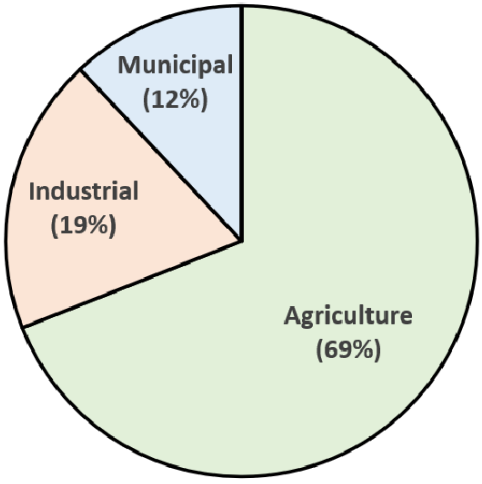


Figure 8.9.2.a: Global water usage by sector. Municipal refers to domestic use, such as by local businesses and households. Data from FAO. Image by Melissa Ha (CC-BY-NC).

Humans require only about 1 gallon per day to survive, but a typical person in a U.S. household uses approximately 80-100 gallons per day, which includes cooking, washing dishes and clothes, flushing the toilet, and bathing. Additionally, we rely on the food, energy, and mineral resources, which all require water to produce. For example, approximately three gallons are needed to produce a tomato, 150 gallons for a loaf of bread, and 1,600 for a pound of beef. (You can learn more about the water footprint of different foods using this interactive website.) Twenty-one gallons are needed to produce one kilowatt-hour (kWh) of electricity from a traditional power plant (about 1 kWh is needed to heat an oven for 30 minutes), and one ton of steel consumes about 63,000 gallons of water. The **water demand** of an area is a function of the population and other uses of water.

In the United States, 281 billion gallons of water were withdrawn per day in 2015, of which 82 billion gallons are fresh groundwater (figure 8.9.2.b). The state of California accounts for 9% of national water withdrawal (figure 8.9.2.c).

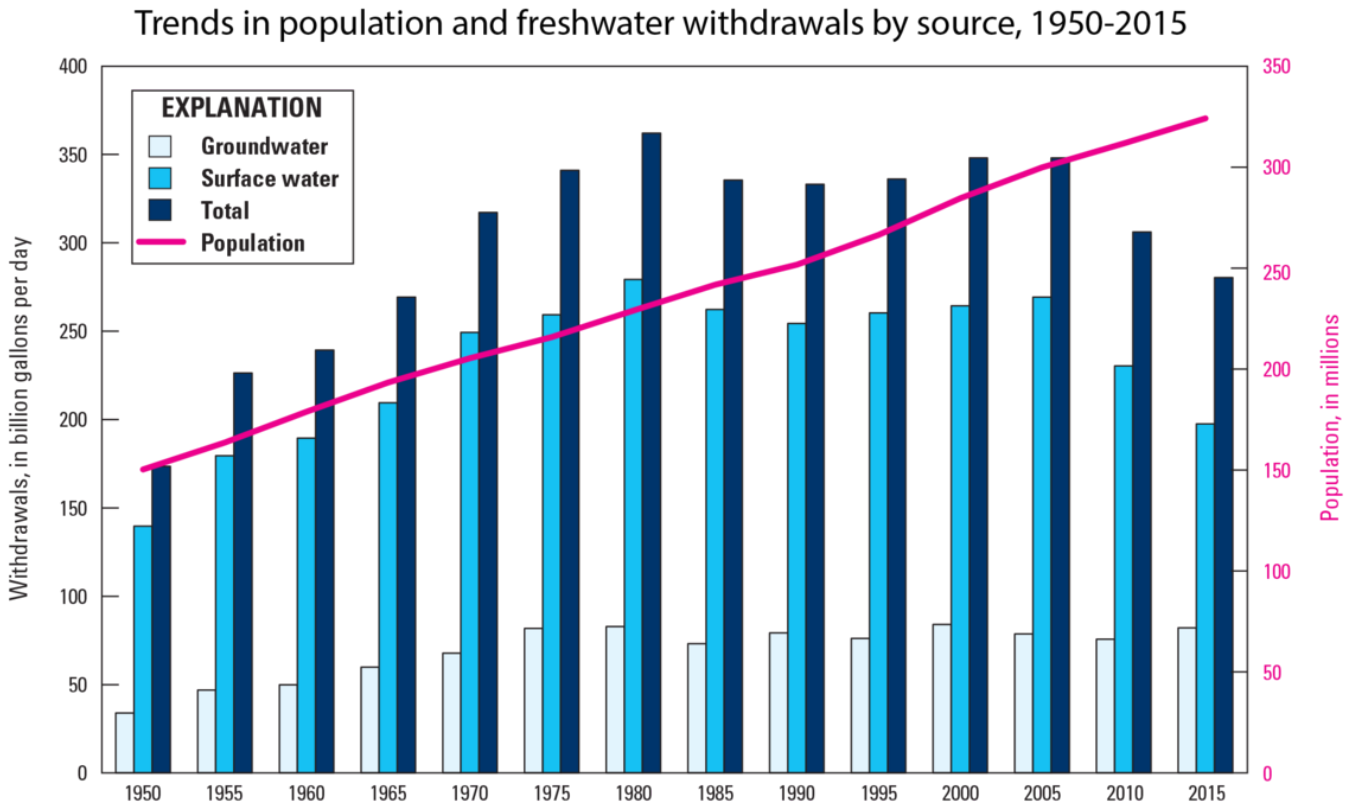


Figure 8.9.2.*b*: Average daily freshwater withdrawals (in billion gallons per day, left y-axis) over time (x-axis) in the United States. Light blue bars (left) represent groundwater withdrawals, medium blue (middle) represents withdrawals from surface water, and dark blue bars (right) represent total withdrawals in billions of gallons per day. In 1950, total withdrawals were around 170 billion gallons per day. This gradually increased until 1980 at just over 350 billion gallons per day. In 1985, total withdrawals dropped to around 325 billion gallons and stayed relatively constant until decreasing in 2010. In 2015, total withdrawals were around 260 billion gallons. Most of the changes in withdrawals overtime were due to changes in surface water withdrawals rather than groundwater withdrawals. Surface water withdrawals have generally been 2-3 times greater than groundwater withdrawals. The pink line represents the United States population size in millions (right y-axis), which has increased from roughly 170 million in 1950 to about 325 million in 2015. Image by USGS (public domain).

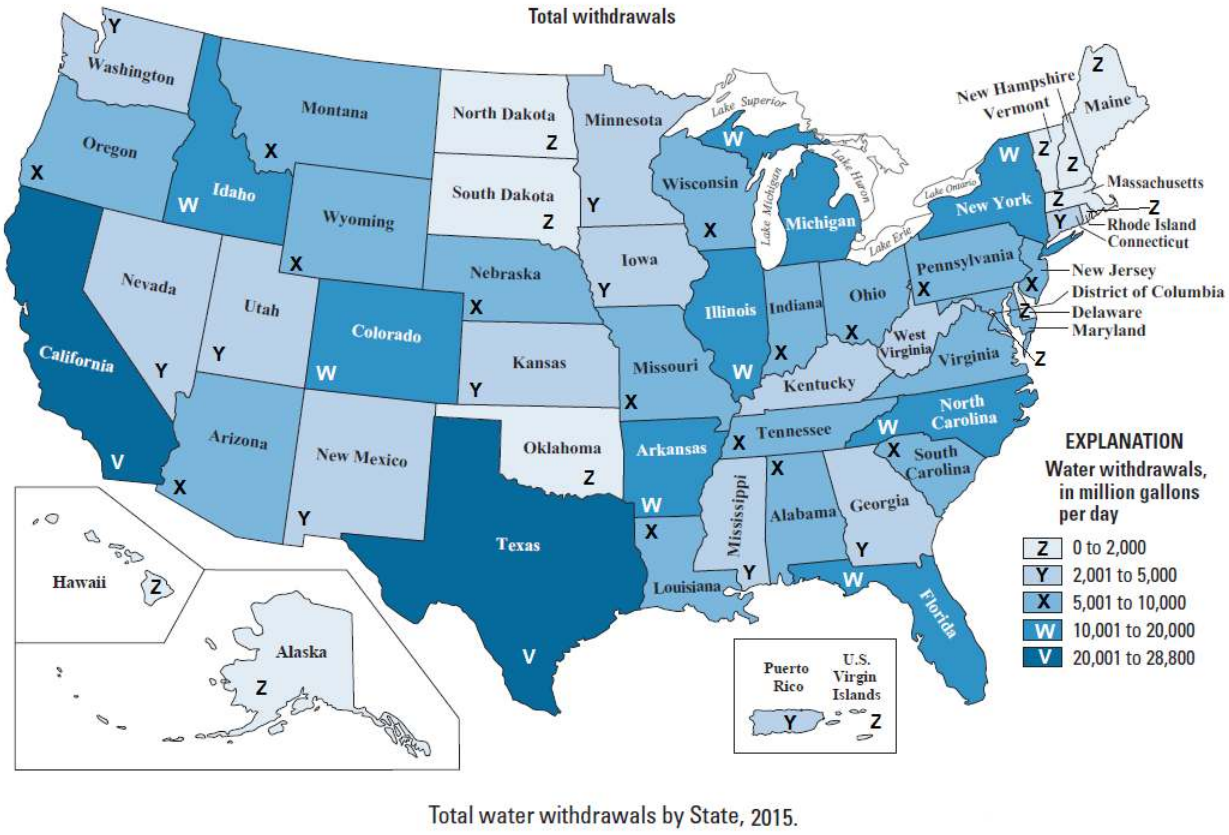


Figure 8.9.2.*c*: Total water withdrawals by state in 2015. The lightest blue (Z) indicates 0-2,000 million gallons per day, and the darkest blue (V) indicates 20,001-28,800 million gallons per day. California and Texas consume the most water, and Alaska, Hawaii, the Dakotas, Oklahoma, and some New England states consume the least. Image modified from USGS (public domain).

Attributions

Modified by Melissa Ha from the following sources

- Water Use and Distribution from *An Introduction to Geology* by Chris Johnson et al. (licensed under CC BY-NC-SA)
- Water Cycle and Fresh Water Supply from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Biogeochemical Cycles and the Flow of Energy in the Earth System from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (licensed under CC-BY). Download for free at CNX.

This page titled 8.9.2: Water Usage is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **13.2: Water Usage** by Melissa Ha and Rachel Schleiger is licensed CC BY-NC-SA 4.0.

8.9.3: Water Scarcity and Solutions

One of the most important environmental goals is to provide clean water to all people. Fortunately, water is a renewable resource and is difficult to destroy. Evaporation and precipitation combine to replenish our fresh water supply constantly; however, water availability is complicated by its uneven distribution over the Earth.

Water Scarcity

The **water crisis** refers to a global situation where people in many areas lack access to sufficient water, clean water, or both. Arid climate and densely populated areas have combined in many parts of the world to create water shortages, which are projected to worsen in the coming years due to population growth, water overuse, water pollution, and climate change. Specifically, climate change shifts precipitation patterns and causes the snow pack that recharges rivers to melt earlier in the year. Furthermore, rising sea levels associated with climate change worsen saltwater intrusion.

Water scarcity refers to water shortages, which can be physical or economic (figure 8.9.3.*a*). **Physical water scarcity** is the lack of sufficient water resources in an area; that is, water is depleted more quickly than it is replenished. Unpredictable precipitation patterns associated with climate change, which increase the risk of flooding and drought, exacerbates physical water scarcity. **Economic water scarcity** occurs when people cannot afford access to water. The United Nations estimates that over half of the global population faces water scarcity for one or more months of the year (see The Sustainable Development Goals Report 2019). According to the World Health Organization and UNICEF, 785 million people lack access to even a basic drinking water service (see Drinking Water) and two billion people lack access to improved sanitation as simple as a pit latrine (see Sanitation), and three billion people lack a facility to wash their hands (see Hand Hygiene for All). As a result, nearly 829,000 people die every year from diarrheal diseases, and 297,000 of those deaths occur among children under the age of five (see Drinking Water).

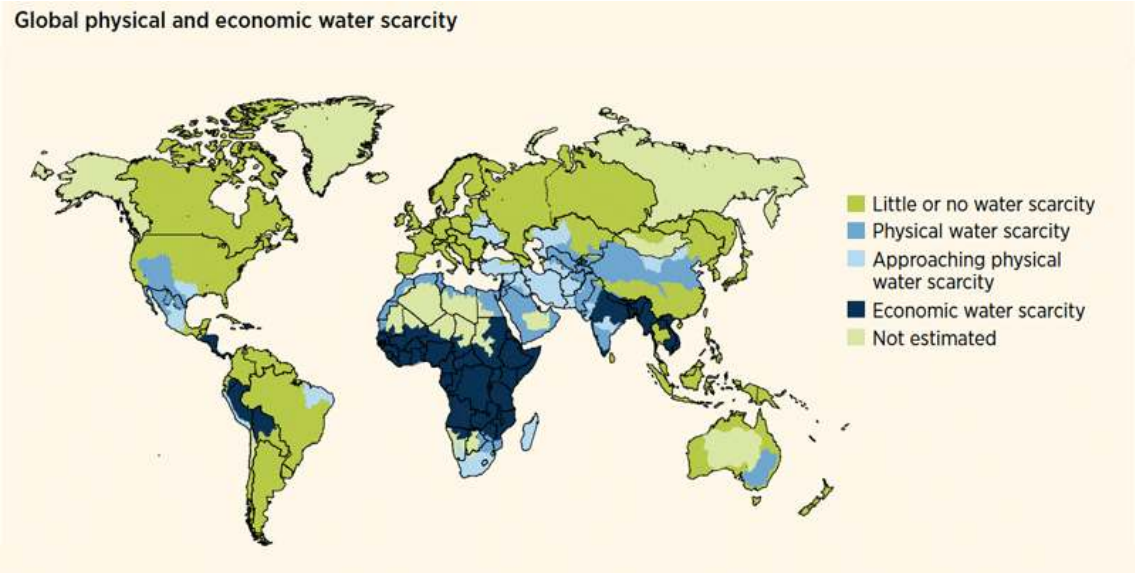


Figure 8.9.3.*a*: Regions facing or approaching water scarcity in 2012. Physical water stress occurs in the southwestern United States, the Middle East, North Africa, southern India, and southeast Australia. Economic water stress occurs in Central and East Africa, western India, and Southeast Asia. Image by World Water Assessment Programme (WWAP; public domain).

Solutions for Addressing Water Shortages

While some human activities have exacerbated the water crisis, humans have also developed technologies to better acquire or conserve freshwater. Solutions to addressing water shortages include dams and reservoirs, rainwater harvesting, aqueducts, desalination, water reuse, and water conservation.

Dams and Reservoirs

Reservoirs (artificial lakes) that form behind dams in rivers can collect water during wet times and store it for use during dry spells (figure 8.9.3.*b*). They also can be used for urban water supplies. Other benefits of dams and reservoirs are hydroelectricity, flood

control, and recreation. Some of the drawbacks are evaporative loss of water in arid climates and downstream river channel erosion. Additionally, dams reduce water flow downstream, which could lead to political conflicts when rivers span states or countries.

The negative ecosystem impacts of dams are another major drawback. For example, dams change a river to a lake habitat and interfere with migration and spawning of fish. Furthermore, warming of the surface water in the reservoir influences the temperature of the water downstream, impacting the fish and aquatic invertebrates that are adapted to colder water. Dams also trap sediments that would otherwise continue to flow down the river, creating habitat and supplying nutrients downstream.



Figure 8.9.3.*b*: The Hoover Dam on the Colorado River in Nevada. Behind the dam is Lake Mead, the largest reservoir in the United States. Image by Dsimic (CC-BY-SA).

Rainwater Harvesting

Rainwater harvesting involves catching and storing rainwater before it reaches the ground. Figure 8.9.3.*c* shows a complex rainwater harvesting system (rain water capture system) proposed for federal buildings, but smaller, simpler systems (sometimes called rain barrels) can be used by individual homeowners (figure 8.9.3.*d*).

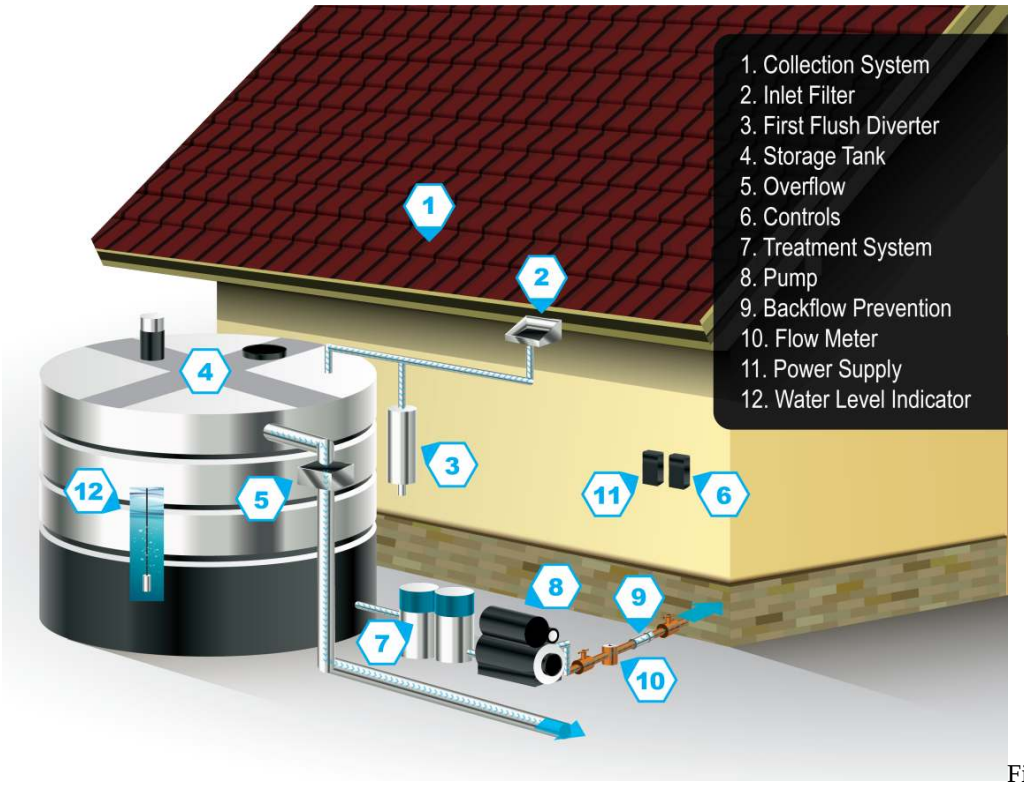


Figure 8.9.3.*c*: The components of a rainwater harvesting system include (1) the collection system consisting of the roof surface and gutters that capture rain water and send it to (2) the inlet filter, a screen that catches large debris. (3) The first flush diverter removes debris not captured by the inlet filter from the initial stream of rainwater. (4) The storage tank is composed of food-grade polyester resin material approved by the U.S. Food and Drug Administration (FDA), which is green in color and helps to reduce bacterial growth. (5) The overflow is a drainage spout that releases excess water if the storage tank is full. The (6) control system monitors the water level and filtration system, and the (7) treatment system is responsible for filtration and disinfection, treating water until it is potable (safe to drink). The (8) pump moves water through the system to where it will be used, and the (9) backflow preventer ensures that water cannot flow backwards through the system in cases of low water pressure (when the storage tank is low). (10) The flow meter (with data logger) measure water production. The (11) power supply may rely on conventional energy sources or solar energy, and the (12) water level indicator monitors the water level in the storage tank. Caption modified from and image from Office of Energy Efficiency and Renewable Energy, Department of Energy (public domain).



Figure 8.9.3.*d*: A rain barrel is a simple rainwater harvesting system. Homeowners commonly use the stored water to irrigate their gardens. Image by Walton LaVonda, USFWS (public domain).

Aqueducts

Aqueducts can move water from where it is plentiful to where it is needed. Aqueducts can be controversial and politically difficult especially if the water transfer distances are large. One drawback is the water diversion can cause drought in the area from where the water is drawn. For example, Owens Lake and Mono Lake in central California began to disappear after their river flow was diverted to the Los Angeles aqueduct (figure 8.9.3.*e*). Without water supply, Owens Lake dried and became a major source of particulate matter, polluting the air during dust storms (see Air Pollution). Owens Lake remains almost completely dry, but Mono Lake has recovered more significantly due to legal intervention. Learn more about the Los Angeles Aqueduct here.



Figure 8.9.3.*e*: The Los Angeles Aqueduct diverts the flow of several eastern California rivers to the city of Los Angeles. Image by Los Angeles (CC-BY-SA).

Desalination

One method that can actually increase the amount of freshwater on Earth is **desalination**, which involves removing dissolved salt and minerals from seawater or saline groundwater (figure 8.9.3.*f*). An advantage of this approach is that there is a virtually unlimited supply of saltwater. There are several ways to desalinate seawater including boiling, filtration, electrodialysis (applying an electric current to removed the ions which comprise salts), and reverse osmosis (figure 8.9.3.*g*). All of these procedures are moderately to very expensive and require considerable energy input, making the water produced much more expensive than freshwater from conventional sources. In addition, the process creates highly saline wastewater, which must be disposed of and creates significant environmental impact. Desalination is most common in the Middle East, where energy from oil is abundant but water is scarce.



Figure 8.9.3.*f*: Port Stanvac Desalination Plant under construction in South Australia. Image by Vmenkov (CC-BY-SA).

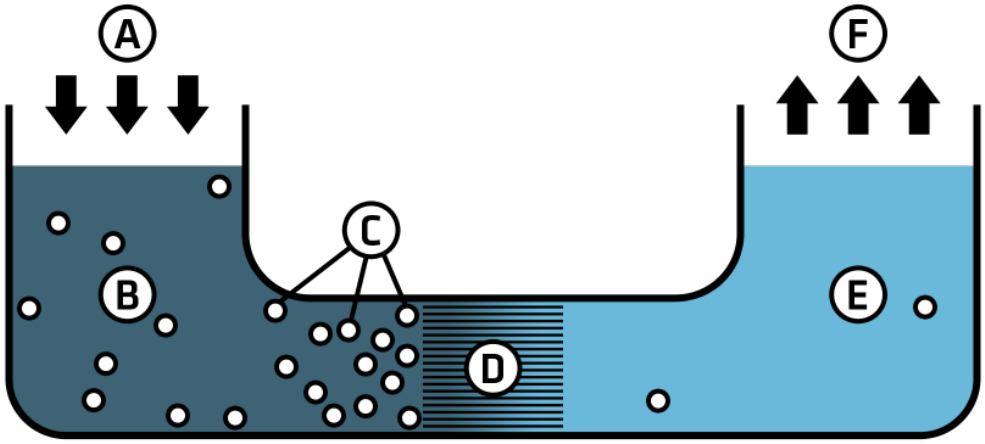


Figure 8.9.3.*g*: Reverse osmosis is one method of desalination. In this process, pressure (A) is applied to salt water, which contains ions (charged atoms or molecules) and other contaminants (C). The water is forced through a semipermeable membrane (D), through which the contaminants cannot pass. Clean freshwater that is safe to drink (E) is then distributed (F). Image by Colby Fisher (CC-BY-SA).

Water Reuse (Water Recycling)

Water recycling refers to reusing water for appropriate purposes such as agriculture, municipal water supply, industrial processes, and environmental restoration (figure 8.9.3.*h*). This could occur at the scale of a single household, for example, installing plumbing that reroutes water drained from the sink to flush the toilet. Water recycling can also occur at large scales. For example, wastewater from the sewage system is regularly treated to an extent, but it can be treated further to produce **potable water** (which is safe to drink) and then pumped into depleted aquifers. This approach limits saltwater intrusion of aquifers near the coast and reduces dependence on precipitation and subsequent infiltration to recharge aquifers. Orange County Water District in California employed this system following an information campaign to explain the purification process and ensure public confidence in the safety of the treated wastewater.

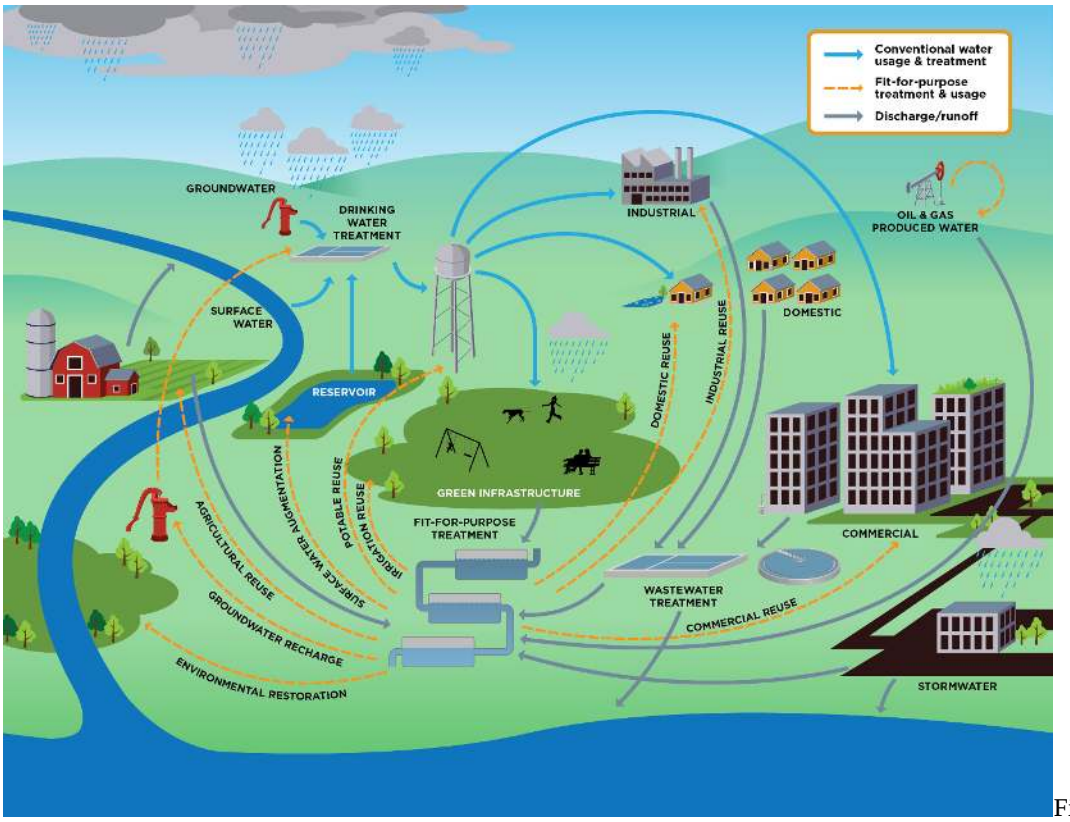


Figure 8.9.3.h: Flowchart of water reuse. Surface water, groundwater, and water from reservoirs undergoes drinking water treatment. From there, it serves commercial, industrial, and domestic purposes and supports green infrastructure. Used water is treated to the extent needed for its purpose (fit-for-purpose treatment). Fit-for-purpose treatment and usage include environmental restoration, ground water recharge, agricultural reuse, surface water augmentation, portable reuse, irrigation reuse, domestic reuse, industrial reuse, and commercial reuse. Discharge from industrial, domestic, and commercial buildings is processed through wastewater treatment. Image by EPA (public domain).

Water Conservation

Water conservation refers to using less water and using it more efficiently. Around the home, conservation can involve both water-saving technologies and behavioral decisions. Examples of water-saving technologies include high-efficiency clothes washers and low-flow showers and toilets. Water-conserving behaviors include turning off the water while you brush your teeth, taking shorter showers and showers instead of baths, and fixing leaky faucets. A dishwasher uses less water than washing dishes by hand, particularly the dishwasher is only run when it is full. Similarly, running fewer, larger loads of laundry conserves water relative to more frequent, smaller loads. Choosing foods with a low water footprint (like eggs) over those with a high water footprint (like beef) can also conserve water.

Gardening offers several water-saving opportunities. If you live in a dry climate, consider growing only native, drought-tolerant vegetation, which requires little irrigation (figure 8.9.3.h). When you do irrigate your garden, do so only as needed and early in the morning, when less water will be lost to evaporation. Drip systems assist in delivering only the needed amount of water in a way that minimizes evaporation. These strategies can also be applied at large scales in agriculture, which is extremely important considering the high agricultural demands on our water supply relative to municipal use. Water conservation strategies in agriculture include growing crops in areas where the natural rainfall can support them, more efficient irrigation systems such as drip systems, and no-till farming, which reduces evaporative losses by covering the soil.

 Figure 8.9.3.*i*: Using water-efficient plants and irrigation not only looks beautiful, but saves resources and supports biodiversity and natural pollinators. Image and caption from CA Department of Water Resources (public domain).

Bottled water is not a sustainable solution to the water crisis. Bottled water is not necessarily any safer than the U.S. public water supply, it costs on average about 700 times more than U.S. tap water, and every year it uses approximately 200 billion plastic and glass bottles that have a relatively low rate of recycling. Compared to tap water, it uses much more energy, mainly in bottle manufacturing and long-distance transportation. (Purchasing a water filter is a more sustainable solution than bottled water if you do not like the taste of tap water.)

References

Drinking Water. 2019. WHO. Accessed 2020-12-29.

Hand Hygeine for All. 2020. UNICEF. Accessed 2020-12-29.

Sanitation. 2019. WHO. Accessed 2020-12-29.

The Sustainable Development Goals Report. 2019. United Nations. Accessed 2020-12-29.

Attribution

Modified by Melissa Ha from the following sources:

- Water Availability and Use from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Basic Information about Water Reuse by United States Environmental Protection Agency (public domain)

This page titled 8.9.3: Water Scarcity and Solutions is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **13.3: Water Scarcity and Solutions** by Melissa Ha and Rachel Schleiger is licensed CC BY-SA 4.0.

8.9.4: Data Dive- Aqueducts of Rome

Overview

All drinking water sources are subject to contamination that require appropriate treatment to eliminate disease-causing pathogens. Sources of water contamination can originate from naturally occurring chemicals and minerals (arsenic, radon, uranium), local land use practices (fertilizers, pesticides), manufacturing processes, and sewage. If left untreated, the presence of contaminants in water can lead to adverse health effects, both acute and chronic. This is why the achievement of the aqueducts of Rome are so astounding! Not only did they water to the residents of Rome, but somehow the Romans were able to identify the cleanest water sources for drinking without the most basic testing equipment. In 1995 Peter Aicher published a book titled, “Guide to the aqueducts of ancient Rome.” The book discusses a lot of the history surrounding the 11 aqueducts while also integrating a more modern analysis of the quality of the water based on where it was sourced. The table below provides a brief overview of the aqueducts along with several notes about their use:

Table 8.9.4.*a***:** Aqueduct characteristics and notes about water volume, source, quality and use. Table by Rachel Schleiger (CC-BY-NC) modified from data in Aicher PJ 1995.

Aqueduct	Construction Complete	Volume (units)	Water Source	Water Quality	Notes
Appia	312 BC	31	Springs	Good	All underground except inside walls 70% for civic/imperial uses
Anio Vetus	272-269 BC	74	River	Poor	Used for baths, gardens, and industry
Marcia	144-240 BC	78	Springs	Best	Pure/cold/hard water Supplied baths
Tepula	126-125 BC	7	Streams	Good	Warm water (60F)
Julia	33 BC	20	Springs	Good	N/A
Virgo	22-19 BC	41	Marsh	Good	Almost all underground, some along channel Supplied baths
Alsietina	2 BC	7	Lake	Poor	Build to supply basin for mock sea battles
Claudia	38-52 AD	76	Springs	Good	Built several branches in the city
Anio Novus	38-52 AD	78	River	Okay	Quality was poor until later improved
Traiana	109 AD	47	Springs	Good	N/A
Alexandria	226 AD	9	Springs	Good	Served baths

Questions

1. How many of the 11 aqueducts do you think have drinkable water?
2. Does the above answer surprise you? Why/why not?
3. How many of the aqueducts with drinkable water come from a higher volume aqueduct?
4. Based on what you can observe in the notes, what did the Romans do when water wasn't drinkable?
5. What is the most surprising thing to you about the data presented in the above table? Why?

Attribution

Rachel Schleiger (CC-BY-NC)

This page titled 8.9.4: Data Dive- Aqueducts of Rome is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **13.4: Data Dive- Aqueducts of Rome** by Melissa Ha and Rachel Schleiger has no license indicated.

8.9.5: Review

Summary

After completing this chapter you should be able to...

- Describe the different water reservoirs and identify which of these reservoirs are accessible to humans.
- Diagram the water cycle and explain how humans interact with the water cycle.
- Explain how groundwater is acquired from aquifers and how aquifers are recharged.
- Identify the main water-consuming sectors and their relative contributions to water usage.
- Compare physical and economic water scarcity.
- Outline strategies for addressing water shortages.
- Provide examples of water conserving technologies and behaviors.

Water is stored in **water reservoirs**. Although there is much water on Earth, only a fraction of this is freshwater that is accessible to humans. The **water cycle** describes the movement of water through each reservoir and involves **sublimination** and **evapotranspiration**, **condensation**, **precipitation**, **infiltration**, and **surface runoff**. Humans alter the water cycle through land use and diverting water flow. Precipitation—a major control of fresh water availability—is unevenly distributed around the globe. More precipitation falls near the equator, and landmasses there are characterized by a tropical rainforest climate. Less precipitation tends to fall near 30 degrees latitude, where the world's largest deserts are located. **Surface waters**, including lakes are rivers, are an important source of freshwater. Additionally, **groundwater** can be extracted from **aquifers**. Agriculture is the greatest consumer of water, followed by industry and municipal use.

The **water crisis** refers to a global situation where people in many areas lack access to sufficient water or clean water or both. **Physical water scarcity** occurs when water resources are insufficient, but people face **economic water scarcity** when they cannot afford access to water. Partial solutions to the water crisis include dams and reservoirs, rainwater harvesting, aqueducts, desalination, water reuse, and water conservation.

Attribution

Modified by Melissa Ha from Water Availability and Use from *Environmental Biology* by Matthew R. Fisher (licensed under CC BY-NC-SA)

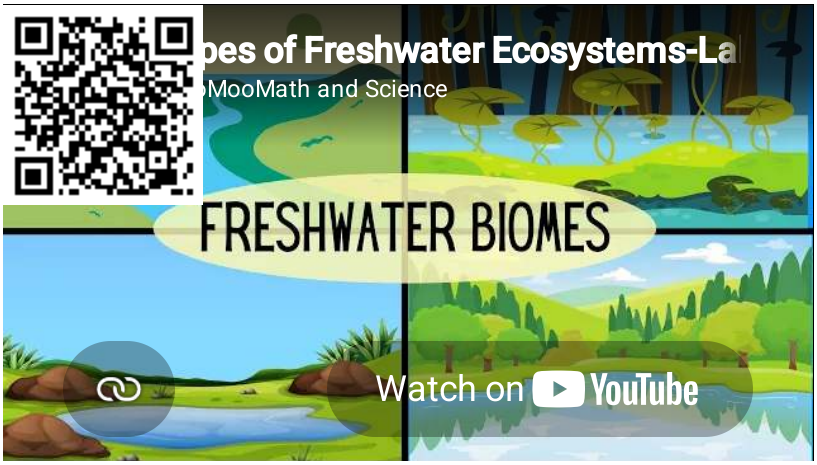
This page titled 8.9.5: Review is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **13.5: Review** by Melissa Ha and Rachel Schleiger is licensed CC BY-NC 4.0.

8.10: Summary

\





8.10: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

9: The Oceans

9.1: Introduction
9.2: Properties of Seawater
9.2.1: Properties of Seawater
9.2.2: Specific Heat and Latent Heat Capacity of Water
9.2.3: Energy Associated With Evaporation and Condensation of Water In the Air
9.2.4: Salinity
9.2.5: Formation of Sea Ice
9.2.6: Relationship of Salinity, Density and Temperature
9.2.7: Salinity and Latitude
9.2.8: "Cline Curves" and Mixing (Surface) Zone
9.2.9: Gases Dissolved In Seawater
9.2.10: Local Conditions in the San Diego Region
9.2.11: Quiz Questions Properties of Seawater
9.3: Biogeochemistry of the surface ocean
9.3.1: Introduction
9.3.2: Biological Pump
9.3.3: Export from the Upper Ocean to the Depths and Sediments
9.3.4: Summary
9.4: 12.4 Biogeochemistry of the Deep Ocean
9.5: Biogeochemistry in ocean sediments
9.5.1: Introduction
9.5.2: Classifying Sediments
9.5.3: Lithogenous Sediments
9.5.4: Biogenous Sediments
9.5.5: Hydrogenous Sediments
9.5.6: Cosmogenous Sediments
9.5.7: Sediment Distribution
9.6: Marine element cycles
9.7: The Oceans and Global Climate Change
9.8: Summary

9: The Oceans is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

9.1: Introduction

Water is the most abundant substance at the earth's surface. Almost all of it is in the oceans, which cover 70% of the surface area of the earth. However, the amounts of water present in the atmosphere and on land (as surface runoff, lakes and streams) is great enough to make it a significant agent in transporting substances between the lithosphere and the oceans. Water interacts with both the atmosphere and the lithosphere, acquiring solutes from each, and thus provides the major chemical link between these two realms. The various transformations undergone by water through the different stages of the hydrologic cycle act to transport both dissolved and particulate substances between different geographic locations.

Reservoir	Volume / 10^6, km^3	Percent of total
Oceans	1370	97.25
Ice caps and glaciers	39	2.05
Deep groundwater	5.3	0.38
Shallow groundwater	4.3	0.30
Lakes	0.125	0.01
Soil moisture	0.065	0.005
Atmosphere	0.013	0.001
Rivers	0.0017	0.0001
Biosphere	0.0006	0.00004
Total	**1408.7**	**100**

Composition of Seawater

The composition of the ocean has attracted the attention of some of the more famous names in science, including Robert Boyle, Antoine Lavoisier and Edmund Halley. Their early investigations tended to be difficult to reproduce, owing to the different conditions under which they crystallized the various salts. As many as 54 salts, double salts and hydrated salts can be obtained by evaporating seawater to dryness. At least 73 elements are now known to be present in seawater.

Table 1: Major ions of seawater: These values, expressed in parts per thousand, are for seawater of 35% salinity.

cations	g/kg	anions	g/kg
Na^+	10.77	Cl^-	19.354
Mg^{2+}	1.29	SO_4^{2-}	2.712
Ca^{2+}	0.412	Br^-	0.087
K^+	0.399		
Sr^{2+}	0.0079		
Al^{3+}	0.005		

The best way of characterizing seawater is in terms of its ionic content, shown above. The remarkable thing about seawater is the constancy of its relative ionic composition. The overall salt content, known as the salinity (grams of salts contained in 1 kg of seawater), varies slightly within the range of 32-37.5%, corresponding to a solution of about 0.7% salt content. The ratios of the concentrations of the different ions, however, are quite constant, so that a measurement of Cl^- concentration is sufficient to determine the overall composition and total salinity. Although most elements are found in seawater only at trace levels, marine organisms may selectively absorb them and make them more detectable. Iodine, for example, was discovered in marine algae

(seaweeds) 14 years before it was found in seawater. Other elements that were not detected in seawater until after they were found in marine organisms include barium, cobalt, copper, lead, nickel, silver and zinc. Si^{32}, presumably deriving from cosmic ray bombardment of Ar, has been discovered in marine sponges.

pH balance. Reflecting this constant ionic composition is the pH, which is usually maintained in the narrow range of 7.8-8.2, compared with 1.5 to 11 for fresh water. The major buffering action derives from the carbonate system, although ion exchange between Na+ in the water and H^+ in clay sediments has recently been recognized to be a significant factor.

Conservative and non-conservative substances

The major ionic constituents whose concentrations can be determined from the salinity are known as conservative substances. Their constant relative concentrations are due to the large amounts of these species in the oceans in comparison to their small inputs from river flow. This is another way of saying that their residence times are very large.

Table 2: Replacement time with respect to river addition for some components of seawater. Concentrations in micromols/L

component	concentration in river water	concentration in seawater	residence time, 10^6 y
Cl^-	250	558,000	87
Na^+	315	479,000	55
Mg^{2+}	150	54,300	13
SO_4^{2-}	120	28,900	8.7
Ca^{2+}	367	10,500	1
K^+	36	10,400	10
HCO_3^-	870	2000	0.083
H_4SiO_4	170	100	0.021
NO_3^-	10	20	0.072
$H_2PO_4^-$	0.7	1	0.080

A number of other species, mostly connected with biological activity, are subject to wide variations in concentration. These include the nutrients NO_3^-, NO_2^-, NH_4^+, and HPO_4^{2-}, which may become depleted near the surface in regions of warmth and light. As was explained in the preceding subsection on coastal upwelling, offshore prevailing winds tend to drive Western coastal surface waters out to sea, causing deeper and more nutrient-rich water to be drawn to the surface. This upwelled water can support a large population of phytoplankton and thus of zooplankton and fish. The best-known example of this is the anchovy fishery off the coast of Peru, but the phenomenon occurs to some extent on the West coasts of most continents, including our own.

Other non-conservative components include Ca^{2+} and SiO_4^{2-}. These ions are incorporated into the solid parts of marine organisms, which sink to greater depths after the organisms die. The silica gradually dissolves, since the water is everywhere undersaturated in this substance. Calcium carbonate dissolves at intermediate depths, but may reprecipitate in deep waters owing to the higher pressure. Thus the concentrations of Ca and of SiO_4^{2-} tend to vary with depth. The gases O_2 and CO_2, being intimately involved with biological activity, are also non-conservative, as are N_2O and CO.

Organic matter

Most of the organic carbon in seawater is present as dissolved material, with only about 1-2% in particulates. The total organic carbon content ranges between 0.5 mg/L in deep water to 1.5 mg/L near the surface. There is still considerable disagreement about the composition of the dissolved organic matter; much of it appears to be of high molecular weight, and may be polymeric. Substances qualitatively similar to the humic acids found in soils can be isolated. The greenish color that is often associated with coastal waters is due to a mixture of fluorescent, high molecular weight substances of undetermined composition known as *"Gelbstoffe"*. It is likely that the significance of the organic fraction of seawater may be much greater than its low abundance

would suggest. For one thing, many of these substances are lipid-like and tend to adsorb onto surfaces. It has been shown that any particle entering the ocean is quickly coated with an organic surface film that may influence the rate and extent of its dissolution or decomposition. Certain inorganic ions may be strongly complexed by humic-like substances. The surface of the ocean is mostly covered with an organic film, only a few molecular layers thick. This is believed to consist of hydrocarbons, lipids, and the like, but glycoproteins and proteoglycans have been reported. If this film is carefully removed from a container of seawater, it will quickly be reconstituted. How significant this film is in its effects on gas exchange with the atmosphere is not known.

Regulation of ocean composition

The salinity of the ocean appears to have been about the same for at least the last 200 million years. There have been changes in the relative amounts of some species, however; the ratio of Na/K has increased from about 1:1 in ancient ocean sediments to its present value of 28:1. Incorporation of calcium into sediments by the action of marine organisms has depleted the Ca/Mg ratio from 1:1 to 1:3.

Table: Mass balance of P, C, and Ca for the oceans

element	input to ocean	dissolved in seawater	in dead organisms	loss to sediments	residence time, y
phosphorus	1 (erosion)	1	1	1 (organic)	10,000
carbon	100 as CO_2 500 as carbonate	1000	125	100 (organic) 600 (carbonate)	165,000
calcium	500 (erosion)	5000	25	500 (as $CaCO_3$)	10^6

If the composition of the ocean has remained relatively unchanged with time, the continual addition of new mineral substances by the rivers and other sources must be exactly balanced by their removal as sediment, possibly passing through one or more biological systems in the process.

Where does the Salt Come From?

In 1715 Edmund Halley suggested that the age of the ocean (and thus presumably of the world) might be estimated from the rate of salt transport by rivers. When this measurement was actually carried out in 1899, it gave an age of only 90 million years.

> *This is somewhat better than the calculation made in 1654 by James Ussher, the Anglican Archbishop of Armagh, Ireland, based on his interpretation of the Biblical book of Genesis, that the world was created at 9 A.M. on October 23, 4004 BC, but it is still far too recent, being about when the dinosaurs became extinct. What Halley actually described was the residence time, which is about right for Na but much to long for some of the minor elements of seawater.*

The commonly stated view that the salt content of the oceans derives from surface runoff that contains the products of weathering and soil leaching is not consistent with the known compositions of the major river waters (See Table). The halide ions are particularly over-represented in seawater, compared to fresh water. These were once referred to as “excess volatiles”, and were attributed to volcanic emissions. With the discovery of plate tectonics, it became apparent that the locations of seafloor spreading at which fresh basalt flows up into the ocean from the mantle are also sources of mineral-laden water. Some of this may be seawater that has cycled through a hot porous region and has been able to dissolve some of the mineral material owing to the high temperature. Much of the water, however, is “juvenile” water that was previously incorporated into the mantle material and has never before been in the liquid phase. The substances introduced by this means (and by volcanic activity) are just the elements that are “missing” from river waters. Estimates of what fraction of the total volume of the oceans is due to juvenile water (most of it added in the early stages of mantle differentiation that began a billion years ago) range from 30 to 90%.

Geochemical processes involving the oceans

The oceans can be regarded as a product of a giant acid-base titration in which the carbonic acid present in rain reacts with the basic materials of the lithosphere. The juvenile water introduced at locations of ocean-floor spreading is also acidic, and is partly neutralized by the basic components of the basalt with which it reacts. Surface rocks mostly contain aluminum, silicon and oxygen combined with alkali and alkaline-earth metals, mainly potassium, sodium and calcium. The CO_2 and volcanic gases in rainwater react with this material to form a solution of the metal ion and HCO_3^-, in which is suspended some hydrated SiO^2. The solid material left behind is a clay such as kaolinite, $Al_2Si_2O_5(OH)_4$. This first forms as a friable coating on the surface of the weathered rock; later it becomes a soil material, then an alluvial deposit, and finally it may reach the sea as a suspended sediment. Here it may undergo a number of poorly-understood transformations to other clay sediments such as illites. Sea floor spreading eventually transports these sediments to a subduction region under a continental block, where the high temperatures and pressures permit reactions that transform it into hard rock such as granite, thus completing the geochemical cycle.

Deep-sea hydrothermal vents are now recognized to be another significant route for both the addition and removal of ionic substances from seawater.

Distribution and cycling of elements in the oceans

Although the relative concentrations of most of the elements in seawater are constant throughout the oceans, there are certain elements that tend to have highly uneven distributions vertically, and to a lesser extent horizontally. Neglecting the highly localized effects of undersea springs and volcanic vents, these variations are direct results of the removal of these elements from seawater by organisms; if the sea were sterile, its chemical composition would be almost uniform.

Plant life can exist only in the upper part of the ocean where there is sufficient light available to drive photosynthesis. These plants, together with the animals that consume them, extract nutrients from the water, reducing the concentrations of certain elements in the upper part of the sea. When these organisms die, they fall toward the lower depths of the ocean as particulate material. On the way down, some of the softer particles, deriving from tissue, may be consumed by other animals and recycled. Eventually, however, the nutrient elements that were incorporated into organisms in the upper part of the ocean will end up in the colder, dark, and essentially lifeless lower part.

Mixing between the upper and lower reservoirs of the ocean is quite slow, owing to the higher density of the colder water; the average residence time of a water molecule in the lower reservoir is about 1600 years. Since the volume of the upper reservoir is only about 1/20 of that of the lower, a water molecule stays in the upper reservoir for only about 80 years.

Except for dissolved oxygen, all elements required by living organisms are depleted in the upper part of the ocean with respect to the lower part. In the case of the major nutrients P, N and Si, the degree of depletion is sufficiently complete (around 95%) to limit the growth of organisms at the surface. These three elements are said to be biolimiting. A few other biointermediate elements show partial depletion in surface waters: Ca (1%), C (15%), Ba (75%).

The organic component of plants and animals has the average composition $C_{80}N_{15}P$. It is remarkable that the ratio of N:P in seawater (both surface and deep) is also 15:1; this raises the interesting question of to what extent the ocean and life have co-evolved.

In the deep part of the ocean the elemental ratio corresponds to $C_{800}N_{15}P$, but of course with much larger absolute amounts of these elements. Eventually some of this deeper water returns to the surface where the N and P are quickly taken up by plants. But since plants can only utilize 80 out of every 800 carbon atoms, 90 percent of the carbon will remain in dissolved form, mostly as HCO_3^-.

> *To work out the balance of Ca and Si used in the hard parts of organisms, we add these elements to the average composition of the lower reservior to get $Ca_{3200}Si_{50}C_{800}N_{15}P$. Particulate carbon falls into the deep ocean in the ratio of about two atoms in organic tissue to one atom in the form of calcite. This makes the overall composition of detrital material something like $C_{120}N_{15}P$; i.e., 80 organic C's and 40 in $CaCO_3$. Accompanying these 40 calcite units will be 40 Ca atoms, but this represents a minor depletion of the 3200 Ca atoms that eventually return to the surface, so this element is only slightly*

> *depleted in the upper waters. Silicon, being far less abundant, is depleted to a much greater extent.*

The oceanic sediments

The particulate shower

A continual rain of particulate material from dead organisms falls through the ocean. This shower is comprised of three major kinds of material: calcite ($CaCO_3$), silica (SiO_2), and organic matter. The first two come from the hard parts of both plants and animals (mainly microscopic animals such as foraminifera and radiolarians). The organic matter is derived mainly from the soft tissues of organisms, and from animal fecal material. Some of this solid material dissolves before it reaches the ocean floor, but not usually before it enters the deep ocean where it will remain for about 1600 years.

The remainder of this material settles onto the floor of the sea, where it forms one component of a layer of sediments that provide important information about the evolution of the sea and of the earth. Over a short time scale of months to years, these sediments are in quasi-equilibrium with the seawater. On a scale of millions of years, the sediments are merely way-stations in the geochemical cycling of material between the earth's surface and its interior.

The oceanic sediments have three main origins:

- **Detrital** material is derived largely from particles deposited in the ocean by rivers and also directly by the wind. These materials are mostly aluminosilicates (clays), along with some quartz. These substances accumulate on the deep ocean floor at a rate of about 0.3 g cm^{-2} per 1000 years.
- **Authigenic** materials are formed by precipitation within the ocean or by crystallization within the sediment itself. These constitute only a small fraction of the total sediment.
- **Biogenic** components consist mainly of the calcium carbonate and silica that make up the hard parts of organisms. Scavenging by other organisms is so efficient that only about 0.3 % of the total deep sea sediment consists of actual organic material.

Our main interest lies with the silica and calcium carbonate, since these substances form a crucial part of the biogeological cycle. Also, their distributions in the ocean are not uniform- a fact that must tell us something.

The skeletons of diatoms and radiolarians are the principal sources of silica sediments. Since the ocean is everywhere undersaturated with respect to silica, only the most resistant parts of these skeletons reach the bottom of the deep ocean and get incorporated into sediments. Silica sediments are less common in the Atlantic ocean, owing to the lower content of dissolved silica.

The parts of the ocean where these sediments are increasing most rapidly correspond to regions of upwelling, where deep water that is rich in dissolved silica rises to the surface where the silica is rapidly fixed by organisms. Where upwelling is absent, the growth of the organisms is limited, and little silica is precipitated. Since deep waters tend to flow from the Atlantic into the Pacific ocean where most of the upwelling occurs, Atlantic waters are depleted in silica, and silica sediments are not commonly found in this ocean.

For calcium carbonate, the situation is quite different. In the first place, surface waters are everywhere supersaturated with respect to both calcite and aragonite, the two common crystal forms of $CaCO_3$. Secondly, Ca^{2+} and HCO_3^- are never limiting factors in the growth of the coccoliths (plants) and forams (animals) that precipitate $CaCO_3$; their production depends on the availability of phosphate and nitrogen. Because these elements are efficiently recycled before they fall into the deep ocean, their supply does not depend on upwelling, and so the production of solid is more uniformly distributed over the world's oceans.

More importantly, however, the chances that a piece of carbonate skeleton will end up as sediment will be highly dependent on both the local CO_3^{2-} concentration and the depth of the ocean floor. These factors give rise to small-scale variations in the production of carbonate sediments that can be quite wide-ranging.

Ocean sediments and continental drift

New crust is being generated and moving away from the crests of the mid-ocean ridges at a rate of a few centimetres per year. Although the crests of these ridges are relatively high points, projecting to within about 3000 m of the surface, the continual injection of new material prevents sediments from accumulating in these areas. Farther from the crests, carbonate sediments do

build up, eventually reaching a depth of about 500 m, but by this time the elevation has dropped off below the saturation horizon, so from this point on the carbonate sediments are overlaid by red clay.

If we drill a hole down through a part of the ocean floor that is presently below the saturation horizon, the top part of the drill core will consist of clay, followed by $CaCO_3$ at greater depths. The core may also contain regions in which silica predominates. Since silica production is very high in equatorial regions, the appearance of such a layer suggests that this particular region of the oceanic crust has moved across the equator.

Contributors and Attributions

- Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook

9.1: Introduction is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

SECTION OVERVIEW

9.2: Properties of Seawater

9.2.1: Properties of Seawater

9.2.2: Specific Heat and Latent Heat Capacity of Water

9.2.3: Energy Associated With Evaporation and Condensation of Water In the Air

9.2.4: Salinity

9.2.5: Formation of Sea Ice

9.2.6: Relationship of Salinity, Density and Temperature

9.2.7: Salinity and Latitude

9.2.8: "Cline Curves" and Mixing (Surface) Zone

9.2.9: Gases Dissolved In Seawater

9.2.10: Local Conditions in the San Diego Region

9.2.11: Quiz Questions Properties of Seawater

This page titled 9.2: Properties of Seawater is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

9.2.1: Properties of Seawater

Properties of Seawater

Seawater is the most abundant resource on Earth! Seawater has evolved to what it is over the billions of years that oceans have existed on Earth. This chapter examines the physical and chemical properties of water and seawater.

Distribution of Water on Earth

97.2% in the world ocean
2.15% frozen in glaciers and ice caps
0.62% in groundwater and soil moisture
0.02% in streams and lakes
0.001% as water vapor in the atmosphere

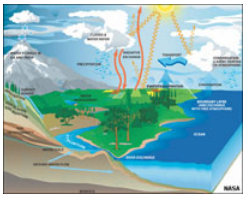


Figure 7.1. The water cycle.

Components of Seawater

Seawater is composed of**:**

• **Water**
• **Dissolved matter: solids and gas** (as **ions**)
• **Suspended matter** (dust and organic residues)

Ions in seawater

• Cl - 55%
• Na - 30.4%
• SO_4 - 7.6%
• Mg - 3.9%
• Ca - 1.2%
• K - 1.1%
• all other dissolved components - <1 %

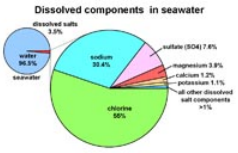


Figure 7.2. Components of seawater

Properties of Water

Water is a **polar substance**. Each molecule of water has a negative charge associated with its oxygen, and a positive charge with it hydrogen (Figure 7.3). This polar character is responsible for its properties cohesion and adhesion.

• **Cohesion:** water has high surface tension because water molecules stick together.
• **Adhesion:** water "sticks" to things.
• **High capillary action**: The cohesion and adhesion properties of water allow water to move upward against gravity in small confined spaces. The smaller the tube, the higher the water will rise. Capillary action helps plants to move water upward from their roots to their leaves.

• **Water is a powerful solvent**: A **solvent** is a substance that dissolves a **solute** (a chemically different liquid, solid or gas), resulting in a **solution**. The polar character of the water molecule allows it to form weak bonds with other polar molecules. Substance held together with ionic bonds will readily dissolve in water. However, the solubility of chemical compounds in water is highly variable. The **solubility** of a chemical compound in water is defined as the maximum amount of the chemical that will dissolve in pure water at a specified temperature.

Seawater is a solution.

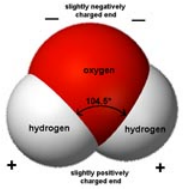


Figure 7.3. Water is a **polar molecule**.

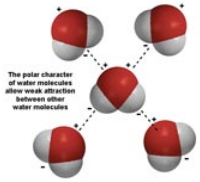


Figure 7.4. Water's polarity allows weak attraction between water molecules.



Figure 7.5. Water droplets illustrate water's high cohesive and adhesive properties.

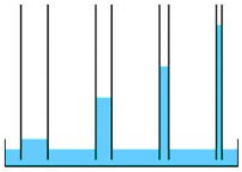


Figure 7.6. Capillary action works because water adheres to material. The smaller the tube the higher the water will rise.

Why don't oil and water mix?

Organic compound containing only carbon and hydrogen (hydrocarbon) are nonpolar and will dissolve in nonpolar solvents (like oil). However, many organic compounds have “functional groups with very electronegative elements” (i.e. oxygen), making the whole molecule polar, allowing them to dissolve in water (ex: sugar and starch can dissolve in water). **Soap** compounds (called **surfactants**) have molecules that are both; they have portions that behave as polar and non-polar ends. One end will stick to hydrocarbons and other non-polar substances whereas the other will stick to water and other polar molecules. This allows polar organic compounds to disperse in water.



Figure 7.7. **Oil** is a **non-polar** substance. **Water** is **polar**.



Figure 7.8. Ship cleaning up after an oil spill.

Properties of Water

pH (acidity and alkalinity)

pH is a measure of the acidity or alkalinity of a solution expressed on a logarithmic scale on which 7 is neutral, lower values are more acid, and higher values more alkaline. pH is an important measurement in seawater. Neutral water is a pH of 7.

Bicarbonate buffering of seawater

A natural buffering system with seawater’s interaction with **carbon dioxide**. **Seawater is generally always within a range of pH of 7.5 to 8.5.** The interactions of dissolved components keep ocean water in a stable range. Organisms living in or near seawater have a limited tolerance for variations in pH and other factors. For instance, **calcite** (as in shell material) is stable within this range,

but will dissolve if exposed to acidic conditions. **Carbonate buffering** keeps pH stable by precipitation (increase pH) or dissolution (decrease pH) of **calcium carbonate - $CaCO_3$**.



Figure 7.9. Salts can **dissolve** in water and then **precipitate** out again as water **evaporates** and concentrates salts in solution.

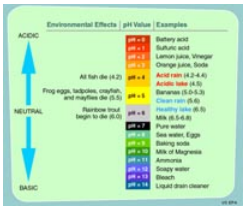


Figure 7.10. **pH** of common household substances compared with pure water (pH=7) and normal seawater (pH=8.2).

This page titled 9.2.1: Properties of Seawater is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.1: Properties of Seawater** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

9.2.2: Specific Heat and Latent Heat Capacity of Water

Specific Heat and Latent Heat Capacity of Water

Specific Heat

Materials vary in their capacity to store thermal energy. For example, a material like copper will heat up much faster than water or wood. **Specific Heat** is a measure of the energy required to heat **1 gram** of substance **1° C**. Specific heat is recorded in "**calories"** for "**mass in grams**" (and "Joules for kg").

Figure 7.11 compares the specific heat of various metals to the specific heat of ice, water, and steam. It takes significantly more energy to warm water than other materials, including both ice and steam. Because of water's high specific heat capacity, the oceans are capable of storing vast quantities of energy from solar heating. Heat absorbed in equatorial regions can be carried long distances and carried by ocean currents before being released in polar regions.

Material	Specific Heat, c C/g*C°	Specific Heat, c, J/kgC°
Silver	0.06	251
Copper	0.09	377
Iron	0.11	461
Steel	0.12	502
Brass	0.22	921
Ice	0.50	2093
Steam	0.50	2093
Water	1.00	4187

Figure 7.11. Comparison of the specific heat of various substances with ice, steam, and liquid water.

High Latent Heat Capacity of Water

When any material is heated to the temperature where it **changes state** (converting from solid to liquid, or liquid to gas), the temperature will remain the same until all the material changes state. Because it takes more energy to convert a substance from one physical state to another (solid to liquid, or liquid to gas), those transitions require a larger amount of energy. **Latent heat** is the heat required (measured in calories burned) to convert a solid into a liquid or vapor, or a liquid into a vapor, **without a change of temperature**. For instance a pot filled with water on the stove will gradually warm up until the water temperature approaches **212° F (or 100° C)**—it will stay at that temperature until all the water has boiled away. The same is true as water freezes. As water cools it will reach **32°F (or 0° C)** is will stay at that temperature until all the water freezes (Figure 7.12).

To convert **1 gram** of **ice** at **0° C** to **1 gram** of **water** at **0° C** requires **80 calories.**
To convert **1 gram** of **water** at **100° C** to **1 gram** of **steam** at **100° C** requires **540 calories**.

When any material is heated to the temperature where it changes state, the temperature will remain the same until all the material changes state. That means ice water will remain at **0° C** (**32° F**) until **all** the ice is melted. The same thing applies when cooling the materials. The reason is that energy must be expended to change the state from solid to liquid or from liquid to gas. Likewise, energy must be withdrawn to change the state when cooling the material. The amount of energy required is call the latent heat of freezing or boiling.

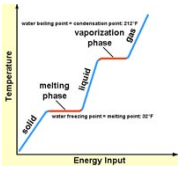


Figure 7.12. Diagram showing heat required to heat water from solid ice to liquid water and to vapor (steam). It take both specific heat and latent heat, released in stages, to convert ice to steam!

Specific Heat and Latent Heat Capacity of Ice, Water, and Steam Illustrated: (using charts listed above)

Example: How much energy would it require to heat **100 grams** of **ice** at **-10° C** to **steam** at **120° C**?

Calculation process:	**Energy required:**
To raise 100 grams of ice at -10° C to 0° C requires: (0.5 cal/g [specific heat of ice] x 10° C x 100 grams)	**500 calories**
To convert 100 grams of ice to water at -0° C requires: (80 cal/g [latent heat conversion] x 100 grams)	**8,000 calories**
To raise 100 grams of water from 0° C to 100° C requires: (1.0 cal/g [specific heat of water] x 100° C x 100 grams)	**10,000 calories**
To convert 100 grams of water to steam at 100° C requires: (540 cal/g [latent heat conversion] x 100 grams)	**54,000 calories**
To raise 100 grams of steam at 100° C to 120° C requires: (0.5 cal/g [specific heat of steam] x 20° C x 100 grams)	**1,000 calories**
Total energy required:	**73, 500 calories**

This page titled 9.2.2: Specific Heat and Latent Heat Capacity of Water is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.2: Specific Heat and Latent Heat Capacity of Water** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

9.2.3: Energy Associated With Evaporation and Condensation of Water In the Air

Energy Associated With Evaporation and Condensation of Water In the Air

• **Evaporation/vaporization** takes a large amount of energy (to break hydrogen bonds). For example, water absorbs energy as it evaporates on your skin. Evaporation on the surface of a swimming pool will cool the water.
• **Condensation** releases a large amount of energy. For example, steam will burn you as it releases energy on your skin.

When water vapor in the air condenses to form water droplets in clouds it releases large amounts of energy.

As water evaporates into in air it cools the air and increases its **humidity**. Air at surface conditions can hold up to about 4% water before it becomes **saturated** and can not absorb more water. As humid air rises it expands and cools and if it reaches the saturation point clouds form as water is forced to condense. The condensation of water releases a lot of energy, heating the air and causing clouds to rise into larger thunderstorms clouds. The water released falls as precipitation.

Heat is absorbed as ice melts and it is released as it freezes. The latent heat of water is an important factor in weather systems and the stability of climates around the world.



Figure 7.13. Cloud formation releases heat into the air.



Figure 7.14. Melting ice absorbs heat from air and water

This page titled 9.2.3: Energy Associated With Evaporation and Condensation of Water In the Air is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.3: Energy Associated With Evaporation and Condensation of Water In the Air** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

9.2.4: Salinity

Salinity

Salinity is a measure of the total amount of solid material (salts) dissolved in water, defined as:

$$\frac{\textbf{Weight (mass) of salt}}{\textbf{Weight (mass) of water}} = \textbf{Salinity}$$

Units are described as:
• **%** is part per hundred (pph)
• **‰** is parts per thousand (ppt)

• Open ocean seawater ranges is 33 to 37 **‰** ppt. Average is about **35 ‰ (ppt)**.



Figure 7.15. Evaporation ponds constructed near the Dead Sea in Jordan are developed to manufacture salt.

Salinity is Measured by:
• electrical conductance (higher salinity = higher conductivity)
• density (higher salinity = higher density)

At room temperature, water can dissolve about 30% of its weight in salt (NaCl). Hot water can hold about 40%. A brine is water that is concentrated with high levels of salt. When evaporation concentrates salty water to it saturation point, salt crystals will precipitate.

Note that difference substances have different saturation levels in water.

Examples:
• Spring 0.3 **‰** ppt (**fresh**)
• Tap water 0.7 **‰** ppt (**fresh**)
• Limit on agriculture irrigation - 2 **‰** ppt
• Baltic Sea 10 **‰** ppt (**brackish**)
• Black Sea - 18 **‰** ppt (**brackish**)
• Average Ocean -35 **‰** ppt
• Mediterranean Sea - 38 **‰** ppt
• Red Sea 42 **‰** ppt (**hypersaline**)
• Great Salt Lake 280 **‰** ppt (**hypersaline**)
• Dead Sea 330 **‰** ppt (**hypersaline**)

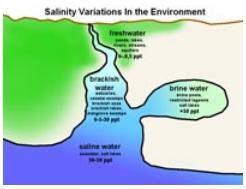


Figure 7.17. Variations in salinity in the natural environment.

Evaporation of Seawater Results In Precipitation Of Mineral Salts

As seawater evaporates in a restricted basin seawater is concentrated becoming a brine. As evaporation proceeds various mineral salts will precipitate out in the reverse order of their solubility. Salty sedimentary deposits produced by evaporation are called **evaporites**. The first to precipitate is **calcite** (if not consumed by organisms first). Next come **$CaSO_4$** (gypsum and anhydrite varieties). This is followed by salt (**NaCl**) (mineral name: **halite**; rock name: **rock salt**). By volume, NaCl is the most abundant salt from seawater. The last to precipitate are potassium salts (**sylvite**: KCl and others) and magnesium salts (**epson salt**: MgSO4 and others). Various other trace salt compounds are concentrated in the last of the brine to evaporate. About 80 different salt minerals have been reportedly found in evaporite deposits.

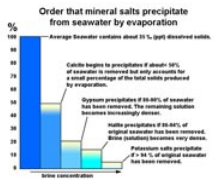


Figure 7.17. Order that salts precipitate from seawater as evaporation proceeds.



Figure 7.18. Salts forming from evaporation in Death Valley, California.

This page titled 9.2.4: Salinity is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.4: Salinity** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

9.2.5: Formation of Sea Ice

Formation of Sea Ice

Fresh water freezes at **32° Fahrenheit (0° Celsius)**. **Seawater** freezes at about **28.4° F (-2 ° C)** because of the salt content. However, when seawater freezes, the ice that forms contains very little salt. Only the water part freezes, the remaining salt is concentrated as **brine** that separates from the sea ice. This process is very important for deep-sea circulation.

This page titled 9.2.5: Formation of Sea Ice is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.5: Formation of Sea Ice** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

9.2.6: Relationship of Salinity, Density and Temperature

Relationship of Salinity, Density and Temperature

Assuming a closed system...

- **Temperature and Density**: **Inverse** (as temperature increases, density decreases)
- **Salinity and Density: Proportional** (as salinity increases, density increases)
- **Temperature and Salinity: None** (as temperature changes, salinity remains the same)

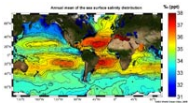


Figure 7.19. Sea Surface Salinity Average 2005

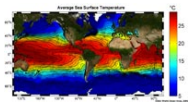


Figure 7.20. Sea Surface Temperature Average 2005

Animations of Satellite and Surface Oceans Data

These websites provide animations that provide stunning views of the Earth created from satellite sensor data. These data are the combined efforts of ocean and atmosphere scientists from many organizations.

NASA animations: **Average Sea Surface Temperature, Salinity and Density**
(This website has links to animations on a globe and a mercator map.) Animations include:
* Average Sea Surface Temperature (SST)
* Average Sea Surface Salinity (SSS)
* Average Sea Surface Density (SSD)

NASA Animations: (mercator) 1 year: December 2011 through December 2012 showing:
Global Sea Surface Salinity
Salinity and ocean circulation
Salinity and global seawater migration

NOAA animations: **Sea Surface Temperature** (World mercator map, last 6 months)
Monthly Isopycnal & Mixed-layer Ocean Climatology (MIMOC): Animations show pycnocline (temperature and salinity) data of different parts of the world.

The NASA animations provide views of how salinity, temperature, and density change over the course of a year.



Figure 7.21. Sea Surface Salinity Average 2009

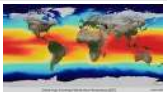


Figure 7.22. Sea Surface Temperature Average 2009



Figure 7.23. Sea Surface Density Average 2009

This page titled 9.2.6: Relationship of Salinity, Density and Temperature is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.6: Relationship of Salinity, Density and Temperature** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

9.2.7: Salinity and Latitude

Salinity and Latitude

Figure 7.24 is a map of the globe comparing the rates of evaporation and precipitation. The map is a compilation of **evaporation minus precipitation (E-P)** values. The data basically shows the regions where there is a net gain of salinity created in surface waters by high evaporation rates. There is also a net loss of salinity where precipitation is higher than evaporation rates (Figure 7.25). In general:

• The **tropics** (equatorial region) is humid and cloudy, and receives much more rain than evaporates.
• The **temperate regions** receive less precipitation, so evaporation dominates.
• The **polar regions** have low evaporation rates relative to the amount of precipitation they receive.

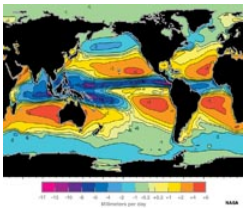


Figure 7.24. Map of net **evaporation minus precipitation (E-P)** on oceans.

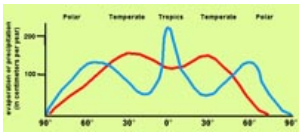


Figure 7.25. Evaporation and precipitation curves compared with latitude.

Variability of ocean salinity: Ocean salinity is stable at depth but can be highly variable at the surface. The upper surface layers of the ocean impacted by wave energy is a mixing zone. Simply this: the more waves, the more mixing. Freshwater is less dense than seawater and without mixing freshwater will float (stratify) on top of seawater.

Factors that decrease salinity:	**Factors that increase salinity**:
• **Precipitation** • **Runoff** • **Melting icebergs/sea ice**	• **Evaporation** • **Freezing sea ice**

This page titled 9.2.7: Salinity and Latitude is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.7: Salinity and Latitude** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

9.2.8: "Cline Curves" and Mixing (Surface) Zone

The "Cline Curves" - changes in temperature, salinity, and density with depth

- **Thermocline** - a steep **temperature** gradient in a body of water marked by a layer above and below which the water is at different temperatures.
- **Halocline** - a vertical zone in the oceanic water column in which **salinity** changes rapidly with depth.
- **Pycnocline** - a layer in an ocean or other body of water in which water **density** increases rapidly with depth.

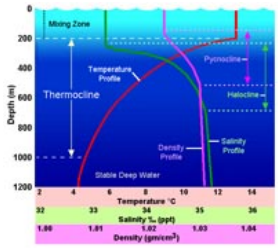


Figure 7.26. The "cline curves"

NOAA Animation: **Annual changes of pychnocline depth**

Mixing (Surface) Zone

Uppermost water where mixing from currents make temperature, salinity, and density mostly constant.

Vertical changes in temperature with latitude:
• **Polar regions** have almost **no thermocline**.
• **Temperate regions** have **weak thermoclines** (moderate in summer, less in winter).
• **Tropical regions** have a **strong thermocline.**

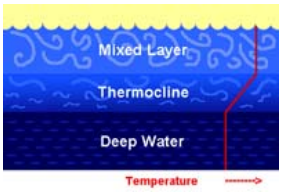


Figure 7.27. Mixing Zone

This page titled 9.2.8: "Cline Curves" and Mixing (Surface) Zone is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.8: "Cline Curves" and Mixing (Surface) Zone** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

9.2.9: Gases Dissolved In Seawater

Gases Dissolved In Seawater

Oxygen concentrations (O_2) in air is about 21% oxygen, in water it is a tiny fraction of 1%. This large difference in oxygen concentrations forces oxygen to dissolve into water along the boundary between air and water. When wind blows creating waves, it increases the surface area, allowing more diffusion to occur.

Carbon dioxide (CO_2) is much more soluble in water than oxygen, but concentrations in the atmosphere are comparatively very low. When dissolved in water it becomes a **bicarbonate ion ($^-HCO_3$)**, so carbon dioxide readily diffuses into the atmosphere if it is not consumed in the production of **calcium carbonate ($CaCO_3$)**. Biological respiration releases $^-HCO_3$ which combines with available dissolved calcium which the organism either excretes or incorporates into its skeletal structure if environmental conditions are warm enough for $CaCO_3$ to persist. One of the gravest concerns about the burning of fossil fuels is that it is increasing the concentration of carbon dioxide in seawater, so organisms that produce carbonate shells and skeletons are negatively impacted.

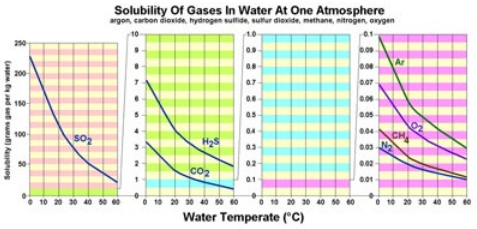


Figure 7.28. Solubility of gases in water is affected by temperature. Gases are much more soluble in cold water than warm water. In contrast, most solid materials (organic and inorganic; examples salt and sugar) are more soluble in hot water. The solubility of elements in seawater is complex and depends on many factors including pH (acid-base), eH (oxidation-reduction), temperature, pressure, and interactions between other compounds dissolved in seawater.

Methane (CH_4) has very low solubility in seawater, however, it is very abundant in sediments rich in organic mater. In cold settings, methane, carbon dioxide, and water form an unusual form of **ice** called a **methane hydrate (Figures 7-29** and **7-30)**.

A **clathrate** is a compound in which molecules of one component are physically trapped within the crystal structure of another, in this case CO_2 and CH_4 are trapped in the crystal structure of ice under certain pressure and temperature condition that exist on the seafloor in cold water, mostly on the outer continental shelves and slopes in polar regions. Global warming of the oceans can cause the release of tremendous amounts of CO_2 and CH_4 from the seafloor, contributing to anoxia conditions, with possible catastrophic consequences. Read about the "**Clathrate Gun Hypothesis**" (Wikipedia).

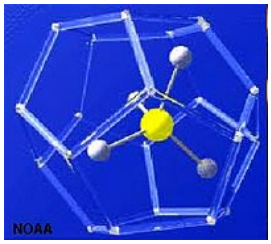


Figure 7.29. Methane-ice clathrate structure



Figure 7.30. Methane-ice clathrates will burn!

Sulfur dioxide (SO_2) is extremely soluble in water. Sulfur dioxide is a gas that **smells like rotten eggs**. It is released in large quantities by volcanic eruptions, forest fires, and by burning coal and petroleum. SO_2 is extremely soluble in water where it combines with water molecules to form **sulfate ions** ($-HSO_4$). When concentrated (when water is removed), the solution becomes **sulfuric acid** (H_2SO_4). When concentrated by **evaporation of seawater** (where **dissolved calcium** and other metallic ions are present) sulfate ions precipitates as the salts **gypsum** ($CaSO_4 \cdot 2H_2O$), **anhydrite** ($CaSO_4$), **epson salt** ($MgSO_4 \cdot 7H_2O$), and other salt minerals.



Figure 7.31. Gypsum

This page titled 9.2.9: Gases Dissolved In Seawater is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.9: Gases Dissolved In Seawater** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

9.2.10: Local Conditions in the San Diego Region

Local Conditions in San Diego region

• In San Diego, we have a **temperate thermocline**, which is weak in the winter and fairly strong in the summer.
• During El Niño years we may get a strong thermocline all year.
• California's weather patterns are cyclical on multi-year periods of rain and drought.

Figure 7.32 shows California's annual precipitation 1895 to 2014. As of fall, 2015, California has currently experienced the worst drought in the period records have been recorded.

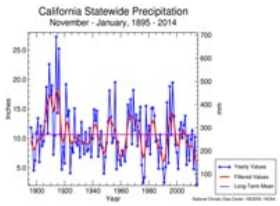


Figure 7.32. California's precipitation history.

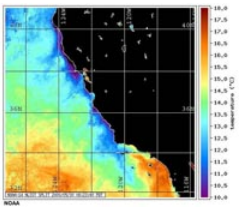


Figure 7.33. SoCal ocean temperatures (5/30/2000).

This page titled 9.2.10: Local Conditions in the San Diego Region is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.10: Local Conditions in the San Diego Region** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

9.2.11: Quiz Questions Properties of Seawater

1. It is currently estimated that 97.2% of water resources on Earth is in the world ocean. **Most** of the remainder is:
a. frozen in glaciers and ice caps.
b. trapped as groundwater and soil moisture.
c. stored in lakes, streams and reservoirs.
d. trapped as water vapor in the atmosphere.

2. Seawater is composed of water, dissolved matter (including solids and gas [as ions], and suspended matter (dust and organic residues). Of the ions in seawater (besides hydrogen and oxygen), the **two** most abundant elements are:
a. sodium and calcium
b. chlorine and sodium
c. calcium and sulphate (SO_4)
d. magnesium and potassium

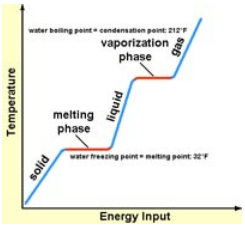
3. Water is a polar substance. Each molecule of water has a negative charge associated with its oxygen, and a positive charge with it hydrogen. This polar character is responsible for its properties to have high surface tension because water molecules stick together. This property is called:
a. cohesion.
b. adhesion.
c. capillary action.
d. a solution.

4. Water is a powerful solvent: A solvent is a substance that dissolves a solute (a chemically different liquid, solid or gas), resulting in a solution. Not everything dissolves equally in water. However, substances that are most easily dissolved are held together by:
a. covalent bonds.
b. metallic bonds.
c. ionic bonds.
d. Van der Waals forces.

5. **pH** is a measure of acidity and alkalinity of a water solution. Because of the natural buffering system of dissolved bicarbonate (+HCO3) dissolve in seawater:
a. seawater is slightly acidic.
b. seawater has a neutral water pH of 7.0.
c. seawater is generally always within a range of pH of 7.5 to 8.5.
d. seawater is highly salty with a pH of about 9.0 or higher.

6. How much energy would it require to heat 20 grams of ice at -20° C to steam at 120° C? **Do the math!**

(Key info)
Specific heat:
ice = 0.5 cal/gm
water = 1.0 cal/gm
steam = 0.5 cal/gm
Latent heat:
ice to water: 80 cal/gm
water to steam: 540 cal/gm



a. 3,400 calories
b. 7,350 calories
c. 12,980 calories
d. 29,600 calories

7. Lots of energy is released when:
a. water is converted to steam.
b. water evaporates on your skin.
c. ice melts to form water.
d. water vapor condenses in the air to form rain droplets.

8. As water evaporates into in air it cools the air and increases its **humidity**. As humid air rises it expands and cools and if it reaches the saturation point clouds form as water is forced to condense. The condensation of water releases a lot of energy, heating the air and causing clouds to rise into larger thunderstorms clouds. The water released falls as precipitation. How much water can the air hold to at at surface conditions before before it becomes **saturated** and can not absorb more water?
a. 2%
b. 4%
c. 8%
d. 12%

9. The salinity of seawater varies slightly from one part of the ocean to another, however, the average salinity of the oceans is about:
a. 4% (pph).
b. 17‰ (ppt).
c. 35‰ (ppt).
d. 40‰ (ppt).

10. Which of the relationships of temperature, salinity, and density of seawater is NOT true?
a. As temperature increases, density decreases.
b. As salinity increases, density increases.
c. As temperature changes, salinity remains the same.
d. As density increases, temperature increases.

11. Global data measurements of precipitation and evaporation show patterns of all EXCEPT which of the following.
a. The temperate regions receive more precipitation than tropical regions.
b. The polar regions have receive more precipitation than evaporates.
c. The temperate regions receive less precipitation than evaporates.
d. The tropics (equatorial region) receives much more rain than evaporates.

12. A thermocline is a layer of water at the ocean surface that prevents the upwelling and mixing of cool nutrient-rich water to ocean surface waters, reducing the production of primary plankton (food and nutrients for marine life). What is the best description of thermoclines based on latitude?
a. Polar regions have well-developed thermoclines in winter months.
b. Temperate regions have strong thermoclines in the winter months.
c. Tropical region have no thermocline in winter months.
d. Temperate regions have weak thermoclines (moderate in summer, less in winter).

13. Comparing oxygen carbon dioxide concentrations in the air compared with seawater, what of the following choices are true?
a. Seawater has higher concentrations of oxygen than the air.
b. Carbon dioxide (CO_2) is much more soluble in water than oxygen, but concentrations of CO_2 in the atmosphere are comparatively very low.
c. Wind and wave action reduces the exchange of oxygen and carbon dioxide (CO_2) with seawater.
d. All of the above are true.

14. Sulfur dioxide (SO_2) smells like rotten eggs. SO_2 is released in large quantities by volcanic eruptions and by burning coal and petroleum. SO_2 is extremely soluble in water where it combines with water molecules to form sulfate ions ($-HSO_4$). When concentrated by evaporation in seawater, what happens?
a. It becomes sulfuric acid ($H2SO_4$).
b. It precipitates as gypsum ($CaSO_4$-2H2O) and anhydrite ($CaSO_4$).

c. It is released back into the air as concentrated SO_2.
d. All of the above.

15. Methane (CH_4) has very low solubility in seawater, however, it is very abundant in sediments rich in organic mater. In cold settings, methane, carbon dioxide, and water form an unusual form of "flammable" ice called:
a. a clathrate.
b. sea ice.
c. permafrost.
d. all of the above.

This page titled 9.2.11: Quiz Questions Properties of Seawater is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Miracosta Oceanography 101 (Miracosta)) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.11: Quiz Questions - Chapter 7 - Properties of Seawater** by Miracosta Oceanography 101 has no license indicated. Original source: https://gotbooks.miracosta.edu/oceans/index.html.

SECTION OVERVIEW

9.3: Biogeochemistry of the surface ocean

9.3.1: Introduction

9.3.2: Biological Pump

9.3.3: Export from the Upper Ocean to the Depths and Sediments

9.3.4: Summary

9.3: Biogeochemistry of the surface ocean is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

9.3.1: Introduction

The oceanic biological carbon pump (BCP) regulates the Earth carbon cycle by transporting part of the photosynthetically fixed CO_2 into the deep ocean. Suppressing this mechanism would result in an important increase of atmospheric CO_2 level. The BCP occurs mainly in the form of (1) organic carbon (OC) particles sinking out the surface ocean, of (2) neutrally buoyant OC (dissolved or particulate) entrained by downward water masses movements and/or mixing, and of (3) active transport of OC by migrating animals such as zooplankton and fishes. These various pools of OC differ in size since their sinking, production and decomposition rates vary spatially and temporally. Moreover, the OC transported to depths via these various export pathways as well as their decomposition pathways all have different ecological origins and therefore may response differently to climate changes. Currently, most ocean biogeochemical models do not resolve these various of OC pathways explicitly; rather, they imply that OC is therein created and destroyed equally. In addition, the organic composition of these various pools is largely unknown, especially at depths below 500 m.

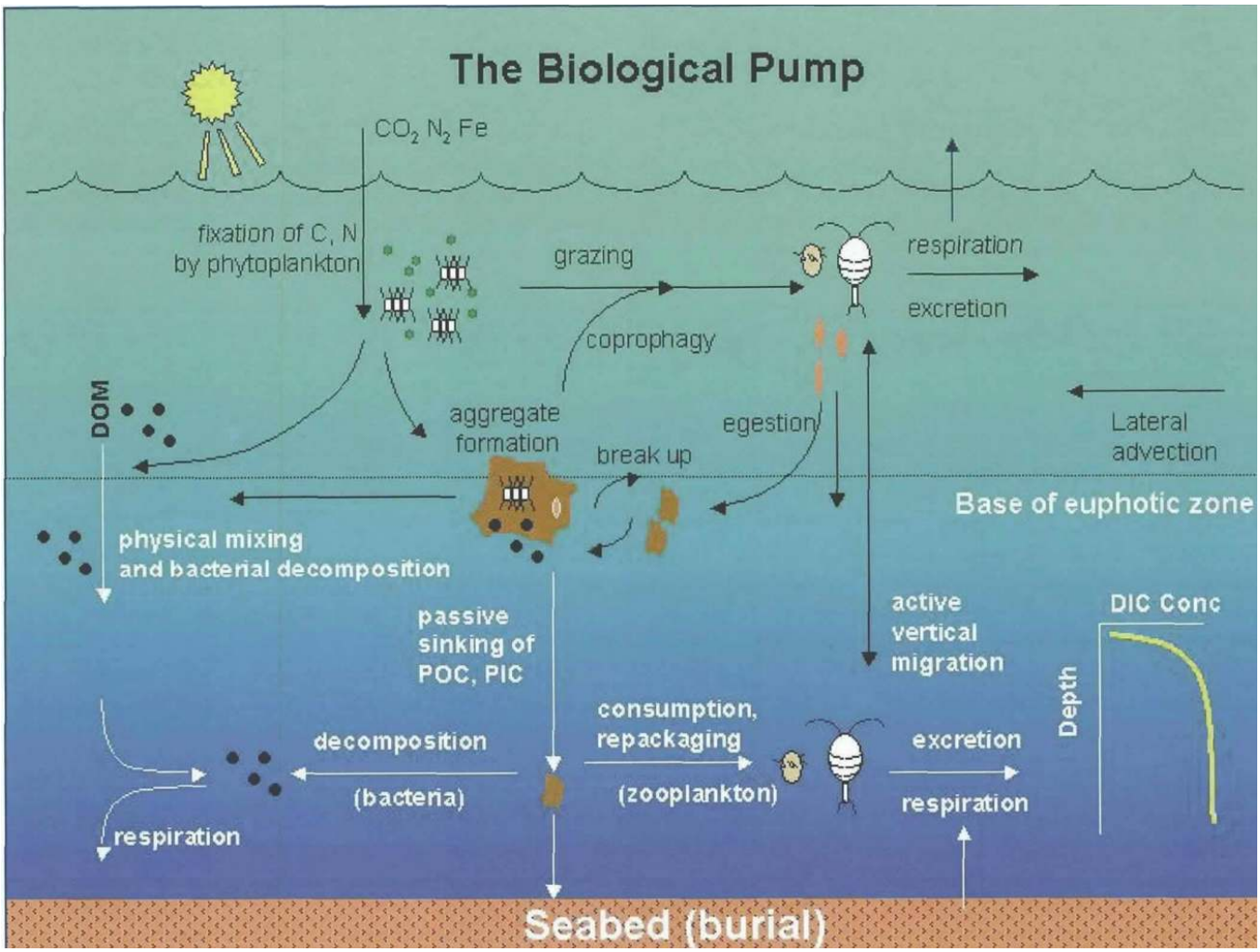


Figure 9.3.1.1: Components of the "biological pump" are responsible for transforming Dissolved Inorganic Carbon (DIC) into organic biomass and pumping it in particulate or dissolved form into the deep ocean. Inorganic nutrients and Carbon Dioxide (CO_2) are fixed during photosynthesis by phytoplankton, which both release Dissolved Organic Matter (DOM) and are consumed by herbivorous zooplankton. Larger zooplankton-such as copepods, egest fecal pellets - which can be reingested, and sink or collect with other organic detritus into larger, more-rapidly-sinking aggregates. DOM is partially consumed by bacteria (black dots) and respired; the remaining refractory DOM is advected and mixed into the deep sea. DOM and aggregates exported into the deep water are consumed and respired, thus returning organic carbon into the enormous deep ocean reservoir of DIC. About 1% of the particles leaving the surface ocean reach the seabed and are consumed, respired, or buried in the sediments. There, carbon is stored for millions of years. The net effect of these processes is to remove carbon in organic form from the surface and return it to DIC at greater depths, maintaining the surface-to-deep ocean gradient of DIC (inset graph at lower right). Thermohaline circulation returns deep-ocean DIC to the atmosphere on millennial timescales.. (Copyright; Ducklow, Steinberg and Busseler, Upper Ocean Carbon Export and the Biological Pump, Oceanography, https://tos.org/oceanography/article...iological-pump)

Excerpted from:

Le Moigne, F. A. (2019). Pathways of organic carbon downward transport by the oceanic biological carbon pump. *Frontiers in Marine Science*, *6*, 634.

9.3.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Frederic A. C. LeMoigne, Frontiers in Marine Science.

9.3.2: Biological Pump

The oceanic biological carbon pump (BCP) is a complex mechanism regulating the Earth carbon (C) cycle by sequestrating part of the photosynthetically fixed CO_2 into the deep ocean and the seafloor. Suppressing this mechanism would increase the current levels of atmospheric CO_2 level by about 50% (Sarmiento et al., 2004; Parekh et al., 2006). Currently, the variability in the proportion of surface primary production (PP) leaving the euphotic zone and the amount of organic carbon (OC) exported to a certain depth via the BCP are challenging to explain. This prevents robust predictions of how C cycle, and hence climate, will change in the future.

Three major forms of the BCP are distinguished: (i) gravitational, (ii) mixing and (iii) migration. The gravitational pump (i) involves downward sinking of the OC contained in various particles (POC) of planktonic origin from the surface ocean to depth due to their large size and/or high density (Siegel et al., 2016). This sinking POC is termed “marine snow” and can be made of single phytoplankton cells (DiTullio et al., 2000; Rembauville et al., 2016; Leblanc et al., 2018), various types of aggregates resulting from coagulation processes (Alldredge and Jackson, 1995; Riley et al., 2012; Laurenceau-Cornec et al., 2015) and zooplankton fecal pellets (FP) (Turner, 2002; Cavan et al., 2015). The mixing pump (ii) occurs when neutrally buoyant OC is transported to depth by downward-moving water masses and/or mixing. This OC is either dissolved (DOC) or entrained in particles which are too small in size or positively buoyant relative to seawater to sink on their own (Hansell et al., 2009; Le Moigne et al., 2016). The physical mechanisms that transport the OC downward can be either diapycnal with DOC diffusing out of the surface waters, or isopycnal, when water mass movements eddies and/or mixed layer bring DOC and/or suspended particles to depth (Hansell et al., 2009; Omand et al., 2015; Dall’Olmo et al., 2016). (iii) Diurnal vertical migration (DVM) of zooplankton and fish is a common feature in oceanic ecosystems (Longhurst, 1991; Steinberg, 2000; Landry and Calbet, 2004). In order to avoid predation, swimmers feed at night in surface waters and defecate deeper in the mesopelagic during the day. Doing so, they actively transport OC into the ocean’s interior.

The presence, absence, and dominance of each of these different export pathways are controlled by biological (Piontek et al., 2014; Agusti et al., 2015; Le Moigne et al., 2016), chemical (Edwards et al., 2015), ecological (Cavan et al., 2015; Guidi et al., 2016), and physical (Levy et al., 2013; Taucher et al., 2014; Dall’Olmo et al., 2016) processes. They all vary in time and space and at depth. Most of these processes, however, are still poorly understood which hinders mechanistic predictions of the magnitude of the downward OC fluxes. At the base of the euphotic zone, the sinking of particles (termed gravitational flux) dominate the amount of OC (Giering et al., 2014). However, in the end of winter, the OC flux is induced by the so-called mixed layer pump which is driven by the seasonal entrainment of surface waters in deeper layers (Gardner et al., 1995) and can supply a substantial amount of OC to the deep ocean. Also, while globally the proportion of OC transported by migrating zooplankton and larger organisms relative to total OC flux is <20%, this active flux can occasionally dominate the total OC flux [(Aumont et al., 2018) and references therein].

References

Agusti, S., González-Gordillo, J. I., Vaqué, D., Estrada, M., Cerezo, M. I., Salazar, J., et al. (2015). Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. *Nat. Commun.* 6:7608. doi: 10.1038/ncomms8608

Alldredge, A. L., and Jackson, G. A. (1995). Aggregation in marine systems - preface. *Deep. Res. Part II Top. Stud. Oceanogr.* 42, 1–7.

Aumont, O., Maury, O., Lefort, S., and Bopp, L. (2018). Evaluating the potential impacts of the diurnal vertical migration by marine organisms on marine biogeochemistry. *Global Biogeochem. Cycles* 32, 1622–1643. doi: 10.1029/2018GB005886

Cavan, E., Le Moigne, F. A. C., Poulton, A. J., Daniels, C. J., Fragoso, G., and Sanders, R. J. (2015). Zooplankton fecal pellets control the attenuation of particulate organic carbon flux in the Scotia Sea, Southern Ocean. *Geophys. Res. Lett.* 42, 821–830. doi: 10.1002/2014GL062744

Dall’Olmo, G., Dingle, J., Polimene, L., Brewin, R. J. W., and Claustre, H. (2016). Substantial energy input to the mesopelagic ecosystem from the seasonal mixed-layer pump. *Nat. Geosci.* 9, 820–823. doi: 10.1038/NGEO2818

DiTullio, G. R., Grebmeier, J. M., Arrigo, K. R., Lizotte, M. P., Robinson, D. H., Leventer, A., et al. (2000). Rapid and early export of Phaeocystis antarctica blooms in the Ross Sea, Antarctica. *Nature* 404, 595–598. doi: 10.1038/35007061

Edwards, B. R., Bidle, K. D., and Van Mooy, B. A. S. (2015). Dose dependent regulation of microbial activity on sinking particles by polyunsaturated aldehydes: implications for the carbon cycle. *PNAS* 112, 5909–5914. doi: 10.1073/pnas.1422664112

Gardner, W. D., Chung, S. P., Richardson, M. J., and Walsh, I. D. (1995). The oceanic mixed-layer pump. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 42, 757–765.

Giering, S. L. C., Sanders, R. J., Lampitt, R. S., Anderson, T. R., Tamburini, C., Boutrif, M., et al. (2014). Reconciliation of the carbon budget in the ocean's twilight zone. *Nature* 507, 480–483. doi: 10.1038/nature13123

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532, 465–470. doi: 10.1038/nature16942

Hansell, D. A., Carlson, C. A., Repeta, D. J., and Schlitzer, R. (2009). Dissolved organic matter in the ocean: a controversy stimulates New Insights. *Oceanography* 22, 202–211. doi: 10.5670/oceanog.2009.109

Landry, M. R., and Calbet, A. (2004). Microzooplankton production in the oceans. *ICES J. Mar. Sci.* 61, 501–507. doi: 10.1016/j.icesjms.2004.03.011

Laurenceau-Cornec, E. C., Trull, T. W., Davies, D. M., Bray, S. G., Doran, J., Planchon, F., et al. (2015). The relative importance of phytoplankton aggregates and zooplankton fecal pellets to carbon export: insights from free-drifting sediment trap deployments in naturally iron-fertilised waters near the Kerguelen Plateau. *Biogeosciences* 12, 1007–1027. doi: 10.5194/bg-12-1007-2015

Leblanc, K., Quéguiner, B., Diaz, F., Cornet, V., Michel-Rodriguez, M., Durrieu De Madron, X., et al. (2018). Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. *Nat. Commun.* 9: 953. doi: 10.1038/s41467-018-03376-9

Le Moigne, F. A. C., Henson, S. A., Cavan, E., Georges, C., Pabortsava, K., Achterberg, E. P., et al. (2016). What causes the inverse relationships between primary production and export efficiency in the southern ocean? *Geophys. Res. Lett.* 43, 4457–4466. doi: 10.1002/2016GL068480

Levy, M., Bopp, L., Karleskind, P., Resplandy, L., Ethe, C., and Pinsard, F. (2013). Physical pathways for carbon transfers between the surface mixed layer and the ocean interior. *Global Biogeochem. Cycles.* 27, 1001–1012. doi: 10.1002/gbc.20092

Longhurst, A. R. (1991). Large marine ecosystems - patterns, processes and yields. *Mar. Policy* 15, 377–378.

Omand, M. M., D'asaro, E., Lee, C., Perry, M. J., Briggs, N., Cetini, I., et al. (2015). Eddy-driven subduction exports particulate organic carbon from the spring bloom. *Science* 348, 222–225. doi: 10.1126/science.1260062

Parekh, P., Dutkiewicz, S., Follows, M. J., and Ito, T. (2006). Atmospheric carbon dioxide in a less dusty world. *Geophys. Res. Lett.* 33:L03610. doi: 10.1029/2005GL025098

Piontek, J., Sperling, M., Nothig, E. M., and Engel, A. (2014). Regulation of bacterioplankton activity in Fram Strait (Arctic Ocean) during early summer: the role of organic matter supply and temperature. *J. Mar. Syst.* 132, 83–94. doi: 10.1016/j.jmarsys.2014.01.003

Rembauville, M., Manno, C., Tarling, G. A., Blain, S., and Salter, I. (2016). Strong contribution of diatom resting spores to deep-sea carbon transfer in naturally iron-fertilized waters downstream of South Georgia. *Deep Sea Res. I* 115, 22–35. doi: 10.1016/j.dsr.2016.05.002

Riley, J., Sanders, R., Marsay, C., Le Moigne, F. A. C., Achterberg, E., and Poulton, A. (2012). The relative contribution of fast and slow sinking particles to ocean carbon export. *Global Biogeochem. Cycles* 26:GB1026. doi: 10.1029/2011GB004085

Sarmiento, J. L., Slater, R., Barber, R., Bopp, L., Doney, S. C., Hirst, A. C., et al. (2004). Response of ocean ecosystems to climate warming. *Global Biogeochem. Cycles* 18:GB3003. doi: 10.1029/2003GB002134

Siegel, D. A., Buesseler, K. O., Behrenfeld, M. J., Benitez-Nelson, C. R., Boss, E., Brzezinski, M. A., et al. (2016). Prediction of the export and fate of global ocean net primary production: the EXPORTS science plan. *Front. Mar. Sci.* 3:22. doi: 10.3389/fmars.2016.00022

Steinberg, D. K. (2000). Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 47, 137–158. doi: 10.1016/s0967-0637(99)00052-7

Taucher, J., Bach, L., Riebesell, U., and Oschlies, A. (2014). The viscosity effect on marine particle flux: a climate relevant feedback mechanism. *Global Biogeochem. Cycles* 28, 415–422. doi: 10.1002/2013GB004728

Turner, J. T. (2002). Zooplankton fecal pellets, marine snow and sinking phytoplankton blooms. *Aquat. Microb. Ecol.* 27, 57–102. doi: 10.3354/ame027057

Excerpted from:

Le Moigne, F. A. (2019). Pathways of organic carbon downward transport by the oceanic biological carbon pump. *Frontiers in Marine Science*, *6*, 634.

9.3.2: Biological Pump is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

9.3.3: Export from the Upper Ocean to the Depths and Sediments

The gravitational flux of OC occurs in the form of various particles of planktonic origin sinking from the surface ocean due to their large size and/or high density. The gravitational pump includes particles that are single phytoplankton cells, various types of aggregates, and zooplankton dejections termed fecal pellets.

Fecal Pellets Flux

Specific OC flux associated with zooplankton fecal pellets is relatively easy to measure. Fecal pellets are compact and have distinctive shapes and thus can easily identified and picked out from samples collected using different types of sediment traps (Refaeli et al., 2008; Turner, 2015). However, some pellets can experience degradation in moored sediment traps cups and are not being counted as such (Wilson et al., 2008). This may result in underestimation of the flux of fecal pellets.

Phytoplankton Cells Flux

Turbulent mixing and buoyancy regulation tend to retain single phytoplankton single cells tend in the surface ocean. However, despite of their small Reynolds number [describes the nature of the surrounding flow and its fall velocity (Stokes, 1851)], some intact cells (not grazed or aggregated) are found in sediment traps. This can happen to single live or dead phytoplankton cells, resting spores or cysts, and colonies.

Phytoplankton cells sink into the ocean's interior as single cells which are live, senescent and/or dead. DiTullio et al. (2000) observed single cells of *Phaeocystis antarctica* in the lower mesopelagic zone (500–600 m) in the Ross sea and concluded that they were exported rapidly and alive following a large *Phaeocystis antarctica* bloom. However, the mechanism delivering these cells down to the mesopelagic zone is unclear. Although physical processes could entrain surface cells to depth (see following sections), the cells observed at 500 m were still photosynthesizing, which suggests that their export to depth was likely too rapid to be induced by a physical process. Phytoplankton cells also sink dead. For example (Salter et al., 2007), observed dead cells of large diatoms *Eucampia antarctica* being selectively exported relative to the surface phytoplankton community in the Southern Ocean. These diatoms were exported rapidly owing to their heavily silicified, and thus dense, frustule (Salter et al., 2007). Phytoplankton resting spores and/or cysts are also exported from the surface ocean down to bathypelagic/abyssal depths (>1000 m). They are formed strategically by some phytoplankton in response to nutrient limitation and contain significant amounts of OC. In sediment traps located within an iron-fertilized bloom in the Southern Ocean near the Kerguelen Islands, Rembauville et al. (2016) measured OC fluxes associated with resting spores of diatoms *Chaetoceros Hyalochaete* or *Thalassiosira Antarctica*. The resting spores of a single species can account for up to 90% of the total gravitational OC flux (Rembauville et al., 2016). Acantharian cysts (unicellular organisms related to Radiolaria producing celestite, the densest biomineral in the ocean) were abundant in deep sediment trap material in the Iceland Basin (Martin et al., 2010) and the Southern Ocean (Belcher et al., 2018). High cyst flux was restricted to only 2 weeks but contributed up to half of the total gravitational OC flux during this period (Martin et al., 2010; Decelle et al., 2013). The Acantarian cysts are made of dense celestite which ensures their fast sedimentation to the bathypelagic depths. Phytoplankton cell colonies and chains are also observed in sediment traps samples. Laurenceau-Cornec et al. (2015) found chains of the pennate diatoms *Fragilariopsis* sp. in the Southern Ocean, while Pabortsava et al. (2017) collected intact colonies of nitrogen-fixing *Trichodesmium* species in the deep traps (3000 m) collecting in the North Atlantic gyre.

Aggregate Flux

Evaluating OC flux associated with aggregates is challenging as their structure is often fragile and erratic (Alldredge, 1998). In moored sediment traps, marine aggregates lose their shapes. Unlike fecal pellets (see section above), aggregates in the bulk sediment trap material are clumped and thus difficult to pick out individually.

The Mixing Flux

Several ocean physical processes transport slow-sinking, neutrally buoyant or dissolved OC (DOC) from the surface where they are produced to the mesopelagic zone. Recently, "mixing fluxes" gained more importance (see following references). This is mainly due the development of semi-automated platforms and devices which enabled high frequency sizing of particles (Picheral et al., 2010; Claustre et al., 2011). Large-scale ocean circulation is also one of the processes transporting particles produced in the surface to ocean's interior. For instance, regions of deep water formation are hotspots for such transport (Levy et al., 2013).

The mixed layer pump is another process that redistributes dissolved OC, neutrally buoyant or slow-sinking organic matter down the water column (Dall'Olmo et al., 2016). This process occurs during transient stratifications at the start of the productive season when suspended POC first accumulates in surface, and then redistributed throughout the water column by deep mixing. Globally, the mixed layer pump is responsible for a flux of 0.5 Gt C yr^{-1}. This process supplies POC mainly to locations with deep winter mixing such as the North Atlantic Ocean or the Southern Ocean (Dall'Olmo et al., 2016).

The water mass subduction induced by eddies also pumps POC into the mesopelagic zone. Omand et al. (2015) quantified the amount of suspended POC sinking following eddy driven subduction in the North Atlantic. In essence, suspended POC produced in surface sink along tilted isopycnals. Globally, eddy-driven subduction of POC is responsible for an export flux of ∼1 Gt C yr^{-1} (Omand et al., 2015). Altogether, these processes may export up to 2 Gt of POC yr^{-1}, nearly a third of the gravitational flux (Boyd et al., 2019).

Large-scale circulation, mixing and eddies subduction also transport DOC beneath the surface ocean. Examining large scale variability of DOC concentrations, Hansell et al. (2009) estimated that DOC export represents 20% that of gravitational POC export flux in the global ocean. However, the amount of DOC transported by processes such as eddy driven subduction and mixed layer pump (see references above) have poor spatial coverage both on local and global scales. Using an artificial neural network, Roshan and DeVries (2017) reconstructed global DOC concentrations distribution. Here, DOC production and export fluxes were estimated by coupling reconstructed DOC concentrations to a global ocean circulation model (DeVries and Primeau, 2011). Large annual DOC export rates were shown in the tropics (50–80 mg of OC m^{-2} d^{-1}) and low in high latitudes (0–40 mg of OC m^{-2} d^{-1}), contrary to findings by Dall'Olmo et al. (2016) and Omand et al. (2015). Their model, however, extrapolated DOC concentrations measured during summer to a global scale. This may have limited the effect of the aforementioned physical processes. More work is required to estimate the annual flux of DOC from the surface ocean.

The Active Migration Flux

The concept of active flux by migrating organisms (also known as "swimmers") in the mesopelagic is based on widespread observations of DVM of zooplankton and fish caught in nets (Longhurst, 1991; Steinberg, 2000; Landry and Calbet, 2004). In essence, to avoid predation swimmers feed at night in surface waters and defecate deeper in the mesopelagic during the day (Steinberg and Landry, 2017). Inverted DVM and desynchronized vertical migration do exist (Neill, 1990; Cohen and Forward, 2016), however, typical DVM as described above are more common in marine systems. This actively supplies OC to depth with minimizing decomposition. To date, estimates of DVM have mostly been focused on mesozooplankton because they are readily caught in multiple opening/closing towed nets (Tarling et al., 2002). DVM is a difficult process to model (Hansen and Visser, 2016), however, this field has made significant progresses recently (Aumont et al., 2018; Archibald et al., 2019; Gorgues et al., 2019).

The active transport of OC by migrating mesozooplankton can contribute up to 40% of the gravitational flux (Stukel et al., 2013) with global estimates ranging from 0.2 to 71 mg m^{-2} d^{-1} (Aumont et al., 2018). Most studies of mesozooplankton associated OC flux were conducted in coastal and upwelling regions, while data from high latitudes and polar waters are scarce. The largest flux is recorded in the Benguela upwelling system while the smallest are observed near the Antarctic peninsula. Large DVM fluxes in upwelling system is counterintuitive given that these waters are productive but often poorly oxygenated below surface. However, some migrating organisms can survive with little oxygen concentrations by adopting adaptive metabolic strategies (Bianchi et al., 2013, 2014; Kiko et al., 2015). Excluding the extreme values of 0.2 and 71 mg m^{-2} d^{-1}, the OC fluxes associated with migratory organisms appear to be relatively homogeneous and in the order of 9.8 ± 7.3 mg C m^{-2} d^{-1}.

References

Alldredge, A. (1998). The carbon, nitrogen and mass content of marine snow as a function of aggregate size. *Deep. Res. I* 45, 529–541. doi: 10.1016/s0967-0637(97)00048-4

Archibald, K. M., Siegel, D. A., and Doney, S. C. (2019). Modeling the impact of zooplankton diel vertical migration on the carbon export flux of the biological pump. *Global Biogeochem. Cycles* 33, 181–199. doi: 10.1029/2018gb005983

Aumont, O., Maury, O., Lefort, S., and Bopp, L. (2018). Evaluating the potential impacts of the diurnal vertical migration by marine organisms on marine biogeochemistry. *Global Biogeochem. Cycles* 32, 1622–1643. doi: 10.1029/2018GB005886

Belcher, A., Manno, C., Thorpe, S., and Tarling, G. (2018). Acantharian cysts: high flux occurrence in the bathypelagic zone of the Scotia Sea. *Southern Ocean. Mar. Biol.* 165:117. doi: 10.1007/s00227-018-3376-1

Bianchi, D., Babbin, A. R., and Galbraith, E. D. (2014). Enhancement of anammox by the excretion of diel vertical migrators. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15653–15658. doi: 10.1073/pnas.1410790111

Bianchi, D., Galbraith, E. D., Carozza, D. A., Mislan, K. A. S., and Stock, C. A. (2013). Intensification of open-ocean oxygen depletion by vertically migrating animals. *Nat. Geosci.* 6, 545–548. doi: 10.1038/ngeo1837

Boyd, P. W., Claustre, H., Levy, M., Siegel, D. A., and Weber, T. (2019). Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature* 568, 327–335. doi: 10.1038/s41586-019-1098-2

Cohen, J. H., and Forward, R. B. (2016). Zooplankton diel vertical migration - a review of proximate control. *Oceanogr. Mar. Biol.* 47, 77–109. doi: 10.1201/9781420094220.ch2

Claustre, H., Antoine, D., Boehme, L., Boss, E., D'Ortenzio, F., Fanton, D., et al. (2011). "Guidelines towards an integrated ocean observation system for ecosystems and biogeochemical cycles," in *Proceedings of the OceanObs'09: Sustained Ocean Observations and Information for Society*, Vol. 1, (Venice: ESA Publication), 21–25.

Dall'Olmo, G., Dingle, J., Polimene, L., Brewin, R. J. W., and Claustre, H. (2016). Substantial energy input to the mesopelagic ecosystem from the seasonal mixed-layer pump. *Nat. Geosci.* 9, 820–823. doi: 10.1038/NGEO2818

Decelle, J., Martin, P., Pabortsava, K., Pond, D. W., Tarling, G., Mahe, F., et al. (2013). Diversity, ecology and biogeochemistry of cyst-forming acantharia (Radiolaria) in the oceans. *PLoS One* 8:e53598. doi: 10.1371/journal.pone.0053598

DeVries, T., and Primeau, F. (2011). Dynamically and observationally constrained estimates of water-mass distributions and ages in the global ocean. *J. Phys. Oceanogr.* 41, 2381–2401. doi: 10.1175/jpo-d-10-05011.1

DiTullio, G. R., Grebmeier, J. M., Arrigo, K. R., Lizotte, M. P., Robinson, D. H., Leventer, A., et al. (2000). Rapid and early export of Phaeocystis antarctica blooms in the Ross Sea, Antarctica. *Nature* 404, 595–598. doi: 10.1038/35007061

Gorgues, T., Aumont, O., and Memery, L. (2019). Simulated changes in the particulate carbon export efficiency due to diel vertical migration of zooplankton in the North Atlantic. *Geophys. Res. Lett.* 46, 5387–5395. doi: 10.1029/2018GL081748

Hansell, D. A., Carlson, C. A., Repeta, D. J., and Schlitzer, R. (2009). Dissolved organic matter in the ocean: a controversy stimulates New Insights. *Oceanography* 22, 202–211. doi: 10.5670/oceanog.2009.109

Hansen, A. N., and Visser, A. W. (2016). Carbon export by vertically migrating zooplankton: an optimal behavior model. *Limnol. Oceanogr.* 61, 701–710. doi: 10.1002/lno.10249

Kiko, R., Hauss, H., Dengler, M., Sommer, S., and Melzner, F. (2015). The squat lobster *Pleuroncodes monodon* tolerates anoxic "dead zone" conditions off Peru. *Mar. Biol.* 162, 1913–1921. doi: 10.1007/s00227-015-2709-6

Landry, M. R., and Calbet, A. (2004). Microzooplankton production in the oceans. *ICES J. Mar. Sci.* 61, 501–507. doi: 10.1016/j.icesjms.2004.03.011

Laurenceau-Cornec, E. C., Trull, T. W., Davies, D. M., Bray, S. G., Doran, J., Planchon, F., et al. (2015). The relative importance of phytoplankton aggregates and zooplankton fecal pellets to carbon export: insights from free-drifting sediment trap deployments in naturally iron-fertilised waters near the Kerguelen Plateau. *Biogeosciences* 12, 1007–1027. doi: 10.5194/bg-12-1007-2015

Levy, M., Bopp, L., Karleskind, P., Resplandy, L., Ethe, C., and Pinsard, F. (2013). Physical pathways for carbon transfers between the surface mixed layer and the ocean interior. *Global Biogeochem. Cycles.* 27, 1001–1012. doi: 10.1002/gbc.20092

Longhurst, A. R. (1991). Large marine ecosystems - patterns, processes and yields. *Mar. Policy* 15, 377–378.

Martin, P., Allen, J. T., Cooper, M. J., Johns, D. G., Lampitt, R. S., Sanders, R., et al. (2010). Sedimentation of acantharian cysts in the Iceland Basin: strontium as a ballast for deep ocean particle flux, and implications for acantharian reproductive strategies. *Limnol. Oceanogr.* 55, 604–614. doi: 10.4319/lo.2009.55.2.0604

Neill, W. E. (1990). Induced vertical migration in copepods as a defence against invertebrate predation. *Nature* 345, 524–526. doi: 10.1038/345524a0

Omand, M. M., D'asaro, E., Lee, C., Perry, M. J., Briggs, N., Cetini, I., et al. (2015). Eddy-driven subduction exports particulate organic carbon from the spring bloom. *Science* 348, 222–225. doi: 10.1126/science.1260062

Pabortsava, K., Lampitt, R. S., Benson, J., Crowe, C., McLachlan, R., Le Moigne, F. A. C., et al. (2017). Carbon sequestration in the deep Atlantic enhanced by Saharan dust. *Nat. Geosci.* 10, 189–194. doi: 10.1038/ngeo2899

Picheral, M., Guidi, L., Stemmann, L., Karl, D. M., Iddaoud, G., and Gorsky, G. (2010). The underwater vision profiler 5: an advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnol. Oceanogr. Methods* 8, 462–473. doi: 10.4319/lom.2010.8.462

Refaeli, Y., Young, R. M., Turner, B. C., Duda, J., Field, K. A., and Bishop, J. M. (2008). The B cell antigen receptor and overexpression of MYC can cooperate in the genesis of B cell lymphomas. *PLoS Biol.* 6:e152. doi: 10.1371/journal.pbio.0060152

Rembauville, M., Manno, C., Tarling, G. A., Blain, S., and Salter, I. (2016). Strong contribution of diatom resting spores to deep-sea carbon transfer in naturally iron-fertilized waters downstream of South Georgia. *Deep Sea Res. I* 115, 22–35. doi: 10.1016/j.dsr.2016.05.002

Roshan, S., and DeVries, T. (2017). Efficient dissolved organic carbon production and export in the oligotrophic ocean. *Nat. Commun.* 8:2036. doi: 10.1038/s41467-017-02227-3

Salter, I., Lampitt, R. S., Sanders, R., Poulton, A., Kemp, A. E. S., Boorman, B., et al. (2007). Estimating carbon, silica and diatom export from a naturally fertilised phytoplankton bloom in the Southern Ocean using PELAGRA: a novel drifting sediment trap. *Deep. Res. Part Ii-Topical Stud. Oceanogr.* 54, 2233–2259. doi: 10.1016/j.dsr2.2007.06.008

Steinberg, D. K. (2000). Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea. *Deep Sea Res. Part I Oceanogr. Res. Pap.* 47, 137–158. doi: 10.1016/s0967-0637(99)00052-7

Steinberg, D. K., and Landry, M. R. (2017). Zooplankton and the ocean carbon cycle. *Ann. Rev. Mar. Sci.* 9, 413–444. doi: 10.1146/annurev-marine-010814-015924

Stokes, G. G. (1851). On the effect of the internal friction of fluids on the motion of pendulums. *Trans. Cambridge Philos. Soc.* 9, 8–106. doi: 10.1017/cbo9780511702266.002

Stukel, M. R., Ohman, M. D., Benitez-Nelson, C. R., and Landry, M. R. (2013). Contributions of mesozooplankton to vertical carbon export in a coastal upwelling system. *Mar. Ecol. Prog. Ser.* 491, 47–65. doi: 10.3354/meps10453

Tarling, G. A., Jarvis, T., Emsley, S. M., and Matthews, J. B. L. (2002). Midnight sinking behaviour in *Calanus finmarchicus*: a response to satiation or krill predation? *Mar. Ecol. Prog. Ser.* 240, 183–194. doi: 10.3354/meps240183

Turner, J. T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump. *Prog. Oceanogr.* 130, 205–248. doi: 10.1016/j.pocean.2014.08.005

Wilson, S. E., Steinberg, D. K., and BUesseler, K. O. (2008). Changes in fecal pellet characteristics with depth as indicators of zooplankton repackaging of particles in the mesopelagic zone of the subtropical and subarctic North Pacific Ocean. *Deep Sea Res. II* 55, 1636–1647. doi: 10.1016/j.dsr2.2008.04.019

Excerpted from

Le Moigne, F. A. (2019). Pathways of organic carbon downward transport by the oceanic biological carbon pump. *Frontiers in Marine Science*, *6*, 634.

9.3.3: Export from the Upper Ocean to the Depths and Sediments is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

9.3.4: Summary

9.3.4: Summary is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

9.4: 12.4 Biogeochemistry of the Deep Ocean

Introduction

The deep sea is an environment that is unfriendly to humankind. As a result, it represents one of the least explored areas on Earth. It has even been said that the deep sea is less explored than the surface of the moon. The deep sea region begins about 1000 meters beneath the ocean's surface, located between the thermocline and the seafloor (Gage 1991). Here, light is limited, and after a depth of 100-500 meters, it disappears completely. Temperatures in the deep sea can significantly drop -- nearing 0°C at the very bottom. However, its temperature typically ranges between 2-3°C. The salinity level is relatively stable, with concentrations ranging between 34-35 parts per thousand (ppt). On the other hand, its pressure steadily rises with depth, increasing approximately 10 atmospheric pressure per meter (Gage 1991; George 2013).

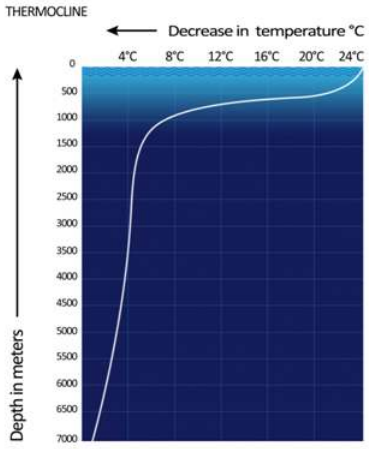


Figure 1. https://upload.wikimedia.org/wikiped...HERMOCLINE.png
This figure depicts the thermocline. Temperature of the surface water remains high until it reaches the thermocline where the temperature drops rapidly as the lack of heat from the sun dissipates. Temperature continues to decline, though much less rapidly, at lower depths.

Life at the Depths

While many scientists believed the deep sea was devoid of life due to its harsh conditions, multiple expeditions have demonstrated the opposite. They found that the deep sea possessed intricate and lively ecosystems, despite being thousands of meters below the surface. In the 1870s, the Challenger expedition revealed the existence of a variety of deep water species. More recently, in 1977, scientists aboard DSV Alvin discovered diverse deep sea communities around hydrothermal vents.

Caused by the absence of light, life in the deep sea faces multiple challenges, such as having difficulty producing energy, attracting prey, and finding mates. At these depths, primary producers can no longer rely on the sun for energy, so they use alternative sources, such as falling detritus from above. These falling particles are known as marine snow, consisting of dead or decaying plant, animal material, fecal matter, and other organic and inorganic particulate matter (Silver 2015). The continuous rain of marine snow provides food for many deep-sea creatures, giving them plenty of carbon and nitrogen to sustain the scavengers of the deep and fuel the rest of the ecosystem.

Physical Adaptations

The animals of the deep sea have adapted to survive in their habitats' extremely harsh conditions: no light, low amounts of food, low amounts of oxygen, high pressure, and cold temperatures. Advantageous physiological changes have developed to enable them to survive in this kind of environment.

Many organisms in the deep sea have adapted a useful physiological trait to aid in the environment's low light conditions: bioluminescence. Bioluminescence is the production and emission of light by a living organism. The ability to produce their own light allows organisms to lure prey, attract mates, and find paths in the water. Additionally, it can be used for defense, mimicry, and

counter-illumination (George 2013). Certain squid species can flash warning lights or secret bioluminescent ink in order to defend themselves and ward off predators. Mimicry can be used to match the light pattern of an otherwise fearsome or undesirable organism in order to fend off predators as well. Counter-illumination employs the use of lights on the underside of an organism to match the light coming from above, allowing an organism to appear dark from above (to match the dark water below it) and light from below (to match the light coming through the water above). A common example of attraction via bioluminescence can be demonstrated by angler fish. Angler fish have an appendage that extends from their head with a small light on the end. As they wave this appendage back and forth, they attract and lure prey towards their mouth (Gage 1991).

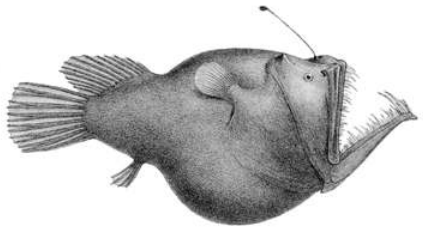


Figure 2. upload.wikimedia.org/wikiped...nglerfish).jpg
This image depicts the anglerfish with a light lure. This is a demonstration of the evolutionary adaptation bioluminescence.

Many deep sea fishes are either completely blind or have developed enlarged eyes, which enable them to detect the faintest of lights. To compensate for the loss of vision, they are well-equipped with a lateral line system, tactile organs, and other sensory structures (e.g. fin rays) (Pandev, 2007).

Body color also has become an adaptation for camouflage. Animals' bodies are often transparent, black, or red. The absence of red light at these depths keeps red-colored animals hidden from predators and prey.

Piezophiles are organisms that thrive in areas with high pressure. In order for them to inhabit such extreme environments, they evolved various mechanisms to counteract the effects of the elevated pressures. They are able to handle such pressure by remaining at a small size and eliminating excess cavities in their bodies that would otherwise collapse.

Many marine organisms have lost appendages and organs to conserve energy output, compensate for the extreme conditions, or because they simply do not use them anymore. The most common organ missing is the swim bladder; this is to escape the high stress of pressure present with extreme depth. In other cases, body parts have become enlarged and exaggerated (sensory organs, eyes, or mouths) in order to adapt to low light or scarce food (Pandey, 2007). An example of exaggerated and reduced body parts is the gulper eel, (*Eurypharynx pelecanoides*) which has tiny fins but also an enormous mouth, much larger than its body. Its mouth is loosely hinged and can open wide enough to swallow animals bigger than itself. Since food can be sparse in the deep sea, being able to eat whatever comes along, despite its size, is a huge advantage. The gulper eel has also adapted to having highly reduced pectoral fins because of their lack of use.



Figure 3. upload.wikimedia.org/wikiped...urypharynx.jpg
This image depicts the gulper eel with a large mouth and reduced fins.

What's Hiding at the Most Solitary Place on Earth? The Deep Sea - Kurzgesagt Video



Marine Snow



Deep Sea Exploration Documentaries 1-3

https://www.youtube.com/watch?v=rCbJMtCdvMY

https://www.youtube.com/watch?v=jOIOXvU0_qk

https://www.youtube.com/watch?v=M9lXaThM5Vo

Angler Fish and Gulper Eel

https://www.youtube.com/watch?v=FswYwyke7cc

Hatchetfish

https://www.youtube.com/watch?v=XmjflLIB9J4

References

1. http://marinebio.org/oceans/deep/
2. http://www.seasky.org/deep-sea/gulper-eel.html
3. www.dosits.org/tutorials/scie...utorial/speed/
4. https://inglestic.wikispaces.com/Anglerfish
5. www.drunkonblue.com/2010/01/the-gulper-eel/
6. http://www.zmescience.com/other/feature-post/remarkably-bizarre-deep-sea-creatures-freak-show/
7. http://www.seasky.org/deep-sea/bioluminescence.html
8. Gage, John D., and Paul A. Tyler. 1991. *Deep-sea biology: a natural history of organisms at the deep-sea floor*. Cambridge University Press.
9. George, Robert Y. 2013 "Deep-Sea Organisms."*Functional Adaptations of Marine Organisms*: 279.
10. Pandey, Kamleshwar and J.P. Shukla. 2007. *Fish and Fisheries.* Meerut India: Rastogi. Print.
11. Silver, Mary. 2015. "Marine Snow: A Brief Historical Sketch."*Limnology and Oceanography Bulletin.”* 24.1: 5-10.

9.4: 12.4 Biogeochemistry of the Deep Ocean is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

SECTION OVERVIEW

9.5: Biogeochemistry in ocean sediments

9.5.1: Introduction

9.5.2: Classifying Sediments

9.5.3: Lithogenous Sediments

9.5.4: Biogenous Sediments

9.5.5: Hydrogenous Sediments

9.5.6: Cosmogenous Sediments

9.5.7: Sediment Distribution

9.5: Biogeochemistry in ocean sediments is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

9.5.1: Introduction

Learning Objectives

After reading this chapter you should:

- know how sediments are classified based on physical characteristics (size, sorting etc.)
- identify the four main sources of marine sediments
- differentiate between organisms that produce different biogenous sediments
- understand the factors that determine the distribution of sediment types in the ocean
- know the different ways sediment samples can be obtained
- understand how biogenous sediments can be used to reconstruct past climate change

Let’s be honest; for the majority of people with an interest in the oceans and oceanography it is not the allure of the sediments that first grabs their attention. At first glance the muddy seafloor may not seem that interesting, but the sediments play a vital role in marine ecosystems and our understanding of ocean and geological processes. The sediments provide habitat for a multitude of marine organisms, and they contain information about past climates, plate tectonics, ocean circulation patterns, and the timing of major extinctions, just to name a few. In this chapter we will examine the major types of sediments, and their distribution on the ocean floor.

9.5.1: Introduction is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

9.5.2: Classifying Sediments

The term “sediment” refers to the tiny particles of rocks and other materials that sink to the ocean floor and eventually settle and accumulate on the bottom. All regions of the seafloor contain some form of sediment, although there are many different types of sediments from a variety of sources, and the amount of accumulated sediment can vary greatly from place to place. Globally, ocean sediments average about 1 km thick, but they can exceed 15 km thick in areas of high accumulation (Figure 9.5.2.1). These areas include regions near the mouths of rivers where there is high sediment discharge, and passive margins near the continents where the seafloor has had millions of years for sediment to accumulate. On the other hand, sediments are sparse along divergent plate boundaries where new oceanic crust is being formed, as the crust is too new for significant accumulation (see section 4.5), and in the central oceans that are far away from any significant sediment sources.

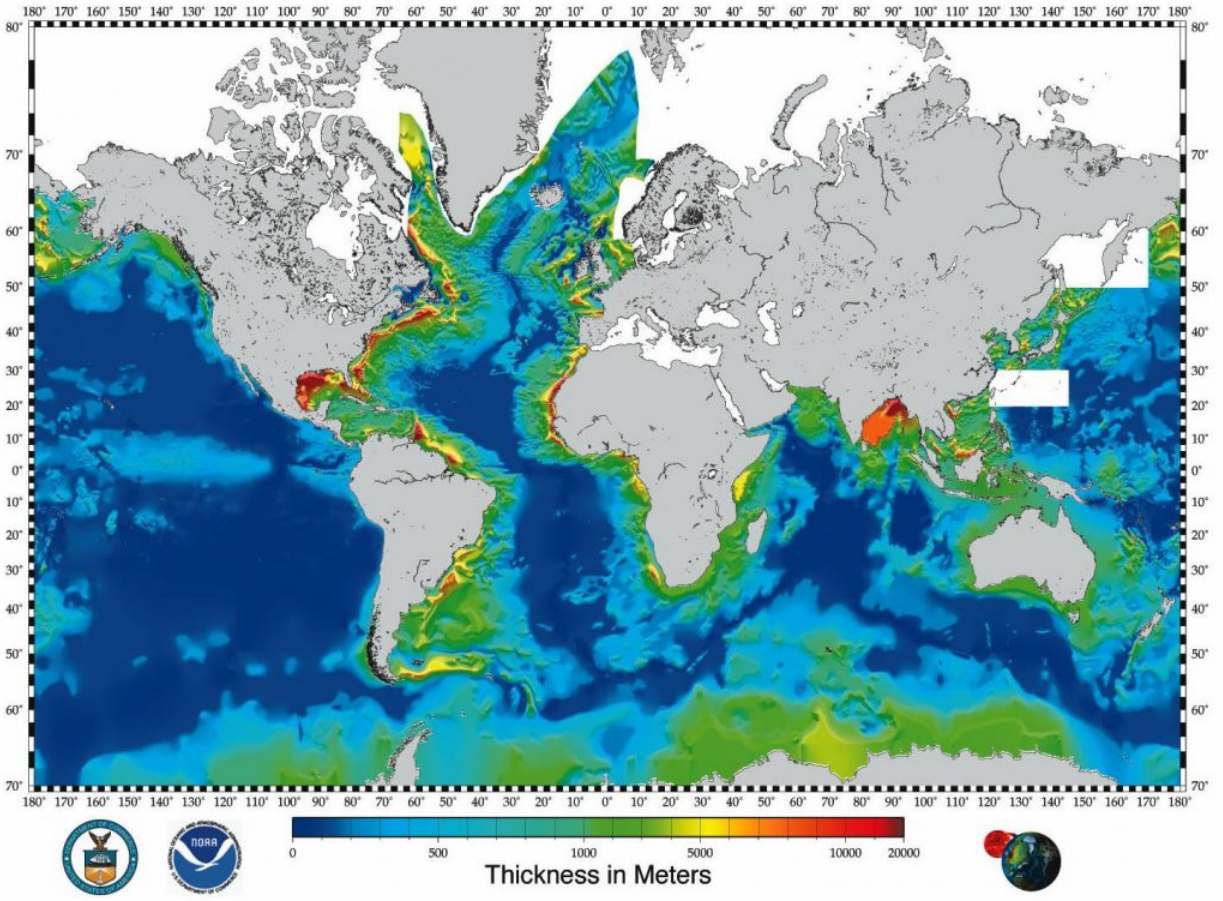


Figure 9.5.2.1 Total sediment thickness of the world ocean (By Divins, D.L., NGDC Total Sediment Thickness of the World’s Oceans & Marginal Seas, http://www.ngdc.noaa.gov/mgg/sedthick/sedthick.html (http://www.ngdc.noaa.gov/mgg/image/sedthick9.jpg) [Public domain], via Wikimedia Commons).

As time passes over millions of years, these sediments can become **lithified** or turned into sedimentary rock. It has been estimated that over half of the exposed rock on the continents is sedimentary rock originally deposited in ancient oceans and uplifted by plate tectonics. Many tall mountains, including Mt. Everest, are composed of rock formations that contain fossils of marine creatures. These rocks were originally formed as ocean sediments which were then lithified and pushed upwards during the process of mountain formation.

Sediment Classification

There are a number of ways that we can classify ocean sediments, and some of the most common distinctions are based on the sediment texture, the sediment composition, and the sediment’s origin.

Texture

Sediment texture can be examined through several variables. The first is **grain size**. Sediments are classified by particle size, ranging from the finest clays (diameter <0.004 mm) to the largest boulders (> 256 mm)(Figure 9.5.2.2). Among other things, grain size represents the conditions under which the sediment was deposited. High energy conditions, such as strong currents or waves, usually results in the deposition of only the larger particles as the finer ones will be carried away. Lower energy conditions will allow the smaller particles to settle out and form finer sediments.

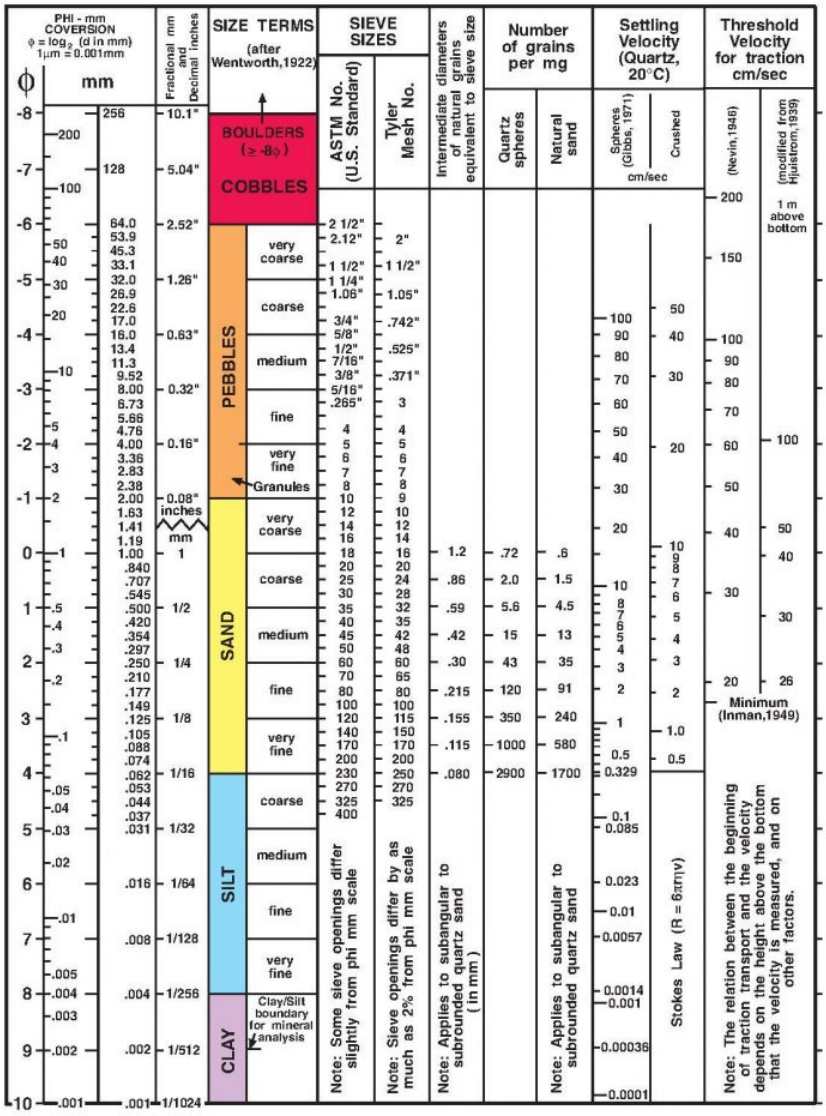


Figure 9.5.2.2 Wentworth grain size chart for classifying sediments (By Jeffress Williams, Matthew A. Arsenault, Brian J. Buczkowski, Jane A. Reid, James G. Flocks, Mark A. Kulp, Shea Penland, and Chris J. Jenkins, USGS [Public domain], via Wikimedia Commons).

Sorting is another way to categorize sediments. Sorting refers to how uniform the particles are in terms of size (Figure 9.5.2.3). If all of the particles are of a similar size, such as in beach sand, the sediment is well-sorted. If the particles are of very different sizes, the sediment is poorly sorted, such as in glacial deposits.

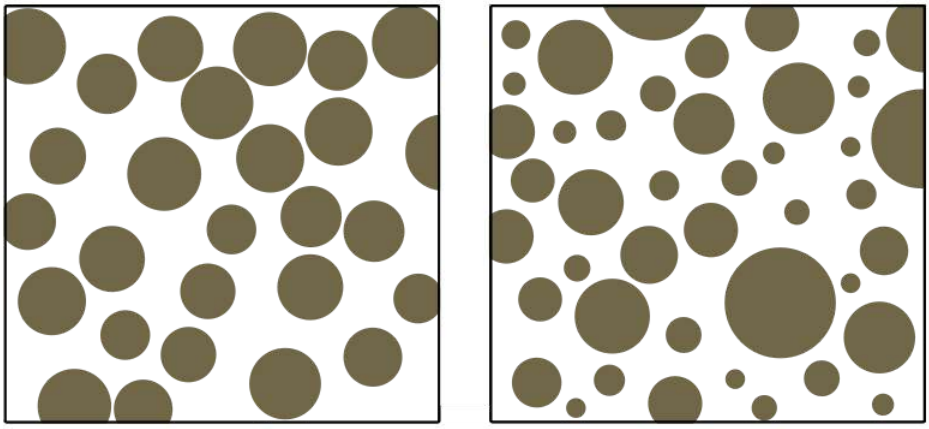


Figure 9.5.2.3 Well-sorted sediments (left) have particles that are all of a similar size. Poorly sorted sediments (right) consist of particles of a wide range of sizes (Woudloper (Own work) [CC BY-SA 3.0], via Wikimedia Commons).

Finally, sediment texture can be described based on its **maturity**, or how long the particles have been transported by water or other vectors. Maturity can be reflected by the degree of rounding of the particles, the amount of sorting, and the composition of the sediment. In the case of **rounding**, the more mature the sediment, the rounder the particles, as a result of the particles being abraded over time. A high degree of sorting indicates maturity, because over time the smaller particles will be washed away, and a given amount of energy will move particles of a similar size over the same distance. Lastly, the older and more mature a sediment the higher the quartz content, at least in sediments derived from rock particles. Quartz is a common mineral in terrestrial rocks, and it is very hard and resistant to abrasion. Over time, particles made from other materials are worn away, leaving only quartz behind. Beach sand is a very mature sediment; it is composed primarily of quartz, and the particles are rounded and of similar size (well-sorted).

Sources of sediments

Sediments are also classified based on their source of origin. There are four main categories for the origin of marine sediments:

- **Lithogenous** sediments are derived from preexisting rock. They are also called **terrigenous** sediments since most of it comes from the land masses and makes its way into the ocean.
- **Biogenous** sediments are composed of the remains of marine organisms.
- **Hydrogenous** sediments are formed when materials that are dissolved in water precipitate out and form solid particles.
- **Cosmogenous** sediments are derived from extraterrestrial sources.

The next few sections will address each of these sediment types in more detail.

This page titled 9.5.2: Classifying Sediments is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Paul Webb via source content that was edited to the style and standards of the LibreTexts platform.

- **12.1: Classifying Sediments** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

9.5.3: Lithogenous Sediments

Lithogenous or terrigenous sediment is primarily composed of small fragments of preexisting rocks that have made their way into the ocean. These sediments can contain the entire range of particle sizes, from microscopic clays to large boulders, and they are found almost everywhere on the ocean floor. Lithogenous sediments are created on land through the process of weathering, where rocks and minerals are broken down into smaller particles through the action of wind, rain, water flow, temperature- or ice-induced cracking, and other erosive processes. These small eroded particles are then transported to the oceans through a variety of mechanisms:

- **Streams and rivers**: Various forms of runoff deposit large amounts of sediment into the oceans, mostly in the form of finer-grained particles (Figure 9.5.3.1). About 90% of the lithogenous sediment in the oceans is though to have come from river discharge, particularly from Asia. Most of this sediment, especially the larger particles, will be deposited and remain fairly close to the coastline, however, smaller clay particles may remain suspended in the water column for long periods of time and may be transported great distances from the source.



Figure 9.5.3.1 River discharge in the Yukon Delta, Alaska. The pale color demonstrates the large amounts of sediment released into the ocean via the rivers (By Jesse Allen and Robert Simmon (NASA Earth Observatory) [Public domain], via Wikimedia Commons).

- **Wind**: Wind-borne (**aeolian**) transport can take small particles of sand and dust and move them thousands of kilometers from the source. These small particles can fall into the ocean when the wind dies down, or can serve as the nuclei around which raindrops or snowflakes form. Aeolian transport is particularly important near desert areas (Figure 9.5.3.2).

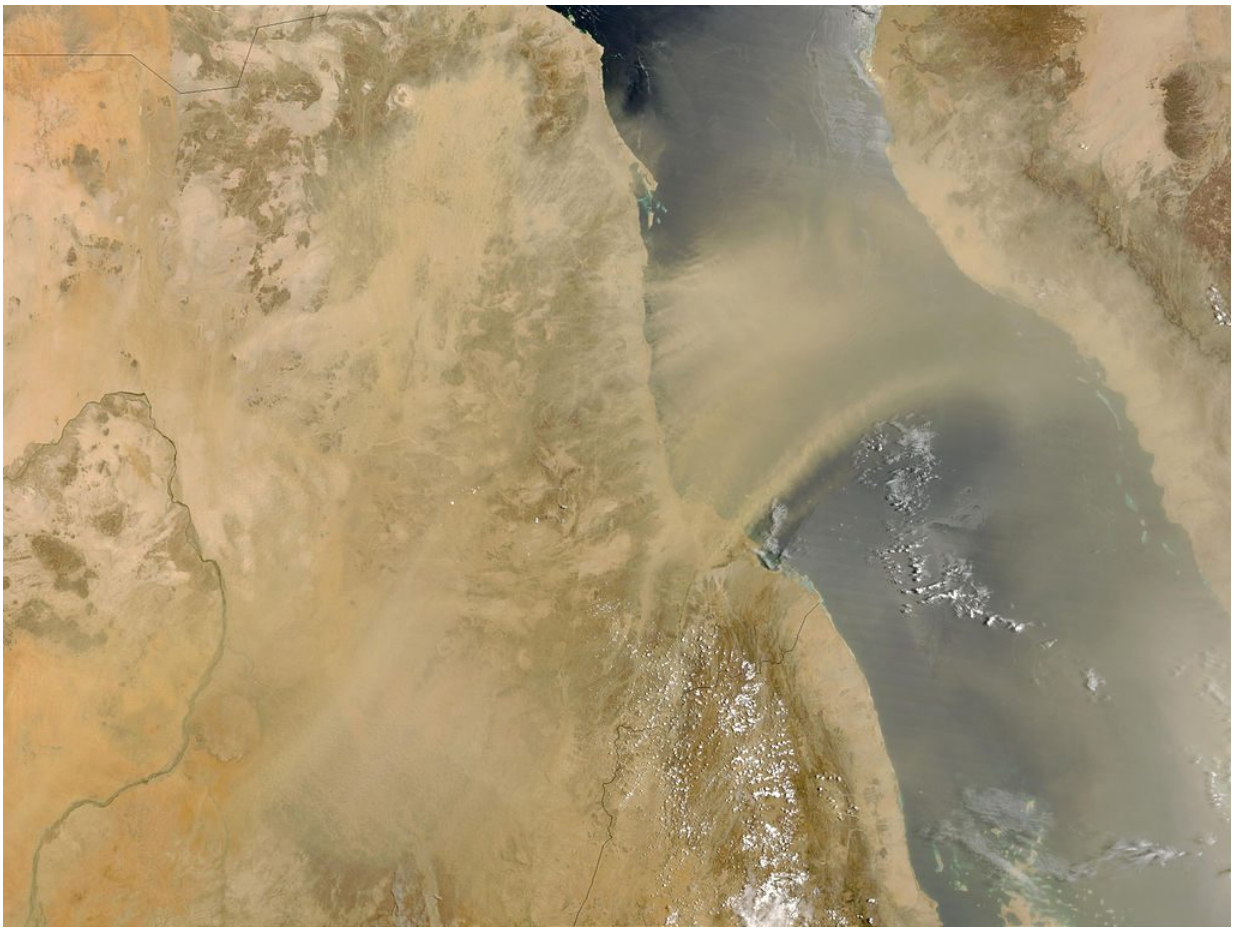


Figure 9.5.3.2 A plume of wind-borne particles from Sudan (left) blow over the Red Sea (By NASA (http://visibleearth.nasa.gov/view_rec.php?id=5645) [Public domain], via Wikimedia Commons).

- **Glaciers and ice rafting**: As glaciers grind their way over land, they pick up lots of soil and rock particles, including very large boulders, that get carried by the ice. When the glacier meets the ocean and begins to break apart or melt, these particles get deposited (Figure 9.5.3.3). Most of the deposition will happen close to where the glacier meets the water, but a small amount of material is also transported longer distances by rafting, where larger pieces of ice drift far from the glacier before releasing their sediment.



Figure 9.5.3.3 When glaciers reach the sea they can break apart, depositing their sediments into the ocean, including very large pieces of rock (By Ianaré Sévi (Own work) [CC BY-SA 3.0], via Wikimedia Commons).

- **Gravity**: Landslides, mudslides, avalanches, and other gravity-driven events can deposit large amounts of material into the ocean when they happen close to shore.

- **Waves**: Wave action along a coastline will erode rocks and will pull loose particles from beaches and shorelines into the water.
- **Volcanoes**: Volcanic eruptions emit vast amounts of ash and other debris into the atmosphere, where it can then be transported by wind to eventually get deposited in the oceans (Figure 9.5.3.4).



Figure 9.5.3.4 Eruption of the Mayon Volcano, Philippines, in 1984. Much of the material spewed from a volcanic eruption may eventually make its way into the oceans (By C.G. Newhall [Public domain], via Wikimedia Commons).

- **Gastroliths**: An interesting, although relatively minor avenue for the transport of lithogenous sediments to the ocean is in the form of gastroliths. Gastrolith means “stomach stones” and comes from the fact that many animals, including seabirds, pinnipeds, and some crocodiles will deliberately swallow stones in one area and regurgitate them in another. Often these stones swallowed on land will be regurgitated at sea. Why swallow stones? Possible explanations include using the stones to help grind up food in the stomach, to act as ballast to aid in buoyancy regulation, or to fill the stomach and reduce feelings of hunger during fasting periods on shore.

Most of these processes deposit lithogenous sediment fairly close to shore. The sediment particles can then be transported farther away by waves and currents, where they may eventually escape the continental shelf and reach the deep ocean floor.

Composition

Lithogenous sediments usually reflect the composition of whatever materials they were derived from, so they are dominated by the major minerals that make up most terrestrial rock. This includes quartz, feldspar, clay minerals, iron oxides, and terrestrial organic matter. Quartz (silicon dioxide, the main component of glass) is one of the most common minerals found in nearly all rocks, and it is very resistant to abrasion (see section 12.1), so it is a dominant component of lithogenous sediments, including sand.

This page titled 9.5.3: Lithogenous Sediments is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Paul Webb via source content that was edited to the style and standards of the LibreTexts platform.

- **12.2: Lithogenous Sediments** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

9.5.4: Biogenous Sediments

Biogenous sediments come from the remains of living organisms that settle out as sediment when the organisms die. It is the “hard parts” of the organisms that contribute to the sediments; things like shells, teeth or skeletal elements, as these parts are usually mineralized and are more resistant to decomposition than the fleshy “soft parts” that rapidly deteriorate after death.

Macroscopic sediments contain large remains, such as skeletons, teeth, or shells of larger organisms. This type of sediment is fairly rare over most of the ocean, as large organisms don’t die in enough of a concentrated abundance to allow these remains to accumulate. One exception is around coral reefs; here there is a great abundance of organisms that leave behind their remains, in particular the fragments of the stony skeletons of corals that make up a large percentage of tropical sand.

Microscopic sediment consists of the hard parts of microscopic organisms, particularly their shells, or **tests**. Although very small, these organisms are highly abundant and as they die by the billions every day their tests sink to the bottom to create biogenous sediments. Sediments composed of microscopic tests are far more abundant than sediments from macroscopic particles, and because of their small size they create fine-grained, mushy sediment layers. If the sediment layer consists of at least 30% microscopic biogenous material, it is classified as a biogenous **ooze**. The remainder of the sediment is often made up of clay.

The primary sources of microscopic biogenous sediments are unicellular algaes and protozoans (single-celled amoeba-like creatures) that secrete tests of either calcium carbonate ($CaCO_3$) or silica (SiO_2). Silica tests come from two main groups, the **diatoms** (algae) and the **radiolarians** (protozoans) (Figure 9.5.4.1).

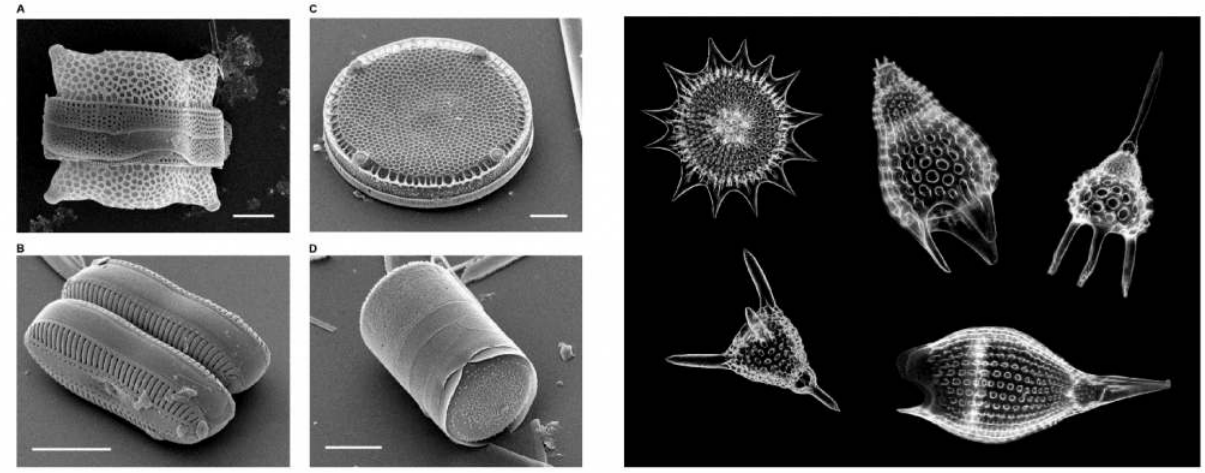


Figure 9.5.4.1 Various diatom (left) and radiolarian (right) tests (Diatom images courtesy of Mary Ann Tiffany, San Diego State University [CC BY 2.5], via Wikimedia Commons; radiolarian images by Andreas Drews, https://pxhere.com/en/photo/239774, [CC by 2.0]).

Diatoms are important members of the phytoplankton, the small, drifting algal photosynthesizers. A diatom consists of a single algal cell surrounded by an elaborate silica shell that it secretes for itself. Diatoms come in a range of shapes, from elongated, pennate forms, to round, or centric shapes that often have two halves, like a Petri dish (Figure 9.5.4.1 *left*). In areas where diatoms are abundant, the underlying sediment is rich in silica diatom tests, and is called **diatomaceous earth** (see box below).

What use are diatoms?

Diatoms are a vital piece of the global ecosystem for their role in oceanic primary production and the creation of much of the oxygen that organisms breathe. But diatoms are also important for many industrial and agricultural applications. Because of the very fine grain size, and the lattice-like structure of the diatom tests, diatomaceous earth has been used as a filtering agent in things like swimming pool filters and beer brewing. The microscopic tests have been added as an abrasive to toothpaste, facial cleansers and household cleaning agents. Alfred Nobel used diatomaceous earth to stabilize nitroglycerine in the production of dynamite. Diatomaceous earth also displays insecticide properties by stimulating dehydration in insects. It is marketed for this purpose in agriculture, as well as for household use to combat ants, cockroaches, and bedbugs. “Food grade” diatomaceous earth has also

entered the market, with proponents touting a range of health benefits arising from its consumption. That's a pretty impressive range of uses from a microscopic algae!

Radiolarians are planktonic protozoans (making them part of the zooplankton), that like diatoms, secrete a silica test. The test surrounds the cell and can include an array of small openings through which the radiolarian can extend an amoeba-like "arm" or pseudopod (Figure 9.5.4.1 *right*). Radiolarian tests often display a number of rays protruding from their shells which aid in buoyancy. Oozes that are dominated by diatom or radiolarian tests are called **siliceous oozes**.

Like the siliceous sediments, the calcium carbonate, or calcareous sediments are also produced from the tests of microscopic algae and protozoans; in this case the **coccolithophores** and **foraminiferans**. Coccolithophores are single-celled planktonic algae about 100 times smaller than diatoms. Their tests are composed of a number of interlocking $CaCO_3$ plates (coccoliths) that form a sphere surrounding the cell (Figure 9.5.4.2 *left*). When coccolithophores die the individual plates sink out and form an ooze. Over time, the coccolithophore ooze lithifies to becomes chalk. The famous White Cliffs of Dover in England are composed of coccolithophore-rich ooze that turned into chalk deposits (Figure 9.5.4.2 *right*).

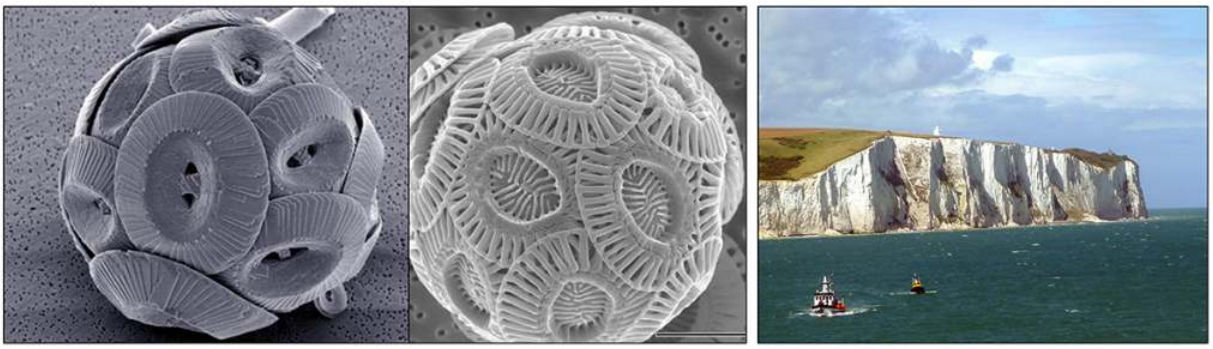


Figure 9.5.4.2 (**Left**) coccolithophore tests (left; By Richard Lampitt, Jeremy Young, The Natural History Museum, London (http://planktonnet.awi.de/); center; by Alison R. Taylor (University of North Carolina Wilmington Microscopy Facility) (PLoS Biology, June 2011, Cover ([1])) [Both images CC BY 2.5], via Wikimedia Commons). (**Right**); the White Cliffs of Dover (Immanuel Giel (Own work) [CC BY-SA 3.0], via Wikimedia Commons).

Foraminiferans (also referred to as "forams") are protozoans whose tests are often chambered, similar to the shells of snails. As the organism grows, is secretes new, larger chambers in which to reside. Most foraminiferans are benthic, living on or in the sediment, but there are some planktonic species living higher in the water column. When coccolithophores and foraminiferans die, they form **calcareous oozes**.



Figure 9.5.4.3 Foraminifera tests collected from a beach in Myanmar (By Psammophile [GFDL (http://www.gnu.org/copyleft/fdl.html) or CC BY-SA 3.0], via Wikimedia Commons).

Older calcareous sediment layers contain the remains of another type of organism, the **discoasters**; single-celled algae related to the coccolithophores that also produced calcium carbonate tests. Discoaster tests were star-shaped, and reached sizes of 5-40 µm across (Figure 9.5.4.4). Discoasters went extinct approximately 2 million years ago, but their tests remain in deep, tropical sediments that predate their extinction.

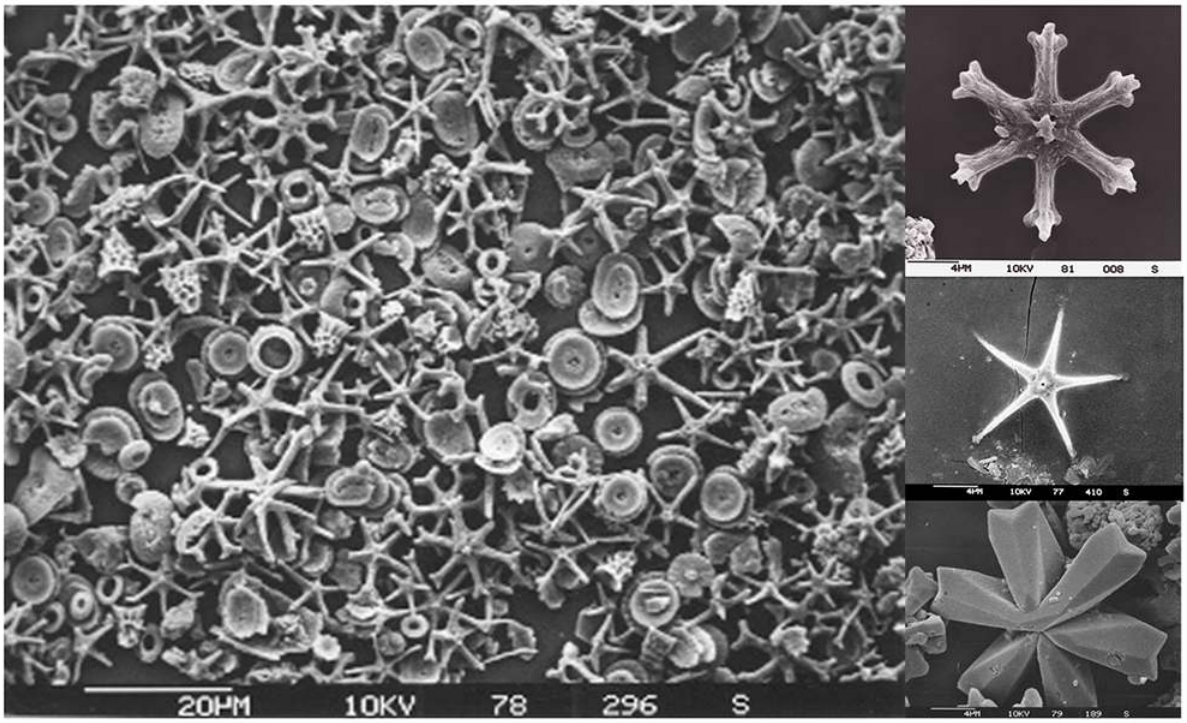


Figure 9.5.4.4 Discoaster tests. Left: discoster tests with assorted coccoliths. Top right *Discoaster surculus;* center right: *Discoaster pentaradiatus;* bottom right: *Discoaster surculus* (All images by Hannes Grobe (Own work) [CC BY 3.0], via Wikimedia Commons).

Because of their small size, these tests sink very slowly; a single microscopic test may take about 10-50 years to sink to the bottom! Given that slow descent, a current of only 1 cm/sec could carry the test as much as 15,000 km away from its point of origin before it reaches the bottom. Yet despite this, we find that the sediments in a particular location are well-matched to the types of organisms and degree of productivity that occurs in the water overhead. This means that the sediment particles must be sinking to the bottom at a much faster rate, so that they accumulate below their point of origin before the currents can disperse them. What is the mechanism for this increased sinking rate? Apparently most of the tests do not sink as individual particles; about 99% of them are first consumed by some other organism, and are then aggregated and expelled as large fecal pellets, which sink much more quickly and reach the ocean floor in only 10-15 days. This does not give the particles as much time to disperse, and the sediment below will reflect the production occurring near the surface. The increased rate of sinking through this mechanism is called the “fecal express.”

This page titled 9.5.4: Biogenous Sediments is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Paul Webb via source content that was edited to the style and standards of the LibreTexts platform.

- **12.3: Biogenous Sediments** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

9.5.5: Hydrogenous Sediments

As we saw in section 5.3 seawater contains many different dissolved substances. Occasionally chemical reactions occur that cause these substances to precipitate out as solid particles, which then accumulate as **hydrogenous sediment**. These reactions are usually triggered by a change in conditions, such as a change in temperature, pressure, or pH, which reduces the amount of a substance that can remain in a dissolved state. There is not a lot of hydrogenous sediment in the ocean compared to lithogenous or biogenous sediments, but there are some interesting forms.

Hydrothermal vents were discussed in section 4.11. Recall that in these systems, seawater percolates into the seafloor, where it becomes superheated by magma before being expelled by the vent. This superheated water contains many dissolved substances, and when it encounters the cold seawater after leaving the vent, these particles precipitate out, mostly as metal sulfides. These particles make up the "smoke" that flows from a vent, and may eventually settle on the bottom as hydrogenous sediment (Figure 9.5.5.1).

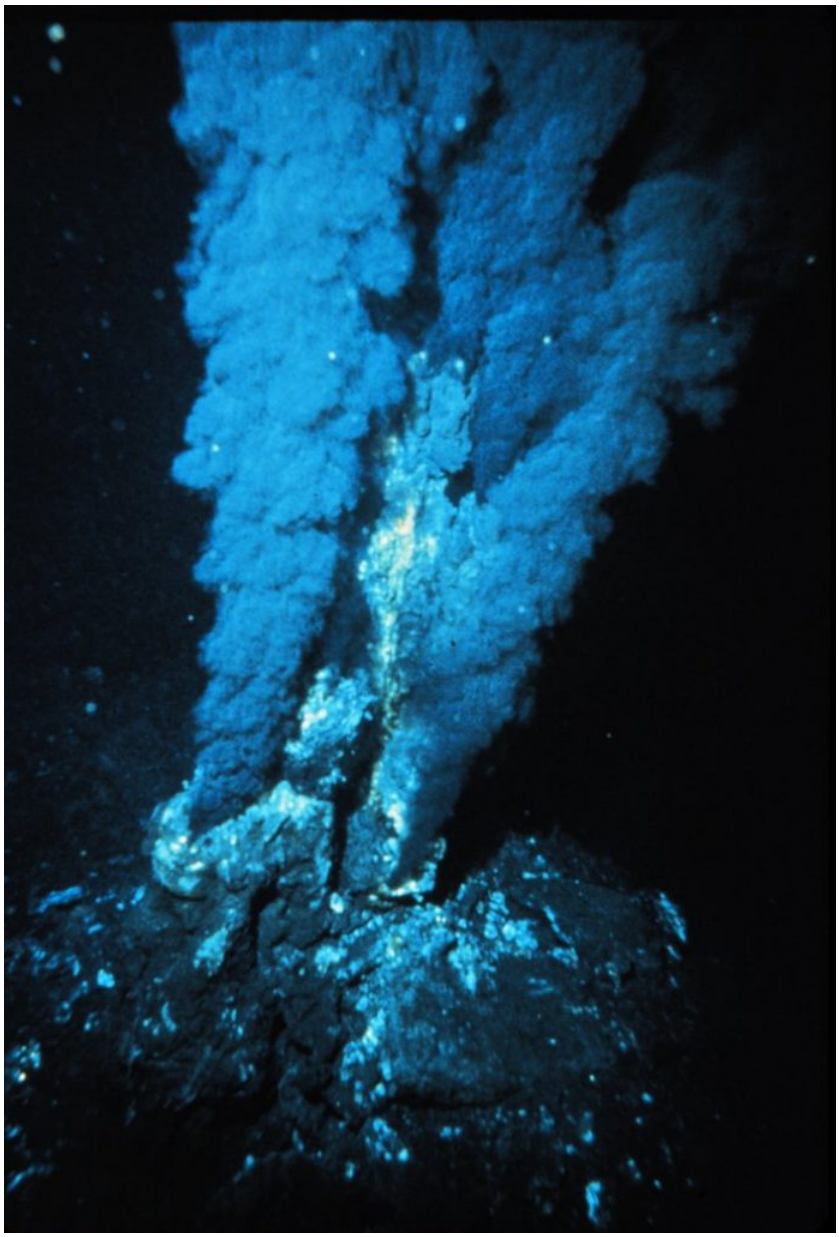


Figure 9.5.5.1 A "black smoker" hydrothermal vent. The "smoke" consists of dissolved particles that precipitate into solids when exposed to colder water (NOAA, http://www.photolib.noaa.gov/htmls/nur04506.htm).

Manganese nodules are rounded lumps of manganese and other metals that form on the seafloor, generally ranging between 3-10 cm in diameter, although they may sometimes reach up to 30 cm (Figure 9.5.5.2). The nodules form in a manner similar to pearls; there is a central object around which concentric layers are slowly deposited, causing the nodule to grow over time. The composition of the nodules can vary somewhat depending on their location and the conditions of their formation, but they are usually dominated by manganese- and iron oxides. They may also contain smaller amounts of other metals such as copper, nickel and cobalt. The precipitation of manganese nodules is one of the slowest geological processes known; they grow on the order of a few millimeters per million years. For that reason, they only form in areas where there are low rates of lithogenous or biogenous sediment accumulation, because any other sediment deposition would quickly cover the nodules and prevent further nodule growth. Therefore, manganese nodules are usually limited to areas in the central ocean, far from significant lithogenous or biogenous inputs, where they can sometimes accumulate in large numbers on the seafloor (Figure 9.5.5.2 *right*). Because the nodules contain a number of commercially valuable metals, there has been significant interest in mining the nodules over the last several decades, although most of the efforts have thus far remained at the exploratory stage. A number of factors have prevented large-scale extraction of nodules, including the high costs of deep sea mining operations, political issues over mining rights, and environmental concerns surrounding the extraction of these non-renewable resources.

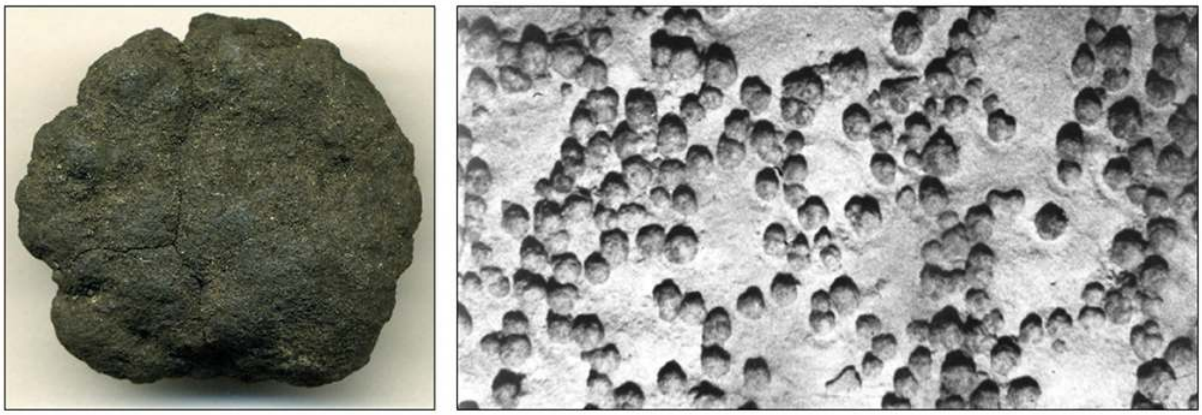


Figure 9.5.5.2 (Left) manganese nodule from the subtropical eastern Pacific Ocean. The nodule is 4.1 cm in diameter (By James St. John, https://www.flickr.com/photos/jsjgeo...n/photostream/ [CC-BY 2.0]). (Right) field of manganese nodules on the seafloor (By United States Geological Survey [Public domain], via Wikimedia Commons).

Evaporites are hydrogenous sediments that form when seawater evaporates, leaving the dissolved materials to precipitate into solids, particularly halite (salt, NaCl). In fact, the evaporation of seawater is the oldest form of salt production for human use, and is still carried out today. Large deposits of halite evaporites exist in a number of places, including under the Mediterranean Sea. Beginning around 6 million years ago, tectonic processes closed off the Mediterranean Sea from the Atlantic, and the warm climate evaporated so much water that the Mediterranean was almost completely dried out, leaving large deposits of salt in its place (an event known as the Messinian Salinity Crisis). Eventually the Mediterranean re-flooded about 5.3 million years ago, and the halite deposits were covered by other sediments, but they still remain beneath the seafloor.



Figure 9.5.5.3 Salt farmers harvesting salt left behind from the evaporation of seawater, Pak Thale, Ban Laem, Phetchaburi, Thailand (By JJ Harrison (Own work) [CC BY-SA 3.0], via Wikimedia Commons).

Oolites are small, rounded grains formed from concentric layers of precipitation of material around a suspended particle. They are usually composed of calcium carbonate, but they may also from from phosphates and other materials. Accumulation of oolites results in oolitic sand, which is found in its greatest abundance in the Bahamas (Figure 9.5.5.4).



Figure 9.5.5.4 Oolites from a beach on Joulter's Cay, The Bahamas (By Wilson44691 (Own work) [Public domain], via Wikimedia Commons).

Methane hydrates are another type of hydrogenous deposit with a potential industrial application. All terrestrial erosion products include a small proportion of organic matter derived mostly from terrestrial plants. Tiny fragments of this material plus other organic matter from marine plants and animals accumulate in terrigenous sediments, especially within a few hundred kilometers of shore. As the sediments pile up, the deeper parts start to warm up (from geothermal heat), and bacteria get to work breaking down the contained organic matter. Because this is happening in the absence of oxygen (a.k.a. anaerobic conditions), the by-product of this metabolism is the gas methane (CH_4). Methane released by the bacteria slowly bubbles upward through the sediment toward the seafloor. At water depths of 500 m to 1,000 m, and at the low temperatures typical of the seafloor (close to 4°C), water and methane combine to create a substance known as methane hydrate. Within a few meters to hundreds of meters of the seafloor, the temperature is low enough for methane hydrate to be stable and hydrates accumulate within the sediment (Figure 9.5.5.5 *left*). Methane hydrate is flammable because when it is heated, the methane is released as a gas (Figure 9.5.5.5 *right*). The methane within seafloor sediments represents an enormous reservoir of fossil fuel energy. Although energy corporations and governments are anxious to develop ways to produce and sell this methane, anyone that understands the climate-change implications of its extraction and use can see that this would be folly.



Figure 9.5.5.5 (Left): Methane hydrate within muddy sea-floor sediment from an area offshore from Oregon (By Wusel007 (Own work) [GFDL (http://www.gnu.org/copyleft/fdl.html) or CC BY-SA 3.0], via Wikimedia Commons). (Right): Methane hydrate on fire (USGS, http://www.usgs.gov/blogs/features/f.../New-Image.jpg).

---

*"Physical Geology" by Steven Earle used under a CC-BY 4.0 international license. Download this book for free at http://open.bccampus.ca

This page titled 9.5.5: Hydrogenous Sediments is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Paul Webb via source content that was edited to the style and standards of the LibreTexts platform.

- **12.4: Hydrogenous Sediments** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

9.5.6: Cosmogenous Sediments

Cosmogenous sediment is derived from extraterrestrial sources, and comes in two primary forms; microscopic spherules and larger meteor debris. Spherules are composed mostly of silica or iron and nickel, and are thought to be ejected as meteors burn up after entering the atmosphere. Meteor debris comes from collisions of meteorites with Earth. These high impact collisions eject particles into the atmosphere that eventually settle back down to Earth and contribute to the sediments. Like spherules, meteor debris is mostly silica or iron and nickel. One interesting form of debris from these collisions are tektites, which are small droplets of glass. They are likely composed of terrestrial silica that was ejected and melted during a meteorite impact, which then solidified as it cooled upon returning to the surface.



Figure 9.5.6.1 Tektite-like glass found near the Volkov River in western Russia (By James St. John [CC BY 2.0], via Wikimedia Commons).

Cosmogenous sediment is fairly rare in the ocean and it does not usually accumulate in large deposits. However, it is constantly being added to through space dust that continuously rains down on Earth. About 90% of incoming cosmogenous debris is vaporized as it enters the atmosphere, but it is estimated that 5 to 300 tons of space dust land on the Earth's surface each day!

This page titled 9.5.6: Cosmogenous Sediments is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Paul Webb via source content that was edited to the style and standards of the LibreTexts platform.

- **12.5: Cosmogenous Sediments** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

9.5.7: Sediment Distribution

Now that we have an understanding of the types of sediments found in the ocean, we can turn our attention to the processes that cause different types of sediments to dominate in different locations. Sediment accumulation will depend on the the amount of material coming from the source, the distance from the source, the amount of time that sediment has had to accumulate, how well the sediments are preserved, and the amounts of other types of sediments that are also being added to the system.

Rates of sediment accumulation are relatively slow throughout most of the ocean, in many cases taking thousands of years for any significant deposits to form. Lithogenous sediment accumulates the fastest, on the order of 1 m or more per thousand years for coarser particles. However, sedimentation rates near the mouths of large rivers with high discharge can be orders of magnitude higher. Biogenous oozes accumulate at a rate of about 1 cm per thousand years, while small clay particles are deposited in the deep ocean at around 1 mm per thousand years. As described in section 12.4, manganese nodules have an incredibly slow rate of accumulation, gaining 0.001 mm per thousand years.

Marine sediments are thickest near the continental margins (refer to figure 9.5.7.1) where they can be over 10 km thick. This is because the crust near passive continental margins is often very old, allowing for a long period of accumulation, and because there is a large amount of terrigenous sediment input coming from the continents. Near mid-ocean ridge systems where new oceanic crust is being formed, sediments are thinner, as they have had less time to accumulate on the younger crust. As you move away from the ridge spreading center the sediments get progressively thicker (see section 4.5), increasing by approximately 100-200 m of sediment for every 1000 km distance from the ridge axis. With a seafloor spreading rate of about 20-40 km/million years, this represents a sediment accumulation rate of approximately 100-200 m every 25-50 million years.

Figure 9.5.7.1 shows the distribution of the major types of sediment on the ocean floor. Cosmogenous sediments could potentially end up in any part of the ocean, but they accumulate in such small abundances that they are overwhelmed by other sediment types and thus are not dominant in any location. Similarly, hydrogenous sediments can have high concentrations in specific locations, but these regions are very small on a global scale. So we will mostly ignore cosmogenous and hydrogenous sediments in the discussion of global sediment patterns.

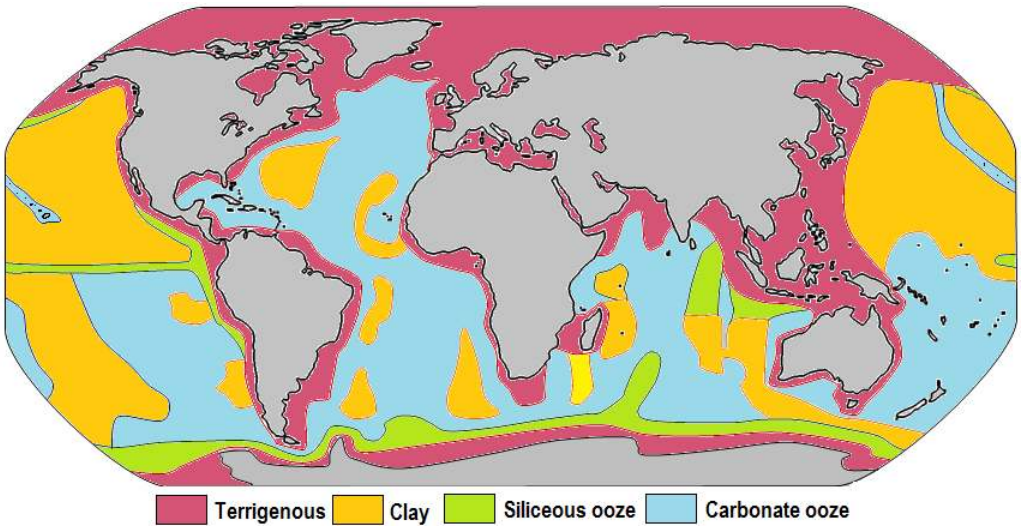


Figure 9.5.7.1 The distribution of sediment types on the seafloor. Within each colored area, the type of material shown is what dominates, although other materials are also likely to be present (Steven Earle, “Physical Geology”).

Coarse lithogenous/terrigenous sediments are dominant near the continental margins as runoff, river discharge, and other processes deposit vast amounts of these materials on the continental shelf (section 12.2). Much of this sediment remains on or near the shelf, while turbidity currents can transport material down the continental slope to the deep ocean floor. Lithogenous sediment is also common at the poles where thick ice cover can limit primary production, and glacial breakup deposits sediments along the ice edge. Coarse lithogenous sediments are less common in the central ocean, as these areas are too far from the sources for these sediments to accumulate. Very small clay particles are the exception, and as described below, they can accumulate in areas that other lithogenous sediment will not reach.

The distribution of biogenous sediments depends on their rates of production, dissolution, and dilution by other sediments. We learned in section 7.4 that coastal areas display very high primary production, so we might expect to see abundant biogenous deposits in these regions. However, recall that sediment must be >30% biogenous to be considered a biogenous ooze, and even in productive coastal areas there is so much lithogenous input that it swamps the biogenous materials, and that 30% threshold is not reached. So coastal areas remain dominated by lithogenous sediment, and biogenous sediments will be more abundant in pelagic environments where there is little lithogenous input.

In order for biogenous sediments to accumulate their rate of production must be greater than the rate at which the tests dissolve. Silica is undersaturated throughout the ocean and will dissolve in seawater, but it dissolves more readily in warmer water and lower pressures; in other words, it dissolves faster near the surface than in deep water. Silica sediments will therefore only accumulate in cooler regions of high productivity where they accumulate faster than they dissolve. This includes upwelling regions near the equator and at high latitudes where there are abundant nutrients and cooler water. Oozes formed near the equatorial regions are usually dominated by radiolarians, while diatoms are more common in the polar oozes. Once the silica tests have settled on the bottom and are covered by subsequent layers, they are no longer subject to dissolution and the sediment will accumulate. Approximately 15% of the seafloor is covered by siliceous oozes.

Biogenous calcium carbonate sediments also require production to exceed dissolution for sediments to accumulate, but the processes involved are a little different than for silica. Calcium carbonate dissolves more readily in more acidic water. Cold seawater contains more dissolved CO_2 and is slightly more acidic than warmer water (section 5.5). Therefore calcium carbonate tests are more likely to dissolve in colder, deeper, polar water than in warmer, tropical, surface water. At the poles the water is uniformly cold, so calcium carbonate readily dissolves at all depths, and carbonate sediments do not accumulate. In temperate and tropical regions calcium carbonate dissolves more readily as it sinks into deeper water. The depth at which calcium carbonate dissolves as fast as it accumulates is called the **calcium carbonate compensation depth**, or **calcite compensation depth**, or simply the **CCD**. The **lysocline** represents the depths where the rate of calcium carbonate dissolution increases dramatically (similar to the thermocline and halocline). At depths shallower than the CCD carbonate accumulation will exceed the rate of dissolution, and carbonate sediments will be deposited. In areas deeper than the CCD, the rate of dissolution will exceed production, and no carbonate sediments can accumulate (Figure 9.5.7.2). The CCD is usually found at depths of 4 – 4.5 km, although it is much shallower at the poles where the surface water is cold. Thus calcareous oozes will mostly be found in tropical or temperate waters less than about 4 km deep, such as along the mid-ocean ridge systems and atop seamounts and plateaus. The CCD is deeper in the Atlantic than in the Pacific since the Pacific contains more CO_2, making the water more acidic and calcium carbonate more soluble. This, along with the fact that the Pacific is deeper, means that the Atlantic contains more calcareous sediment than the Pacific. All told, about 48% of the seafloor is dominated by calcareous oozes.

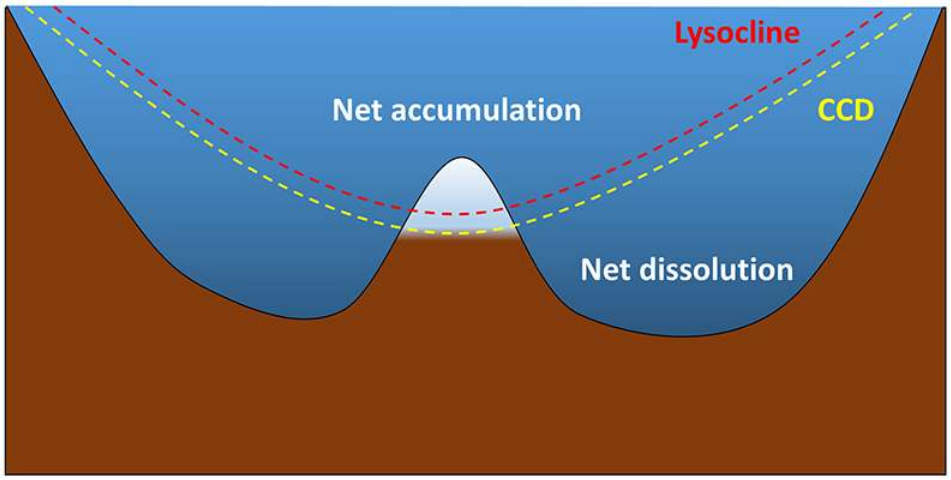


Figure 9.5.7.2 Calcareous sediment can only accumulate in depths shallower than the calcium carbonate compensation depth (CCD). Below the CCD, calcareous sediments dissolve and will not accumulate. The lysocline represents the depths where the rate of dissolution increases dramatically (PW).

Much of the rest of the deep ocean floor (about 38%) is dominated by abyssal clays. This is not so much a result of an abundance of clay formation, but rather the lack of any other types of sediment input. The clay particles are mostly of terrestrial origin, but

because they are so small they are easily dispersed by wind and currents, and can reach areas inaccessible to other sediment types. Clays dominate in the central North Pacific, for example. This area is too far from land for coarse lithogenous sediment to reach, it is not productive enough for biogenous tests to accumulate, and it is too deep for calcareous materials to reach the bottom before dissolving. Because clay particles accumulate so slowly, the clay-dominated deep ocean floor is often home to hydrogenous sediments like manganese nodules. If any other type of sediment was produced here it would accumulate much more quickly and would bury the nodules before they had a chance to grow.

9.5.7: Sediment Distribution is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

- **12.6: Sediment Distribution** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

9.6: Marine element cycles



9.6: Marine element cycles is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

9.7: The Oceans and Global Climate Change

If one thing has been constant about Earth's climate over geological time, it is its constant change. In the geological record, we can see this in the evidence of glaciations in the distant past, and we can also detect periods of extreme warmth by looking at the isotope composition of seafloor sediments. Not only has the climate changed frequently, the temperature fluctuations have been very significant. Today's mean global temperature is about 15°C. However, during its coldest periods, the global mean was as cold as -50°C, while at various times during the Paleozoic and Mesozoic and during the Paleocene–Eocene thermal maximum, it was close to 30°C.

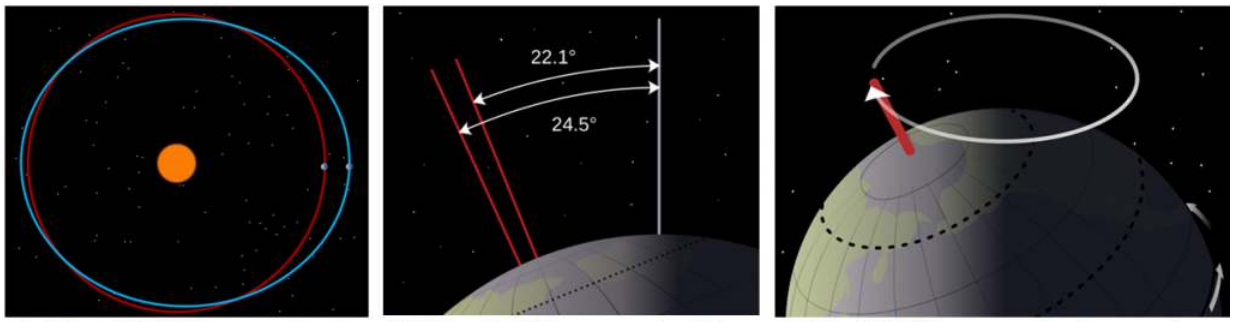
There are two parts to climate change, the first one is known as **climate forcing**, which is when conditions change to give the climate a little nudge in one direction or the other. The second part of climate change, and the one that typically does most of the work, is what we call a **feedback**. When a climate forcing changes the climate a little, a whole series of environmental changes take place, many of which either exaggerate the initial change (**positive feedback**), or suppress the change (**negative feedback**).

An example of a climate forcing mechanism is the increase in the amount of carbon dioxide (CO_2) in the atmosphere that results from our use of fossil fuels. CO_2 traps heat in the atmosphere and leads to climate warming. Warming changes vegetation patterns; contributes to the melting of snow, ice, and permafrost; causes sea level to rise; reduces the solubility of CO_2 in sea water; and has a number of other minor effects. Most of these changes contribute to more warming. Melting of permafrost, for example, is a strong positive feedback because frozen soil contains trapped organic matter that is converted to CO_2 and methane (CH_4) when the soil thaws. Both these gases accumulate in the atmosphere and add to the warming effect. On the other hand, if warming causes more vegetation growth, that vegetation should absorb CO_2, thus reducing the warming effect, which would be a negative feedback. Under our current conditions — a planet that still has lots of glacial ice and permafrost — most of the feedbacks that result from a warming climate are positive feedbacks and so the climate changes that we cause get naturally amplified by natural processes.

Natural Climate Forcing

Natural climate forcing has been going on throughout geological time. A wide range of processes has been operating at widely different time scales, from a few years to billions of years. The longest-term natural forcing variation is related to the evolution of the Sun. Like most other stars of a similar mass, our Sun is evolving. For the past 4.6 billion years, its rate of nuclear fusion has been increasing, and it is now emitting about 40% more energy (as light) than it did at the beginning of geological time. A difference of 40% is big, so it's a little surprising that the temperature on Earth has remained at a reasonable and habitable temperature for all of this time. The mechanism for that relative climate stability has been the evolution of our atmosphere from one that was dominated by CO_2, and also had significant levels of CH_4 — both greenhouse gasses — to one with only a few hundred parts per million of CO_2 and just under 1 part per million of CH_4. Those changes to our atmosphere have been no accident; over geological time, life and its metabolic processes have evolved (such as the evolution of photosynthetic bacteria that consume CO_2) and changed the atmosphere to conditions that remained cool enough to be habitable.

The position of the Earth relative to the Sun is another important component of natural climate forcing. Earth's orbit around the Sun is nearly circular, but like all physical systems, it has natural oscillations. First, the shape of the orbit changes on a regular time scale (close to 100,000 years) from being close to circular to being very slightly elliptical. But the circularity of the orbit is not what matters; it is the fact that as the orbit becomes more elliptical, the position of the Sun within that ellipse becomes less central or more **eccentric** (Figure 9.7.1a). **Eccentricity** is important because when it is high, the Earth-Sun distance varies more from season to season than it does when eccentricity is low.



(a) The 100,000 year cycle of eccentricity of the Sun within the Earth's orbit. The blue orbit is more elliptical than the red one, and the Sun is offset from the centre of the ellipse

(b) The 41,000 year cycle of the angle of tilt (obliquity) of the Earth's axis (red lines) compared to a line perpendicular to the plane of the Earth's orbit (light blue line).

(c) the 20,000 year cycle of precession of the Earth's rotational axis

Figure 9.7.1 Components of the Milankovitch cycles, which influence global climate over thousands of years (Steven Earle, "Physical Geology").

Second, Earth rotates around an axis through the North and South Poles, and that axis is at an angle to the plane of Earth's orbit around the Sun (Figure 9.7.1b). The angle of tilt (also known as **obliquity**) varies on a time scale of 41,000 years. When the angle is at its maximum (24.5°), Earth's seasonal differences are accentuated. When the angle is at its minimum (22.1°), seasonal differences are minimized. The current hypothesis is that glaciation is favored at low seasonal differences as summers would be cooler and snow would be less likely to melt and more likely to accumulate from year to year. Third, the direction in which Earth's rotational axis points also varies, on a time scale of about 20,000 years (Figure 9.7.1c). This variation, known as **precession**, means that although the North Pole is presently pointing to the star Polaris (the pole star), in 10,000 years it will point to the star Vega. The importance of eccentricity, tilt, and precession to Earth's climate cycles (now known as **Milankovitch Cycles**) was first pointed out by Yugoslavian engineer and mathematician Milutin Milankovitch in the early 1900s. Milankovitch recognized that although the variations in the orbital cycles did not affect the total amount of insolation (light energy from the Sun) that Earth received, it did affect where on Earth that energy was strongest.

Volcanic eruptions don't just involve lava flows and exploding rock fragments; various particulates and gases are also released, the important ones being sulphur dioxide and CO_2. Sulphur dioxide is an aerosol that reflects incoming solar radiation and has a net cooling effect that is short lived (a few years in most cases, as the particulates settle out of the atmosphere within a couple of years), and doesn't typically contribute to longer-term climate change. Volcanic CO_2 emissions can contribute to climate warming but only if a greater-than-average level of volcanism is sustained over a long time (at least tens of thousands of years). It is widely believed that the catastrophic end-Permian extinction (at 250 Ma) resulted from warming initiated by the eruption of the massive Siberian Traps over a period of at least a million years.

Ocean currents are important to climate, and currents also have a tendency to oscillate. Glacial ice cores show clear evidence of changes in the Gulf Stream that affected global climate on a time scale of about 1,500 years during the last glaciation. The east-west changes in sea-surface temperature and surface pressure in the equatorial Pacific Ocean, known as the El Niño Southern Oscillation or ENSO (see section 9.6) varies on a much shorter time scale of between two and seven years. These variations tend to garner the attention of the public because they have significant climate implications in many parts of the world. The strongest El Niños in recent decades were in 1983, 1998, and 2015 and those were very warm years from a global perspective. During a strong El Niño, the equatorial Pacific sea-surface temperatures are warmer than normal and heat the atmosphere above the ocean, which leads to warmer-than-average global temperatures.

Climate Feedbacks

As already stated, climate feedbacks are critically important in amplifying weak climate forcings into full-blown climate changes. Since Earth still has a very large volume of ice, mostly in the continental ice sheets of Antarctica and Greenland, but also in alpine glaciers and permafrost, melting is one of the key feedback mechanisms. Melting of ice and snow leads to several different types of feedbacks, an important one being a change in albedo, or the reflectivity of a surface. Earth's various surfaces have widely differing albedos, expressed as the percentage of light that reflects off a given material. This is important because most solar energy that hits

a very reflective surface is not absorbed and therefore does little to warm Earth. Water in the oceans or on a lake is one of the darkest surfaces, reflecting less than 10% of the incident light, while clouds and snow or ice are among the brightest surfaces, reflecting 70% to 90% of the incident light. When sea ice melts, as it has done in the Arctic Ocean at a disturbing rate over the past decade, the albedo of the area affected changes dramatically, from around 80% down to less than 10%. Much more solar energy is absorbed by the water than by the pre-existing ice, and the temperature increase is amplified. The same applies to ice and snow on land, but the difference in albedo is not as great. When ice and snow on land melt, sea level rises. (Sea level is also rising because the oceans are warming and that increases their volume; see section 13.7). A higher sea level means a larger proportion of the planet is covered with water, and since water has a lower albedo than land, more heat is absorbed and the temperature goes up a little more. Since the last glaciation, sea-level rise has been about 125 m; a huge area that used to be land is now flooded by heat-absorbent seawater. During the current period of anthropogenic climate change, sea level has risen only about 20 cm, and although that doesn't make a big change to albedo, sea-level rise is accelerating.

Most of northern Canada, Alaska, Russia, and Scandinavia has a layer of permafrost that ranges from a few centimeters to hundreds of meters in thickness. Permafrost is a mixture of soil and ice and it also contains a significant amount of trapped organic carbon that is released as CO_2 and CH_4 when the permafrost breaks down. Because the amount of carbon stored in permafrost is in the same order of magnitude as the amount released by burning fossil fuels, this is a feedback mechanism that has the potential to equal or surpass the forcing that has unleashed it. In some polar regions, including northern Canada, permafrost includes methane hydrate, a highly concentrated form of CH_4 trapped in solid form. Breakdown of permafrost releases this CH_4. Even larger reserves of methane hydrate exist on the seafloor, and while it would take significant warming of ocean water down to a depth of hundreds of meters, this too is likely to happen in the future if we don't limit our impact on the climate. There is strong isotopic evidence that the Paleocene–Eocene thermal maximum was caused, at least in part, by a massive release of sea-floor methane hydrate.

There is about 45 times as much carbon in the ocean (as dissolved bicarbonate ions, HCO_3^-) as there is in the atmosphere (as CO_2), and there is a steady exchange of carbon between the two reservoirs (see section 5.5). But the solubility of CO_2 in water decreases as the temperature goes up. In other words, the warmer it gets, the more oceanic bicarbonate that gets transferred to the atmosphere as CO_2. That makes CO_2 solubility another positive feedback mechanism. Vegetation growth responds positively to both increased temperatures and elevated CO_2 levels, and so in general, it represents a negative feedback to climate change because the more the vegetation grows, the more CO_2 is taken from the atmosphere. But it's not quite that simple, because when trees grow bigger and more vigorously, forests become darker (they have lower albedo) so they absorb more heat. Furthermore, climate warming isn't necessarily good for vegetation growth; some areas have become too hot, too dry, or even too wet to support the plant community that was growing there, and it might take centuries for something to replace it successfully. All of these positive (and negative) feedbacks work both ways. For example, during climate cooling, growth of glaciers leads to higher albedos, and formation of permafrost results in storage of carbon that would otherwise have returned quickly to the atmosphere.

Anthropogenic Climate Change

When we talk about anthropogenic climate change, we are generally thinking of the industrial era, which really got going when we started using fossil fuels (coal to begin with, and later oil and natural gas) to drive machinery and trains, and to generate electricity. That was around the middle of the 18th century. The issue with fossil fuels is that they involve burning carbon that was naturally stored in the crust over hundreds of millions of years as part of Earth's process of counteracting the warming Sun.

A rapidly rising population, the escalating level of industrialization and mechanization of our lives, and an increasing dependence on fossil fuels have driven the anthropogenic climate change of the past century. The trend of mean global temperatures since 1880 is shown in Figure 9.7.2. For approximately the past 55 years, the temperature has increased at a relatively steady and disturbingly rapid rate, especially compared to past changes. The average temperature now is approximately 0.8°C higher than before industrialization, and two-thirds of this warming has occurred since 1975.

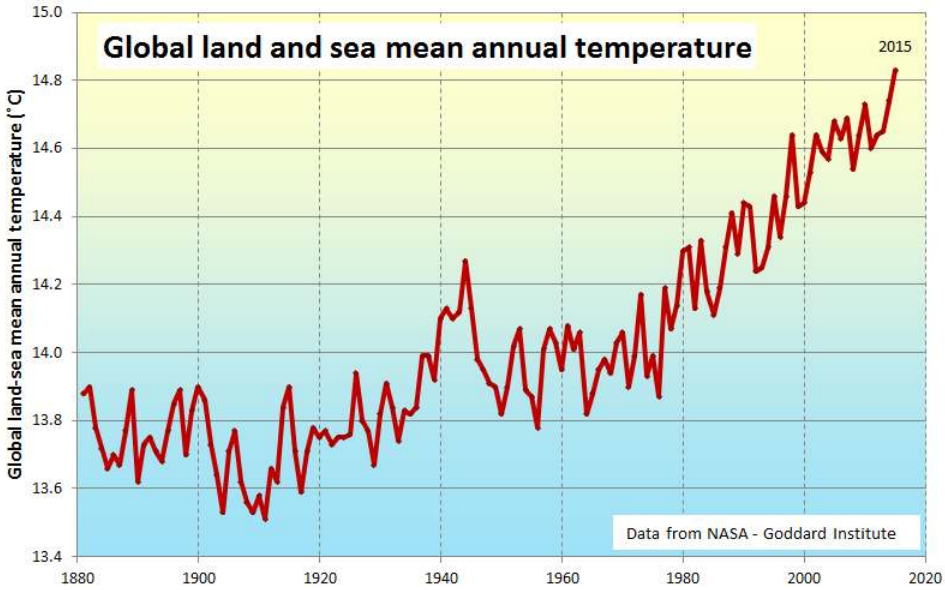


Figure 9.7.2 Global mean annual temperatures for the period from 1880 to 2015 (Steven Earle, "Physical Geology", by SE from data at NASA at: http://data.giss.nasa.gov/gistemp/ta...LB.Ts+dSST.txt).

The Intergovernmental Panel on Climate Change (IPCC), established by the United Nations in 1988, is responsible for reviewing the scientific literature on climate change and issuing periodic reports on several topics, including the scientific basis for understanding climate change, our vulnerability to observed and predicted climate changes, and what we can do to limit climate change and minimize its impacts. Figure 9.7.3, from the fifth report of the IPCC, issued in 2014, shows the relative contributions of various greenhouse gases and other factors to current climate forcing, based on the changes from levels that existed in 1750.

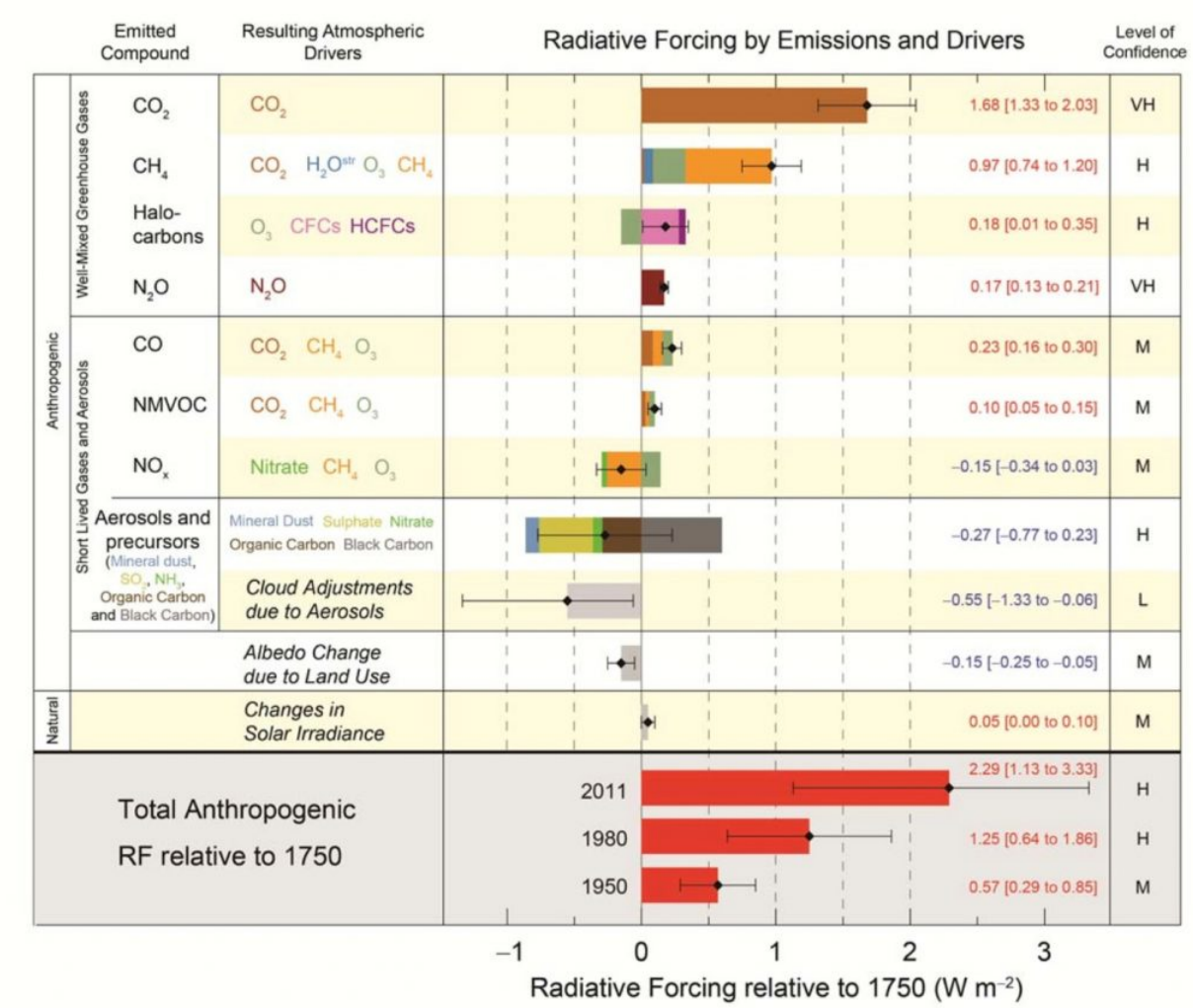


Figure 9.7.3 The relative importance of factors that are contributing to anthropogenic warming (from http://www.ipcc.ch/report/graphics/i...0-%20WG1&f=SPM).

The biggest anthropogenic contributor to warming is the emission of CO_2, which accounts for 50% of positive forcing. CH_4 and its atmospheric derivatives (CO_2, H_2O, and O_3) account for 29%, and the halocarbon gases (mostly leaked from air-conditioning appliances) and nitrous oxide (N_2O) (from burning fossils fuels) account for 5% each. Carbon monoxide (CO) (also produced by burning fossil fuels) accounts for 7%, and the volatile organic compounds other than methane (NMVOC) account for 3%. CO_2 emissions come mostly from coal- and gas-fired power stations, motorized vehicles (cars, trucks, and aircraft), and industrial operations (e.g., smelting), and indirectly from forestry. CH_4 emissions come from production of fossil fuels (escape from coal mining and from gas and oil production), livestock farming (mostly beef), landfills, and wetland rice farming. N_2O and CO come mostly from the combustion of fossil fuels. In summary, close to 70% of our current greenhouse gas emissions come from fossil fuel production and use, while most of the rest comes from agriculture and landfills. Figure 9.7.4 shows the IPCC's projections for temperature increases over the next 100 years as a result of these increasing greenhouse gases.

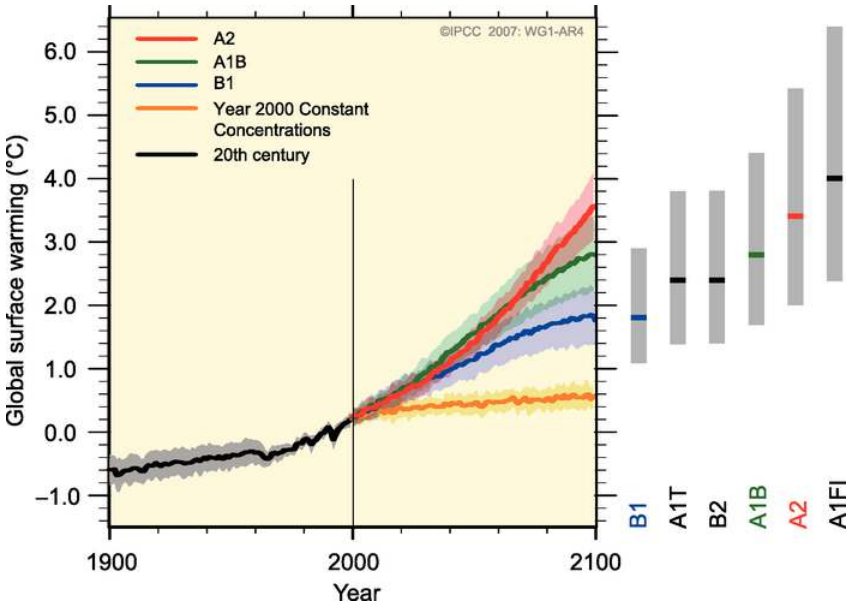


Figure 9.7.4 Projected global temperature increases for the 21st century based on a range of different IPCC scenarios of future political and technological variables (from https://www.ipcc.ch/publications_and...re-spm-5-l.png).

Impacts of Climate Change

We've all experienced the effects of climate change over the past decade. However, it's not straightforward for climatologists to make the connection between a warming climate and specific weather events, and most are justifiably reluctant to ascribe any specific event to climate change. In this respect, the best measures of climate change are those that we can detect over several decades, such as the temperature changes shown in Figure 9.7.2, or the sea level rise shown in Figure 9.7.5. As already stated, sea level has risen approximately 20 cm since 1750, and that rise is attributed to both warming (and therefore expanding) seawater and melting glaciers and other land-based snow and ice (melting of sea ice does not contribute directly to sea level rise as it is already floating in the ocean, see section 13.7).

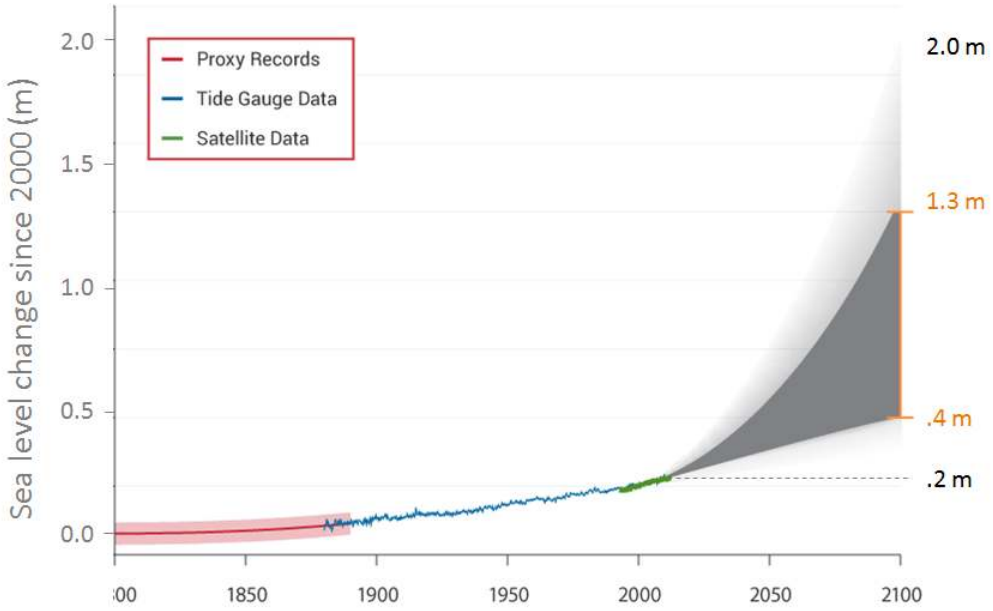


Figure 9.7.5 Projected sea-level increases to 2100, showing likely range (grey) and possible maximum (Stephen Earle, "Physical Geology", adapted by SE from: http://nca2014.globalchange.gov/repo...ntro-section-2 based on data in Parris et al., 2012, NOAA).

Projections for sea level rise to the end of this century vary widely. This is in large part because we do not know which of the above climate change scenarios (Figure 9.7.4) we will most closely follow, but many are in the range from 0.5 m to 2.0 m. One of the problems in predicting sea level rise is that we do not have a strong understanding of how large ice sheets, such as Greenland and

Antarctica, will respond to future warming. Another issue is that the oceans don't respond immediately to warming. For example, with the current amount of warming, we are already committed to a future sea level rise of between 1.3 m and 1.9 m, even if we could stop climate change today. This is because it takes decades to centuries for the existing warming of the atmosphere to be transmitted to depth within the oceans and to exert its full impact on large glaciers. Most of that committed rise would take place over the next century, but some would be delayed longer. And for every decade that the current rates of climate change continue, that number increases by another 0.3 m. In other words, if we don't make changes quickly, by the end of this century we'll be locked into 3 m of future sea level rise. In a 2008 report, the Organization for Economic Co-operation and Development (OECD) estimated that by 2070 approximately 150 million people living in coastal areas could be at risk of flooding due to the combined effects of sea level rise, increased storm intensity, and land subsidence. The assets at risk (buildings, roads, bridges, ports, etc.) are in the order of $35 trillion ($35,000,000,000,000). Countries with the greatest exposure of population to flooding are China, India, Bangladesh, Vietnam, U.S.A., Japan, and Thailand. Some of the major cities at risk include Shanghai, Guangzhou, Mumbai, Kolkata, Dhaka, Ho Chi Minh City, Tokyo, Miami, and New York.

One of the other risks for coastal populations, besides sea level rise, is that climate warming is also associated with an increase in the intensity of tropical storms (e.g., hurricanes or typhoons; see section 8.4), which almost always bring serious flooding from intense rain and storm surges. Some recent examples are New Orleans in 2005 with Hurricane Katrina, and New Jersey and New York in 2012 with Hurricane Sandy. Tropical storms get their energy from the evaporation of warm seawater in tropical regions. In the Atlantic Ocean, this takes place between 8° and 20° N in the summer. Figure 9.7.6 shows the variations in the sea-surface temperature (SST) of the tropical Atlantic Ocean (in blue) versus the amount of power represented by Atlantic hurricanes between 1950 and 2008 (in red). Not only has the overall intensity of Atlantic hurricanes increased with the warming since 1975, but the correlation between hurricanes and sea-surface temperatures is very strong over that time period.

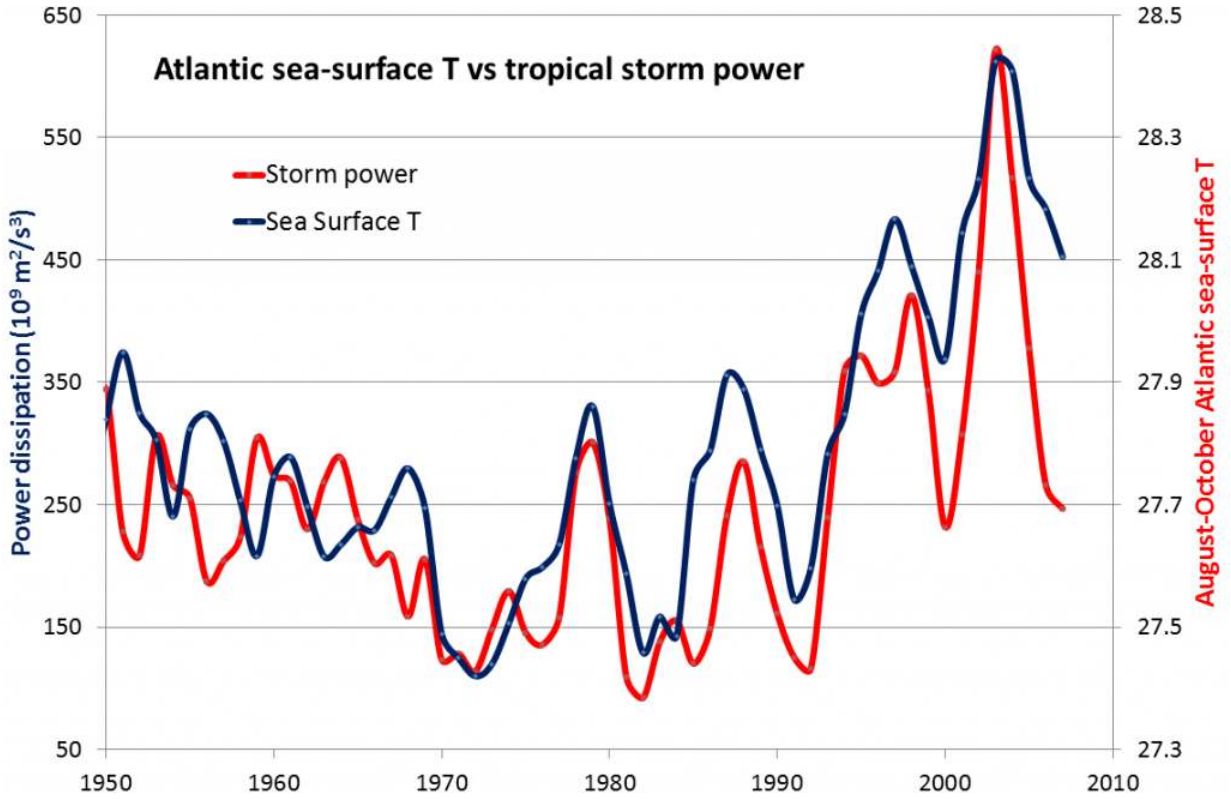


Figure 9.7.6 Relationship between Atlantic tropical storm cumulative annual intensity and Atlantic sea-surface temperatures (Steven Earle, "Physical Geology", by SE from data at: http://wind.mit.edu/~emanuel/Papers_data_graphics.htm).

The geographical ranges of diseases and pests, especially those caused or transmitted by insects, have been shown to extend toward temperate regions because of climate change. West Nile virus and Lyme disease are two examples that already directly affect North Americans, while dengue fever could be an issue in the future (dengue became a "nationally notifiable condition" in the United States in 2010). For several weeks in July and August of 2010, a massive heat wave affected western Russia, especially the area southeast of Moscow, and scientists have stated that climate change was a contributing factor. Temperatures soared to over 40°C, as

much as 12°C above normal over a wide area, and wildfires raged in many parts of the country. Over 55,000 deaths are attributed to the heat and to respiratory problems associated with the fires. A summary of the impacts of climate change on natural disasters is given in Figure 9.7.7. The major types of disasters related to climate are floods and storms, but the health implications of extreme temperatures are also becoming a great concern. In the decade 1971 to 1980, extreme temperatures were the fifth most common natural disasters; by 2001 to 2010, they were the third most common.

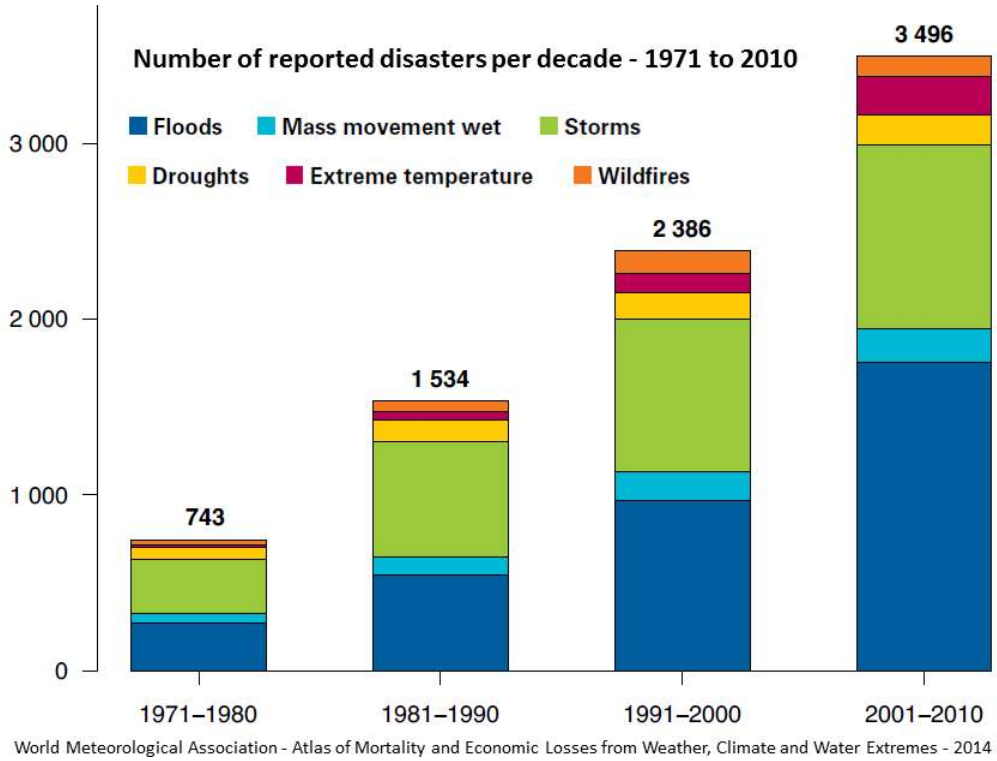


Figure 9.7.7 Numbers of various types of disasters between 1971 and 2010 (From WMO atlas of mortality and economic Losses from weather, climate and water extremes, 2014).

---

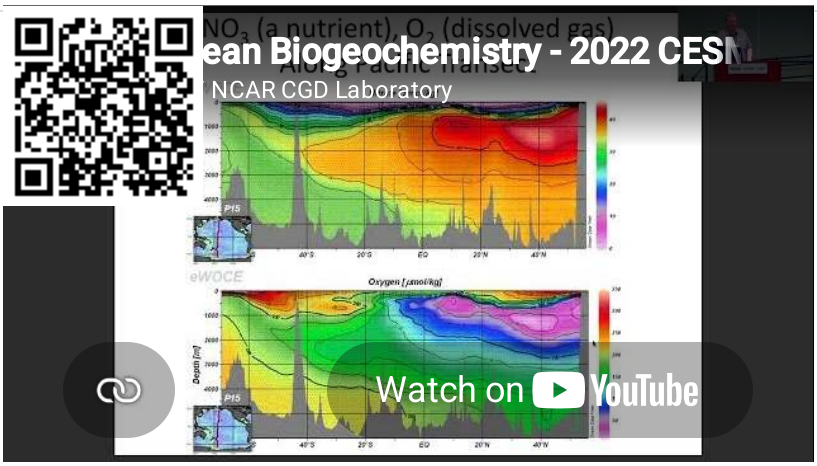
*"Physical Geology" by Steven Earle used under a CC-BY 4.0 international license. Download this book for free at http://open.bccampus.ca

---

9.7: The Oceans and Global Climate Change is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

- **8.5: Climate Change** by Paul Webb is licensed CC BY 4.0. Original source: https://rwu.pressbooks.pub/webboceanography.

9.8: Summary



NASA is thinking about the biogeochemistry of the Earth and . . . other ocean worlds'



9.8: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

10: The Global Water Cycle

10.1: Introduction
10.2: Models of the hydrologic cycle
10.3: The history of the water cycle
10.4: The water cycle and climate change
10.5: Summary

10: The Global Water Cycle is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

10.1: Introduction

The water cycle, often called the hydrologic cycl,e involves the continuous circulation of water in the Earth-Atmosphere system.

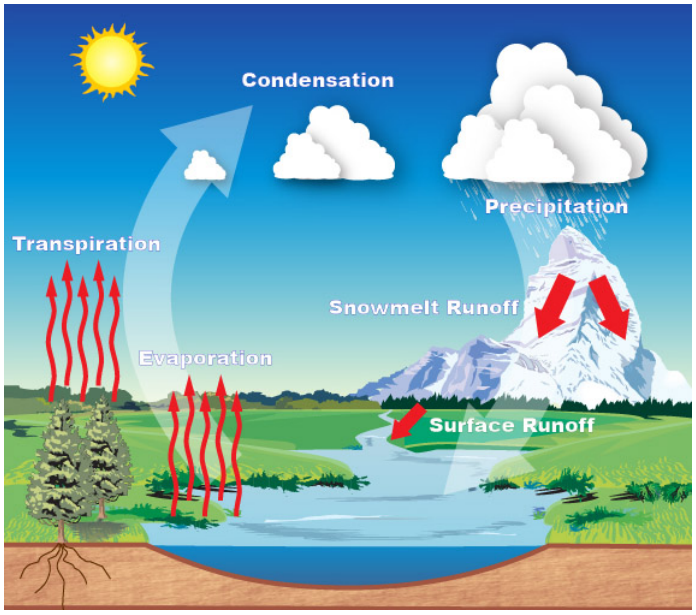


Figure 10.1.1: The water cycle

At its core, the water cycle is the motion of the water from the ground to the atmosphere and back again. Of the many processes involved in the hydrologic cycle, the most important are:

- evaporation
- transpiration
- condensation
- precipitation
- runoff

Evaporation

Evaporation is the change of state in a substance from a liquid to a gas. In meteorology, the substance we are concerned about the most is water.

For evaporation to take place, energy is required. The energy can come from any source: the sun, the atmosphere, the Earth, or objects on the Earth such as humans.

Everyone has experienced evaporation personally. When the body heats up due to the air temperature or through exercise, the body sweats, secreting water onto the skin. The purpose is to cause the body to use its own heat to evaporate the liquid, thereby removing heat and cooling the body. The same effect can be seen when you step out of a shower or swimming pool. The coolness you feel is the removal of body heat through evaporation of the water on your skin.

Transpiration

Transpiration is the evaporation of water from plants through stomata. Stomata are small openings found on the underside of leaves that are connected to vascular plant tissues. In most plants, transpiration is a passive process largely controlled by the humidity of the atmosphere and the moisture content of the soil. Of the transpired water passing through a plant only 1% is used in the growth process of the plant. The remaining 99% is passed into the atmosphere.

Condensation

Condensation is the process whereby water vapor is changed into a liquid state. In the atmosphere, condensation may appear as clouds or dew. This is also the process whereby water appears on the side of an uninsulated cold drink can or bottle.

Condensation is not a matter of one particular temperature but of a difference between two temperatures: the air temperature and the dewpoint temperature. The dew point is the temperature at which dew can form - it is the point at which air becomes saturated and can not hold any more water vapor. Any additional cooling causes water vapor to condense. Foggy conditions often occur when air temperature and dew point are equal.

Condensation is the opposite of evaporation. Since water vapor has a higher energy level than that of liquid water, when condensation occurs, the excess energy in the form of heat energy is released. This release of heat aids in the formation of hurricanes.

Precipitation

Precipitation results when tiny condensation particles, through collision and coalescence, grow too large for the rising air to support, and thus fall to the Earth. Precipitation can be in the form of rain, hail, snow, or sleet.

Precipitation is the primary way we receive fresh water on Earth. On average, the world receives about 38½" (980 mm) each year over both the oceans and land masses.

Runoff



Figure 10.1.2: Runoff

Runoff occurs when there is excessive precipitation and the ground is saturated (cannot absorb any more water). Rivers and lakes are results of runoff. Some runoff evaporates into the atmosphere, but most water in rivers and lakes returns to the oceans.

If runoff water flows into a lake only with no outlet for water to flow out, then evaporation is the only means for water to return to the atmosphere. As water evaporates, impurities or salts are left behind. As a result, the lake becomes salty, as in the case of the Great Salt Lake in Utah or Dead Sea in Israel.

Evaporation of this runoff into the atmosphere begins the hydrologic cycle over again. Some of the water percolates into the soil and into the ground water only to be drawn into plants again for transpiration to take place.

Excerpted from

The Hydrologic Cycle, NOAA, https://www.noaa.gov/jetstream/atmos...,transpiration Accessed December 2023

10.1: Introduction is shared under a Public Domain license and was authored, remixed, and/or curated by NOAA https://www.noaa.gov/jetstream/atmosphere/hydro#:~:text=The%20hydrologic%20cycle%20involves%20the,transpiration.

10.2: Models of the hydrologic cycle

Key Points

- Water cycling affects the climate, transports minerals, purifies water, and replenishes the land with fresh water.
- Water with a longer residence time, such as water in oceans and glaciers, is not available for short-term cycling, which occurs via evaporation.
- Surface water evaporates (water to water vapor) or sublimates (ice to water vapor), which deposits large amounts of water vapor into the atmosphere.
- Water vapor in the atmosphere condenses into clouds and is eventually followed by precipitation, which returns water to the earth's surface.
- Rain percolates into the ground, where it may evaporate or enter bodies of water.
- Surface runoff enters oceans directly or via streams and lakes.

The Water (Hydrologic) Cycle

Water is essential for all living processes. The human body is more than one-half water and human cells are more than 70 percent water. Thus, most land animals need a supply of fresh water to survive. However, when examining the stores of water on earth, 97.5 percent of it is non-potable salt water (Figure 10.2.1 – 2). Of the remaining water, 99 percent is locked underground as water or as ice but this water is inconveniently located, mostly in Antarctica and Greenland. Shallow groundwater is the largest reservoir of usable fresh water. Less than one percent of fresh water is present in lakes and rivers, the most heavily used water resources. If all of world's water was shrunk to the size of 1 gallon, then the total amount of fresh water would be about 1/3 cup, and the amount of readily usable fresh water would be 2 tablespoons.

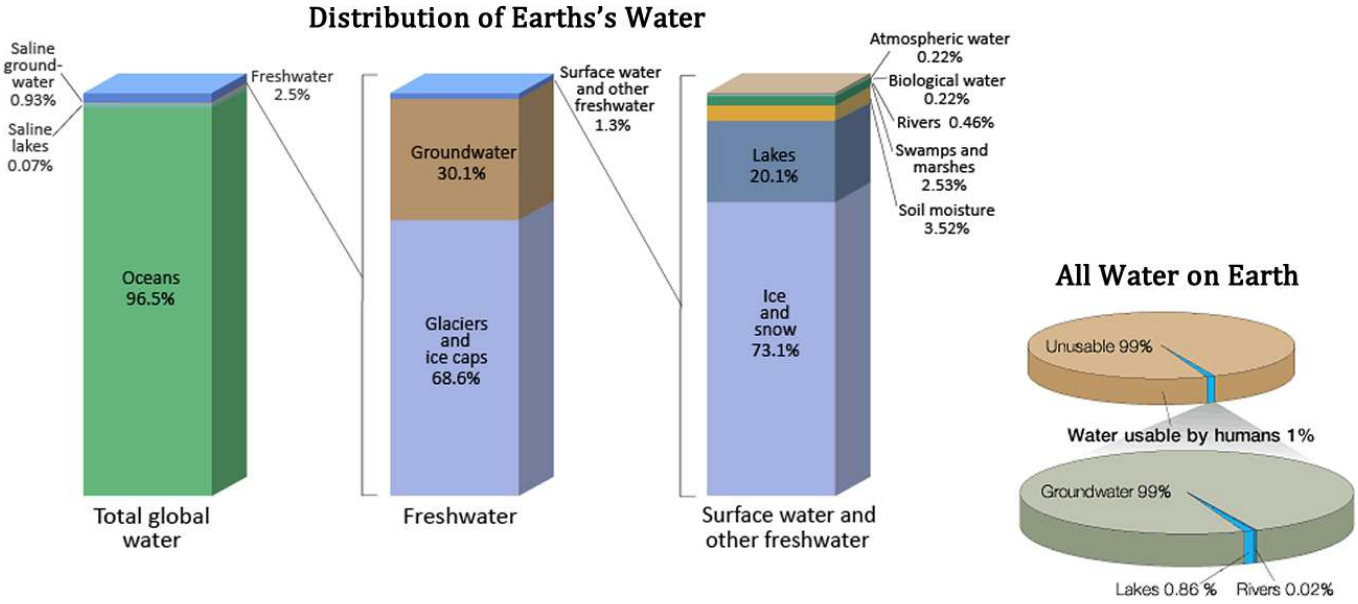


Figure 10.2.1**: Earth's Water Reservoirs** Bar chart Distribution of Earth's water including total global water, fresh water, and surface water and other fresh water and Pie chart Water usable by humans and sources of usable water. Only 2.5 percent of water on Earth is fresh water, and less than 1 percent of fresh water is easily accessible to living things. *Source: United States Geographical Survey Igor Skiklomanov's chapter "World fresh water resources" in Peter H. Gleick (editor), 1993, Water in Crisis: A Guide to the World's Fresh Water Resources*

Many living things, such as plants, animals, and fungi, are dependent on the small amount of fresh surface water supply, a lack of which can have massive effects on ecosystem dynamics. Humans, of course, have developed technologies to increase water availability, such as digging wells to harvest groundwater, storing rainwater, and using desalination to obtain drinkable water from the ocean. Although this pursuit of drinkable water has been ongoing throughout human history, the supply of fresh water is still a major issue in modern times.

Water is the only common substance that occurs naturally on Earth in three forms: solid, liquid and gas. The **hydrosphere** is the area of Earth where water movement and water storage occurs. **Water reservoirs** are the locations where water is stored. (Note that this term can also refer to artificial lakes created by dams.) Water is found as a liquid on the surface (rivers, lakes, oceans) and beneath the surface (groundwater), as ice (polar ice caps and glaciers), and as water vapor in the atmosphere. Figure 10.2.2 illustrates the average time that an individual water molecule may spend in the Earth's major water reservoirs. **Residence time** is a measure of the average time an individual water molecule stays in a particular reservoir.

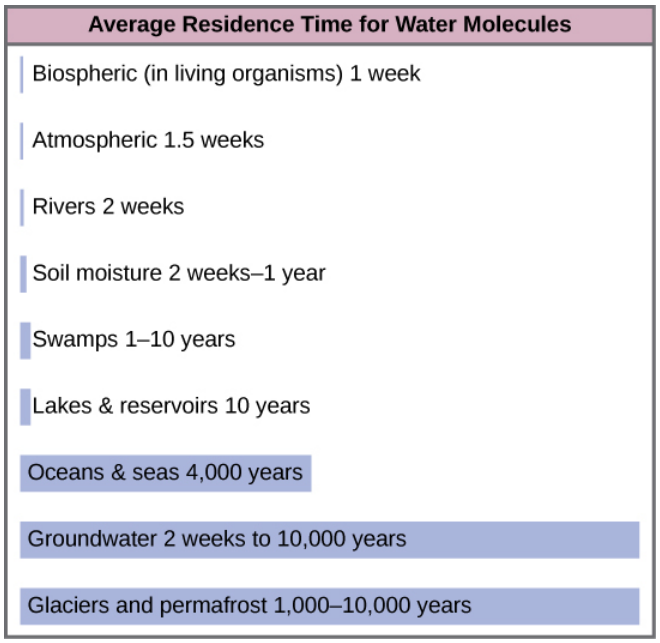


Figure 10.2.2**:** Average residence time that water remains in each reservoir. Water remains in organisms for about one week, in the atmosphere for 1.5 weeks, in rivers for two weeks, as soil moisture from two weeks to a year, in swamps for 1-10 years, in lakes for 10 years, in oceans and seas for 4,000 years, as groundwater for 2 weeks to 10,000 years, and in glaciers or as permafrost for 1,000-10,000 years. Image from OpenStax (CC-BY).

The Water Cycle

Water cycling is extremely important to ecosystem dynamics as it has a major influence on climate and, thus, on the environments of ecosystems. For example, when water evaporates, it takes up energy from its surroundings, cooling the environment. When it condenses, it releases energy, warming the environment. The evaporation phase of the cycle purifies water, which then replenishes the land with fresh water. The flow of liquid water and ice transports minerals across the globe. It is also involved in reshaping the geological features of the earth through processes including erosion and sedimentation. The various processes that occur during the cycling of water are illustrated in Figure 10.2.4. The processes include the following:

- **evaporation** and **sublimation**
- **transpiration**
- **condensation** and **precipitation**
- subsurface water flow
- surface **runoff** and snowmelt
- streamflow

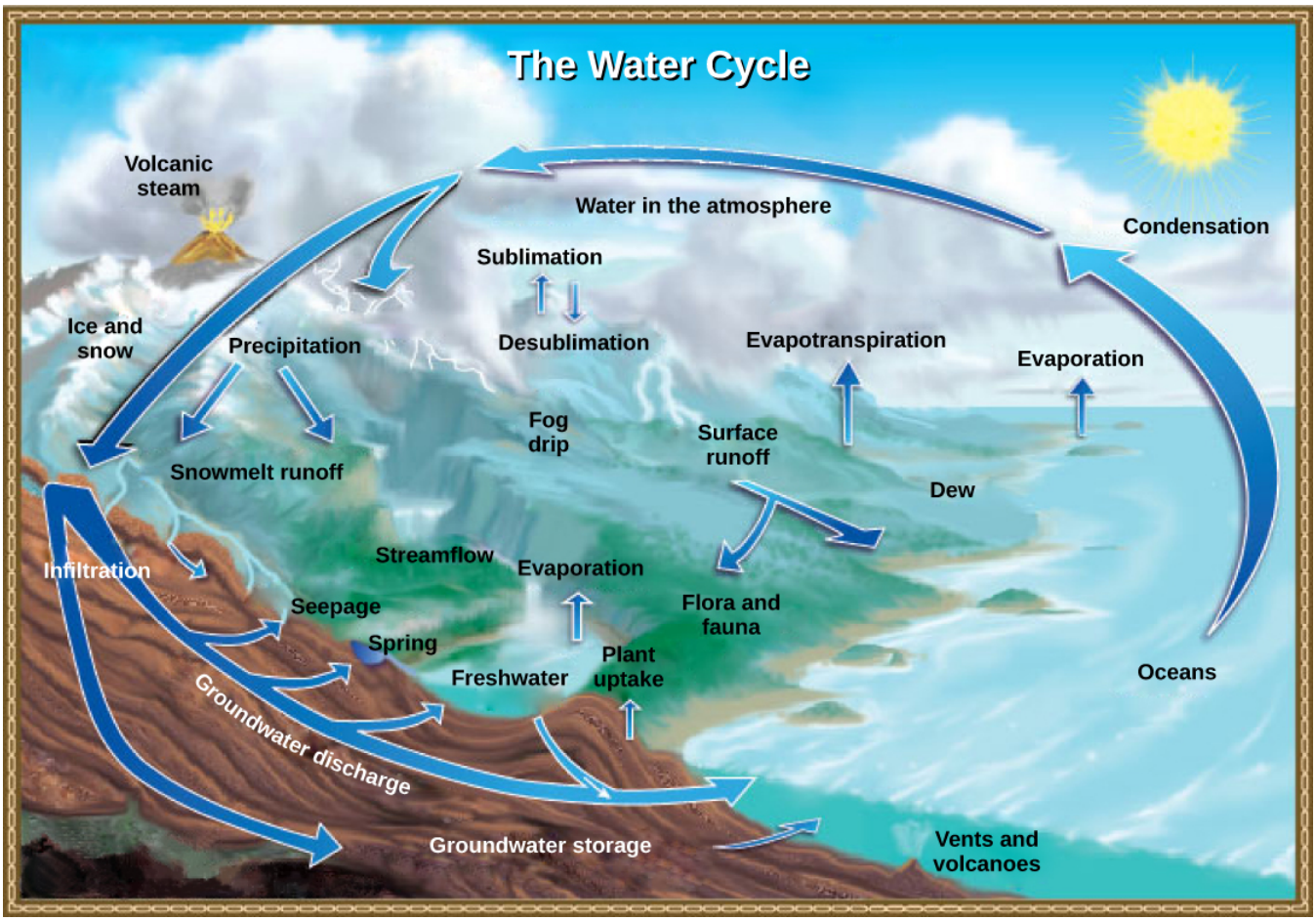


Figure 10.2.4: Water from the land and oceans enters the atmosphere by evaporation or sublimation, where it condenses into clouds and falls as rain or snow. Precipitated water may enter freshwater bodies or infiltrate the soil. The cycle is complete when surface or groundwater reenters the ocean. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

The water cycle is driven by the Sun's energy as it warms the oceans and other surface waters. This leads to evaporation (water to water vapor) of liquid surface water and sublimation (ice to water vapor) of frozen water, thus moving large amounts of water into the atmosphere as water vapor. As the water vapor rises in the atmosphere, it cools and condenses. **Condensation** is the process in which water vapor changes to tiny droplets of liquid water. Over time, this water vapor condenses into clouds as liquid or frozen droplets and eventually leads to precipitation (rain or snow), which returns water to Earth's surface. Most precipitation falls into the ocean. Some frozen precipitation becomes part of ice caps and glaciers. These masses of ice can store frozen water for hundreds of years or longer. Rain reaching Earth's surface may evaporate again, flow over the surface, or percolate into the ground. Most easily observed is surface runoff: the flow of fresh water either from rain or melting ice. Runoff can make its way through streams and lakes to the oceans or flow directly to the oceans themselves.

In most natural terrestrial environments rain encounters vegetation before it reaches the soil surface. A significant percentage of water evaporates immediately from the surfaces of plants. What is left reaches the soil and begins to move down. Surface runoff will occur only if the soil becomes saturated with water in a heavy rainfall. **Infiltration** is the process through which water sinks into the ground and is determined by the soil or rock type through which water moves. Most water in the soil will be taken up by plant roots. The plant will use some of this water for its own metabolism, and some of that will find its way into animals that eat the plants, but much of it will be lost back to the atmosphere through a process known as **evapotranspiration**. Water enters the vascular system of the plant through the roots and evaporates, or transpires, through the stomata of the leaves. Water in the soil that is not taken up by a plant and that does not evaporate is able to percolate into the subsoil and bedrock. Here it forms groundwater.

Groundwater is a significant reservoir of fresh water. It exists in the pores between particles in sand and gravel, or in the fissures in rocks. Shallow groundwater flows slowly through these pores and fissures and eventually finds its way to a stream or lake where it becomes a part of the surface water again. Streams do not flow because they are replenished from rainwater directly; they flow

because there is a constant inflow from groundwater below. Some groundwater is found very deep in the bedrock and can persist there for millennia. Most groundwater reservoirs, or **aquifers**, are the source of drinking or irrigation water drawn up through wells. In many cases these aquifers are being depleted faster than they are being replenished by water percolating down from above.

Rain and surface runoff are major ways in which minerals, including carbon, nitrogen, phosphorus, and sulfur, are cycled from land to water. More precipitation falls near the equator, and landmasses there are characterized by a tropical rainforest climate (Figure 10.2.5). Less precipitation tends to fall near 20–30° north and south latitude, where the world's largest deserts are located. These rainfall and climate patterns are related to global wind circulation cells. The intense sunlight at the equator heats air, causing it to rise and cool, which decreases the ability of the air mass to hold water vapor and results in frequent rainstorms. Around 30° north and south latitude, descending air conditions produce warmer air, which increases its ability to hold water vapor and results in dry conditions. Both the dry air conditions and the warm temperatures of these latitude belts favor evaporation. Global precipitation and climate patterns are also affected by the size of continents, major ocean currents, and mountains.

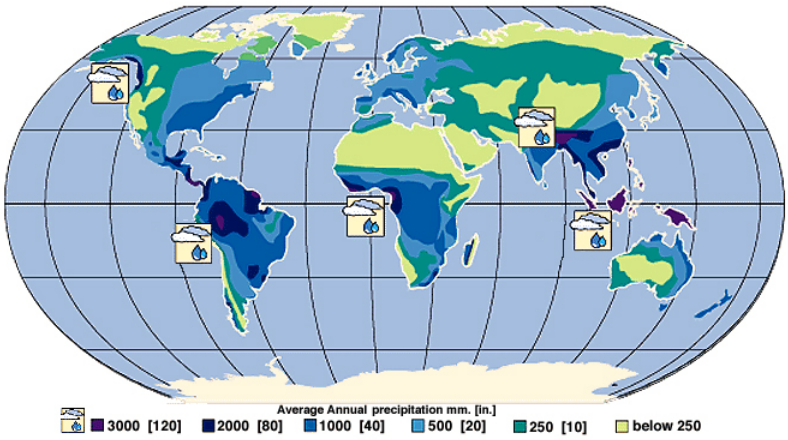
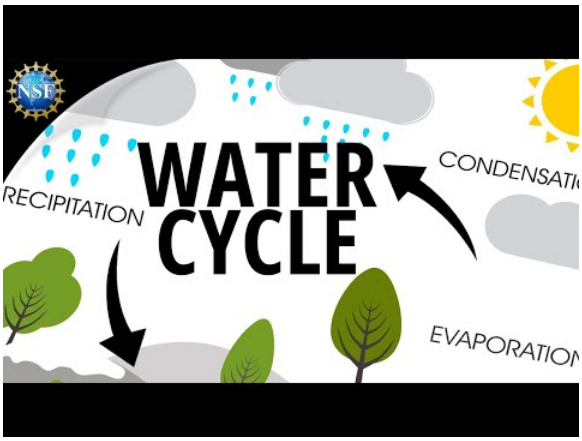


Figure 10.2.5**: World Rainfall Map** The false-color map above shows the amount of rain that falls around the world. Areas of high rainfall include Central and South America, western Africa, and Southeast Asia. Since these areas receive so much rainfall, they are where most of the world's rainforests grow. Areas with very little rainfall usually turn into deserts. The desert areas include North Africa, the Middle East, western North America, and Central Asia. *Source: United States Geological Survey Earth Forum, Houston Museum Natural Science.*

An important part of the water cycle is how water varies in salinity, which is the abundance of dissolved ions in water. The saltwater in the oceans is highly saline, with about 35,000 mg of dissolved ions per liter of seawater. Evaporation is a distillation process that produces nearly pure water with almost no dissolved ions. As water vaporizes, it leaves the dissolved ions in the original liquid phase. Eventually, condensation forms clouds and sometimes precipitation. After rainwater falls onto land, it dissolves minerals in rock and soil, which increases its salinity. Rain and surface runoff are major ways in which minerals, including phosphorus and sulfur, are cycled from land to water. Freshwater (such as lakes, rivers, and near-surface groundwater) has a relatively low salinity.

The steps of the water cycle are also explained in the video below.





Human Interactions with The Water Cycle

Humans alter the water cycle by extracting large amounts of freshwater from surface waters as well as groundwater. Freshwater supply is one of the most important provisioning ecosystem services on which human well-being depends. By 2000, the rate of our water extraction from rivers and aquifers had risen to almost 4000 cubic kilometers per year. The greatest use of this water is for irrigation in agriculture, but significant quantities of water are also extracted for public and municipal use, as well as industrial applications and power generation.

Other major human interventions in the water cycle involve changes in land cover and infrastructure development of river networks. As we have deforested areas for wood supply and agricultural development we have reduced the amount of vegetation, which naturally acts to trap precipitation as it falls and slow the rate of infiltration into the ground. As a consequence, surface runoff has increased. This, in turn, means flood peaks are greater and erosion is increased. Erosion lowers soil quality and deposits sediment in river channels, where it can block navigation and harm aquatic plants and animals. Where agricultural land is also drained these effects can be magnified. Urbanization also accelerates streamflow by preventing precipitation from filtering into the soil and shunting it into drainage systems.

Additional physical infrastructure has been added to river networks with the aim of altering the volume, timing, and direction of water flows for human benefit. This is achieved with reservoirs, weirs, aqueducts and diversion channels (figure 10.2.6). For example, so much water is removed or redirected from the Colorado River in the western United States that, despite its

considerable size, in some years it is dry before reaching the sea in Mexico. In an extreme example, the Aral Sea in Central Asia has decreased to only 10% of its initial size after water was diverted for agriculture (see this case study for more details).



Figure 10.2.6**:** The California Aqueduct carries water needed for agriculture from Northern California to Southern California. Image by USGS (public domain).

We also exploit waterways through their use for navigation, recreation, hydroelectricity generation and waste disposal. These activities, especially waste disposal, do not necessarily involve removal of water, but do have impacts on water quality and water flow that have negative consequences for the physical and biological properties of aquatic ecosystems.

The water cycle is key to the ecosystem service of climate regulation as well as being an essential supporting service that impacts the function of all ecosystems. Consider the widespread impacts on diverse natural and human systems when major droughts or floods occur. Consequently, human disruptions of the natural water cycle have many undesirable effects and challenge sustainable development. There are two major concerns. First, the need to balance rising human demand with the need to make our water use sustainable by reversing ecosystem damage from excess removal and pollution of water. Traditionally, considerable emphasis has been on finding and accessing more supply, but the negative environmental impacts of this approach are now appreciated, and improving the efficiency of water use is now a major goal. Second, there is a need for a safe water supply in many parts of the world, which depends on reducing water pollution and improving water treatment facilities.

Although **glaciers** represent the largest reservoir of fresh water, they generally are not used as a water source because they are located too far from most people (Figure 10.2.7). Melting glaciers do provide a natural source of river water and groundwater. During the last Ice Age there was as much as 50% more water in glaciers than there is today, which caused sea level to be about 100 m lower. Over the past century, sea level has been rising in part due to melting glaciers. If Earth's climate continues to warm, the melting glaciers will cause an additional rise in sea level.



Figure 10.2.7: *Figure 7. Mountain Glacier in Argentina Glaciers are the largest reservoir of fresh water but they are not used much as a water resource directly by society because of their distance from most people. Source: Luca Galuzzi – www.galuzzi.it*

Further "Reading"

For more information on the water cycle you might want to watch this water cycle video from USGS.

Contributors and Attributions

Modified by Kyle Whittinghill and Melissa Ha from the following sources

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd).

  20.2: Biogeochemical Cycles by OpenStax, is licensed CC BY from *Biology 2e* by OpenStax (CC-BY). Access for free at openstax.org.
- **6.6: Water Cycle** by CK-12: Biology Concepts, is licensed CC BY-NC
- Biogeochemical Cycles and the Flow of Energy in the Earth System and Water Cycle and Fresh Water Supply from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (licensed under CC-BY). Download for free at CNX.
- Water Cycle and Fresh Water Supply, Water Supply Problems and Solutions, and Biogeochemical Cycles from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- The Water Cycle, Water Use and Distribution, and Groundwater from *An Introduction to Geology* by Chris Johnson et al. (licensed under CC-BY-NC-SA)
- General Microbiology **Provided by**: Boundless.com. **License**: ***CC BY-SA: Attribution-ShareAlike***

10.2: Models of the hydrologic cycle is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **20.2: Biogeochemical Cycles** by OpenStax is licensed CC BY 4.0.

10.3: The history of the water cycle

Hydrology has a long history dating back to several millennia (Biswas 1970). However, the birth of hydrologic modeling can be traced to the 1850s when Mulvany (1850) developed a method for computing the time of concentration and hence the rational method for computing peak discharge which is still used for urban drainage design, Darcy (1856) who conducted experiments on flow-through sands and developed what is now referred to as Darcy's law which laid the foundation of quantitative groundwater hydrology, and Fick's first law which states that under steady-state conditions the diffusive flux is proportional to the concentration gradient (spatial) which laid the foundation of water quality hydrology. About half a century earlier, Dalton (1802) formulated the law of evaporation which states that the rate of evaporation is directly proportional to the difference between saturation vapor pressure at the water surface and the actual vapor pressure in the air. This law constituted the foundation for developing the physics of evaporation. For a period of over a century until the 1960s, many groundbreaking advances in modeling different components of the hydrologic cycle were made. Some of these advances were based on the laws of mathematical physics and some had their basis in laboratory and/or field experiments. The current state of hydrologic science and engineering owes a great deal to the pre-1960 advances. The handbook of applied hydrology edited by Chow (1964) provided an up-to-date account of hydrologic advances until the 1960s, whereas the handbook of hydrology edited by Maidment (1993) and the encyclopedia of hydrology and water resources edited by Hershey and Fairbridge (1998) dealt with advances that occurred during the intervening period. Singh and Woolhiser (2002) provided a historical account of developments that occurred in modeling different components of the hydrologic cycle.

The decade of the 1960s witnessed the birth of computer revolution and hydrologic modeling took a giant leap forward. The computer provided the power for doing computations that was not available before. As a result, a new branch of hydrology, called digital or numerical hydrology, was born. Another branch that came into being was statistical or stochastic hydrology that often required analyses of large volumes of data. Then, several major advances ensued. First, simulation of the entire hydrologic cycle became a reality, as illustrated by the development of the Stanford Watershed Model (Crawford and Linsley 1966) which was followed in the decades to come by umpteen watershed models that were developed all over the world (Singh 1995; Singh and Frevert 2002a, b, 2006). Second, optimization or operations research techniques were developed, which formed the basis for reservoir management and operation as well as river basin simulation. Some of these techniques were also used for calibrating hydrologic models (Beven 2001; Duan et al. 2003). Third, two- and three-dimensional modeling was made possible because of advances in numerical mathematics. Consequently, two- and three-dimensional models of groundwater as well as of infiltration and soil water flow were developed (Bear 1979; Pinder and Celia 2006; Remson et al. 1971). Fourth, simultaneous simulation of water flow and sediment and pollutant transport was undertaken; likewise, simultaneous simulation of different phases of flow, such as liquid and gaseous, was done (Bear and Verruijt 1987; Charbeneau 2000). Fifth, modeling at large spatial scales, such as a large river basin like the Mississippi, and that at small temporal scales, such as seconds or minutes, was undertaken (Molley and Wesse 2009; Sorooshian et al. 2008). Sixth, integration of hydrology with allied sciences became possible. For example, it was possible to couple hydrology with climatology for precipitation modeling and forecasting (Sorooshian et al. 2008), with geomorphology for river basin geometric representation (Baker et al. 1988; Bates and Lane 2002; Beven and Kirkby 1993), with hydraulics for describing flow characteristics (Singh 1996), with soil physics for quantifying soil texture and structure (Bohne 2005; Guymon 1994; Miyazaki 2006; Smith et al. 2002; Singh 1997), and with geology for aquifer characterization (Delleur 1999; Fetter 1980; Singh 2017a, b, c). The coupling of hydrology with ecosystems gave rise to ecohydrology (Eagleson 2002; Gordon et al. 2006; Rodriguez-Iturbe and Porporato 2004). Climate change and global warming became part of hydrologic analysis (Arnell 1997). A more detailed account of developments in different components of the hydrologic cycle is given in Singh (2013, 2014, 2015, 2017a).

In the decades that followed, computing prowess increased exponentially and hydrology began maturing and expanding in both depth (vertically) and breadth (horizontally). Tools from fluid mechanics, statistics, information theory, and mathematics were employed and became part of hydrology (Bras and Rodriguez-Iturbe 1985; Clarke 1998; Gelhar 1993; Mays and Tung 1992; Singh et al. 2007; Tung and Yen 2005). Further, computer also made possible the development of user-friendly software, and tools for date acquisition, storage, retrieval, processing, and dissemination (Croley 1980; Hoggan 1989). Remote sensing tools, such as radar and satellites, came into being that made possible to acquire spatial data for large areas (Engman and Gurney 1991; Hogg et al. 2017; Lakshmi 2017; Lakshmi et al. 2015). Likewise, geographical information systems (GIS) were developed for processing huge quantities of raster and vector data (Maidment 2002). The past two decades witnessed the development of artificial neural networks, fuzzy logic, genetic programming, and wavelet models (Kumar et al. 2006; Ross 2010; Sen 2010; Tayfur 2012). New

theories borrowed from other areas were introduced in hydrology. Examples of these theories are entropy theory (Singh 2013, 2014, 2015, 2016, 2017b), copula theory (Singh and Zhang 2018), chaos theory (Sivakumar 2017), network theory (Sivakumar et al. 2017), and catastrophe theory (Poston and Stewart 1978; Zeeman 1978). These theories will find increasing place in hydrologic modeling in the years ahead.

Another area that mushroomed subsequent to the pre-computer era is instrumentation. New instruments which were more accurate and sophisticated were developed for measuring all kinds of hydrologic variables, such as velocity, soil moisture, water and air quality parameters, fluxes in porous media, energy fluxes, and so on. Further, instrumentation for data transmission from place of measurement to place of storage, processing, storage, retrieval, and dissemination became highly robust and accessible (Liang et al. 2013; Sivakumar and Berndtsson 2010).

Source

Singh, V.P. Hydrologic modeling: progress and future directions. *Geosci. Lett.* **5**, 15 (2018). https://doi.org/10.1186/s40562-018-0113-z

10.3: The history of the water cycle is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

10.4: The water cycle and climate change

Changes to the patterns of current precipitation and those projected for the future on account of climate change are discussed in Chapter 6, "Future Geographies: Global Precipitation Patterns". The kind of precipitation is already changing, as northern regions are experiencing more rain than snow. Total snow cover has on average decreased in both hemispheres.The variability of precipitation is changing as well. Many subtropical regions have been or are gripped in devastating droughts. Precipitation is but one element of the hydrologic cycle, that will be affected by climate. The impact of climate change on water resources will have profound effects on ecosystems and humans.

Evapotranspiration

As noted earlier in this chapter, potential evapotranspiration is a measure of water need based primarily on the energy input. Potential evapotranspiration is expected to increase in nearly all regions. As temperature rises the saturation point of air increases, but relative humidity is not expected to change significantly. As a result, the evaporation of water increases yielding a larger amount of water vapor in the air. The increase evaporation of water from the soil places a greater strain on plants as their need for water to maintain growth processes increases. Actual evaporation is projected to increase over most bodies of water, the amount depending on the geographic variation in surface warming.

Plants take in carbon dioxide and release water vapor through their stomata. The increase in atmospheric carbon dioxide has competing roles in the affect of evapotranspiration from vegetation. At higher CO_2 concentrations, stomata need to open less to take in the same amount of carbon dioxide for photosynthesis. However, increased carbon dioxide can act a gaseous fertilizer to spur plant growth and leaf area. Increased leaf area results in more transpiration. The degree to which higher CO_2 concentration have depend on the type of vegetation, nutrient availability, temperature changes, and water availability

Surface water and Runoff

The impact of climate change on river flows and lake levels resulting depend on changes in the amount, type, and timing of precipitation. Changes in evaporation rates also affect river and lake levels. Global scale climate and hydrological models project runoff increases in the wet tropics and high latitudes, and decreases in the mid-latitudes and some parts of the dry tropics. Most studies show that rising temperatures lead to changes in the seasonality of river flow due to changes in the type of precipitation. Higher winter flows occurring in the European Alps, Scandinavia, the Himalayas, western, central, and eastern North America will result from more precipitation falling as rain and earlier snow melt. The frequency of heavy precipitation events are expected to increase in an ever warming world for most regions. This will likely increase the threat for flash flooding and urban flooding.

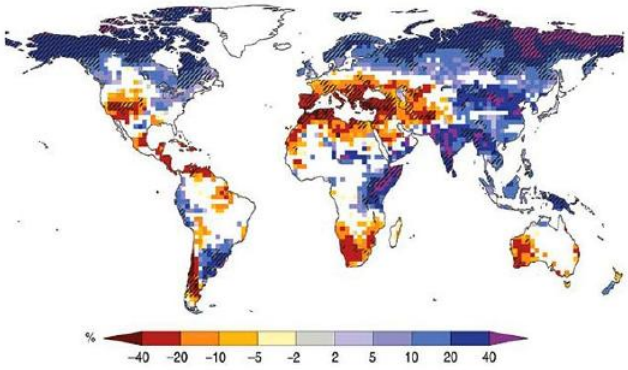


Figure 10.4.1: Runoff model projections of large-scale relative changes in annual runoff for the period 2090-2099 relative to 1980-1999. (Courtesy IPCC)

Ground water

Ground water levels are maintained when aquifers are adequately recharged. How this occurs varies with soil conditions, vegetation and precipitation intensity and timing. Generally speaking, if precipitation intensity is greater than the infiltration capacity of the soil, most water will runoff and not percolate down into the ground water zone. However, more frequent and larger precipitation events projected for some humid region may result in decreased groundwater recharge as the infiltration capacity of

the soil is exceeded more often. In semi-arid and arid regions, groundwater level may rise with high-intensity storms as rainfall is ability to infiltrate downward before evaporating.

Soil moisture

The amount of moisture held in the soil depends on when and how much precipitation and evaporation occurs at a place. The resulting changes to the spatial patterns of soil moisture is expected to be quite similar to that of changes in precipitation. Though the amount of change is not well established, projections generally agree on the kind of change that will occur. A reduction in soil moisture in the subtropics and Mediterranean regions will accompany a decrease in precipitation. Projected decreases in snow cover at high latitudes will also result in less soil moisture. Soil moisture will likely increase in East Africa and central Asia where increases precipitation is expected to occur.

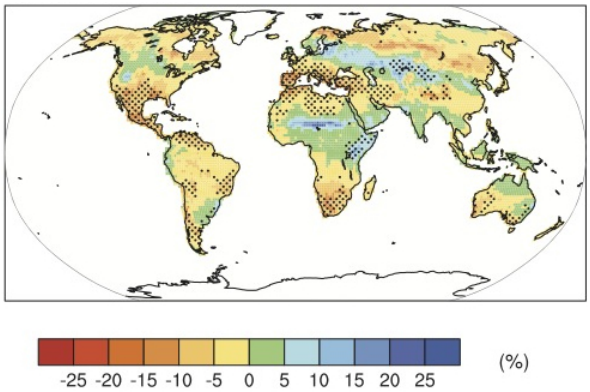


Figure 10.4.2: Fifteen-model mean changes in soil moisture. Stippled areas are those where 80% of models agree on sign of change. (Source: IPCC)

Water supply

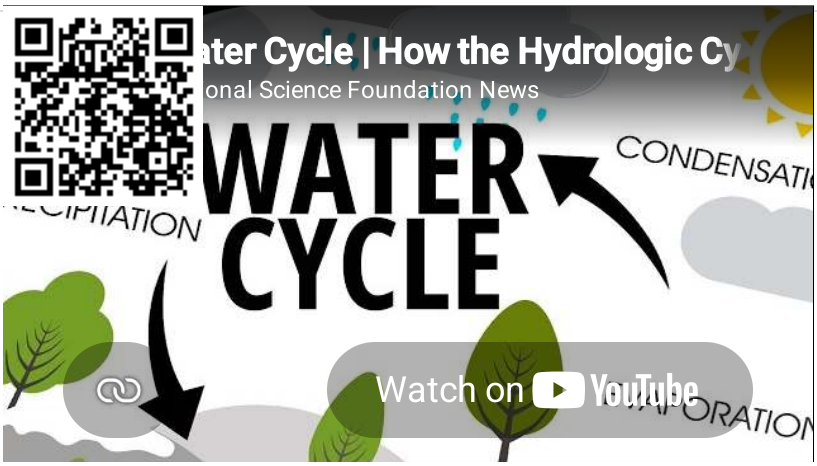
Population growth, economic development and land use change largely determine the stress put on water supplies. Climate change will be an additional factor that affects stressed water resources. Glacier meltwater provides a source of water for many living in mountainous underdeveloped countries. Recent loss of glacier ice has or will place these people at risk. Peruvian glaciers have lost more than 20% of their mass in the past 35 years, resulting in a 12 percent decrease in runoff to the country's coastal region, where 60 percent of Peru's population live.

Agriculture, industry, and urban populations across the earth depend on major river systems to transpiration, irrigation, and drinking water. Millions of acres of farmland and tens of million people rely on the Colorado River system for water. Impoundment of water behind large dams regulate the flow and delivery of water supplies. The timing of when regulated flows deliver water is based on the experience of the past 100 years. Scheduled delivery of water may be missed 60 to 90 percent of the time as human-induced climate change continues to make the Colorado River basin drier.

This page titled 10.4: The water cycle and climate change is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Michael E. Ritter (The Physical Environment) via source content that was edited to the style and standards of the LibreTexts platform.

- **10.4: Future Geographies - Water Resources and Climate Change** by Michael E. Ritter is licensed CC BY-SA 4.0. Original source: https://www.thephysicalenvironment.com.

10.5: Summary



10.5: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

11: The Global Carbon and Oxygen Cycles

11.1: Introduction
11.2: What might stimulate net primary production?
11.3: Temporal perspectives of the carbon cycle
11.4: Atmospheric methane
11.5: Carbon monoxide
11.6: Synthesis- Linking the carbon and oxygen cycles
11.7: Summary

11: The Global Carbon and Oxygen Cycles is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

11.1: Introduction

Now it's time to talk about the other key atom, **C**, the ratio of ^{13}**C**/^{12}**C** and corresponding **δ**^{13}**C** values, not only in Ca**C**O_3 but also in **C**O_2 and the organic molecules it transforms into through photosynthesis and the heterotrophic organisms that consume them. ^{13}**C** partitions not only into inorganic carbon but also into organic molecules throughout life. Hence we need a more detailed understanding of the **carbon cycle**.

The carbon cycle is likely discussed in introductory chapters in biology textbooks, but probably never in chemistry texts. It is fundamental to an understanding of climate, its control and change, and human processes that alter it. Figure 11.1.1 shows a representation of the carbon cycle. Calculated amounts of carbon found in the **lithosphere** (the solid part of the earth), the **atmosphere** (specifically the lower part, the troposphere), and the **hydrosphere** are shown. (The **cryosphere**, the frozen ice found in Greenland and Antarctica, is not shown). The **biosphere** includes part of each of these "spheres" that harbor life. Since life has been shown to exist 10 km down in the crust, we'll refer to the entire region in the diagrams as the biosphere. Figure 11.1.1 presents carbon stores in petagrams (10^{15} g) or gigatons of carbon (GtC), as 1 petagram equals 1 billion metric tons (or approximately 1.1 billion US tons).

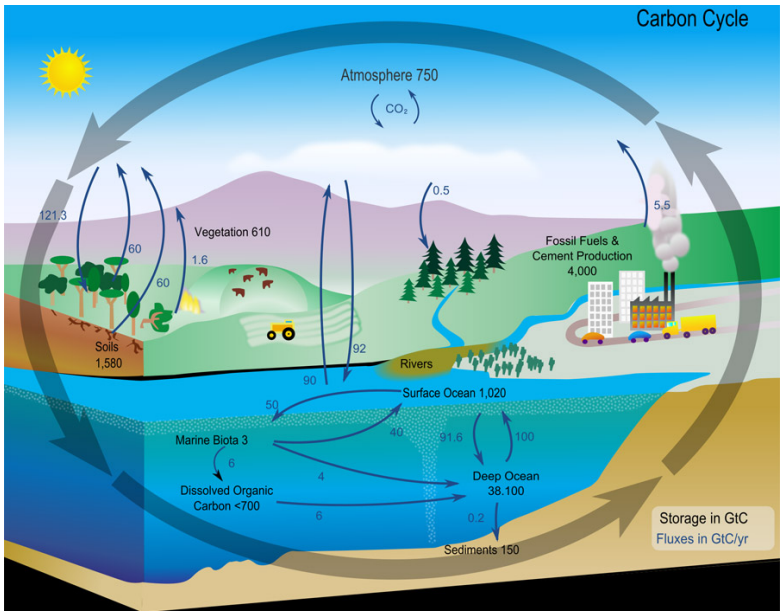


Figure 11.1.1: The carbon cycle. https://commons.wikimedia.org/wiki/F...te_diagram.svg

In addition to the total amount of carbon stored in each region, (GtC), the net changes in carbon per year as it moves into and out of reserves (GtC/yr) are shown in blue arrows with attached numbers.

The exchanges of carbon in the cycle occur at different time scales. Geologically "fast" exchange, on a time scale up to 1000s of years, occurs among the oceans, atmosphere, and land, while a slow exchange (over hundreds of thousands to millions of years) occurs in deep soils, deep ocean sediments, and rocks. We will mostly consider exchanges among the atmosphere, land, and oceans.

CO_2 in the terrestrial biosphere is removed by photosynthesis and returned by respiration by autotrophs like plants, and heterotrophs like microbes that consume soil carbon and plant remains. CO_2 in the atmosphere is also removed by ocean autotrophs like ocean phytoplankton and through partitioning into dissolved inorganic carbons (DIC) molecules like HCO_3^- and CO_3^{2-} into the oceans.

Before we probe some relevant reactions within it, let's look at the big picture and perhaps the most relevant to our climate crisis - the factors causing our increasing CO_2 atm and global warming. To do that, we must put numbers on the cycle to quantify it.

Quantitating the carbon cycle

We use parts per million (ppm) as a unit for expressing the amount of CO_2 in the atmosphere. Table 11.1.1 below shows how to translate the percentage (parts per 100) for each component gas in the atmosphere (with which you are familiar) into ppm.

Gas	% (parts per 100) in atm	part per million
N_2	78.09	780,900
O_2	20.94	209,400
Ar	0.93	9300
CO_2	0.0415	415

Table 11.1.1: Unit conversion - % to ppm

Climate scientists use ppm instead of concentration (in molecules/m^3) since they wish to know the relative percent or ppm increase with time, which does not depend on temperature and pressure. In contrast, concentration does depend on T and P, as you will remember from the ideal gas law, PV=nRT or n/V=P/RT that you studied in introductory chemistry.

It is important to use dimensional analyses to interconvert units as well. Table 11.1.2 below shows conversion factors to switch between GtC, Gt CO_2, and ppm.

Convert from	to	conversion factor
GtC (Gigatons of carbon)	ppm CO_2	divide by 2.124
GtC (Gigatons of carbon)	PgC (Petgrams of carbon)	multiply by 1
Gt CO_2 (Gigatons of carbon)	GtC (Gigatons of C)	divide by 3.664 = 44.01/12.01)
GtC (Gigatons of carbon)	MtC	multiply by 10000

Table 11.1.2: Unit conversion - GtC and Gt CO_2

To be more technical, atmospheric CO_2 concentrations are expressed in mol fraction of CO_2 in the dry air atmosphere. The ppm for CO_2 is hence µmol CO_2 per mole of dry air.

We have to put numbers on the components of the carbon cycle to quantitatively analyze changes in its components, otherwise, we can't know what is presently happening nor will we be able to predict with some certainty the future. Stoichiometry and reaction kinetics are probably the least liked parts of chemistry for many, but they are critical in understanding climate change We have to apply them on a global scale. Two key terms are important:

Stocks or reserves: How much carbon (mass in Gigatons or petagrams) is stored in given locations in the biosphere. This allows us to understand what % of all carbon stocks are in the ocean, for example. Stocks are usually reported as gigatons of carbon (GtC), not gigatons of carbon dioxide, since many stocks (like fossil fuels) consist of mostly C and H without oxygen. As in stoichiometry calculations in introductory chemistry courses, GtC in the atmosphere can be converted to gigatons of CO_2 by using dimensional analysis.

Fluxes (rates): How much carbon is transferred from one reserve to another per year (Gigatons/yr). Climate scientists are simply applying the Law of Mass Conservation that you learned in introductory chemistry, to the entire biosphere.

Figure 11.1.2 shows the reserves/stock (GtC) for reserves and **decadal (2012-2021) average** fluxes (large and small arrows, GtC/yr) for individual or aggregated stocks.

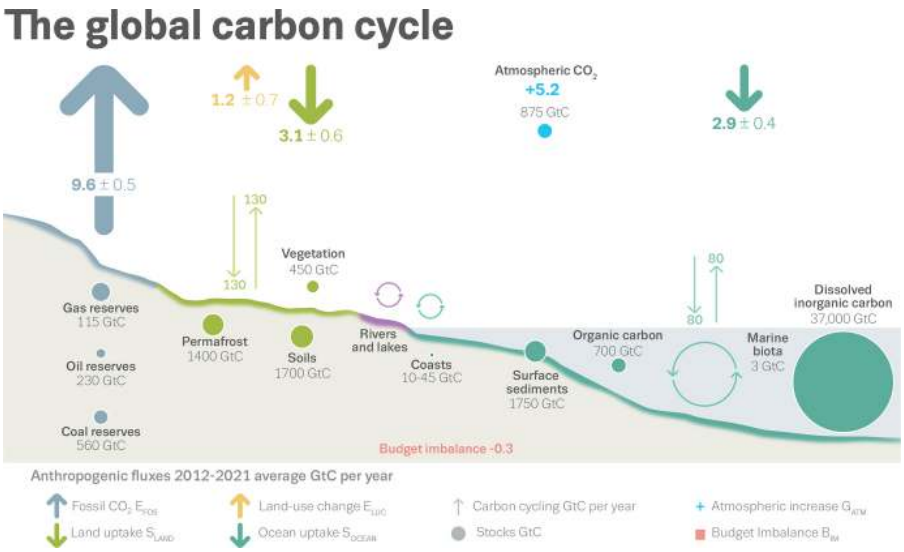


Figure 11.1.2: Schematic representation of the overall perturbation of the global carbon cycle caused by anthropogenic activities **averaged globally for the decade 2012–2021**. E represents emission and S "sink". Earth Syst. Sci. Data, 14, 4811–4900, 2022. https://doi.org/10.5194/essd-14-4811-2022. © Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.

The abbreviations used are: E_{FOS} (emissions, fossil fuels), E_{LUC} (emissions land use changes - mostly deforestation), S_{LAND} (terrestrial CO_2 sink), S_{OCEAN} (ocean CO_2 sink), G_{ATM} (Growth Rate CO_2 atm), B_{IM} (carbon budget imbalance). Uncertainties are also shown except for the atmospheric CO_2 growth rate which is known precisely and accurately through modern measurements. (It's also the easiest to measure). Human (anthropogenic changes) occurs on top of the carbon cycle.

The upward arrows indicate release into the atmosphere and the downward arrows the absorption in the oceans and land. The thickness of the arrows gives a relative measure of the size of emission or absorption. The thickest arrow and highest value (9.6 GtC/yr) is for the anthropogenic emission of carbon from our use of fossil fuels. Think about that! **Humans are presently the biggest contributor to the carbon cycle**. Before the industrial revolution, human contributions were minimal.

It's worse than that!

Figure x above shows that on average, 9.6 GtC/yr was released from fossil fuel use between 2012-2021. The actual flux of anthropogenic carbon release **in 2021 was 9.9 GtC**, equivalent to 36.4 Gt CO_2.

If you add the up arrows and subtract from that sum the down ones, you get +4.8 GtC/yr. This represents the net average increase in GtC in atmosphere CO_2 per year for 2012-2021. That is close to the accurately, and precisely known value of +5.2 GtC/yr increase from the CO_2 we pour into the atmosphere through our use of fossil fuels. Hence the figure above is a bit out of balance (about 0.3 GtC too low - the B_{IM} carbon budget imbalance), but given the difficulty in calculating these values, it is remarkably close to "mass balance" as you learned in introductory chemistry classes. In general, before the industrial revolution, the sum of the fluxes leading to the addition of CO_2 into the atmosphere was equal to the sum of the fluxes that removed it. That is, the system was in a **steady state.** That is no longer the case.

Figure 11.1.3 shows a breakdown of the factors contributing to annual (left) and cumulative (right) fluxes of carbon (GtC/yr), a metric for CO_2 flux, over time since 1850.

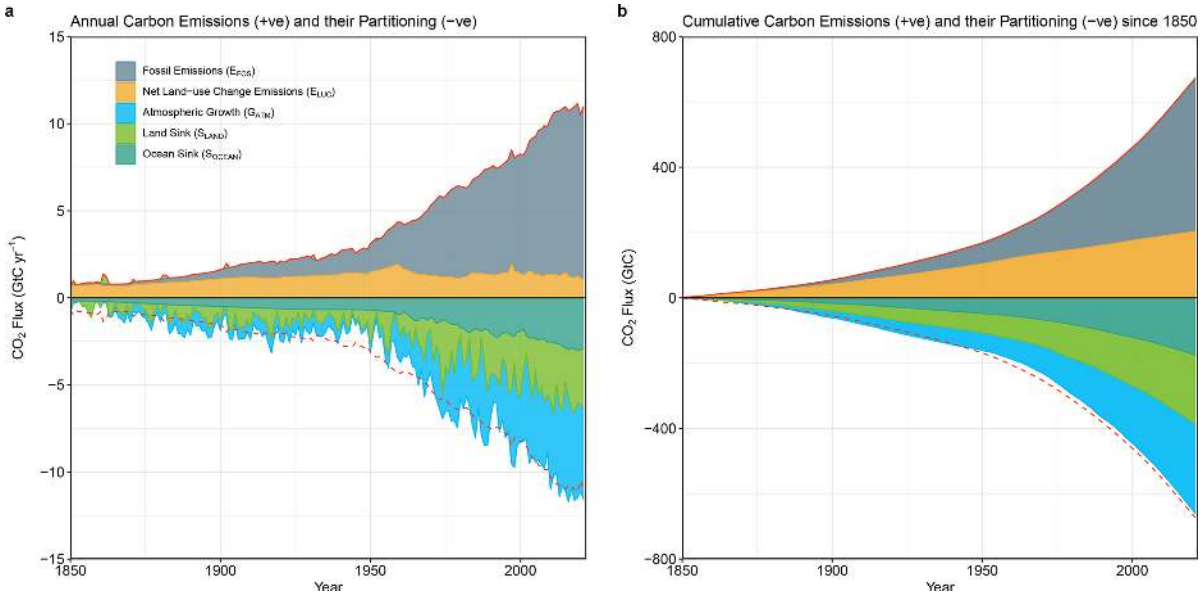


Figure 11.1.3: Combined components of the global carbon budget as a function of time for fossil CO_2 emissions (E_{FOS}, including a small sink from cement carbonation; grey) and emissions from land-use change (E_{LUC}; brown), as well as their partitioning among the atmosphere (G_{ATM}; cyan), ocean (S_{OCEAN}; blue), and land (S_{LAND}; green). Panel **(a)** shows annual estimates of each flux (GtC yr^{-1}, and panel **(b)** shows the cumulative flux (the sum of all prior annual fluxes) since the year 1850. Again, the graph shows GtC not Gt CO_2. . © Author(s), ibid

You might ask why the atmospheric growth in CO_2 (shown in green) is negative. We'll answer that question below.

Lastly, let's think about the total cumulative changes in GtC released and absorbed since 1850 (pre-US civil war and before the big release of CO_2 in modern times). Those data are shown in a bar graph in panel A of Figure 11.1.4. The bar graph in the right panel shows the mean decadal averages that are shown in Figure 2 above.

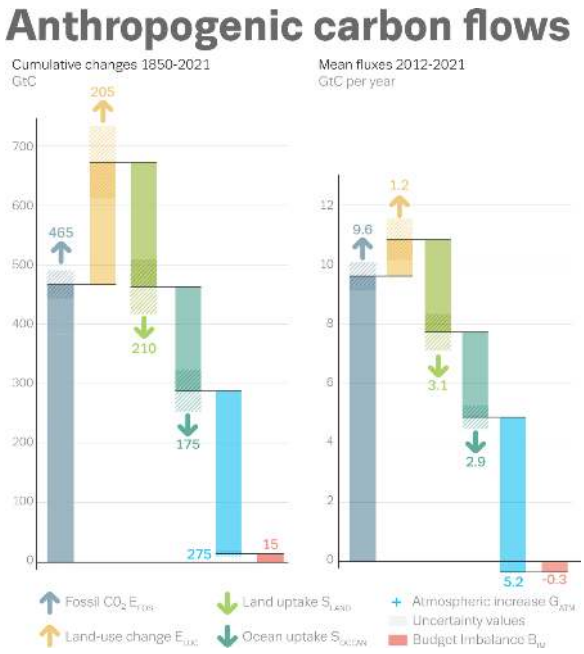


Figure 11.1.4: Total cumulative changes in GtC released and absorbed since 1850 (panel A) and mean decadal fluxes (panel B). E_{FOS} (emissions, fossil fuels), E_{LUC} (emissions land use changes - mostly deforestation), S_{LAND} (terrestrial CO_2 sink), S_{OCEAN} (ocean CO_2 sink), G_{ATM} (Growth Rate CO_2 atm), B_{IM} (carbon budget imbalance). © Author(s), ibid

The positive emission and negative absorption contributions are easy to see in the bar graph. The blue bar represents the net emission of carbon from fossil fuels and fills the gap to complete mass balance as we discussed above. It also explains the negative blue region in Figure 3. Just keep in mind that the blue net flux from fossil fuels is positive.

The cumulative contributions from fossil fuel emissions required to close the gap and fulfill mass balance is is +275 GtC, which when multiplied by the conversion factor (1ppm/2.124 GtC) translates into a 129.5 ppm increase in atmospheric CO_2 over that time. This is very close to independent measurements of a rise of 129.3 ppm (14.7-284.7) over that time.

The data from Figure 11.1.4 has been entered in the first four columns of Table 11.1.3 below.

Source	Subtype	Stock reserves (GtC)	J (Fluxes) GtC/yr (avg 2012/2021) + emission - absorption	$J=k_{app}$[stock]
Atmosphere	-	875		
Buried Fossil Fuels	Coal	560	+9.6	J=+9.6=k[905] k=0.0106
	Oil	230		
	Gas	115		
Terrestrial	Permafrost	1,400	+1.2 (Land use Δ) -3.1 (Land uptake)	J_{use}=+1.2=k_{use}[3550]; k_{use}=0.000338 J_{up}=-3.1=k_{up}[875]; k_{up}=0.00354
	Soil	1,700		
	Vegetation	450		
Oceans	Coasts	10-45	-2.9	J=-2.9=k[875] k=0.00331
	Ocean Surface Sediments	1,750		
	Organic carbon	700		
	Marine Biota	3		
	Dissolve Inorganic Carbon (DIC)	37,000		

We can use this data to develop our own crude computational model predicting future CO_2 emissions using Vcell, the program we used to produce time course (concentration vs time) graphs for both simple and coupled signal transduction reaction pathways.

Vcell can be used to calculate fluxes (J) in reaction pathways, where J is the change in concentrations of a species with time, given the initial concentration or amount of a reactant, and the rate constants affecting its production or removal. If we use the amount of carbon (GtC) in each reservoir in the biosphere and crust as a relative measure of "concentration" and the known fluxes (GtC/yr) for the transfer of carbon to and from the atmosphere as given in Table 3, we could calculate an "apparent rate constant for each flux using this equation:

$J_{stock} = k_{app}$[stock] (where stock is given in GtC).

$$J_{stock} = k_{app}[\text{ stock in GtC}] \tag{11.1.1}$$

These "apparent" rate constants are needed to run the Vcell simulation that can reproduce the actual fluxes shown in Figures 2 and 4). The simulation can be run over time

A simple four-term model based on Figures 2 and 4 is shown below. Run the simulation and see how atmospheric CO_2 changes with time. This model is offered only to show how climate models are made and used, and also for fun. The graphs are valid and sound based on the input parameters, but the outputs are based on many assumptions that vastly simplify the model.

The simulation shows CO_2 atm levels peaking at about 982 **GtC** in 51 years (2073) from its average decadal (2011-2021) value of 875. That is an increase of 107 **GtC** over now (50 ppm CO_2 rise from the present 414 to 463 ppm). Over a 50-year period, this gives an average annual rise of 2.14 GtC/yr or about 1 ppm CO_2/yr. A comparison of the predicted atmospheric CO_2 (ppm) levels through 2100 for the IPCC SSP1-2.6 scenario (**blue**) and simple Vcell model (**red**) is shown in Figure 11.1.5.

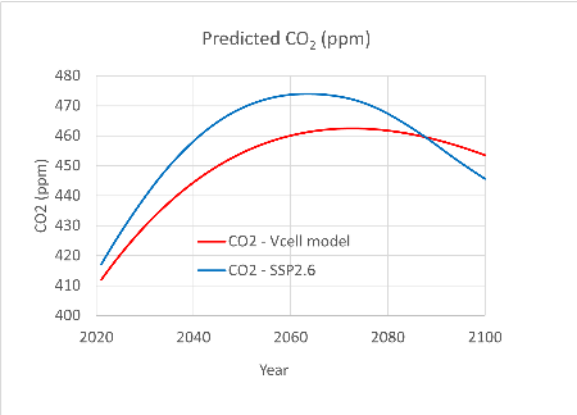


Figure 11.1.5: Predicted atmospheric CO_2 (ppm) for SSP1-2.6 scenario (**blue**) and simple Vcell model (**red**)

SSP1-2.6 data - History: Meinshausen et al. GMD 2017 (https://doi.org/10.5194/gmd-10-2057-2017); Future: Meinshausen et al., GMD, 2020 (https://doi.org/10.5194/gmd-2019-222). https://climateanalytics.org/media/g...-3571-2020.pdf. https://gmd.copernicus.org/articles/13/3571/2020/

However imperfect the Vcell model is (incorrect assumptions, lack of complexity and feedback mechanisms, etc), the results shown above are remarkably close to the projected increases in carbon dioxide in ppm described in IPCC reports for the SSP1-2.6 socioeconomic pathways, shown in the right panel (dark blue line) of Figure 11.1.6. This pathway predicts a rise of approximately 1.8^0 C in average global temperatures.

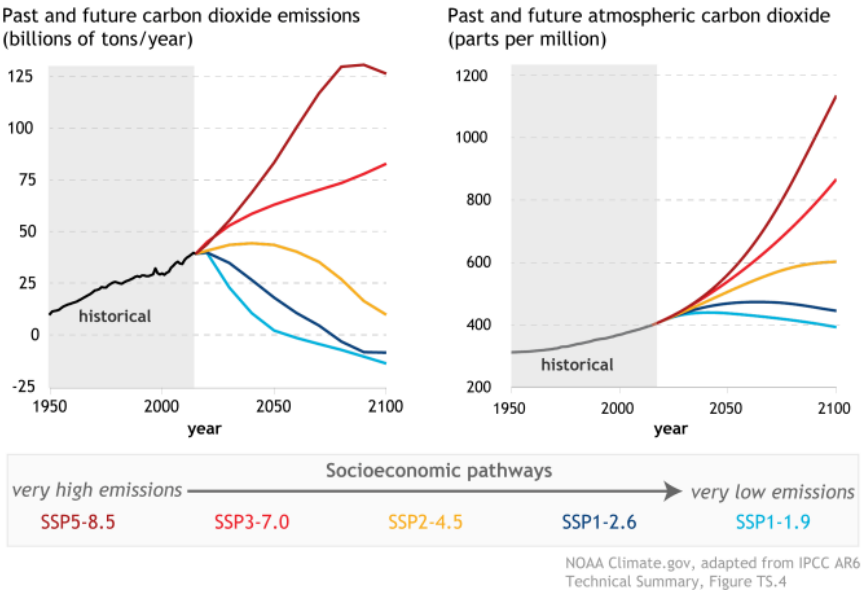


Figure 11.1.6: IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)].

Again, remember that the model is based on a ten-year average of CO_2 emissions. Think of all the other assumptions in this model (other than the stock reserves and fluxes) that would give higher or lower values of future CO_2 levels. One major one is that flux values are all held constant to allow calculations of the apparent rate constants for Vcell use. The model depletes much of the fossil fuel reserve. In addition, CO_2 emissions in 2021 were actually 9.9 GtC/yr and going up!

In addition, a change in one parameter can affect the others. For example, the net uptake of atmospheric CO_2 into the land and oceans has increased from 1960-2010, which makes sense given increased CO_2 in the air forcing additional uptake (think LaChatelier's Principle). The oceans have taken up nearly 40% of the CO_2 from fossil fuel use since the industrial revolution. If the rate of uptake decreases (i.e. if we start to saturate the uptake into oceans), CO_2 accumulation in the atmosphere would accelerate. Data also suggest that if we successfully decrease CO_2 in the atmosphere, the oceans would respond by decreasing uptake as well, which would slow the progress in reducing temperatures.

An interesting example relating atmospheric and ocean CO_2 occurred from 1990-2000 when it has been shown that the ocean acted as a weaker sink. This occurred because of a decreasing gradient (the Δ or"effective concentration differences") between atmospheric CO_2 and ocean "CO_2", which decreased the ability of the ocean to act as a sink for CO_2. You can decrease the Δ in two ways:

- decreasing the rate of entry of CO_2 into the atmosphere from fossil fuel use. There indeed was a temporary slowdown in this decade.
- paradoxically, by briefly making the ocean in a shorter term a better sink. This happened in 1991 after the eruption of Mt. Pinatubo, which led to decreased air and ocean temperatures. CO_2 is a nonpolar gas, which has higher solubility in water at lower temperatures (think about soda). This was a short-term and more minor effect than the decreased rate of fossil fuel emissions.

More complex models with more terms for emissions and absorptions of CO_2 can be made. One is shown in Figure 11.1.7. This model adds CO_2 release from the soil through respiration by microorganisms, as well as from plant respiration (CHO to CO2atm). Another term has been added for release by oceans.

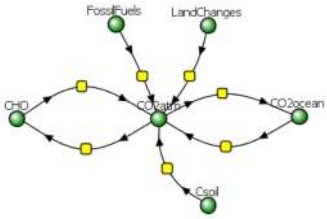


Figure 11.1.7: More complicated Vcell climate model.

Fortunately, we don't have to rely on these simple models to predict future trends in temperature and CO_2. A complex dynamic model simulator that is in accord with many different climate models is available at your fingertips. Developed at MIT and Climate Interactive, and available for free from any web browser, the EN-ROADS program allows users to change sliders for key inputs and see future predicted temperature and CO_2 levels. In accordance with RCP and SSP IPCC pathways that tie future emissions to socio-economic policies (discussed in Chapter 31.1), the program allows users to change variables such as carbon pricing and incentives to move to clean energy in transportation, building and energy supplies sectors. Access the program directly from this page by clicking the Close icon in the program window in Figure 11.1.8 below.

Figure 11.1.8: EN-ROADS global climate simulator

Here is also an external link to the En-Roads global climate simulator (Developed by Climate Interactive, the **MIT Sloan Sustainability Initiative**, and **Ventana Systems**)

Move the interactive sliders and see the effect on greenhouse gas emissions and global temperatures. Here is a link to a one-page tutorial on its use.

The Skeptic's Corner: Climate Change Misinformation

We should all be skeptical of models, especially ones that predict changes 80 or more years into the future. We gain confidence in a model if it accurately fits data going back in time and into the future data as well. We mentioned in Chapter 31.1 that oil company scientists knew of the likely climatic effects arising from fossil fuel emissions, but the company executives did not act on their models. Their models were startlingly accurate as shown in Figure 11.1.9 below, which shows their predictions for both CO_2 levels and the associated increases in temperature caused by them.

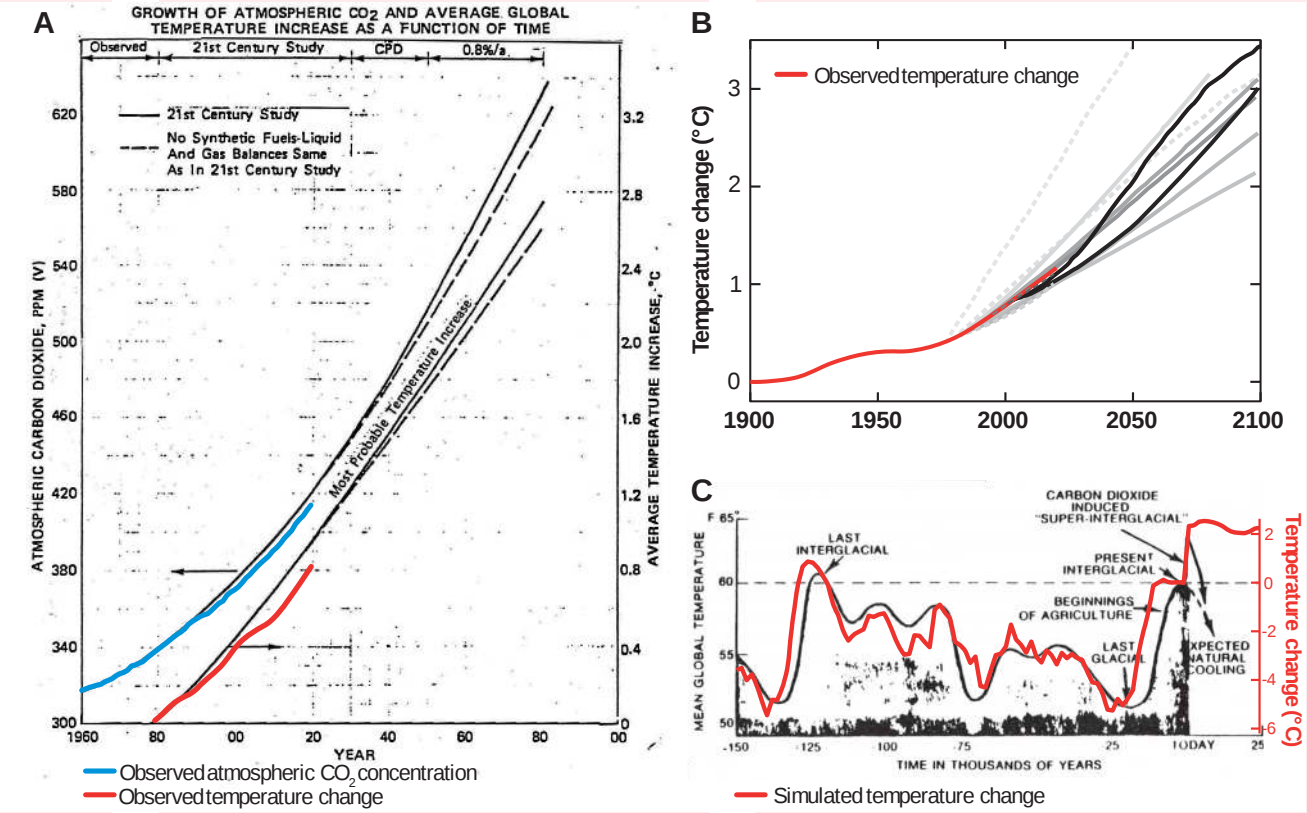


Figure 11.1.9: Historically observed temperature change (red) and atmospheric carbon dioxide concentration (blue) over time, compared against global warming projections reported by ExxonMobil scientists. Supran, G., Rahmstorf, S., and Oreskes, N. Assessing ExxonMobil's global warming projections. Science (2023). https://www.science.org/doi/abs/10.1...cience.abk0063. Reprinted with permission from AAAS. Not for reuse.

Panel (A) shows “Proprietary” 1982 Exxon-modeled projections.

Panel (B) shows a summary of projections in seven internal company memos and five peer-reviewed publications between 1977 and 2003 (gray lines).

Panel (C) shows a 1977 internally reported graph of the global warming “effect of CO2 on an interglacial scale.” (A) and (B) display averaged historical temperature observations, whereas the historical temperature record in (C) is a smoothed Earth system model simulation of the last 150,000 years.

As these graphs clearly show, oil companies knew since the late 1970s, over 40 years ago, of the climatic effects of CO_2 emissions. They could even predict the temperatures since the last ice age. In the 70s, solar and wind energy were much more expensive to produce and use than now, but if we had subsidized their development back then as we have done for decades for the fossil fuel industries, our climatic situation now would be much less precarious. Figure 11.1.10 below shows worldwide fossil fuel subsidies in US $billion and in % global GDP from 2015 to 2020 and projections after that.

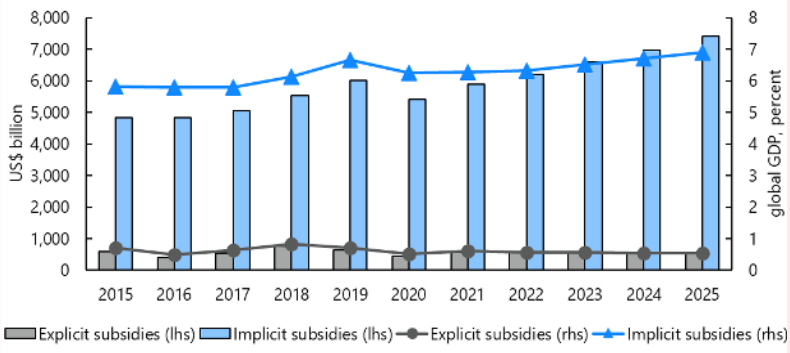


Figure 11.1.10: Worldwide subsidizes in US $billion and in % global GDP. Bar graphs are for US$biillons and the circles and triangles for % global GDP. IMF. ttps://www.imf.org/en/Publications/W...bsidies-466004

The subsidies are broken down into explicit subsidies (tax breaks or direct payments to help fossil fuel companies to fund their uncompensated costs) and implicit ones (undercharging for environmental costs of fossil fuel use that the oil companies don't pay). These latter "hidden" costs are passed down to countries, states, and individuals. In 2020, global subsidies were $5.9 trillion or 6.8% of the world's GDP. The explicit subsidies given to fossil fuel companies, about 8% of the total, amounted to $472 billion just in 2020!

"Company executives chose to publicly denigrate climate models, insist there was no scientific consensus on anthropogenic climate change, and claim the science was highly uncertain when their own scientists were telling them the opposite" (ref). They also propagated the myth that the global climate was actually cooling. This is a powerful and unsettling example of disinformation with enormous consequences.

Now that we have seen the big picture, let's look at how carbon moves through various pools of carbon-containing molecules. We have already discussed photosynthesis in great detail in Chapter 5, so we fill focus our attention more on dissolved inorganic carbons (DIC) including species such as HCO_3^- and CO_3^{2-}. Another view of the carbon cycle that includes weathering of rocks to produce silicates and bicarbonates, and the formation of shells in the ocean from HCO_3^-, CO_3^{2-} and silicates, is shown in Figure 11.1.10.

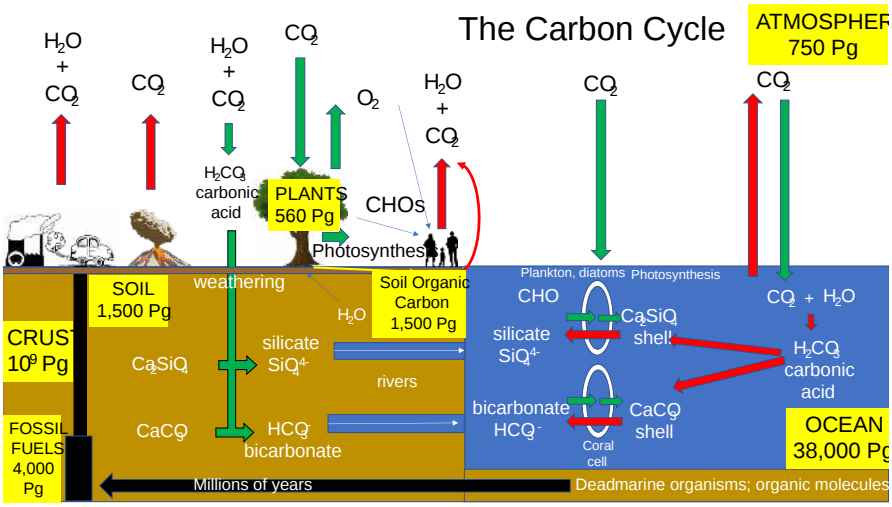


Figure 11.1.10:

Let's focus on the oceans first. The reversible movement of CO_2 from the atmosphere to the oceans, $CO_{2\ atm} \leftrightarrow CO_{2\ ocean}$, depends on the difference in the partial pressures of CO_2 (ΔpCO_2) in the atmosphere and surface waters. The reaction is also driven to the right by the removal of CO_2 (aq) as it forms carbonic acid (H_2CO_3), which then forms bicarbonate (HCO_3^-) and carbonate (CO_3^{2-}). These coupled reactions chemically buffer ocean water, thus regulating ocean pCO_2 and pH.

pCO_2 is not homogenous in ocean surface waters and varies with different conditions of current and temperature. CO_2 can be more readily released from upwellings of nutrient-rich and warm waters, especially in the tropics. In cold Northern waters and also in the Southern Ocean, where water sinks, it is taken up from the atmosphere (again CO_2 is more soluble in cold water).

As we discussed in Chapter 31.1, the **ocean chemistry of CO_2 determines in large part the levels of atmospheric CO_2.** The coupled reactions of CO_2 in the oceans are shown below.

$$CO_2(\ g,\ atm) \leftrightarrow CO_2(aq,\ ocean) \tag{11.1.2}$$

$$CO_2(aq,\ ocean\ ) + H_2O(I,\ ocean\ ) \leftrightarrow H_3O^+(aq) + HCO_3^-(aq) \tag{11.1.3}$$

$$H_2O(l) + HCO_3^-(aq) \leftrightarrow H_3O^+(aq) + CO_3^{2-}(aq, \text{ sparingly soluble }) \quad (11.1.4)$$

These reactions should be familiar to all chemistry students and were presented previously in Chapter 31.1 and in Chapter 2. A significant contributor to ocean bicarbonate is **weathering** of rocks like limestone. and marble, which are both forms of $CaCO_3$. The relevant reactions are shown below.

$$\begin{aligned} CaCO_3(s) + H_2O &\leftrightarrow Ca^{2+}(aq) + CO_3^{2-}(aq) \\ CO_3^{2-}(aq) + H_2O &\leftrightarrow HCO_3^-(aq) + OH^-(aq) \end{aligned} \quad (11.1.5)$$

CO_2 is nonpolar and not very soluble in water. Either is CO_3^{2-} in the presence of divalent cations like Ca^{2+}. However HCO_3^- is and can be considered a "soluble" form of carbon. This soluble form from terrestrial weatherings ends up in rivers and eventually enters the ocean. It is also the form of carbonate that is transferred into cells by anion transporters for eventual shell formation. HCO_3^- is also a chief regulator of both blood and ocean pH. Weathering is slow compared to anthropogenic emissions of CO_2 from fossil fuel use, but it is nevertheless a key player in the carbon cycle and the regulation of ocean pH.

The same weathering reactions on silicate rocks lead to the transfer of silicate ions into rivers and then into the ocean, where they can be taken up by diatoms in the formation of $CaSiO_4$ shells. As the oceans take up more CO_2, they become more acidic, which leads to the equivalent of "weathering" of shells of living organisms, leading to their potential death. Silicon is directly underneath carbon in the periodic table so the following simplified reaction is analogous to those we seen with CO_2 and its inorganic ions.

$$H_4SiO_4 = SiO_2 + 2H_2O \quad (11.1.6)$$

H_4SiO_4 is silicic acid.

$^{13}C/^{12}C$ ratios in ice core and ocean sediments

We are now in the position to explore how isotopes of carbon can be used for more than radio- ^{14}C dating, which is quite limited in climate studies. ^{13}C, a stable isotope of carbon, however, is extremely useful because C-^{13}C bond dynamics are influenced by it. Reaction rates are affected by the presence of ^{13}C when C-C bonds are made or cleaved. This isotope effect leads to different $^{13}C/^{12}C$ ratios in reactants and products, and hence different $\delta^{13}C$ values.

Isotopes have a long history in the study of biochemical reactions. The k_{cat} and k_{cat}/K_M values for enzyme-catalyzed reactions can be affected if the rate-limiting step involves cleavage or the creation of a C-^{13}C, C-D (deuterium) or C-T (tritium) bond. Substrates labeled with the isotopes have similar transition state energies for the formation/cleavage of a bond involving an isotope, but the ground state vibrational energy for the isotope-substituted atom are proportionately lower, as illustrated in Figure 11.1.11.

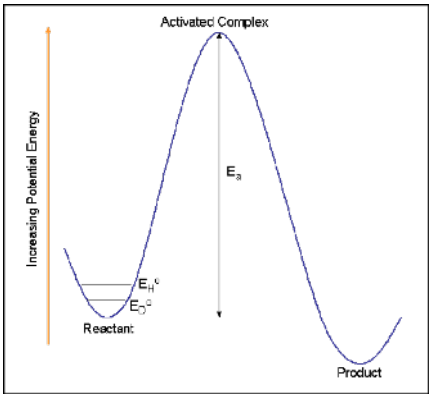


Figure 11.1.11: Kinetic Isotope Effects.

This primary **kinetic isotope effect** leads to higher activation energy for the formation/cleavage of a bond with the isotope. For C-D and C-T bond cleavages that are rate-limiting, the rates are 7X and 16X slower than the cleavage of a C-H bond, respectively. Cleavage or formation of bonds to heavy isotopes of carbon, oxygen, nitrogen, sulfur, and bromine have much smaller isotope effects (between 1.02 and 1.10). The difference in the magnitude of the kinetic isotope effect is directly related to the percentage change in mass. Large effects are seen when hydrogen is replaced with deuterium because the percentage mass change is very large (mass is being doubled). .

Hence the kinetic isotope effect is at play in carbon fixation in photosynthesis, for example. This is evidenced by the observation that the $^{13}C/^{12}C$ ratios are lower in plants than in the atmosphere, showing that $^{12}CO_2$ is preferentially "fixed" in the ribulose bisphosphate carboxylase/oxygenase reaction in plants and other photosynthetic organisms. Also, $^{12}CO_2$, a lighter molecule, has a faster rate of diffusion through the stromata, regulated pores in leaves that facilitate the passage of CO_2, O_2 and H_2O.

In Chapter 31.2, we discussed the use of $\delta^{18}O$ values in ice core and ocean core sediments for measuring past CO_2 and temperatures.

$$\delta^{18}O = \left[\frac{\left(\frac{^{18}O}{^{16}O}\right)_{\text{sample}}}{\left(\frac{^{18}O}{^{16}O}\right)_{\text{reference}}} - 1 \right] * 1000$$

$\delta^{18}O$ values for ice core water samples werer easier to interpret than $\delta^{18}O$ values for $CaCO_3$ sampls, since the deposition of ice is a simple physical process compared to the complexity of the deposition of $CaCO_3$ in ocean sediments, which depends on chemical reactions and nonequilbrium processes (as described in Chapter 31).

Climate scientists can determine and use $\delta^{13}C$ values. An analogous equation for it is shown below.

$$\delta^{13}C = \left[\frac{\left(\frac{13}{12}C\right)_{\text{sample}}}{\left(\frac{13}{12}C\right)_{\text{reference}}} - 1 \right] * 1000$$

As for using $\delta^{18}O$ in carbonate samples, using $\delta^{13}C$ is more difficult as well. The shells of ocean sediment foraminifera were made from dissolved inorganic carbon (DIC) in the ocean at the time so their $\delta^{13}C$ values reflect that. However, shell formation is not a simple equilibrium process since biological shells are formed rapidly so kinetic effects in carbonate and hence isotope fractionation are important. In addition, the biochemistry of shell formation is complicated.

In the open ocean, planktic foraminifera are perhaps the most important marine organism that forms shells given that they produce and export into the ocean about 2.9 Gt $CaCO_3$/yr. Their shells form in a process involving many metastable calcite phases. It starts with a soft template that contains Mg^{2+} and Na^{+} ions which play a key role in crystallization. Growth occurs by successive additions of "chambers" to the shell. An F-actin mesh, which forms microtubular structures, leads to the formation of protective envelopes for chamber formation. The layered templates sequester and help control the mineralization of shells and separate bulk sea water for a more intracellular vs extracellular process for biomineralization. Seawater containing minerals becomes vacuolized in a process which for some foraminfera excludes a competing cation, Mg^{2+}. In addition, both Ca^{2+} and HCO_3^{-} transporters are required. This all combines to form an environment low in Mg^{2+} and supersaturated in Ca^{2+} and CO_3^{2-} for calcite formation. The kinetic fractionation of ^{13}C isotopes into shells is also different than for ^{18}O isotopes since the "pool" of oxygen in the oceans is much greater than carbon. Likewise, the $\delta^{13}C$ is more location-dependent that the $\delta^{18}O$.

Buried organic matter can also be studied. The $\delta^{13}C$ value for buried organic matter depends on primary productivity on land and in the oceans. As mentioned above, autotrophs preferentially take up $^{12}CO_2$. Heterotrophs that eat them also become enriched in ^{12}C. Hence organisms have negative $\delta^{13}C$ values, typically around −25‰, with the number depending on pathways of incorporation and metabolism. Methane in hydrates in the ocean can be either biogenic, made by methanogens, for example, at low temperatures, or thermogenic, made during high-temperature reactions. Biogenic methane has a $\delta^{13}C$ of around - 60‰, while thermogenic methane has a value of around −40‰. Terrestrial plants have different δ^{13} values. $\delta^{13}C$ in C_4 plants range from -16 to −10‰ while for C_3 plants they range from −33 to −24‰.

> Changes in $\delta^{13}C$ in ice cores and ocean sediments are used in climate studies. Sometimes it's confusing to understand the cause and effect of the changes. This following explanation for changes in the already negative values of $\delta^{13}C$ might offer help to those with a chemistry-centric view of biochemistry who struggle with mass balance outside of simple chemical equations.
>
> Under climatic conditions, when there is an abundance of **terrestrial** plants that lock in and sequester atmospheric $^{12}CO_2$, the atmosphere becomes depleted in $^{12}CO_2$ and correspondingly enriched in $^{13}CO_2$. Hence primary production (fixing of carbon and anabolic metabolism) by photoplankton in the oceans, under robust growing conditions, would sequester more ^{13}C, causing an **increase in $\delta^{13}C$** (i.e. more positive) values for buried organic and calcite sediments.
>
> During times of mass extinction, when terrestrial plant primary production drops precipitously, the $\delta^{13}C$ becomes more negative with the decrease in primary production and release of plant carbon, leaving more $^{12}CO_2$ in the atmosphere. This drop is called a **negative $\delta^{13}C$ excursion**. When life is robustly favored and carbon is fixed by autotrophs, and the organic carbon resulting from them is eventually buried in sedimentary rocks, the rise in $\delta^{13}C$ is called a positive **$\delta^{13}C$ excursion**.

Examples of climatic events accompanied by changes in $\delta^{13}C$.

Late Devonian period

Fossil evidence from the late Devonian, when large terrestrial plants evolved and expanded, is characterized by increases in $\delta^{13}C$.

Paleocene/Eocene Thermal Maximum

We saw in Chapter 31.1 that around 55 MYA, sediment records indicate a spike in temperatures of about 5^0 F occurring over about a 100K year timeframe. This was accompanied by a dramatic spike in CO_2 and a dramatic drop in ocean pH as measured by the loss of deep sea $CaCO_3$ (chalk). This very short time frame is called the Paleocene/Eocene thermal maximum (PETM), which shows very quick spikes (on the geological time scale) can and do occur. Sediment records for this time indicate a large negative $\delta^{13}C$ excursion, consistent with a loss of plants with their preferential uptake of $^{12}CO_2$, leading to an accompanying increase in $^{12}CO_2$ in the atmosphere.

1500-1650 CE

We examined $\delta^{18}O$ values during the Little Ice Ages in Chapter 31.2. What about $\delta^{13}C$ values? CO_2 and $\delta^{13}C$ values from 1000 to 1900 are shown in Figure 11.1.12.

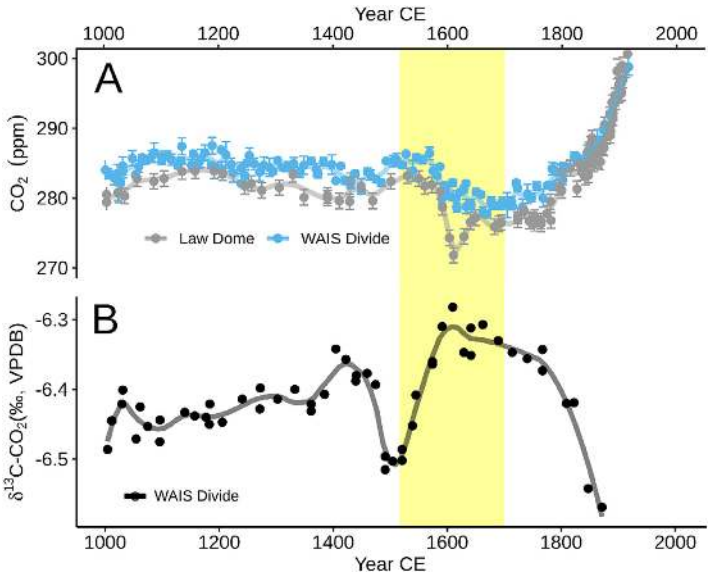


Figure 11.1.12: CO_2 and $\delta^{13}C$ values from 1000 to 1900. Koch et al. Quaternary Science Reviews, 207, 2019, 13-36. https://doi.org/10.1016/j.quascirev.2018.12.004. CC BY license (http://creativecommons.org/licenses/by/4.0/).

Panel (A) shows the CO2 concentrations recorded in two Antarctic ice cores: Law Dome (grey, MacFarling Meure et al., 2006) and West Antarctic Ice Sheet (WAIS) Divide (blue, Ahn et al., 2012).

Panel (B) shows the carbon isotopic ratios recorded in CO2 from the WAIS Divide ice core (black, Bauska et al., 2015) showing an increased terrestrial carbon uptake over the 16th century (B). The yellow box is the span of the major indigenous depopulation event (1520e1700 CE). Loess smoothed lines for visual aid.

Koch et al have strong evidence to suggest that the cooling after 1510 (area in the yellow box in the above figure) was associated with a dip in CO_2 caused by the reforestation of indigenous peoples' land in Meso and South American after epidemics of European disease killed upwards of 90% (around 55 million) of the indigenous peoples. The open and agricultural land reverted back to forests. The diseases included smallpox, measles, influenza, the bubonic plague, and eventually malaria, diphtheria, typhus and cholera. Domesticated farm animals brought from Europe to the Americans led to most of the disease. Along with the death of so many people was a concomitant return of cleared and agricultural lands (about 56 million hectares or 212,000 mi^2) to forest and plant growth. This may have led to a 7-10 ppm drop in CO_2 in the late 1500s and early 1600s, peaking in 1601 (middle of the yellow box). This decrease in temperature was associated with a small rise (small positive excursion) in the $\delta^{13}C$ values, as $^{12}CO_2$ was preferentially removed from the atmosphere. Global surface air temperatures decreased by around 0.15ºC. This "Great Dying" of Indigenous peoples shows the power of humankind to globally alter climate in calamitous ways, even before the use of fossil fuels. The decrease in $\delta^{13}C$ values before 1500 was unexplained.

1800-the present

$\delta^{13}C$ values can also be used to unequivocally prove that the increase in CO_2 since the industrial revolution is from the burning of fossil fuels, which is of biogenic origin and hence have more negative $\delta^{13}C$ values. Figure 11.1.13 shows atmospheric CO_2 levels in ppm plotted along with $\delta^{13}C$ values. There is a perfect correlation between the rise in atmospheric CO_2 starting with the industrial revolution with the decrease in the $\delta^{13}C$ values over the same time.

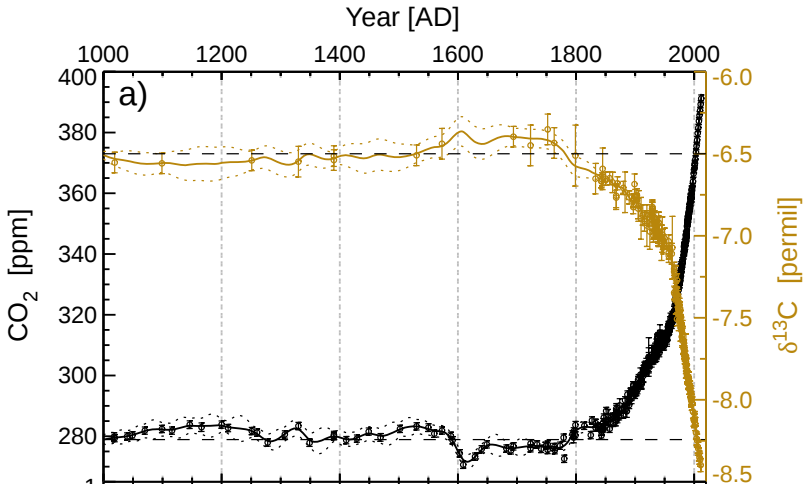


Figure 11.1.13: CO_2 concentration (black circles) and the $\delta^{13}C$ (brown circles) from 1000 to 2010. Rubino et al. Journal of Geophysical Research: Atmospheres. https://doi.org/10.1002/jgrd.50668. With permission (Copyright Clearance Center)

Summary

In the first three sections of Chapter 31 (31.1, 31.2, and this one), we have introduced the basics of climate change, and how climate scientists obtain, analyze and interpret climate data. We emphasized the scientific rigor by which they do that and offered a detailed analysis of the use of isotopes to document past and present changes in climate, Finally, we offered models to predict and mitigate future climate changes. After reading this material, you should be enabled to discuss climate change with others from a sound and valid position. More importantly for this book, you will have a better knowledge base and understanding for the rest of the chapter sections, which will address specific topics in "biochemistry and climate change".

Key Points - Beta version from Chat.openai

1. The carbon cycle is the process by which carbon moves through the Earth's systems, including the atmosphere, oceans, and biosphere.
2. The carbon cycle is driven by the exchange of carbon between different reservoirs, such as the atmosphere, oceans, and living organisms.
3. The main processes involved in the carbon cycle include photosynthesis, respiration, and the formation and weathering of rocks.
4. Human activities, such as burning fossil fuels and deforestation, have significantly increased the amount of carbon dioxide (CO2) in the atmosphere, disrupting the natural balance of the carbon cycle.
5. The increase in atmospheric CO2 is the primary driver of climate change, as it causes the greenhouse effect, trapping heat in the atmosphere and warming the Earth's surface.
6. The ocean also plays a critical role in the carbon cycle, as it acts as a sink for CO2, absorbing about 25% of the CO2 emitted by human activities.
7. The acidification of the ocean caused by the uptake of CO2 is having a significant impact on marine ecosystems, altering the chemistry of seawater and making it more difficult for some organisms to build and maintain their shells and skeletons.
8. Understanding the carbon cycle and carbon chemistry is crucial for understanding the causes and impacts of climate change and for developing strategies to mitigate and adapt to its effects.

11.1: Introduction is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

11.2: What might stimulate net primary production?

Terrestrial ecosystems are dependent on the sun's energy to support growth and maintenance. Plants are primary producers that, through photosynthesis, manufacture organic molecules such as carbohydrates and lipids from raw inorganic materials (CO_2, water, mineral nutrients). Primary productivity is thus a fundamental determinant of both the structure and functioning of terrestrial biomes. The energy and carbon of primary production supplies consumers, including humans, with the necessary fuel to support their metabolism while providing essential carbon compounds that form the bricks and mortar of living cells. In addition to solar radiation, the main abiotic factors that affect rates of photosynthesis and NPP are water, temperature, carbon dioxide concentration, and nutrients. Globally, there is broad equilibrium relationship between NPP, temperature and precipitation that is strongly impacted by nutrient limitations and, hence, knowledge of the global distribution of NPP is important for understanding vegetation dynamics in biomes, patterns of biodiversity, potential agricultural yield, and predicting global climatic changes.



Figure 11.2.1: Erosion on hillsides. Extreme topsoil erosion in Iceland. An extreme example of near total loss of topsoil having huge impacts on NPP. In Iceland, thick Andosoils have been removed by the forces of wind and water leaving shallow and poor soils with limited vegetation cover.
Source: CC-BY Zinneke. Wikimedia Commons.

Some of the key factors that lead to long-term reductions in NPP are associated with various forms of land degradation. This includes soil erosion (wind, water), nutrient depletion, salinisation, soil compaction and crusting, topsoil losses and nutrient depletion, alterations of vegetation composition and structure, and water depletion. In fact, most types of soil disturbance can have devastating impacts on a region’s productivity. One of the most extreme examples of soil loss’s impacts on productivity is Iceland. Icelandic agriculture is primarily based on sheep farming, dairy and poultry production. Due to numerous factors, but especially poor grazing management and wood harvesting, the entire topsoil has been lost over vast areas of the island. Numerous biogeochemical processes affect the carbon balance of terrestrial biomes, including photosynthesis, plant respiration, microbial respiration, leaching losses, erosion, herbivory, fire, and rates of rock weathering. Human appropriation of NPP and modification of

the Earth's surface over the past several centuries has altered many of these processes. Net biome production (NBP), which applies to changes in carbon stocks over large spatial areas and long time periods, is helpful to describe changes in carbon balances after losses due to natural or anthropogenic disturbances. NBP balances carbon emissions with non-respiratory losses such as fire, deforestation, insect infestation, droughts, agricultural harvests, and ecosystem respiration fluxes with NPP, and indicates the carbon source-sink strength and possible positive feedback flux to atmospheric CO2 concentrations. Climate change and human-induced degradation may well result in an increased or decreased NBP. Global environmental change is rapidly altering the dynamics of terrestrial biomes. This has major consequences for the functioning and structure of the Earth system, including the provision of ecosystem services. Long-term satellite measurements have identified a widespread greening of the Earth. The four key drivers of this change are the fertilisation effects of atmospheric CO2 (this explains 70 % of the observed greening trend), nitrogen deposition (9 %), climate change (8 %) and land cover change (4 %). Global CO2 concentrations have risen from about 280 ppm at the start of the industrial revolution to about 406 ppm in 2017.

Elevated carbon dioxide concentrations have numerous effects on plants, such as acting as a fertiliser that stimulates increases in photosynthesis. Elevated CO2 concentrations also tend to reduce water loss in plants, which may be more important than the direct effect of increased photosynthesis rates due to global trends in changing aridity. There is evidence that the CO2 fertilisation effect can change plant-species mix (such as enhancing woody plant growth over grass growth), lower the carbon-to-nitrogen ratio of plants (making grazing and browse material less palatable), and evoke long-term evolutionary responses. It is therefore likely that increased NPP due to increased CO2 concentrations may be changing useful ecosystem services in some areas, despite the higher plant productivity. However, generalisations are difficult to make because there are many feedbacks and interactions with other variables, such as temperature, nutrients, water availability, and plant-plant competition.
The direct CO2 effect on plants should be most strongly expressed in warm, arid environments where water is the dominant limitation to vegetation growth and where land degradation is widespread. Indeed, it has been shown that the 14 % increase in atmospheric CO2 (covering the period 1982–2010) led to a 5-10 % increase in green foliage in warm, arid environments. Can global increases in NPP mask the impacts of degradation? Long-term change in NPP is potentially a useful indicator of land degradation, but interpreting short-term changes in NPP as degradation can be misleading because it can be a reflection of climatic fluctuations. Scientists have attempted to use rain use efficiency (NPP per unit of water) to better understand degradation tends, which have been most successful over long time periods. Also, in combination with climate impact, anthropogenic land use can increase NPP but mask other forms of degradation. Highly productive cultivation systems may increase NPP but affect other ecosystem services, such as water and nutrient supply. Grazing pressure can cause a species change to, for example, woody shrubs showing an increase in NPP but reduce palatability and biodiversity. Hence, it is important to include NPP changes in land degradation assessment but these changes can only be interpreted correctly when considering the dynamics of other ecosystem services and socio-economic situations.

Biomass production is the most important process of the biosphere. It directly impacts many ecosystem services, such as the global carbon cycle, which in turn affects the water cycle and climate. In fact, many ecosystem services are positively correlated with net primary production (NPP), including food production, climate regulation, purification of water, maintenance of nutrients, healthy soils, carbon sinks, biodiversity, and aesthetic landscape function. NPP dictates the amount of carbon synthesised within an ecosystem, which is ultimately available to consumers, including humans. In fact, associated with increased population growth over the last millennium, a disproportionate amount of the world’s NPP is now consumed by humans.





Figure 11.2.1: A flux tower in Kruger National Park South Africa, measures the "breathing" of an African savanna ecosystem. During summer days, CO_2 is taken up by the vegetation leading to an increase in overall biomass. Some of this biomass is lost to herbivory and microbial decay, but most is lost to the frequent fires that are an integral component of the ecosystem. An important question in the context of climate change is whether this system is in long-term equilibrium, or if the carbon store is increasing or decreasing. The flux tower helps scientists understand the dynamics of the system, and trends in carbon dynamics, as well as being a useful tool in calibrating satellite-based NPP products.
Source: photo from and text based on: https://www.csir.co.za/eddy-covariance-flux-towers

Humans have major impacts on NPP through the use of irrigation and fertilisers. It is the loss of NPP through actions such as increased soil erosion, deforestation and soil salinisation that forms the basis for many forms of land degradation. There are also forms of degradation where NPP may stay constant, or even increase, but where important ecosystem services change. Examples include plant species, compositional changes in response to grazing pressure where palatable grasses are replaced by less palatable ones, or in some cases where palatable grasses are replaced by unpalatable woody shrubs. In such circumstances the grazing capacity of the rangeland may be greatly reduced, biodiversity can be lost, but carbon sequestration and other regulating ecosystem services may be maintained or even enhanced. CO_2 fertilisation is having an impact on Leaf Area Index (LAI) (and hence NPP) over vast areas of the Earth. 70 % of the observed impacts on LAI to CO_2 fertilisation effects, with factors such as nitrogen deposition, climate change-induced rainfall and temperature as well as land-cover change being responsible for the remaining observed changes.

Source

European Commission, Joint Research Center, World Atlas of Desertification, Limits to Sustainability, Net Primary Production. Accessed December 2023 https://wad.jrc.ec.europa.eu/primaryproduction

11.2: What might stimulate net primary production? is shared under a Public Domain license and was authored, remixed, and/or curated by European Commission, Joint Research Center.

11.3: Temporal perspectives of the carbon cycle

Left unperturbed, the fast and slow carbon cycles maintain a relatively steady concentration of carbon in the atmosphere, land, plants, and ocean. But when anything changes the amount of carbon in one reservoir, the effect ripples through the others.

In Earth's past, the carbon cycle has changed in response to climate change. Variations in Earth's orbit alter the amount of energy Earth receives from the Sun and leads to a cycle of ice ages and warm periods like Earth's current climate. (See Milutin Milankovitch.) Ice ages developed when Northern Hemisphere summers cooled and ice built up on land, which in turn slowed the carbon cycle. Meanwhile, a number of factors including cooler temperatures and increased phytoplankton growth may have increased the amount of carbon the ocean took out of the atmosphere. The drop in atmospheric carbon caused additional cooling. Similarly, at the end of the last Ice Age, 10,000 years ago, carbon dioxide in the atmosphere rose dramatically as temperatures warmed.

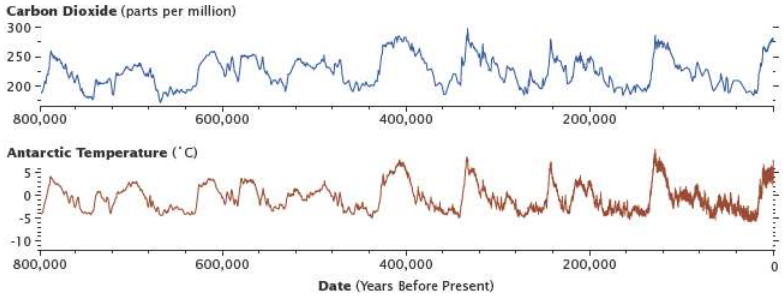


Figure 11.3.1: Levels of carbon dioxide in the atmosphere have corresponded closely with temperature over the past 800,000 years. Although the temperature changes were touched off by variations in Earth's orbit, the increased global temperatures released CO_2 into the atmosphere, which in turn warmed the Earth. Antarctic ice-core data show the long-term correlation until about 1900. (Graphs by Robert Simmon, using data from Lüthi et al., 2008, and Jouzel et al., 2007.)

Shifts in Earth's orbit are happening constantly, in predictable cycles. In about 30,000 years, Earth's orbit will have changed enough to reduce sunlight in the Northern Hemisphere to the levels that led to the last ice age.

Today, changes in the carbon cycle are happening because of people. We perturb the carbon cycle by burning fossil fuels and clearing land.

When we clear forests, we remove a dense growth of plants that had stored carbon in wood, stems, and leaves—biomass. By removing a forest, we eliminate plants that would otherwise take carbon out of the atmosphere as they grow. We tend to replace the dense growth with crops or pasture, which store less carbon. We also expose soil that vents carbon from decayed plant matter into the atmosphere. Humans are currently emitting just under a billion tons of carbon into the atmosphere per year through land use changes.

Without human interference, the carbon in fossil fuels would leak slowly into the atmosphere through volcanic activity over millions of years in the slow carbon cycle. By burning coal, oil, and natural gas, we accelerate the process, releasing vast amounts of carbon (carbon that took millions of years to accumulate) into the atmosphere every year. By doing so, we move the carbon from the slow cycle to the fast cycle. In 2009, humans released about 8.4 billion tons of carbon into the atmosphere by burning fossil fuel.

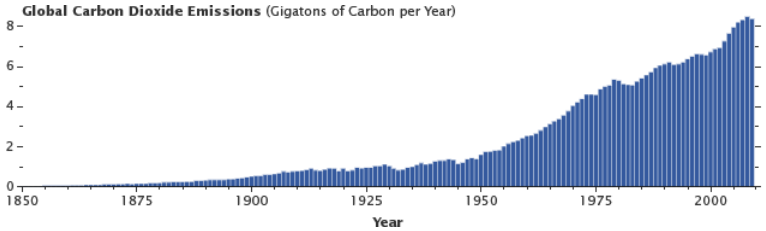


Figure 11.3.2: Emissions of carbon dioxide by humanity (primarily from the burning of fossil fuels, with a contribution from cement production) have been growing steadily since the onset of the industrial revolution. About half of these emissions are removed by the fast carbon cycle each year, the rest remain in the atmosphere. (Graph by Robert Simmon, using data from the Carbon Dioxide Information Analysis Center and Global Carbon Project.)

Since the beginning of the Industrial Revolution, when people first started burning fossil fuels, carbon dioxide concentrations in the atmosphere have risen from about 280 parts per million to 387 parts per million, a 39 percent increase. This means that for every million molecules in the atmosphere, 387 of them are now carbon dioxide—the highest concentration in two million years. Methane concentrations have risen from 715 parts per billion in 1750 to 1,774 parts per billion in 2005, the highest concentration in at least 650,000 years.

Effects of Changing the Carbon Cycle

All of this extra carbon needs to go somewhere. So far, land plants and the ocean have taken up about 55 percent of the extra carbon people have put into the atmosphere while about 45 percent has stayed in the atmosphere. Eventually, the land and oceans will take up most of the extra carbon dioxide, but as much as 20 percent may remain in the atmosphere for many thousands of years.

The changes in the carbon cycle impact each reservoir. Excess carbon in the atmosphere warms the planet and helps plants on land grow more. Excess carbon in the ocean makes the water more acidic, putting marine life in danger.

Atmosphere

It is significant that so much carbon dioxide stays in the atmosphere because CO_2 is the most important gas for controlling Earth's temperature. Carbon dioxide, methane, and halocarbons are greenhouse gases that absorb a wide range of energy—including infrared energy (heat) emitted by the Earth—and then re-emit it. The re-emitted energy travels out in all directions, but some returns to Earth, where it heats the surface. Without greenhouse gases, Earth would be a frozen -18 degrees Celsius (0 degrees Fahrenheit). With too many greenhouse gases, Earth would be like Venus, where the greenhouse atmosphere keeps temperatures around 400 degrees Celsius (750 Fahrenheit).

Because scientists know which wavelengths of energy each greenhouse gas absorbs, and the concentration of the gases in the atmosphere, they can calculate how much each gas contributes to warming the planet. Carbon dioxide causes about 20 percent of Earth's greenhouse effect; water vapor accounts for about 50 percent; and clouds account for 25 percent. The rest is caused by small particles (aerosols) and minor greenhouse gases like methane.

Water vapor concentrations in the air are controlled by Earth's temperature. Warmer temperatures evaporate more water from the oceans, expand air masses, and lead to higher humidity. Cooling causes water vapor to condense and fall out as rain, sleet, or snow.

Carbon dioxide, on the other hand, remains a gas at a wider range of atmospheric temperatures than water. Carbon dioxide molecules provide the initial greenhouse heating needed to maintain water vapor concentrations. When carbon dioxide concentrations drop, Earth cools, some water vapor falls out of the atmosphere, and the greenhouse warming caused by water vapor drops. Likewise, when carbon dioxide concentrations rise, air temperatures go up, and more water vapor evaporates into the atmosphere—which then amplifies greenhouse heating.

So while carbon dioxide contributes less to the overall greenhouse effect than water vapor, scientists have found that carbon dioxide is the gas that sets the temperature. Carbon dioxide controls the amount of water vapor in the atmosphere and thus the size of the greenhouse effect.

Rising carbon dioxide concentrations are already causing the planet to heat up. At the same time that greenhouse gases have been increasing, average global temperatures have risen 0.8 degrees Celsius (1.4 degrees Fahrenheit) since 1880.

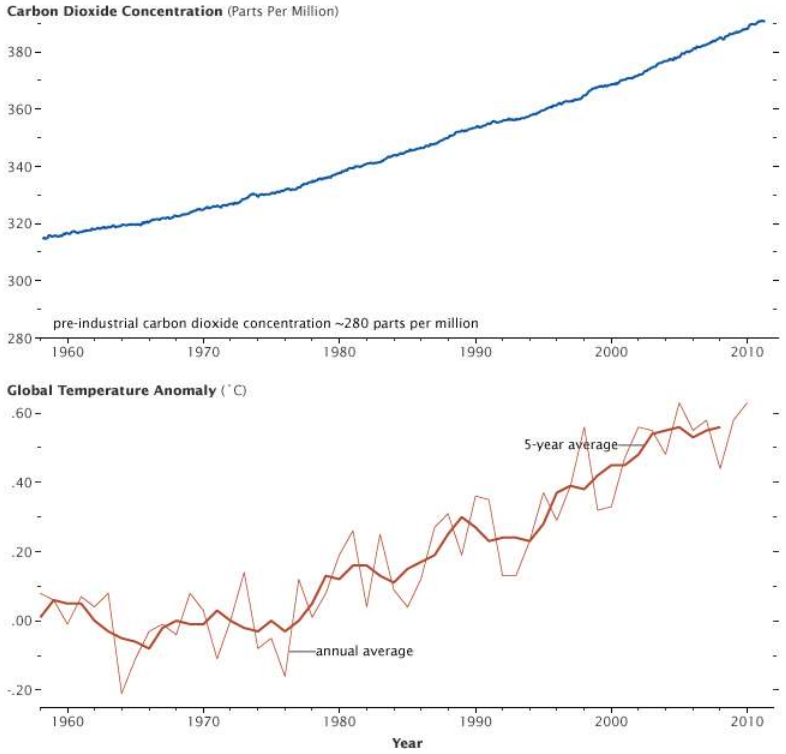


Figure 11.3.3: With the seasonal cycle removed, the atmospheric carbon dioxide concentration measured at Mauna Loa Volcano, Hawaii, shows a steady increase since 1957. At the same time global average temperatures are rising as a result of heat trapped by the additional CO_2 and increased water vapor concentration. (Graphs by Robert Simmon, using CO_2 data from the NOAA Earth System Research Laboratory and temperature data from the Goddard Institute for Space Studies.)

This rise in temperature isn't all the warming we will see based on current carbon dioxide concentrations. Greenhouse warming doesn't happen right away because the ocean soaks up heat. This means that Earth's temperature will increase at least another 0.6 degrees Celsius (1 degree Fahrenheit) because of carbon dioxide already in the atmosphere. The degree to which temperatures go up beyond that depends in part on how much more carbon humans release into the atmosphere in the future.

Ocean

About 30 percent of the carbon dioxide that people have put into the atmosphere has diffused into the ocean through the direct chemical exchange. Dissolving carbon dioxide in the ocean creates carbonic acid, which increases the acidity of the water. Or rather, a slightly alkaline ocean becomes a little less alkaline. Since 1750, the pH of the ocean's surface has dropped by 0.1, a 30 percent change in acidity.

Ocean acidification affects marine organisms in two ways. First, carbonic acid reacts with carbonate ions in the water to form bicarbonate. However, those same carbonate ions are what shell-building animals like coral need to create calcium carbonate shells. With less carbonate available, the animals need to expend more energy to build their shells. As a result, the shells end up being thinner and more fragile.

Second, the more acidic water is, the better it dissolves calcium carbonate. In the long run, this reaction will allow the ocean to soak up excess carbon dioxide because more acidic water will dissolve more rock, release more carbonate ions, and increase the ocean's capacity to absorb carbon dioxide. In the meantime, though, more acidic water will dissolve the carbonate shells of marine organisms, making them pitted and weak.

Warmer oceans—a product of the greenhouse effect—could also decrease the abundance of phytoplankton, which grow better in cool, nutrient-rich waters. This could limit the ocean's ability to take carbon from the atmosphere through the fast carbon cycle.

On the other hand, carbon dioxide is essential for plant and phytoplankton growth. An increase in carbon dioxide could increase growth by fertilizing those few species of phytoplankton and ocean plants (like sea grasses) that take carbon dioxide directly from the water. However, most species are not helped by the increased availability of carbon dioxide.

Land

Plants on land have taken up approximately 25 percent of the carbon dioxide that humans have put into the atmosphere. The amount of carbon that plants take up varies greatly from year to year, but in general, the world's plants have increased the amount of carbon dioxide they absorb since 1960. Only some of this increase occurred as a direct result of fossil fuel emissions.

With more atmospheric carbon dioxide available to convert to plant matter in photosynthesis, plants were able to grow more. This increased growth is referred to as carbon fertilization. Models predict that plants might grow anywhere from 12 to 76 percent more if atmospheric carbon dioxide is doubled, as long as nothing else, like water shortages, limits their growth. However, scientists don't know how much carbon dioxide is increasing plant growth in the real world, because plants need more than carbon dioxide to grow.

Plants also need water, sunlight, and nutrients, especially nitrogen. If a plant doesn't have one of these things, it won't grow regardless of how abundant the other necessities are. There is a limit to how much carbon plants can take out of the atmosphere, and that limit varies from region to region. So far, it appears that carbon dioxide fertilization increases plant growth until the plant reaches a limit in the amount of water or nitrogen available.

Some of the changes in carbon absorption are the result of land use decisions. Agriculture has become much more intensive, so we can grow more food on less land. In high and mid-latitudes, abandoned farmland is reverting to forest, and these forests store much more carbon, both in wood and soil, than crops would. In many places, we prevent plant carbon from entering the atmosphere by extinguishing wildfires. This allows woody material (which stores carbon) to build up. All of these land use decisions are helping plants absorb human-released carbon in the Northern Hemisphere.

Changes in land cover—forests converted to fields and fields converted to forests—have a corresponding effect on the carbon cycle. In some Northern Hemisphere countries, many farms were abandoned in the early 20th century and the land reverted to forest. As a result, carbon was drawn out of the atmosphere and stored in trees on land.

In the tropics, however, forests are being removed, often through fire, and this releases carbon dioxide. As of 2008, deforestation accounted for about 12 percent of all human carbon dioxide emissions.

The biggest changes in the land carbon cycle are likely to come because of climate change. Carbon dioxide increases temperatures, extending the growing season and increasing humidity. Both factors have led to some additional plant growth. However, warmer temperatures also stress plants. With a longer, warmer growing season, plants need more water to survive. Scientists are already seeing evidence that plants in the Northern Hemisphere slow their growth in the summer because of warm temperatures and water shortages.

Dry, water-stressed plants are also more susceptible to fire and insects when growing seasons become longer. In the far north, where an increase in temperature has the greatest impact, the forests have already started to burn more, releasing carbon from the plants and the soil into the atmosphere. Tropical forests may also be extremely susceptible to drying. With less water, tropical trees slow their growth and take up less carbon, or die and release their stored carbon to the atmosphere.

The warming caused by rising greenhouse gases may also "bake" the soil, accelerating the rate at which carbon seeps out in some places. This is of particular concern in the far north, where frozen soil—permafrost—is thawing. Permafrost contains rich deposits of carbon from plant matter that has accumulated for thousands of years because the cold slows decay. When the soil warms, the organic matter decays and carbon—in the form of methane and carbon dioxide—seeps into the atmosphere.

Current research estimates that permafrost in the Northern Hemisphere holds 1,672 billion tons (Petagrams) of organic carbon. If just 10 percent of this permafrost were to thaw, it could release enough extra carbon dioxide to the atmosphere to raise temperatures an additional 0.7 degrees Celsius (1.3 degrees Fahrenheit) by 2100.

Excerpted from

Holli Riebeek, *Design by Robert Simmon*, The Carbon Cycle, NASA, Earth Observatory Accessed December 2023 https://earthobservatory.nasa.gov/features/CarbonCycle/page1.php, Public domain

11.3: Temporal perspectives of the carbon cycle is shared under a Public Domain license and was authored, remixed, and/or curated by Holli Riebeek, Design by Robert Simmon, The Carbon Cycle, NASA, Earth Observatory.

11.4: Atmospheric methane

Total global CH_4 emissions are approximately 550 teragrams (Tg) of CH_4 per year (1 Tg CH_4 per year = 1012 grams of CH_4 per year; Saunois et al., 2016). Of this, roughly 40% comes from natural sources. The largest (and most uncertain) natural emissions of CH_4 are from wetlands, defined as regions that are permanently or seasonally waterlogged. Natural wetlands include high-latitude bogs and fens, tropical swamps, and temperate wetlands. Saturated soils in warm tropical environments tend to produce the most CH_4. However, warming Arctic temperatures raise concerns of increasing emissions from high-latitude wetlands and future decomposition of carbon currently stored in frozen Arctic soils (e.g., Schaefer et al., 2011; Schuur et al., 2015). Figure 11.4.1, provides a pictorial representation of the main components of the global methane cycle.

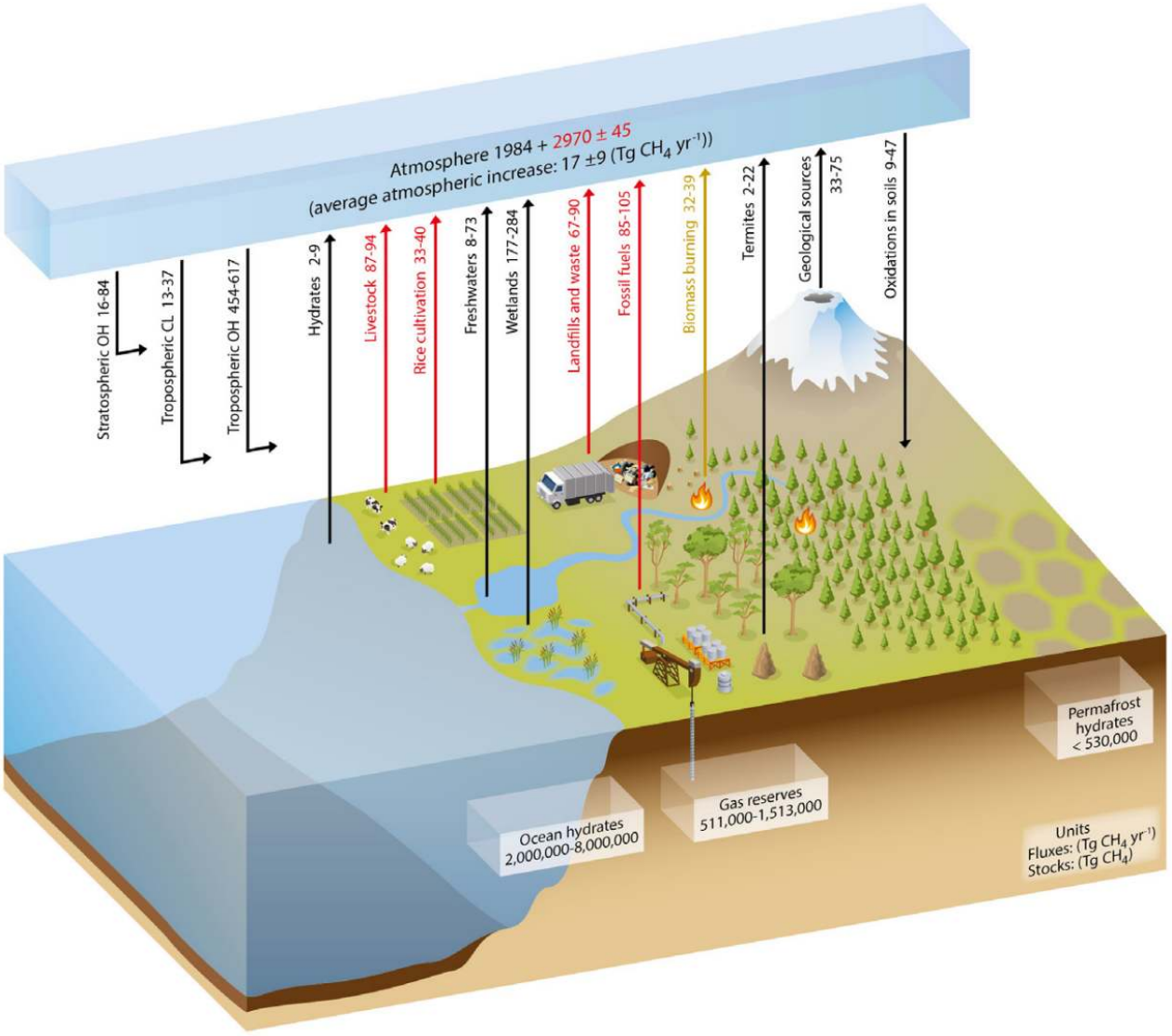


Figure 11.4.1: A Pictorial Illustration of the Global Methane (CH4) Cycle. The arrows and boxed numbers represent annual fluxes in teragrams (Tg) of CH4 per year estimated from 2000 to 2009 and CH4 reservoirs in Tg CH4. Reservoirs include the atmosphere and three geological reservoirs (i.e., hydrates on land and in the ocean floor and gas reserves). The black arrows show natural emissions, while red arrows show anthropogenic fluxes. The brown arrow represents total anthropogenic and natural emissions. [Figure source: Reprinted from Ciais et al., 2013, Figure 6.2. Copyright IPCC, used with permission.]

Estimates of global CH_4 emissions from wetlands range from 127 to 227 Tg CH_4 per year (Saunois et al., 2016), with most probable values between 167 and 185 Tg CH_4 per year. Most emissions occur in tropical regions (Matthews 1989; Melton et al., 2013; Saunois et al., 2016). Currently, only about 25 Tg CH_4 per year (i.e., 4% of global emissions) are thought to be emitted from

high northern latitudes (AMAP 2015; Saunois et al., 2016). Because emissions are sensitive to temperature and precipitation they exhibit significant seasonal cycles, especially at high latitudes, as well as interannual variability caused by moisture and temperature variability. Smaller amounts of CH4 are emitted from fires, the ocean, and enteric fermentation in termites and wild animals (20 Tg CH4 per year or less for each). In addition, up to 60 Tg CH4 per year may be emitted from geological sources, such as seeps, clathrates, mud volcanoes, and geothermal systems (Etiope et al., 2008; Schwietzke et al., 2016).

Table 11.4.1: Historic[a] and Decadal[b] Global Mean Emissions and Their Partitioning to the Carbon Reservoirs of Atmosphere, Ocean, and Land 1750–2011 Cumulative Pg Cc 1980–1989 Pg C per Year 1990–1999 Pg C per Year 2000–2009 Pg C per Year 2007–201

	1750–2011 Cumulative Pg Cc	1980–1989 Pg C per Year	1990–1999 Pg C per Year	2000–2009 Pg C per Year	2007–2016 Pg C per Year	2016 Pg C per Year
Emissions						
Fossil Fuels and Industry	375 ± 30	5.5 ± 0.3	6.3 ± 0.3	7.8 ± 0.4	9.4 ± 0.5	9.9 ± 0.5
Land-Use Change	180 ± 80	1.2 ± 0.7	1.3 ± 0.7	1.2 ± 0.7	1.3 ± 0.7	1.3 ± 0.7
Partitioning to Carbon Reservoir						
Partitioning to Carbon Reservoir	240 ± 10	3.4 ± 0.1	3.1 ± 0.1 4	4.0 ± 0.1	4.7 ± 0.1	6.0 ± 0.2
Ocean Uptake	160 ± 80	1.7 ± 0.5	1.9 ± 0.5	2.1 ± 0.5 2	2.4 ± 0.5	2.6 ± 0.5
Land Uptake	155 ± 30	2.0 ± 0.6	2.5 ± 0.5	2.9 ± 0.8 3.0	3.0 ± 0.8	2.7 ± 0.9

a. Historic cumulative emissions and partitioning from the Intergovernmental Panel on Climate Change Fifth Assessment Report (Ciais et al., 2013).
b. Decadal means from the Global Carbon Project (Le Quéré et al., 2017).
c. Pg C, petagrams of carbon; CO_2, carbon dioxide.

Unlike CO2, CH4 has an atmospheric chemical sink that nearly balances total global emissions. Removal of atmospheric CH4 by reaction with the hydroxyl radical (OH) results in a CH4 atmospheric lifetime of about 9 to 10 years. Observationally constrained estimates of CH4 lifetime suggest either small decreases of about 2% from 1980 to 2005 (Holmes et al., 2013) or stable CH4 lifetimes with the possibility of interannual variability of about 2% (Montzka et al., 2011). CH4 is a much more powerful greenhouse gas than CO2 (on a per mass basis and over 100 years, CH4 is about 25 times more effective at trapping heat than CO2).

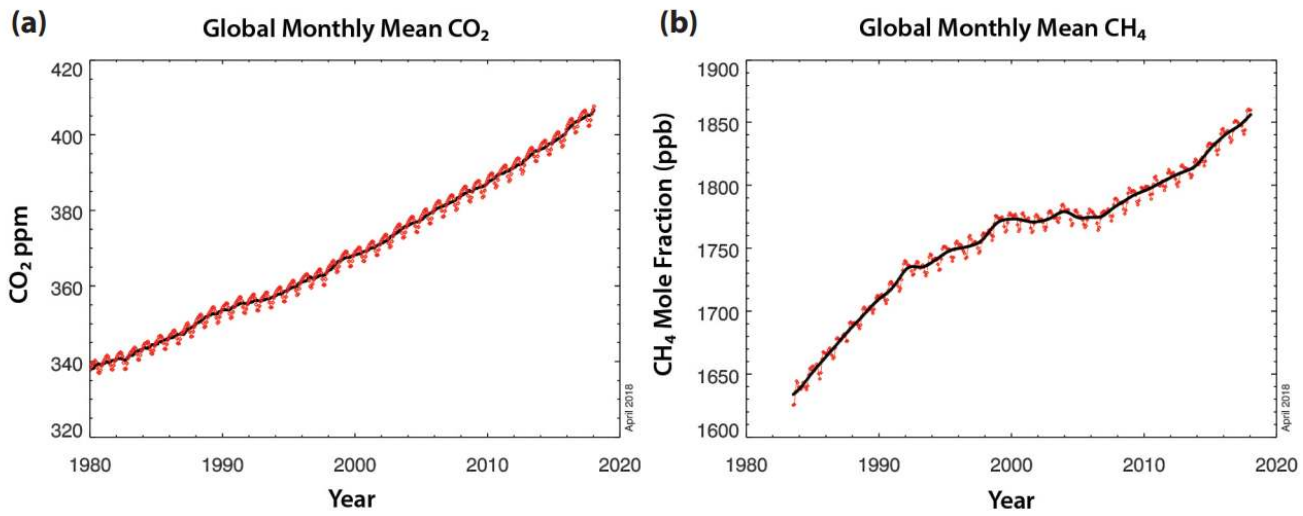


Figure 11.4.2: . Global Averages of Atmospheric Gases Derived from Surface Air Samples. (a) Carbon dioxide (CO2) in parts per million (ppm). (b) Methane (CH4) in parts per billion (ppb). [Figure source: Redrawn from NOAA-ESRLGMD 2017.]

As shown in Figure 11.4.2b, this page, atmospheric CH4 increased rapidly during the 1980s and early 1990s before its growth leveled off between the mid-1990s and early 2000s. Methane has resumed its increase in the atmosphere since 2006, and observations show that this growth has even accelerated since 2014. The changing atmospheric CH4 growth rate has been the subject of much debate, questioning why growth rate slowed for a decade starting in the mid-1990s. Several studies suggested that this slower rate was due to decreases in fugitive emissions from fossil fuel production (Aydin et al., 2011; Simpson et al., 2012) or to decreased emissions from anthropogenic microbial sources, such as rice agriculture (Kai et al., 2011). On the other hand, Dlugokencky et al. (1998, 2003) proposed that the slowing of CH4 growth in the atmosphere was due to an approach to a quasi–steady state, reached when global sources and sinks are in balance. Consistent with this view, the study of Schwietzke et al. (2016) found that emissions from oil and gas production have remained stable over the past several decades, implying increasing efficiency in fossil fuel production industries while their production was increasing over time.

Dlugokencky et al. (2003) predicted that CH4 would approach a steady state in the atmosphere of about 1,780 ppb by the 2010s if there were no major changes in its budget. The methane budget did change, however, because the atmospheric growth of CH4 resumed its rise in 2006. The cause of the recent increase in CH4 growth also has been much debated. Based on global observations of the CH4 isotope, 13CH4, the global growth in CH4 appears likely to have been dominated by microbial sources in the tropics (wetlands or agriculture and waste), rather than fossil fuel production (Nisbet et al., 2016; Schaefer et al., 2016), as suggested by some studies (e.g., Rice et al., 2016). Other studies have argued that 13CH4 may not be a very strong constraint on the global methane budget and that changes in the atmospheric CH4 chemical sink are responsible for the global methane changes (Rigby et al., 2017; Turner et al., 2017). However, plausible chemical mechanisms that could explain the changes in the CH4 sink have not been identified. Using space-based retrievals of carbon monoxide, Worden et al. (2017) argued that the isotopic data record also can be consistent with increased fossil fuel emissions if global biomass-burning emissions have decreased twice as much as estimates based on space-based observations of burned areas. If the recent rise of global atmospheric CH4 is indeed due to increases in microbial emissions, then the question becomes whether anthropogenic or natural microbial sources are responsible. Some studies have suggested that anthropogenic microbial sources, such as livestock, are behind the increased atmospheric growth of CH4 (Schaefer et al., 2016; Saunois et al., 2016). If the increase is due to emissions from wetlands, especially in the tropics, then this raises the possibility that changing climate could be changing natural emissions.

Methane Perturbations in the Global Carbon Cycle

The carbon cycle undergoes perturbations caused by a variety of natural processes such as wildfires, droughts, insect infestations, and disease. These processes can themselves be affected by human activities, for example through GHG emissions that change climate, wildfire suppression, and land-use change. During longer periods, variations in the Earth's orbit also drive significant perturbations to the global carbon cycle. Over the recent several centuries, human activity has resulted in perturbations to the

carbon cycle that have no precedent in geological records. Anthropogenic emissions also can directly alter the chemistry of the atmosphere, possibly affecting its ability to remove pollutants.

Atmospheric CH4 also is influenced by diverse human activities, ranging from food production (e.g., ruminants and rice) to waste (e.g., sewage and landfills) to fossil fuel production (e.g., coal, oil, and gas). Future increases in population likely will increase CH4 emissions from agriculture and waste as demand rises for more food production. Furthermore, the current boom in shale oil and gas exploitation has focused attention on leakage from drilling, storage, and transport of fossil fuel (e.g., Peischl et al., 2015; Pétron et al., 2014). Chemical reaction with OH accounts for about 90% of the total CH4 sink (Ehhalt 1974). These OH radicals, produced through the photolysis of ozone (O3) in the presence of water vapor, are destroyed by reactions with CH4 and other compounds. Uncertainty in the sink due to chemical loss by OH is 10% to 20%, because the OH distribution remains uncertain at regional to global scales (Saunois et al., 2016).

Current estimates reported by Saunois et al. (2016) for anthropogenic emissions average 328 Tg CH4 per year (ranging from 259 to 370 Tg CH4 per year). Extraction and processing of fossil fuels account for 32% to 34% of all anthropogenic emissions. Livestock, agriculture, landfills, and sewage together account for another 55% to 57%, with the remainder due to biomass and biofuel burning. A recent study using observations of the isotopic composition of CH4 suggests that emissions from fossil fuel production and geological emissions may be 20% to 60% higher than previously thought. This increase would require a compensating reduction in microbial emissions from natural and anthropogenic sources (Schwietzke et al., 2016) for the atmosphere to be in balance with the observed global average CH4 abundance. Current CH4 levels are unprecedented in over at least 800,000 years (Loulergue et al., 2008). Recent National Oceanic and Atmospheric Administration atmospheric network observations have shown that global CH4 increased rapidly through the late 1990s, leveled off during the early 2000s, and began to increase again in 2007 (Dlugokencky et al., 2009; Rigby et al., 2008). These changes in global CH4 are not well understood and are under debate. Although Dlugokencky et al. (1998, 2003) suggested that the plateau in CH4 growth resulted from an approximate balance between global sources and sinks, some studies suggested that decreases in anthropogenic emissions (Aydin et al., 2011; Kai et al., 2011; Simpson et al., 2012) led to the period of slow CH4 growth. Isotopic evidence points toward increased emissions from microbial sources as an explanation for the recent rise in global CH4 (Nisbet et al., 2016; Schaefer et al., 2016; Schwietzke et al., 2016). However, increases in anthropogenic emissions also have been proposed (Rice et al., 2016), as well as decreases in the chemical loss (Rigby et al., 2017; Turner et al., 2017). Worden et al. (2017) have recently suggested a significant role for fossil fuel emissions in the recent growth of atmospheric CH4 based on decreases in biomass burning that could change the interpretation of methane isotope observations. This result is based on space-based observations of atmospheric CO, which itself may be responding to changes in other sources besides biomass burning. Figure 1.1, p. 45, shows that CH4 contributed just over 0.5 W/m2 in 2017 to global total anthropogenic radiative forcing, an amount which is about one-fourth of that from CO2. Although CH4 is much more effective at absorbing infrared radiation (Hofmann et al., 2006; Myhre et al., 2013),3 it is about a hundred times less abundant in the atmosphere than CO2.

Future Changes in Atmospheric Methane

Coupled carbon cycle–climate models forced with future “business as usual” emissions scenarios suggest that the changing carbon cycle will be a net positive feedback on climate, reinforcing warming, but the size of the projected feedback is highly uncertain (Friedlingstein et al., 2014). Besides the uncertain trajectories of human factors such as fossil fuel emissions, land use, or significant mitigation efforts, various natural processes can lead to the carbon cycle being a positive feedback. For example, a warming climate can lead to increased fires and droughts and less storage of carbon in the terrestrial biosphere. In particular, warming is expected to decrease carbon uptake in the tropics and midlatitudes. In the high latitudes, a warmer climate is expected to lead to a more productive biosphere and more uptake but also may result in increased respiration and release of stored CO2 and CH4 in soils and lakes. Negative feedbacks also are possible, such as increased atmospheric CO2, leading to increased carbon storage in the terrestrial biosphere (e.g., Schimel et al., 2015), although the relative roles of this effect relative to land-use change, nitrogen deposition, and temperature increases on the cumulative land carbon sink over the last century are not fully understood (Huntzinger et al., 2017). Human impacts on land use can directly impact climate. Deforestation and agriculture can affect carbon storage in soil and biomass. Fertilizer use also affects the global nitrogen budget and can increase carbon storage. Large-scale drainage of wetlands and conversion to agricultural land can reduce CH4 emissions from anaerobic respiration while potentially increasing faster soil carbon loss through aerobic respiration.

Frozen Arctic soils compose another potential carbon cycle–climate feedback (see Ch. 11: Arctic and Boreal Carbon, p. 428, and Ch. 19: Future of the North American Carbon Cycle, p. 760). An estimated 1,460 to 1,600 Pg C are frozen in Arctic soils, and warming has proceeded in the Arctic faster than in any other region. Current understanding suggests that approximately 146 to 160 Pg C, primarily as CO2, could be vulnerable to thaw and release to the atmosphere over the next century (Schuur et al., 2015; see Ch. 11: Arctic and Boreal Carbon). This release of carbon from permafrost is likely to be gradual and occur on century timescales (Schuur et al., 2015). If the amount of carbon estimated to enter the atmosphere by Schuur et al. (2015) were released annually at a constant rate, emissions would be far lower than annual fossil fuel emissions (about 9 Pg C per year) but comparable to land-use change (0.9 Pg C per year).

References

AMAP, 2015: AMAP Assessment 2015: Methane as an Arctic Climate Forcer. Arctic Monitoring and Assessment Programme, 146 pp

Aydin, M., K. R. Verhulst, E. S. Saltzman, M. O. Battle, S. A. Montzka, D. R. Blake, Q. Tang, and M. J. Prather, 2011: Recent decreases in fossil-fuel emissions of ethane and methane derived from firn air. Nature, 476(7359), 198-201, doi: 10.1038/ nature10352.

Dlugokencky, E. J., L. Bruhwiler, J. W. C. White, L. K. Emmons, P. C. Novelli, S. A. Montzka, K. A. Masarie, P. M. Lang, A. M. Crotwell, J. B. Miller, and L. V. Gatti, 2009: Observational constraints on recent increases in the atmospheric CH4 burden. Geophysical Research Letters, 36(18), doi: 10.1029/2009GL039780.

Dlugokencky, E. J., 2003: Atmospheric methane levels off: Temporary pause or a new steady-state? Geophysical Research Letters, 30(19), doi: 10.1029/2003gl018126.

Dlugokencky, E. J., K. A. Masarie, P. M. Lang, and P. P. Tans, 1998: Continuing decline in the growth rate of the atmospheric methane burden. Nature, 393(6684), 447-450, doi: 10.1038/30934.

Ehhalt, D. H., 1974: The atmospheric cycle of methane. Tellus A: Dynamic Meteorology and Oceanography, 26(1-2), 58-70, doi: 10.1111/j.2153-3490.1974.tb01952.x.

Etiope, G., K. R. Lassey, R. W. Klusman, and E. Boschi, 2008: Reappraisal of the fossil methane budget and related emission from geologic sources. Geophysical Research Letters, 35(9), doi: 10.1029/2008gl033623.

Hofmann, D. J., J. H. Butler, T. J. Conway, E. J. Dlugokencky, J. W. Elkins, K. Masarie, S. A. Montzka, R. C. Schnell, and P. Tans, 2006: Tracking climate forcing: The annual greenhouse gas index. Eos, Transactions American Geophysical Union, 87(46), 509, doi: 10.1029/2006eo460002.

Holmes, C. D., M. J. Prather, O. A. Søvde, and G. Myhre, 2013: Future methane, hydroxyl, and their uncertainties: Key climate and emission parameters for future predictions. Atmospheric Chemistry and Physics, 13(1), 285-302, doi: 10.5194/acp-13-285-2013.

Kai, F. M., S. C. Tyler, J. T. Randerson, and D. R. Blake, 2011: Reduced methane growth rate explained by decreased Northern Hemisphere microbial sources. Nature, 476(7359), 194-197, doi: 10.1038/nature10259.

Loulergue, L., A. Schilt, R. Spahni, V. Masson-Delmotte, T. Blunier, B. Lemieux, J. M. Barnola, D. Raynaud, T. F. Stocker, and J. Chappellaz, 2008: Orbital and millennial-scale features of atmospheric CH4 over the past 800,000 years. Nature, 453(7193), 383-386, doi: 10.1038/nature06950.

Matthews, E., 1989: Global Data Bases on Distribution, Characteristics and Methane Emission of Natural Wetlands: Documentation of Archived Data Tape. National Aeronautics and Space Administration. NASA TM-4153.

Melton, J. R., R. Wania, E. L. Hodson, B. Poulter, B. Ringeval, R. Spahni, T. Bohn, C. A. Avis, D. J. Beerling, G. Chen, A. V. Eliseev, S. N. Denisov, P. O. Hopcroft, D. P. Lettenmaier, W. J. Riley, J. S. Singarayer, Z. M. Subin, H. Tian, S. Zürcher, V. Brovkin, P. M. van Bodegom, T. Kleinen, Z. C. Yu, and J. O. Kaplan, 2013: Present state of global wetland extent and wetland methane modelling: Conclusions from a model inter-comparison project (WETCHIMP). Biogeosciences, 10(2), 753-788, doi: 10.5194/ bg-10-753-2013.

Montzka, S. A., M. Krol, E. Dlugokencky, B. Hall, P. Jockel, and J. Lelieveld, 2011: Small interannual variability of global atmospheric hydroxyl. Science, 331(6013), 67-69, doi: 10.1126/ science.1197640.

Myhre, G., D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura, and H. Zhang, 2013: Anthropogenic and natural radiative forcing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. [T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA.

Nisbet, E. G., E. J. Dlugokencky, M. R. Manning, D. Lowry, R. E. Fisher, J. L. France, S. E. Michel, J. B. Miller, J. W. C. White, B. Vaughn, P. Bousquet, J. A. Pyle, N. J. Warwick, M. Cain, R. Brownlow, G. Zazzeri, M. Lanoisellé, A. C. Manning, E. Gloor, D. E. J. Worthy, E. G. Brunke, C. Labuschagne, E. W. Wolff, and A. L. Ganesan, 2016: Rising atmospheric methane: 2007-2014 growth and isotopic shift. Global Biogeochemical Cycles, 30(9), 1356-1370, doi: 10.1002/2016gb005406.

Peischl, J., T. B. Ryerson, K. C. Aikin, J. A. de Gouw, J. B. Gilman, J. S. Holloway, B. M. Lerner, R. Nadkarni, J. A. Neuman, J. B. Nowak, M. Trainer, C. Warneke, and D. D. Parrish, 2015: Quantifying atmospheric methane emissions from the Haynesville, Fayetteville, and northeastern Marcellus shale gas production regions. Journal of Geophysical Research: Atmospheres, 120(5), 2119-2139, doi: 10.1002/2014jd022697.

Pétron, G., A. Karion, C. Sweeney, B. R. Miller, S. A. Montzka, G. J. Frost, M. Trainer, P. Tans, A. Andrews, J. Kofler, D. Helmig, D. Guenther, E. Dlugokencky, P. Lang, T. Newberger, S. Wolter, B. Hall, P. Novelli, A. Brewer, S. Conley, M. Hardesty, R. Banta, A. White, D. Noone, D. Wolfe, and R. Schnell, 2014: A new look at methane and nonmethane hydrocarbon emissions from oil and natural gas operations in the Colorado Denver-Julesburg Basin. Journal of Geophysical Research: Atmospheres, 119(11), 6836-6852, doi: 10.1002/2013jd021272.

Rice, A. L., C. L. Butenhoff, D. G. Teama, F. H. Roger, M. A. Khalil, and R. A. Rasmussen, 2016: Atmospheric methane isotopic record favors fossil sources flat in 1980s and 1990s with recent increase. Proceedings of the National Academy of Sciences USA, 113(39), 10791-10796, doi: 10.1073/pnas.1522923113.

Rigby, M., R. G. Prinn, P. J. Fraser, P. G. Simmonds, R. L. Langenfelds, J. Huang, D. M. Cunnold, L. P. Steele, P. B. Krummel, R. F. Weiss, S. O'Doherty, P. K. Salameh, H. J. Wang, C. M. Harth, J. Mühle, and L. W. Porter, 2008: Renewed growth of atmospheric methane. Geophysical Research Letters, 35(22), doi: 10.1029/2008gl036037.

Saunois, M., P. Bousquet, B. Poulter, A. Peregon, P. Ciais, J. G. Canadell, E. J. Dlugokencky, G. Etiope, D. Bastviken, S. Houweling, G. Janssens-Maenhout, F. N. Tubiello, S. Castaldi, R. B. Jackson, M. Alexe, V. K. Arora, D. J. Beerling, P. Bergamaschi, D. R. Blake, G. Brailsford, V. Brovkin, L. Bruhwiler, C. Crevoisier, P. Crill, K. Covey, C. Curry, C. Frankenberg, N. Gedney, L. Höglund-Isaksson, M. Ishizawa, A. Ito, F. Joos, H.-S. Kim, T. Kleinen, P. Krummel, J.-F. Lamarque, R. Langenfelds, R. Locatelli, T. Machida, S. Maksyutov, K. C. McDonald, J. Marshall, J. R. Melton, I. Morino, V. Naik, S. O'Doherty, F.-J. W. Parmentier, P. K. Patra, C. Peng, S. Peng, G. P. Peters, I. Pison, C. Prigent, R. Prinn, M. Ramonet, W. J. Riley, M. Saito, M. Santini, R. Schroeder, I. J. Simpson, R. Spahni, P. Steele, A. Takizawa, B. F. Thornton, H. Tian, Y. Tohjima, N. Viovy, A. Voulgarakis, M. van Weele, G. R. van der Werf, R. Weiss, C. Wiedinmyer, D. J. Wilton, A. Wiltshire, D. Worthy, D. Wunch, X. Xu, Y. Yoshida, B. Zhang, Z. Zhang, and Q. Zhu, 2016: The global methane budget 2000–2012. Earth System

Schaefer, H., S. E. Mikaloff Fletcher, C. Veidt, K. R. Lassey, G. W. Brailsford, T. M. Bromley, E. J. Dlugokencky, S. E. Michel, J. B. Miller, I. Levin, D. C. Lowe, R. J. Martin, B. H. Vaughn, and J. W. White, 2016: A 21st-century shift from fossil-fuel to biogenic methane emissions indicated by 13CH4. Science, 352(6281), 80-84, doi: 10.1126/science.aad2705.

Schaefer, K., T. Zhang, L. Bruhwiler, and A. P. Barrett, 2011: Amount and timing of permafrost carbon release in response to climate warming. Tellus B: Chemical and Physical Meteorology, 63(2), 165-180, doi: 10.1111/j.1600-0889.2011.00527.x.

Schuur, E. A., A. D. McGuire, C. Schadel, G. Grosse, J. W. Harden, D. J. Hayes, G. Hugelius, C. D. Koven, P. Kuhry, D. M. Lawrence, S. M. Natali, D. Olefeldt, V. E. Romanovsky, K. Schaefer, M. R. Turetsky, C. C. Treat, and J. E. Vonk, 2015: Climate change and the permafrost carbon feedback. Nature, 520(7546), 171-179, doi: 10.1038/nature14338.

Schwietzke, S., O. A. Sherwood, L. M. Bruhwiler, J. B. Miller, G. Etiope, E. J. Dlugokencky, S. E. Michel, V. A. Arling, B. H. Vaughn, J. W. White, and P. P. Tans, 2016: Upward revision of global fossil fuel methane emissions based on isotope database. Nature, 538(7623), 88-91, doi: 10.1038/nature19797.

Simpson, I. J., M. P. Sulbaek Andersen, S. Meinardi, L. Bruhwiler, N. J. Blake, D. Helmig, F. S. Rowland, and D. R. Blake, 2012: Long-term decline of global atmospheric ethane concentrations and implications for methane. Nature, 488(7412), 490-494, doi:

10.1038/nature11342.

Turner, A. J., C. Frankenberg, P. O. Wennberg, and D. J. Jacob, 2017: Ambiguity in the causes for decadal trends in atmospheric methane and hydroxyl. Proceedings of the National Academy of Sciences USA, 114(21), 5367-5372, doi: 10.1073/pnas.1616020114.

Worden, J. R., A. A. Bloom, S. Pandey, Z. Jiang, H. M. Worden, T. W. Walker, S. Houweling, and T. Rockmann, 2017: Reduced biomass burning emissions reconcile conflicting estimates of the post-2006 atmospheric methane budget. Nature Communications, 8(1), 2227, doi: 10.1038/s41467-017-02246-0.

Excerpted from

USGCRP, 2018: Second State of the Carbon Cycle Report (SOCCR2): A Sustained Assessment Report [Cavallaro, N., G. Shrestha, R. Birdsey, M. A. Mayes, R. G. Najjar, S. C. Reed, P. RomeroLankao, and Z. Zhu (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, 878 pp., https://doi.org/10.7930/SOCCR2.2018 Accessed in December 2023 at https://carbon2018.globalchange.gov/ Public Domain

11.4: Atmospheric methane is shared under a Public Domain license and was authored, remixed, and/or curated by U.S. Global Change Research Program.

11.5: Carbon monoxide

CO Fate in the Atmosphere

CO is one of the most common air pollutants, and its concentration in contaminated continental air masses reaches up to several ppmv (Badr and Probert, 1994). CO is emitted to the atmosphere through anthropogenic processes, such as biomass combustion, fossil fuels, waste incineration, industrial processes, and transport. Additional contributors are natural sources (oceans, soils, plants, forest fires), atmospheric oxidation of CH_4, and other non-CH_4 hydrocarbons (NMHC) (Badr and Probert, 1995; Tarr et al., 1995; Schade and Crutzen, 1999; Bruhn et al., 2013). CH_4 oxidation has the largest share of these sources, producing approximately 700 Tg-CO·yr^{-1} (Bergamaschi et al., 2000; Monson and Holland, 2001); fossil fuel combustion together with biofuel use and other industrial emissions are responsible for 500–900 Tg-CO·yr^{-1}, while biomass burning–for 400–800 Tg (Duncan et al., 2007). Photochemical CO production due to the oxidation of naturally emitted and anthropogenic NMHC equals 450 and 110 Tg-CO·yr^{-1}, respectively, (Rozante et al., 2017). In general, global CO levels rose from the Industrial Revolution until 1980; then a gradual decrease in its concentration, especially in the Northern Hemisphere, was observed in measurements from the global surface network from the National Oceanic and Atmospheric Administration (NOAA), caused by both the use of catalytic converters in cars and technological advances in combustion since 2000 (Bakwin et al., 1994; Voiland, 2015; Gaubert et al., 2017). More recently, the downward trends in CO observed for both the Northern and Southern Hemispheres have shown good consistency with long-term trends in bottom-up emissions in Europe, the United States, and China; where an improvement in combustion efficiency and a reduction in emissions from anthropogenic sources was observed (Gaubert et al., 2017). It is worth noting, however, that despite its main global sources, atmospheric CO levels show spatial as well as seasonal variability. While the oxidation of CH_4, a gas that is evenly distributed around the world, provides a similarly constant CO background of around 25 ppb, the remaining emission groups depend on space-time aspects. Thus, inter alia, CO from fossil fuels shows higher levels in the northern mid-latitudes, mainly in winter, while biomass combustion in tropical continents contributes to higher CO concentrations in the summer, during the dry season, along with the rainforests (Andreae et al., 2012). Additionally, the spatial variation in CO concentration is also characteristic on a smaller scale, e.g., in urban areas, where it depends not only on meteorological conditions or thermal inversion but also directly on atmospheric turbulence and traffic intensity (Oliveira et al., 2003).

CO is mainly utilized by the tropospheric reaction with the OH hydroxyl radical (Logan et al., 1981; Khalil and Rasmussen, 1984; Badr and Probert, 1995). CO is oxidized to CO_2 in the stratosphere, where it migrates *via* convection, turbulence, and mixing (Seiler and Warneck, 1972; Seiler, 1974). The reaction is fast and independent of temperature; thanks to it, the residence time of CO in the atmosphere is relatively short, from 2 weeks to 3 months (Rozante et al., 2017; Rakitin et al., 2021). Soils and oceans are also involved in CO capture (Ingersoll et al., 1974; Conrad et al., 1982) and higher plants and algae (Krall and Tolbert, 1957; Chappelle, 1962).

Taking into account the characteristics of greenhouse gases, CO is not considered one of them as it is not capable of absorbing infrared radiation (Rozante et al., 2017). However, because of the primary mechanism for removing atmospheric CO by reaction with the OH radical, CO is recognized as an essential trace gas that controls the oxidative ability of the atmosphere (Bruhn et al., 2013). An increase in CO concentration in the troposphere causes changes in the distribution and amount of OH; reactions of the radical with CO and CH_4 constitute about 97% of its destruction (Logan et al., 1981; Levine et al., 1985; Badr and Probert, 1995). The change in the atmospheric OH affects the concentration of other gases, including CH_4 and O_3 (Hameed et al., 1980). Thus, CO indirectly affects the energy budget of the atmosphere (Evans and Puckrin, 1995), increasing the concentration of GHGs and the time of their utilization in the troposphere, as well as controlling the transfer to the stratosphere, which in turn has an impact on stratospheric O_3 (Ramanathan et al., 1985; Thompson, 1992; Bruhn et al., 2013). The radiative forcing from CO is estimated more than from N_2O and halogenated hydrocarbons (Rakitin et al., 2021).

Further detailed information at Sobieraj, K., Stegenta-Dąbrowska, S., Luo, G., Koziel, J. A., & Białowiec, A. (2022). Carbon monoxide fate in the environment as an inspiration for biorefinery industry: A review. *Frontiers in Environmental Science*, *10*, 822463. https://www.frontiersin.org/articles/10.3389/fenvs.2022.822463/full

CO Fate in Soil

CO is constantly supplied to the atmosphere, which results from both natural and anthropogenic sources. Despite the significant amount of gas emitted, its concentration in the atmosphere does not seem to increase (Bartholomew and Alexander, 1979). It's

because natural processes are responsible for utilizing CO shortly after its release (Inman et al., 1971).

Soil is considered one of the main sinks of atmospheric CO, responsible for 40% of total consumption (Seiler, 1978). However, soils can also be a CO source in the global CO cycle, as noted, especially in the savannas and deserts (Conrad and Seiler, 1985a; Kuhlbusch et al., 1998). For this reason, CO uptake by soils is a net flux consisting simultaneously of consumption and production (Figure 1) (Seiler, 1978; King and Crosby, 2002; Bruhn et al., 2013; van Asperen et al., 2015; Pihlatie et al., 2016). Simultaneous chemical, physical and microbiological processes (Kuhlbusch et al., 1998) depend on many climatic, biological, and physical soil factors, making the equilibrium CO vary between a few ppbv up to hundreds of ppbv (parts per billion by volume) (Conrad and Seiler, 1979; Conrad and Seiler, 1980b; Conrad and Seiler, 1982a). The most important soil parameters include water content, temperature, organic matter content, pH, soil type, the depth of CO consumption horizon, and CO concentration in the gas phase (Potter et al., 1996). Even small changes in this balancing between CO production and soil uptake can severely impact tropospheric chemistry (Moxley and Smith, 1998b).

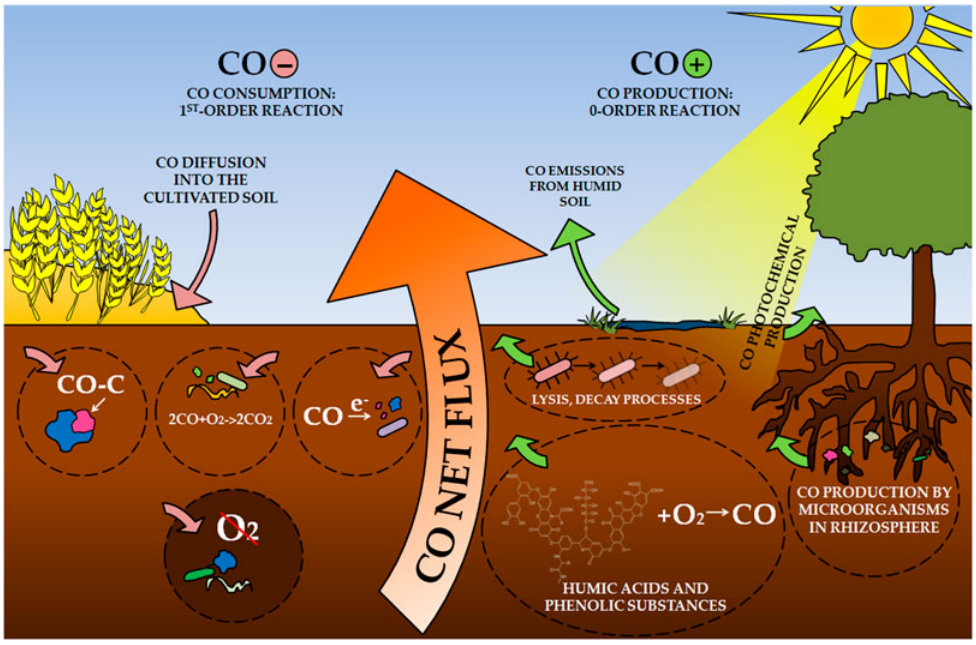


Figure 11.5.1: Carbon monoxide cycle in soil. CC-BY, Frontiers, Sobieraj et al.

Further detailed information at Sobieraj, K., Stegenta-Dąbrowska, S., Luo, G., Koziel, J. A., & Białowiec, A. (2022). Carbon monoxide fate in the environment as an inspiration for biorefinery industry: A review. *Frontiers in Environmental Science*, *10*, 822463. https://www.frontiersin.org/articles/10.3389/fenvs.2022.822463/full

CO Fate in Water

The ocean has been recognized as a source of CO released into the atmosphere since the early 1970s (Swinnerton et al., 1970; Lamontagne et al., 1971). Despite the low ocean share (0.4–9%) among all sources of CO in the atmosphere (Bates et al., 1995), it can constitute up to 50% of the load in the marine boundary layer (Erickson and Taylor, 1992; Stubbins et al., 2006a). The southern hemisphere is particularly important here, in which CO production accounts for almost 60% of the total CO flux from the ocean surface (Erickson, 1989). Surface ocean waters are saturated with CO compared to atmospheric equilibrium, which causes a net flux of this gas at the ocean-atmosphere interface (Linnenbom et al., 1973; Logan et al., 1981; Zuo et al., 1998). The CO emissions to air are controlled mainly by the concentration of this gas in water (Bates et al., 1995). It depends on several factors such as photochemical production, consumption by microorganisms, exchange between air and water, and physical mixing (Figure 2) (Wilson et al., 1970; Conrad et al., 1982; Butler et al., 1987; Jones, 1991). Due to the impact of these factors, CO concentration in waters shows diurnal, seasonal, and regional diversification (Bates et al., 1995).

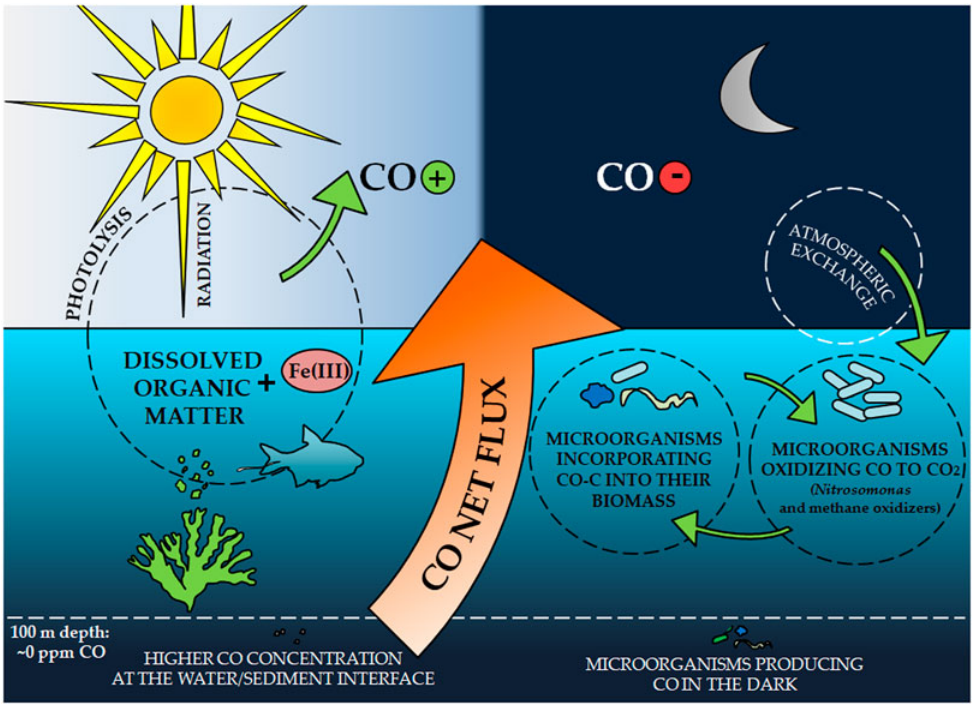


Figure 11.5.2: Carbon monoxide cycle in water. CC-BY, Frontiers, Sobieraj et al.

CO was identified as the second most important product of dissolved organic matter (DOM) photolysis in water bodies (Mopper and Kieber, 2000; Stubbins et al., 2006b). The rate of CO formation is one order of magnitude higher compared to other low molecular weight carbon photoproducts produced under aqueous conditions (Mopper et al., 1991; Zuo and Jones, 1995). It is the photodegradation of the DOM by part of the UV solar radiation that is indicated as the main source of CO from both ocean and sea waters, as well as from the surface of lakes, rivers, wetlands, and coastal waters (Zuo and Jones, 1997; Pos et al., 1998; Zuo et al., 1998; Stubbins et al., 2006a; Blomquist et al., 2012).

Further detailed information at Sobieraj, K., Stegenta-Dąbrowska, S., Luo, G., Koziel, J. A., & Białowiec, A. (2022). Carbon monoxide fate in the environment as an inspiration for biorefinery industry: A review. *Frontiers in Environmental Science*, *10*, 822463. https://www.frontiersin.org/articles/10.3389/fenvs.2022.822463/full

CO Impact on Plants

Biosynthesis and photoproduction of CO in plants were observed in the second half of the 20th century (Wilks, 1959; Schade et al., 1999). This compound is formed during oxidative heme catabolism due to the activity of the enzyme heme oxygenase (HO). The result is three products: CO, biliverdin, and free iron Fe^{2+} (Figure 3). The second one is immediately transformed into bilirubin, while the iron is involved in ferritin induction (Bilban et al., 2008). Among the three isoforms of HO discovered so far, HO-1, HO-2 and HO-3 have been distinguished, the last two of which are characterized by low activity (Maines, 1997).

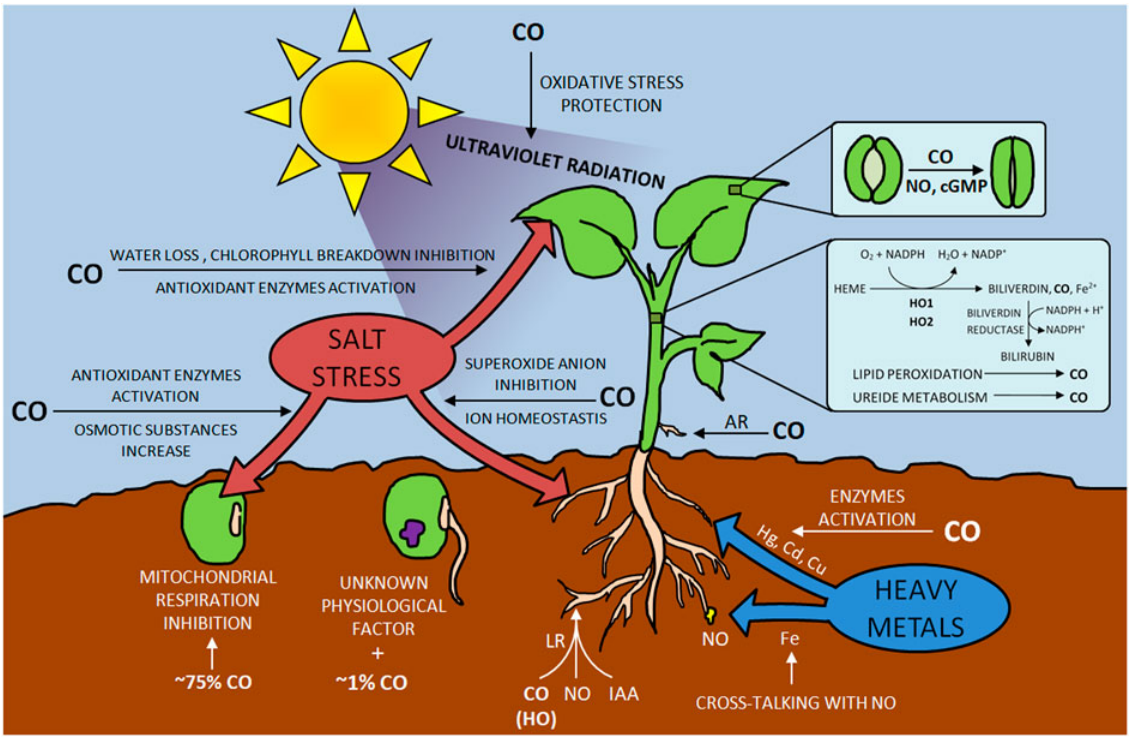


Figure 11.5.3: Impact of CO on plants development. CC-BY, Frontiers, Sobieraj et al.

Further detailed information at Sobieraj, K., Stegenta-Dąbrowska, S., Luo, G., Koziel, J. A., & Białowiec, A. (2022). Carbon monoxide fate in the environment as an inspiration for biorefinery industry: A review. *Frontiers in Environmental Science*, *10*, 822463. https://www.frontiersin.org/articles/10.3389/fenvs.2022.822463/full

Source

Sobieraj, K., Stegenta-Dąbrowska, S., Luo, G., Koziel, J. A., & Białowiec, A. (2022). Carbon monoxide fate in the environment as an inspiration for biorefinery industry: A review. *Frontiers in Environmental Science*, *10*, 822463. CC-BY accessed December 2023 https://www.frontiersin.org/articles/10.3389/fenvs.2022.822463/full

11.5: Carbon monoxide is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

11.6: Synthesis- Linking the carbon and oxygen cycles

Plants use photosynthesis to capture energy from the sun, building tissue. In the distant past the same was true for primordial cyanobacteria and phytoplankton in the oceans. Collectively, these organisms have been the most powerful agents of geochemical change: they extracted CO_2 from the atmosphere, shucking off its oxygen as waste, and retaining the carbon, first in the form of glucose. In the far past much of the carbon was buried, sequestered from the atmosphere, and transformed into fossil fuels. Today burning fossil fuels is reversing that. In the words of Roger Revelle

> *". . . human beings are now carrying out a large scale geophysical experiment of a kind that could not have happened in the past nor be reproduced in the future. Within a few centuries we are returning to the atmosphere and oceans the concentrated organic carbon stored in sedimentary rocks over hundreds of millions of years. This experiment, if adequately documented, may yield a far-reaching insight into the processes determining weather and climate."*

Carbon burial

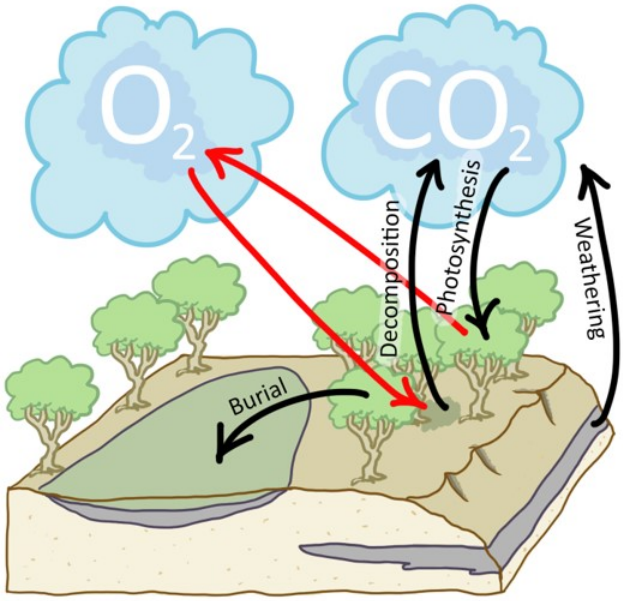


Figure 11.6.1: Carbon and oxygen fluxes between plants, the atmosphere, and the geosphere. Callan Bentley cartoon.

Consider this diagram, which highlights some of the fluxes (flows) between the atmosphere, plants, and layers of sedimentary rock. Plants pull CO_2 from the atmosphere and release O_2 as a waste product. The O_2 will eventually react with the plant carbon through decomposition unless the plant matter is buried. Once buried, it is isolated from free oxygen; **it cannot decompose**. Its carbon then becomes part of sedimentary rocks, and is locked in the geosphere until uplift (often, mountain-building) brings it back into contact with atmospheric oxygen. This allows it to be weathered (re-oxidized). This occurs on timescales of **tens to hundreds of millions of years**.

Impacts on the cryosphere and biosphere

Let us consider the impacts of carbon burial on three phenomena: fire, climate, and animal physiology.

Fire



Figure 11.6.2: A 2007 wildfire in Idaho. US Fish & Wildlife Service photo.

Fire is a natural part of the modern Earth system. **Wildfires** can be sparked by lightning strikes or volcanic eruptions. They are more likely to ignite and to spread in areas that are temporarily warm and dry, but have seen wet enough conditions in the past to build up a supply of plant fuel. Combustion occurs when the plant tissues react with atmospheric oxygen, and release the carbon in the plants back into the atmosphere as CO_2.

So **oxygen is a pre-requisite** for wildfire, and oxygen levels have changed throughout Earth history. How much oxygen is necessary? It turns out that fire requires at least **13% O_2** in the surrounding atmosphere. Over the long-term, our planet has bulked up the free oxygen content of its atmosphere, from an initial 0% to the current ~21%. It is estimated that the critical 13% threshold was crossed in the Silurian. Two lines of evidence support this interpretation:

First, the oldest terrestrial plant fossils are spores that date to the Ordovician, suggesting colonization of the land surface by organisms akin to club mosses. These plants provided fuel to burn, but also shelter for animals. The oldest terrestrial animal fossils date to shortly thereafter. As plants colonized the land, they led to more of the planet's surface being devoted to (incidental) oxygen production. It is thought that oxygenation of the atmosphere would also have led to the development of an ozone (O_3) layer, which would have enabled animals to spread across the land due to lowering the influx of harmful ultraviolet radiation from the Sun.

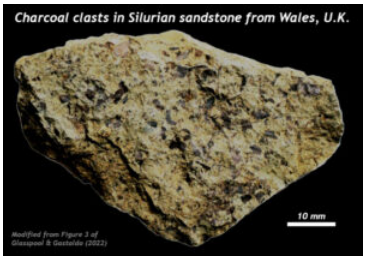


Figure 11.6.3: Charcoal clasts in Silurian sandstone from Wales. Modified from Figure 3 of Glasspool & Gastaldo (2022).

Second, by 420 Ma (**Silurian**), **charcoal** starts showing up. Charcoal is a carbon-rich version of what used to be wood, but all the water and volatile components have been driven off through heating it in the presence of limited amounts of oxygen. It is our first evidence that fires had started burning on Earth.

With the exception of a "gap" in the mid-Devonian, charcoal has been a common constituent of terrestrial sediments for the rest of the Phanerozoic Eon. Thanks to the rise in oxygen levels, fire has become a standard phenomenon on the surface of Earth. So the burial of plant biomass preserves some carbon for the future, but leaves behind oxygen at the surface, and that oxygen that makes it more likely for unburied plants to burn.

Ice

In the mid- to late-Paleozoic, plants evolved and spread across the land area of our planet. These plants extracted CO_2 from the atmosphere, but if they burned or rotted over short timescales, that carbon went straight back into the atmospheric reservoir as CO_2. That's an example of the short-term carbon cycle. On the other hand, if the plants **were buried** in swampy sediments, then the plant carbon could no longer can interact with the oxygen in the atmosphere. It had been **sequestered**.

There is evidence of the reduction of atmospheric CO_2 at several points in Earth history. During the late Devonian period, excessive formation of black shale (full of carbon captured by marine plankton) led to an ice age and one of the big five mass extinctions. During the ensuing Carboniferous period, repeated transgressive burials of coastal swamps led to cyclothems including the coal seams that powered the Industrial Revolution. The Carboniferous gets its name from the carbon-rich nature of the rocks

laid down at that time in England. In the Neoproterozoic Era, the planet froze over when CO_2 levels moved close to zero due to excessive weathering of the continental crust. It wasn't organic carbon burial in that case, but reaction of CO_2 with silicate minerals that caused the CO_2 drawdown and cooling.

In summary, high rates of carbon burial lead to lower CO_2 levels, which causes a reduction in the greenhouse effect, and thus deep ice ages. Sometimes this global cooling is severe enough to cause a mass extinction.

Big bugs

But burial of organic carbon doesn't just lower global CO_2 levels. It also means that the waste product of photosynthesis, oxygen, builds up and has a lesser amount of carbon available to react with. With less organic carbon available for oxidation, atmospheric oxygen levels would be expected to go up.

In the planet's earliest years, there was no free oxygen in its atmosphere. Gradually it built up through time, largely thanks to the photosynthetic work of cyanobacteria. By the Paleoproterozoic/ Mesoproterozoic boundary, it had accumulated to something like 1% of the atmosphere by volume. This has been dubbed the Great Oxygenation "Event." Currently, the atmosphere has an atmosphere that consists mainly of nitrogen (N_2) but also includes ~21% free oxygen (O_2).

However, it hasn't been a straight path from the GOE to our present oxygenation. During the Carboniferous period, rates of carbon burial were high, so that means atmospheric CO_2 was low and therefore atmospheric O_2 was high. Estimates are that it got as high as **35% of the atmosphere**. In such an atmosphere, you wouldn't have to inflate your lungs as often in order to get the oxygen you need.

Insects don't have lungs. Instead, they have a network of branching tubes that runs throughout their body, and through which air circulates. The maximum size of insects today is controlled by the efficiency of this tube network in delivering sufficient oxygen to the tissues that need it. The Goliath beetle is the current record-holder as "biggest bug." It can get up to 4.5 inches (11 cm) long. While that's certainly impressive, it doesn't hold a candle to this:

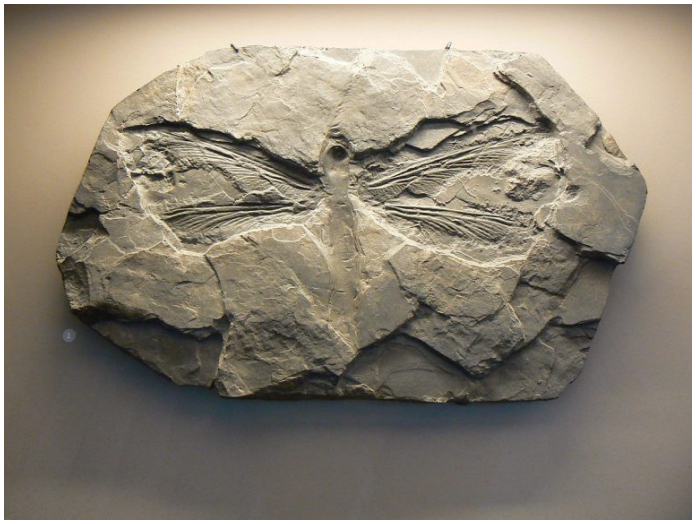


Figure 11.6.4: Large fossilized Meganeura dragonfly. Creative Commons Attribution-Share Alike 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic license.

That's a dragonfly ***the size of a hawk***. Its wingspan is 28 inches (70 cm). It could never survive today, but during the Carboniferous, with an atmosphere of 35% oxygen, such a huge insect was plausible. Other examples of giant arthropods include *Arthopleura* and *Megarachne*.

Fossil fuels

Buried carbon does not necessarily stay buried. By oxidizing that carbon vis combustion, energy originally bound up in chemical bonds as much as ~350 million years ago is released. We refer here to and coal, petroleum, and natural gas: the "fossil fuels." They get this name because all the carbon in each fuel came originally from living things extracting CO_2 from the atmosphere.

Coal

Coal is a solid fossil fuel that forms from compressed plant matter called **peat**. Frequently coal layers ("seams") still retain the impressions of leaves, stems, seeds, spores, and other botanical details. Coal comes in several "grades" that reflect varying degrees of compression. The lowest grade is **lignite** (brown coal), which is just a step up from peat, the original mix of plant bits. If it is more compressed still, it will become **bituminous coal**. And the highest grade of coal, **anthracite**, forms only when a coal seam is subject to metamorphism under the influence of mountain-building.

The plants that are the progenitor of coal live on the land, in wet, aquatic, or swampy settings and thus are poorly oxygenated. Carbon-rich plant fragments falling into that water are therefore more likely to avoid oxygenation long enough to be buried. The rate of new carbon being added as plants shed debris outstrips the rate at which oxygen diffuses across the surface of the water.

But peat can also form in cool, wet climates such as Ireland or Scotland. There, a layer of vegetation clings to the topography of the land. The damp weather means that sphagnum and heather, the main peat-building vegetation, grow on slopes and hilltops as well as in the valley bottoms, with underlying peat layers averaging a meter or two thick. The entire landscape is essentially blanketed in bog. It grows upward at about 1 cm/year.

Oil

Oil has a different origin story. Here, the photosynthesis was conducted by marine or lake algae, **phytoplankton** floating as single-celled organisms in the uppermost, sunlit portion of the water column. When they died, their bodies rained down to the bottom. This transferred the carbon they contained to the dark depths. Often, the bottoms of lakes or the ocean have far less dissolved oxygen than waters closer to the surface. As a result, carbon entering these anoxic waters tends to be buried before it can oxidize. As with swamp coal, it's a case of the carbon burial rate exceeding the flow of dissolved oxygen to the same place.

Once buried, the algae have to be warmed up to a critical temperature – **about 100°C**. If it's too cold, their dead cells won't transform into petroleum. If it's too hot, the mixture will turn into natural gas (described below) and graphite.

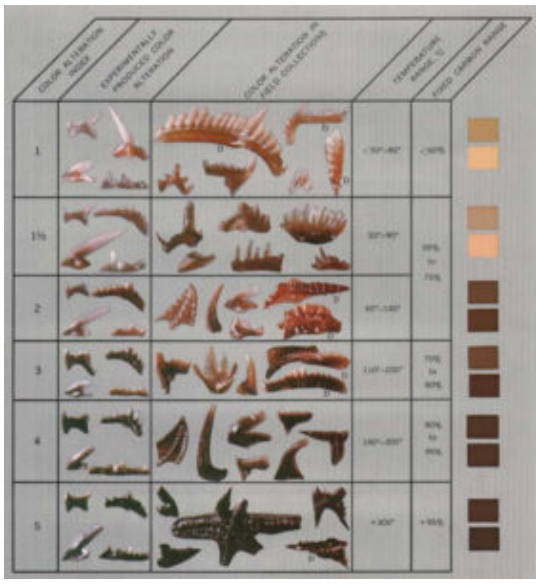


Figure 11.6.5:The Conodont Alteration Index of Anita Harris (then Anita Epstein), et al., 1977. USGS image.

Interestingly, one way geologists can make a quick but accurate assessment of the temperature to which potentially-oil-bearing rocks have been warmed is by examining fossils. Specifically, the spiky "elements" of the conodont animal. These are made of hydroxylapatite, a phosphate mineral which **changes color when heate**. Geologist Anita Harris developed a tool, the Conodont Alteration Index, a 6-point color scale that uses the various colors of conodonts to see if their host rocks warmed up to the appropriate temperature for oil generation.

Conodonts can be collected from numerous sites in a region, their color assessed using the Index, and then the presence of each color can be contoured. Here's an example for the sedimentary strata of the Appalachian Plateaus + the Valley & Ridge provinces:

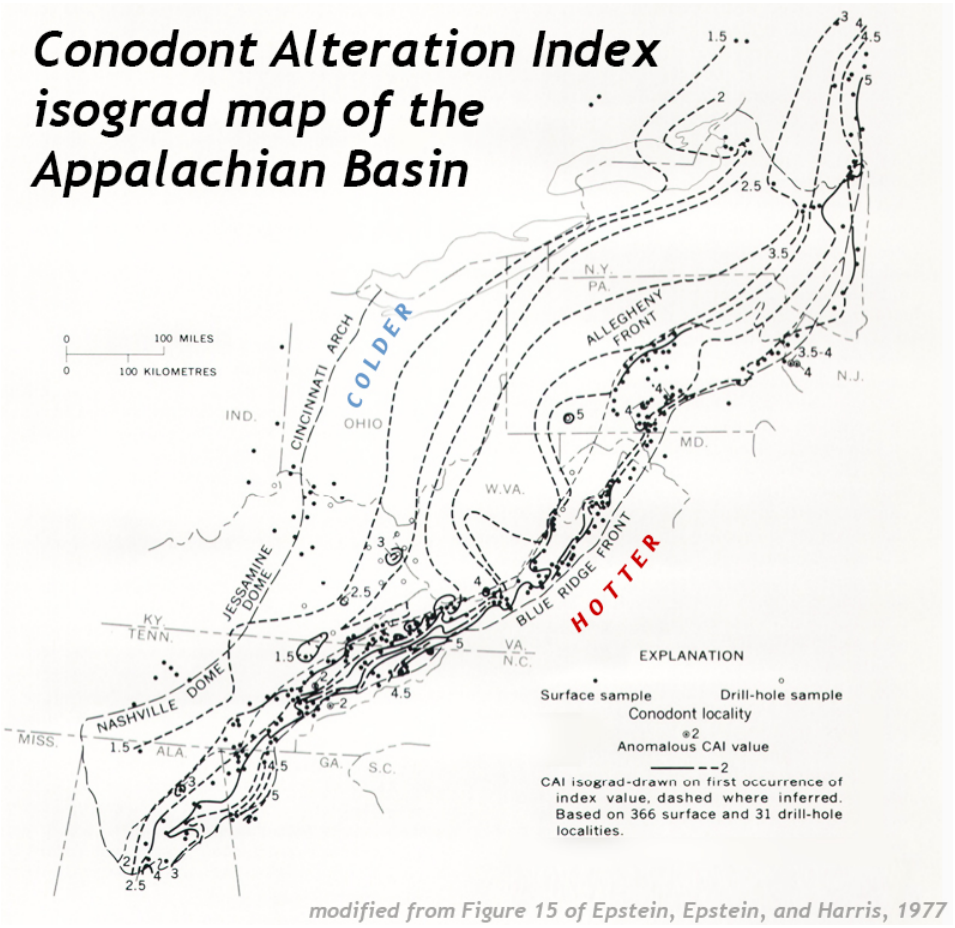


Figure 11.6.6: Conodont Alteration Index color distribution across the Appalachian Basin. Modified from a USGS figure.

Once produced, oil must migrate to a spot where it can collect in amounts where humans will find it economically worthwhile to drill it out and extract it. There are places where oil naturally seeps out at Earth's surface, such as the legendary tar pits of Rancho La Brea in Los Angeles, California.

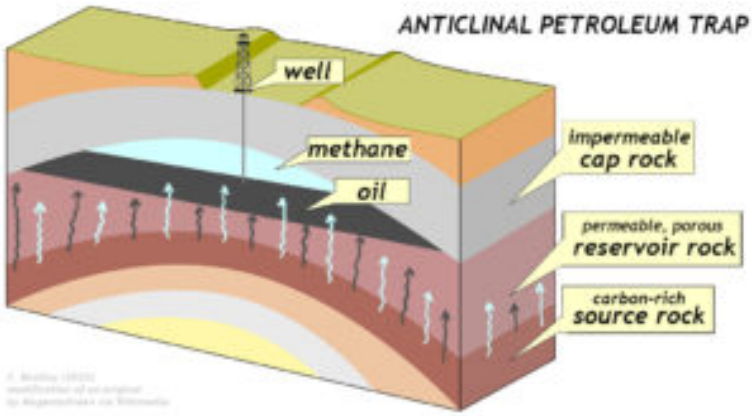


Figure 11.6.7:An anticlinal petroleum trap. C.Bentley modification of an original diagram by GreenMagenta via Wikimedia.

Most of the oil pumped today found a stable residence underground in some variety of subterranean oil trap. These have various shapes due to various geological circumstances, but one of the easiest to grasp is that of an anticline. If the arch-like shape of an anticline consists of a stack of layers with carbonaceous source rocks at the bottom (such as black shale), impermeable cap rock strata (such as shale) at the top, and a spongey reservoir layer (such as sandstone) in between, that makes for an ideal **petroleum trap**. Petroleum geologists are charged with finding these traps and directing the company's engineers to tap into them.

Natural gas



Figure 11.6.8: Outcrop of the Devonian-aged Marcellus Shale in Pennsylvania. Yellow field notebook for scale. Modified from a Terry Engelder photo.

The third and final fossil fuel is neither a solid like coal, nor a liquid like oil. Instead, it's a mixture of gases, principally methane (CH_4). Like oil, its progenitor is phytoplankton. While these algae live, they photosynthesize. When they die, their bodies rain down into anoxic deep waters, imparting a high carbon content to the sediments deposited there. Typically, the resulting rock is a **black shale**.

Black shales are more common in the geologic record during times of sluggish ocean circulation or widespread stagnation, such as the end-Permian or the mid-Cretaceous. Black shales can also be deposited in lakes; they are not necessarily marine in origin.

The Marcellus Shale of the Appalachians and the Bakken Shale of the Dakotas are both Devonian-aged black shales with lots of natural gas trapped inside them. In the past decade, the advent of horizontal drilling has allowed fossil fuel companies to drill into these shales and intentionally shatter them, allowing the gas to flow out, up the drill hole, where it can be collected, transported, and sold. The shattering process is achieved by (a) injecting fluids under very high pressures, along with (b) sand, which props open the fractured shale once the cracks open up. This process is called **hydraulic fracturing**, or "fracking."

Natural gas is also produced in coal seams. This variety is referred to as "coal-bed methane." Coal miners have been very conscious of coal-bed methane since coal mining began – because it is invisible gaseous reduced carbon, one spark from mining equipment could trigger an explosion. Far too many miners lost their lives to this hazard. They called it (and any other inflammable gas) "firedamp."

Anthropogenic climate change

Because of the rich energy content of fossil fuels, humans are burning them, and depending intensely on the energy released in the combustion. The modern global economy and the lifestyles of modern humans are powered in large part by the combustion of fossil fuel — solids, liquids, and gases that derive their energy content from sunlight that shone on Earth tens to hundreds of millions of years ago. As plants do with oxygen waste, the CO_2 waste goes into the atmosphere. Because it's invisible, we've been able to ignore it for a long time. Out of sight, out of mind.

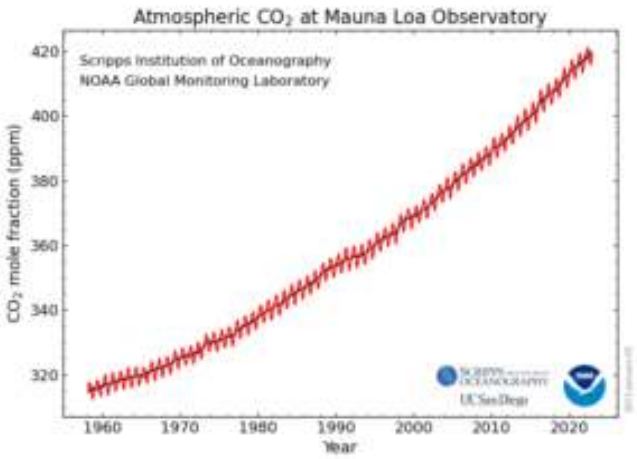


Figure 11.6.9: The record of the past half-century of changing CO_2 levels in Earth's atmosphere. Scripps / NOAA graph.

Since humanity began burning fossil fuels in earnest (the start of the Industrial Revolution), CO_2 levels have increased in the atmosphere of our world. The seasonal signal of photosynthesis is still there, but every year, CO_2 ends up at a higher level than the previous year. From a pre-Industrial level of ~280 parts per million (ppm) to the current ~420 ppm, it has increased by about 50%.

It would have been an even larger increase, had not the Earth system been so intimately inter-connected. Much of the CO_2 that humanity has emitted through the burning of fossil fuels has been re-extracted from the atmosphere: some has been taken out by plants doing photosynthesis more efficiently, and some has been soaked up by the oceans.

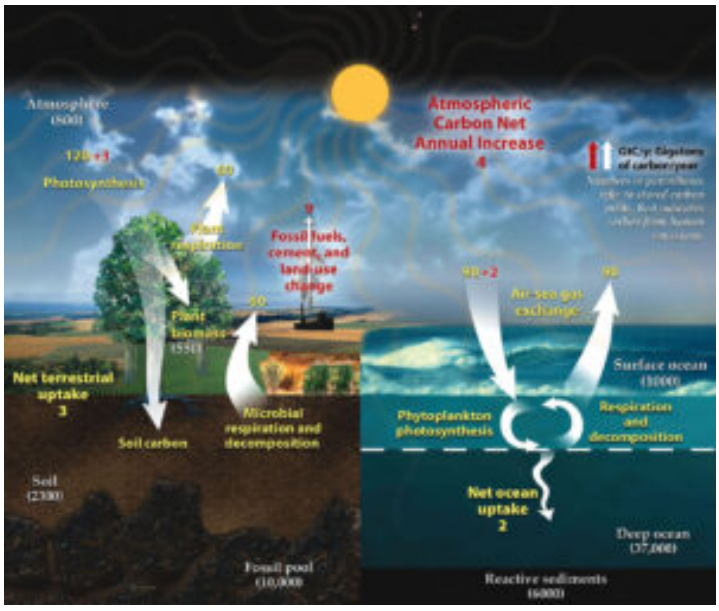


Figure 11.6.10: Thanks to CO_2 absorption by the oceans and photosynthesis by plants, the human-induced carbon loading of the planet's atmosphere isn't as severe as it might otherwise have been. NASA graphic.

Estimates of the flow of carbon between reservoirs in the Earth system suggest that human activities release about nine billion tons (gigatons) of carbon into the atmosphere per year, of which three is then pulled out again by plant photosynthesis, and two is dissolved in the oceans. That leaves a **net atmospheric gain of four gigatons per year**. Examine the graphic at left to explore the complexities of carbon fluxes between reservoirs, but follow the red numbers, which show the changes that result from the human economy's dependence on burning ancient trees and algae. This imbalance is why the graph above shows a relentless climb: the rates of carbon **absorption** can't match the rates of carbon **emission**.

Oceanographers are well aware of one of the consequences of this natural carbon sequestration: those extra two gigatons of carbon soaked into the oceans each year produce carbonic acid, which is making the ocean overall more acidic. Already surface ocean pH has dropped by 0.1 unit, with a forecast total of -0.7 pH units change over the next 750 years. Remember that pH is -log[H^+] so 0.1pH units is roughly a 20% increase in [H^+]. This potentially has enormous effects for the viability of life in the oceans. Many of our greatest episodes of mass extinction are correlated with times of intense **ocean acidification**.

But the more significant issue is probably **global warming**. Because of the selective transparency of the CO_2 molecule, the higher level of this greenhouse gas has resulted in Earth retaining more energy than previously. This has caused the average global temperature to increase.

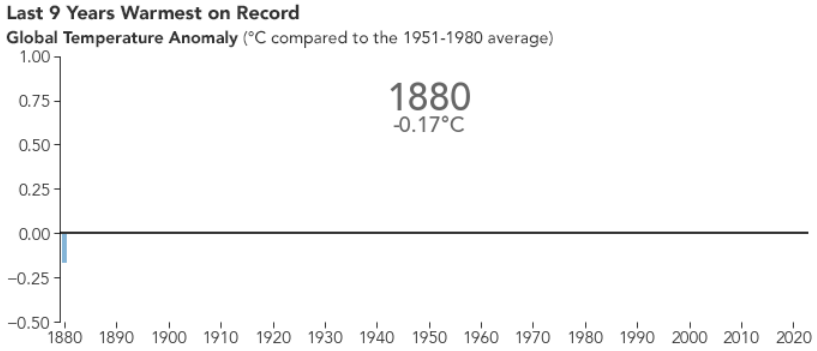


Figure 11.6.11:Ranking the last 142 years of global average temperatures in chronological order. NASA animation of NASA data.

There are many **consequences** of this change in the planet's climate, including: melting glacial ice and permafrost, rising sea levels, intensification of tropical storms, increasing desertification, and a spur to the migration of species toward more pole-ward latitudes and higher elevations. It is an interesting time to be alive and scientifically aware on planet Earth.

O_2 fall

There is another aspect of the burning of fossil fuels which is also noteworthy. The basic combustion reaction of $C + O_2 \rightarrow CO_2$ generates CO_2 of course, but it also consumes O_2. For every atom of carbon we oxidize to power our energy grid or heat our homes or propel our vehicles, we lose one molecule of free oxygen from the atmosphere. Thus, as the past six decades of measurements show a rise in CO_2, we should also expect a commensurate **decline in O_2**.

Indeed, that is what we have observed:

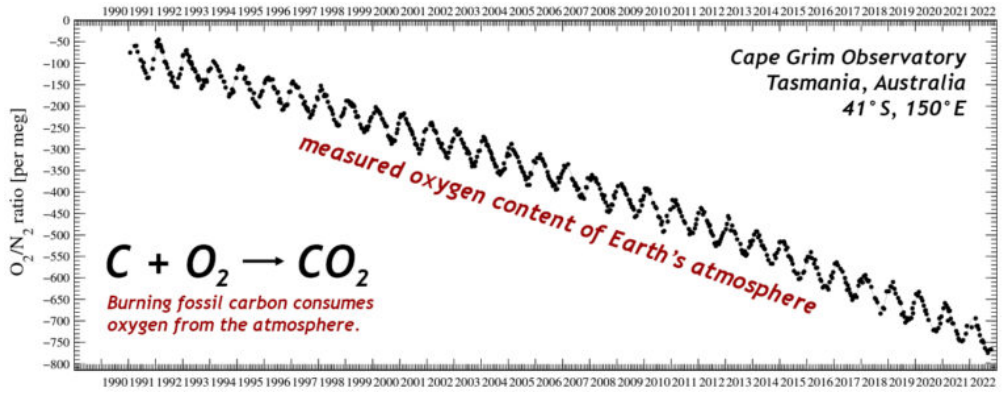


Figure 11.6.12:Measurements of atmospheric oxygen levels show a long-term decline. Graph by Scripps O_2 program, annotated by C. Bentley.

At first glance, this graph might seem alarming, but remember that this is a parts-per-million level decline in a very large reservoir of oxygen: ~21%. It represents a drop of a few ten-thousandths of the total amount of oxygen. So **you don't have to worry** about running out of breathable air. We will exhaust our supply of fossil fuels long before atmospheric oxygen levels change by even 1%.

Instead, save your worrying for the impacts of global warming.

Further reading

Wally Broecker, 1996. "Et tu, O_2?" *21st C*, Columbia University.

Ken Caldeira and Michael Wickett, 2003. "Anthropogenic carbon and ocean pH," *Nature* 425, p.365. https://doi.org/10.1038/425365a

Anita Epstein, Jack Epstein, and Leonard Harris, 1977. "Conodont color alteration — an index to organic metamorphism." USGS Professional Paper 995. 27 pages.

Ian Glasspool and Robert Gastaldo, 2022. "A baptism by fire: fossil charcoal from eastern Euramerica reveals the earliest (Homerian) terrestrial biota evolved in a flammable world," *Journal of the Geological Society* pre-print. Accessed online 14 January 2023.

Iman Gosh, 2021. "All the Biomass of Earth, in One Graphic," *Visual Capitalist*. (20 August 2021) Retrieved 10 January 2023.

Elizabeth Kolbert, 2006. "The Darkening Sea," *The New Yorker.* November 20, 2006, p.66.

Museum of the Earth, 2023. "Anita Harris," *Daring to Dig* Paleontologist Profiles. Retrieved 12 January 2023.

Holly Riebeek, 2011. "The Carbon Cycle," *NASA's Earth Observatory*. Retrieved 12 January 2023.

A.A. Venn, J.E. Loram, A.E. Douglas, "Photosynthetic symbioses in animals," *Journal of Experimental Botany*, Volume 59, Issue 5, March 2008, Pages 1069–1080, https://doi.org/10.1093/jxb/erm328

Original Source

Callan Bentley, Karen Layou, Russ Kohrs, Shelley Jaye, Matt Affolter, and Brian Ricketts.**Historical Geology** A free online textbook for Historical Geology courses, https://opengeology.org/historicalge...gen-interplay/ Accessed December 2023 CC-BY-SA-NC

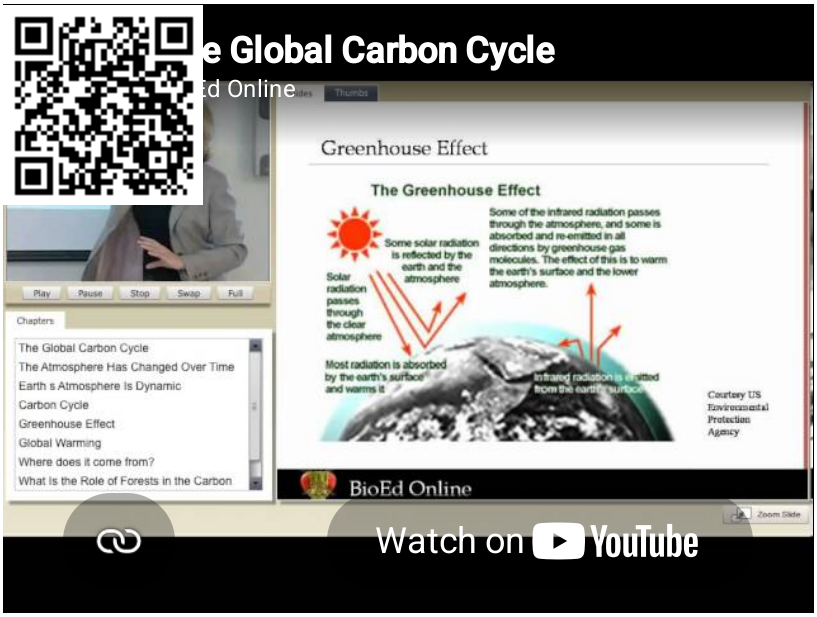
with editing and additions by LibreTexts

11.6: Synthesis- Linking the carbon and oxygen cycles is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Callan Bentley, Karen Layou, Russ Kohrs, Shelley Jaye, Matt Affolter, and Brian Ricketts.Historical Geology.

11.7: Summary



Under Construction



11.7: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

12: The Global Cycles of Nitrogen, Phosphorus and Potassium

12.1: Introduction

12.2: The global nitrogen cycle

12.2.1: Chemistry of the nitrogen cycle

12.3: Temporal variations in the global nitrogen cycle

12.4: Nitrous Oxide (N2O)

12.4.1: Terrestrial

12.4.2: Marine

12.4.3: Human Sources

12.5: The global phosphorus cycle

12.6: Potassium- An overlooked limiting element?

12.7: Linking global biogeochemical cycles

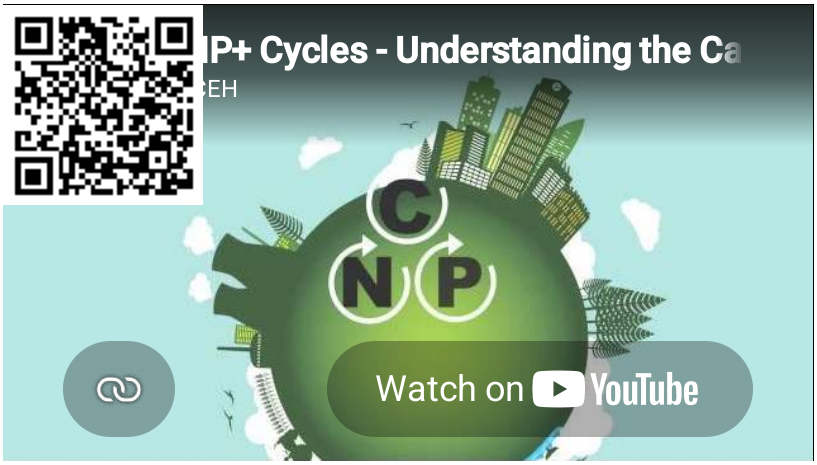
12.8: Summary

12: The Global Cycles of Nitrogen, Phosphorus and Potassium is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

12.1: Introduction

Under development, until it's ready you can watch the video





12.1: Introduction is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

SECTION OVERVIEW

12.2: The global nitrogen cycle

Key Points

- Bacteria, such as cyanobacteria, convert nitrogen into nitrogen gas via nitrogen fixation.
- Nitrogen fixation occurs in three steps: ammonification, nitrification, and denitrification.
- Human activity can release nitrogen into the environment by the combustion of fossil fuels and by the use of artificial fertilizers in agriculture.
- Atmospheric nitrogen is responsible for acid rain, the release of greenhouse gasses, and eutrophication.
- Nitrogen fixation can be performed by marine bacteria; nitrogen falls to the ocean floor as sediment and is then moved to land, becoming incorporated into terrestrial rock.

The Nitrogen Cycle

All organisms require nitrogen because it is an important component of nucleic acids, proteins, and other organic molecules. Getting nitrogen into the living world is difficult. Plants and phytoplankton are not equipped to incorporate nitrogen from the atmosphere (which exists as tightly bonded, triple covalent N_2) even though this molecule comprises approximately 78 percent of the atmosphere (Figure 12.2.1). Nitrogen enters the living world through **nitrogen fixation (**Figure $12.2.1-2$, the process of converting nitrogen gas into ammonia (NH_3), which spontaneously becomes ammonium (NH_4^+). Ammonium is found in bodies of water and in the soil (figure $12.2.1-2$).

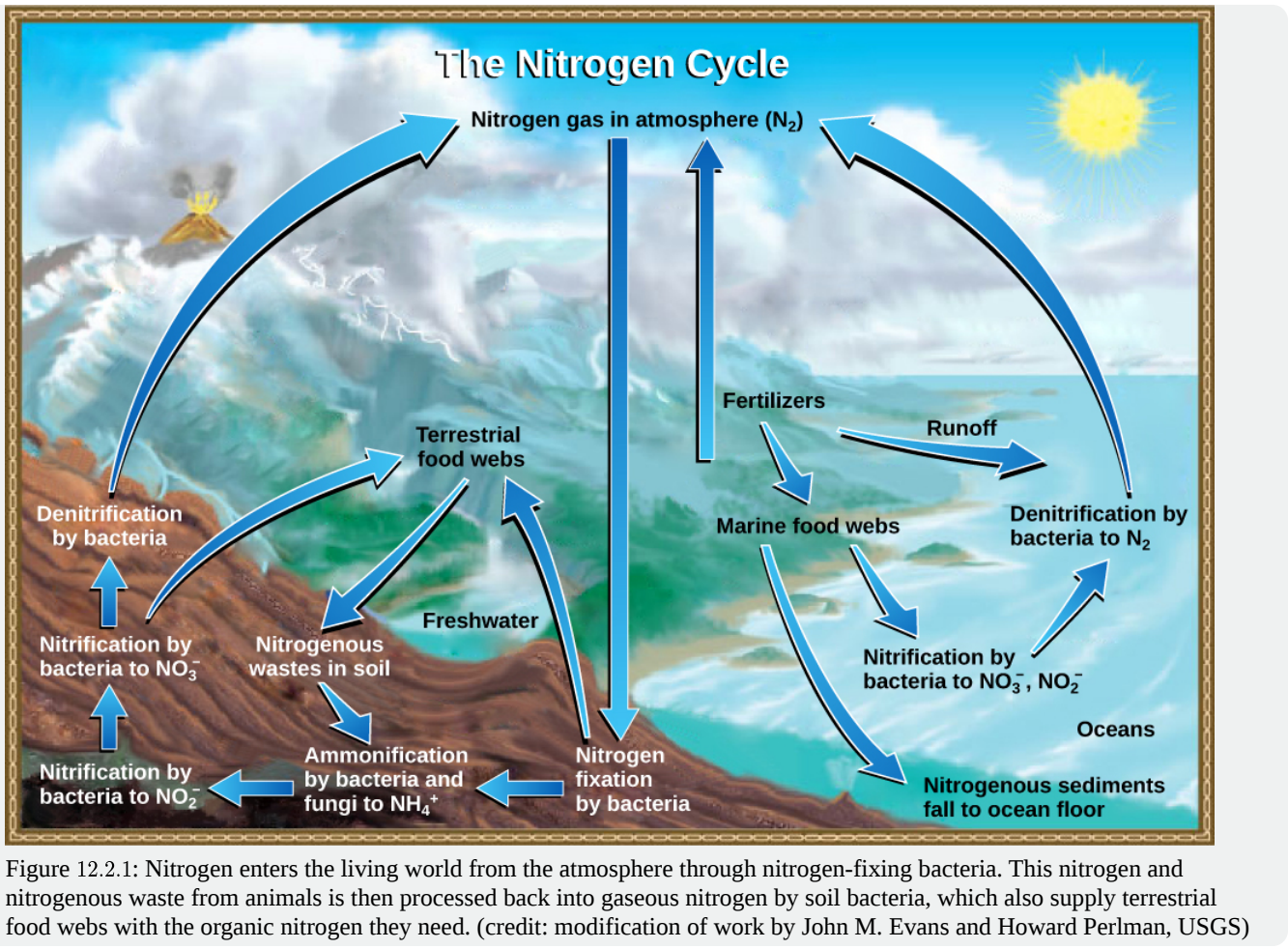


Figure 12.2.1: Nitrogen enters the living world from the atmosphere through nitrogen-fixing bacteria. This nitrogen and nitrogenous waste from animals is then processed back into gaseous nitrogen by soil bacteria, which also supply terrestrial food webs with the organic nitrogen they need. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

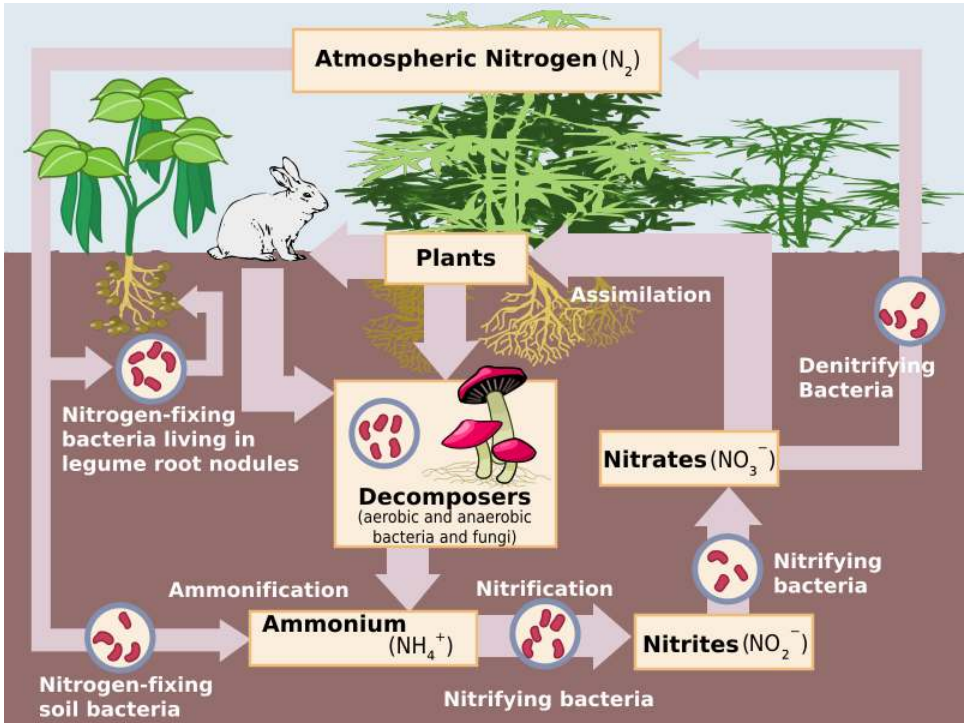


Figure 12.2.2: In the nitrogen cycle, nitrogen-fixing bacteria in the soil or legume root nodules convert nitrogen gas (N_2) from the atmosphere to ammonium (NH_4^+). Nitrification occurs when bacteria convert ammonium to nitrites (NO_2^-) and then to nitrates (NO_3^-). Nitrates re-enter the atmosphere as nitrogen gas through denitrification by bacteria. Plants assimilate ammonium and nitrates, producing organic nitrogen, which is available to consumers. Decomposers, including aerobic and anaerobic bacteria and fungi, break down organic nitrogen and release ammonium through ammonification. (credit: "Nitrogen cycle" by Johann Dréo & Raeky is licensed under CC BY-SA 3.0)

Three processes are responsible for most of the nitrogen fixation in the biosphere. The first is **atmospheric fixation** by lightning. The enormous energy of lightning breaks nitrogen molecules and enables their atoms to combine with oxygen in the air forming nitrogen oxides. These dissolve in rain, forming nitrates, that are carried to the earth. Atmospheric nitrogen fixation probably contributes some 5-8% of the total nitrogen fixed. The second process is **industrial fixation.** Under great pressure, at a temperature of 600°C (1112°F), and with the use of a **catalyst** (which facilitates chemical reactions), atmospheric nitrogen and hydrogen can be combined to form ammonia (NH3). Ammonia can be used directly as fertilizer, but most of it is further processed to urea and ammonium nitrate (NH4NO3).

The third process is **biological fixation** by certain free-living or symbiotic bacteria, which incorporate nitrogen into their macromolecules. Cyanobacteria live in most aquatic ecosystems where sunlight is present; they play a key role in nitrogen fixation. Nitrogen-fixing cyanobacteria are essential to maintaining the fertility of semi-aquatic environments like rice paddies. Free-living bacteria, such as *Azotobacter*, are also important nitrogen fixers. Some nitrogen fixing bacteria form a symbiotic relationship with plants in the legume family, which includes beans, peas, soybeans, alfalfa, and clovers (figure 12.2.3). *Rhizobium* bacteria live symbiotically in the root nodules of legumes (such as peas, beans, and peanuts) and provide them with the organic nitrogen they need. Some nitrogen-fixing bacteria even establish symbiotic relationships with animals, e.g., termites and "shipworms" (wood-eating bivalves). Although the first stable product of the process is ammonia, this is quickly incorporated into protein and other organic nitrogen compounds.



Figure 12.2.3: Nitrogen-fixing bacteria live in the spherical nodules of this soybean root. Image by United Soybean Board (CC-BY).

Organic nitrogen is especially important to the study of ecosystem dynamics since many ecosystem processes, such as primary production and decomposition, are limited by the available supply of nitrogen. Plants and other producers directly use ammonium and nitrates to make organic molecules through the process of **assimilation (**Figure 12.2.2). This nitrogen is now available to consumers. Consumers excrete organic nitrogen compounds that return to the environment. Additionally dead organisms at each trophic level contain organic nitrogen.

As shown in Figure 12.2.2, the nitrogen that enters living systems by nitrogen fixation is eventually converted from organic nitrogen back into nitrogen gas by bacteria. This process occurs in three steps in terrestrial systems: ammonification, nitrification, and denitrification. First, the **ammonification** or **nitrogen mineralization** process converts nitrogenous waste from living organisims or the remains of dead organisms into ammonium (NH_4^+) by certain bacteria and fungi. Second, this ammonium is then converted to nitrites (NO_2^-) and then nitrates (NO_3^-) by nitrifying bacteria and archaea, such as Nitrosomonas or *Nitrobacter*, through the process of **nitrification**. In addition, both soil and the ocean contain **archaeal** microbes, assigned to the Crenarchaeota, that convert ammonia to nitrites. They are more abundant than the nitrifying bacteria and may turn out to play an important role in the nitrogen cycle.. Like ammonium, nitrites and nitrates are found in water and the soil. Some nitrates are converted back into nitrogen gas, which is released into the atmosphere. The process, called **denitrification**, is conducted by bacteria, such as *Pseudomonas* and *Clostridium*, which use nitratewhen decomposing organic matterin the absence of oxygen. In the process of denitrification several intermediates are formed and may be released to the atmosphere including nitric oxide (NO) and nitrous oxide (N_2O, a greenhouse gas). Under anaerobic conditions in marine and freshwater systems other species of bacteria are able to oxidize ammonia with nitrite forming nitogen gas in a process called **anammox** (anaerobic ammonia oxidation).

In marine ecosystems, nitrogen compounds created by bacteria, or through decomposition, collects in ocean floor sediments. It can then be moved to land in geologic time by uplift of Earth's crust and thereby incorporated into terrestrial rock. Although the movement of nitrogen from rock directly into living systems has been traditionally seen as insignificant compared with nitrogen fixed from the atmosphere, a recent study showed that this process may indeed be significant and should be included in any study of the global nitrogen cycle.[1]

Human Alteration of the Nitrogen Cycle

Human activity can release nitrogen into the environment by two primary means: the combustion of fossil fuels, which releases different nitrogen oxides, and by the use of artificial fertilizers (which contain nitrogen and phosphorus compounds) in agriculture, which are then washed into lakes, streams, and rivers by surface runoff. Humans are also increasing the amount of reactive nitrogen in the environment by the cultivation of nitrogen fixing crops, such as soybeans. If the nitrogen fixation from leguminous crops (e.g. beans, alfalfa) is included, then the anthropogenic flux of nitrogen from the atmosphere to the land exceeds natural fluxes to the land. Atmospheric nitrogen (other than N_2) is associated with several effects on Earth's ecosystems including the production of

acid deposition (as nitric acid, HNO_3), also known as acid rain. Acid deposition damages healthy trees, destroys aquatic systems and erodes building materials such as marble and limestone. Like carbon dioxide, nitrous oxide (N_2O) is a greenhouse gas, potentially causing climate change when released during denitrification.

Humans are primarily dependent on the nitrogen cycle as a supporting ecosystem service for crop and forest productivity. Nitrogen fertilizers are added to enhance the growth of many crops and plantations (figure 12.2.4). The enhanced use of fertilizers in agriculture was a key feature of the green revolution that boosted global crop yields in the 1970s. The industrial production of nitrogen-rich fertilizers has increased substantially over time and now matches more than half of the input to the land from biological nitrogen fixation (90 megatons = 1 million tons of nitrogen each year). If the nitrogen fixation from legume crops is included, then the anthropogenic flux of nitrogen from the atmosphere to the land exceeds natural fluxes to the land. Fertilizers are washed into lakes, streams, and rivers by surface runoff, resulting in saltwater and freshwater **eutrophication**, a process whereby nutrient runoff causes the overgrowth of algae, the depletion of oxygen, and death of aquatic fauna. Excess nitrates in water supplies have also been linked to human health problems.



Figure 12.2.4**:** Fertilizer containing nitrogen is conventionally applied at large scales in agriculture. Image by Bob Nichols, USDA Natural Resources Conservation Service (public domain).

Efforts to reduce nitrogen pollution focus on increasing the efficiency of synthetic fertilizer use, altering feeding of animals to reduce nitrogen content in their excreta, and better processing of livestock waste and sewage sludge to reduce ammonia release. At the same time, increasing demand for food production from a growing global population with a greater appetite for meat is driving greater total fertilizer use, so there is no guarantee that better practices will lead to a reduction in the overall amount of nitrogen pollution.

Footnotes

1. 1 Scott L. Morford, Benjamin Z. Houlton, and Randy A. Dahlgren, “Increased Forest Ecosystem Carbon and Nitrogen Storage from Nitrogen Rich Bedrock,” *Nature* 477, no. 7362 (2011): 78–81.

Contributors and Attributions

Modified by Kyle Whittinghill (University of Pittsburgh) and Melissa Ha from the following sources:

- Biogeochemical Cycles, Energy, and Energy Enters Ecosystems Through Photosynthesis from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Nitrogen Cycle from *Biology* by John W. Kimball (licensed under CC-BY)

- Soil and Sustainability and Biogeochemical Cycles and the Flow of Energy in the Earth System from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (licensed under CC-BY). Download for free at CNX.
- Cycling of Matter from *AP Environmental Science* by University of California College Prep, University of California (licensed under CC-BY). Download for free at CNX.
- Nutrient Cycles from *Life Sciences Grade 10* by Siyavula (CC-BY)
- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd).

  20.2: Biogeochemical Cycles by OpenStax, is licensed CC BY
- Biogeochemical Cycles and the Flow of Energy in the Earth System from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin
- General Microbiology **Provided by**: Boundless.com. **License**: ***CC BY-SA: Attribution-ShareAlike***

12.2.1: Chemistry of the nitrogen cycle

12.2: The global nitrogen cycle is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

12.2.1: Chemistry of the nitrogen cycle

The nitrogen cycle is presented in Figure 12.2.1.1. Atmospheric N_2 is fixed into ammonia (NH_3) by free-living and symbiotic bacteria and archaea (diazotrophs), using the nitrogenase enzyme, the universal catalyst, to break the N_2 triple bond. In soil, NH_3 can be converted into ammonium ion (NH_4^+), which can be oxidized to nitrate ion (NO_3^-), in a three-step process called nitrification. Nitrite (NO_2^-) and NO_3^- ions are generated during nitrification and they may be reduced during the denitrification process, that is the stepwise reduction of NO_3^- to N_2 by four enzymes, generating intermediate products: NO_2^-, NO and N_2O. During the NO_3^- ammonification to NH_4^+, via NO_2^-, it can also occur NO_3^- reduction, thus producing N_2O (Thomson et al. 2012).

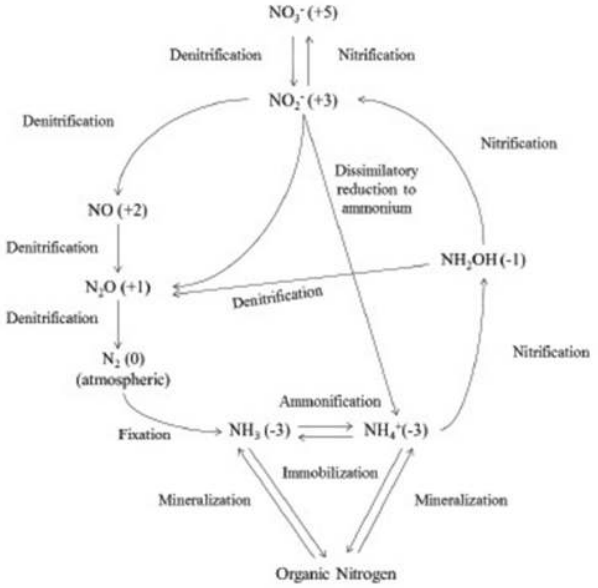


Figure 12.2.1.1: Nitrogen cycle. The oxidation states of N are given in parentheses and the names of the processes are indicated

Several concurrent processes are responsible for nitrogen gases emissions in agricultural soils (Bockman & Olfs 1998, Stevens & Laughlin 1998). Most of the N_2O is produced through the biological processes of nitrification and denitrification. Autotrophic aerobic nitrification (by ammonia-oxidizing bacteria and nitrite-oxidizing bacteria) and anaerobic denitrification, mediated by denitrifying bacteria, are the main microbial processes in the nitrogen cycle in the soil. Other microorganisms are involved in nitrification and denitrification processes: anammox bacteria can convert NH_4^+ and NO_2^- into N_2, under anaerobic conditions, while some fungi can produce N_2 and N_2O by denitrification and codenitrification, and archeae mediate nitrification in marine ecosystems and are capable of promoting denitrification in soils (Hayatsu et al. 2008)

A small fraction of N_2O is produced in non-biological processes: the chemical decomposition of nitrite (chemidenitrification) and hydroxylamine oxidation (NH_2OH) (Bremner 1997). Chemidenitrification is the decomposition of NO_2^- that occurs in neutral and acidic soils, causing volatilization and fixation of NO_2^- in the soil organic matter (Bremner et al. 1980, Bremner 1997). The amount of N_2O produced this way is almost negligible (Bremner et al. 1980, Bremner 1997). Hydroxylamine is an intermediate compound in the oxidation of NH_4^+ to NO_3^- that can produce much more N_2O than the chemidenitrification process (Bremner 1997). In neutral and acidic soils, N_2O is the main product of the NH_2OH oxidation, due to its reaction with Mn and Fe, while in calcareous soils (pH from 7.8 to 8.2) the NH_2OH reacts with $CaCO_3$ and the main product is N_2 (Bremner et al. 1980).

Nitrification and denitrification processes are further discussed, since they are the most important ones to the N_2O formation in soils. Nitrification is the aerobic oxidation of NH_4^+ to NO_3^- caused by chemoautotrophic bacteria in two stages: nitritation, in which the NH_4^+ is oxidized to NO_2^- by *Nitrosomonas* sp., *Nitrosospira* sp. and *Nitrosococcus* sp.; and nitratation, in which the NO_2^- is oxidized to NO_3^- by *Nitrobacter* sp., *Nitrosospira* sp. and *Nitrococcus* sp. (Moreira & Siqueira 2006). In general, nitrification can be summarized by the following reactions: nitritation: $2NH_4^+ + 3O_2 \rightarrow 2NO_2^- + H_2O + 4H^+ +$ energy; nitratation: $2NO_2^- + O_2 \rightarrow 2NO_3^-$ + energy. During this process, the NO_2^- concentration increases as NH_4^+ is being oxidized and then decreases as NO_3^- is being formed.

Denitrification is the reduction process of NO_3^- to N_2, mediated by facultative anaerobic bacteria, which correspond to 0.1-5.0% of the total bacteria population in the soil (Moreira & Siqueira 2006). This process can be complete, resulting in N_2, or can be incomplete, and a variable fraction of N can be emitted as NO and N_2O (Figure 1).

Despite denitrification is responsible for the most part of the N_2O produced in the soil, nitrification can also produce N_2O, when O_2 is limited. Bremner (1997) presented results from many researches which showed that nitrifying microorganisms can significantly contribute to N_2O emissions from soils. According to this author, the N_2O production during nitrification is increased when the soil pH and organic matter content increase and with increase in soil moisture (from air dried to field capacity) and in soil temperature (5-40ºC), by addition of nitrifiable N forms, animal manures and plant residues. N_2O produced by nitrification can be decreased when nitrification inhibitors are used.

Under anaerobic condition, the concentration of NO_2^-, which is a toxic compost, increases in the soil (Khalil et al. 2004), and it may be alternatively used by the nitrifying microorganisms as an electron final acceptor, resulting in N_2O and NO during nitrification (Snyder et al. 2009), as it follows:

References

BOCKMAN, O. C.; OLFS, H. W. Fertilizers, agronomy, and N2 O. Nutrient Cycling in Agroecosystems, Dordrecht, v. 52, n. 2/3, p. 165-170, 1998.

BREMNER, J. M. et al. Formation of nitrous oxide and dinitrogen by chemical decomposition of hydroxylamine in soils. Soil Biology and Biochemistry, Oxford, v. 12, n. 3, p. 263-269, 1980.

BREMNER, J. M. Sources of nitrous oxide in soils. Nutrient Cycling in Agroecosystems, Dordrecht, v. 49, n. 1-3, p. 7-16, 1997.

HAYATSU, M.; TAGO, K.; SAITO, M. Various players in the nitrogen cycle: diversity and functions of the microorganisms involved in nitrification and denitrification. Soil Science and Plant Nutrition, Tokyo, v. 54, n. 1, p. 33-45, 2008.

MOREIRA, F. M. S.; SIQUEIRA, J. O. Microbiologia e bioquímica do solo. 2. ed. Lavras: UFLa, 2006.

SNYDER, C. S. et al. Review of greenhouse gas emissions from crop production systems and fertilizer management effects. Agriculture, Ecosystems & Environment, Amsterdam, v. 133, n. 3-4, p. 247-266, 2009.

STEVENS, R. J.; LAUGHLIN, R. J. Measurement of nitrous oxide and di-nitrogen emissions from agricultural soils. Nutrient Cycling in Agroecosystems, Dordrecht, v. 52, n. 2-3, p. 131-139, 1998.

THOMSON, A. J. et al. Biological sources and sinks of nitrous oxide and strategies to mitigate emissions. Philosophical Transactions of The Royal Society Biological Sciences, London, v. 367, n. 1593, p. 1157- 1168, 2012.

Excerpted from:

Signor, D., & Cerri, C. E. P. (2013). Nitrous oxide emissions in agricultural soils: a review. *Pesquisa Agropecuária Tropical*, *43*, 322-338. Accessed December 2023 at https://www.scielo.br/j/pat/a/yyfWQ9zyWJVdknyBfYFP3wD/, CC-BY-4.0

12.2.1: Chemistry of the nitrogen cycle is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

12.3: Temporal variations in the global nitrogen cycle

Denitrification and N_2 fixation are N transformation processes which can both control and be controlled by environmental factors (Vitousek et al., 2002; Seitzinger et al., 2006). Together with anammox, N_2 fixation and denitrification are the primary biological processes that control N_2 flux from ecosystems (Fox et al., 2014; Nifong et al., 2020). Denitrification is a form of anaerobic respiration in which microbes break down organic carbon (C) for energy and use nitrate (NO_3^-)in the place of oxygen (O_2) as an electron acceptor, so supply of organic C and $NO_3^-$are important limitations on denitrification rates (Wall et al., 2005). Along with the required reactant availability, environmental factors including oxygen concentration and temperature can alter denitrification rates. Biological N_2 fixation is the microbially-mediated conversion of N_2 gas into ammonium (NH_4^+); it is an energetically expensive process so high availability of $NH_4^+$and/or other forms of dissolved inorganic N (DIN) may cause microbes to reduce the process rate to conserve energy (Burris and Roberts, 1993; Bandyopadhyay et al., 2013). Light may be especially important for controlling N_2 fixation rates in shallow aquatic ecosystems like streams, where it is often carried out by cyanobacteria who obtain their energy via photosynthesis (Burris and Roberts, 1993; Scott and Marcarelli, 2012). Light intensity along with these other factors vary through time, and therefore they may lead to temporal variation in rates of both of these N transformation processes in aquatic ecosystems. Yet, the mechanisms by which environmental factors affect rates of denitrification and N_2 fixation may differ depending on the time scale.

In streams and rivers in temperate climate zones, seasonal changes in denitrification and N_2 fixation rates in response to changes in environmental conditions may be gradual over the course of days to weeks, caused by the abundance and/or activity of microbes increasing or decreasing, or the identity of the microbes within the community changing. Environmental characteristics that show strong seasonal changes in temperate streams and rivers include nutrient availability, temperature, and light availability (Coble et al., 2019). Low biotic uptake rates in winter can lead to increased inorganic *N* concentrations in temperate forested streams (Stottlemyer and Toczydlowski, 1999a,b) which could provide reactants for denitrification and reduce rates of N_2 fixation. Temperate forests have large inputs of allochthonous organic matter during the fall and high dissolved organic C (DOC) flux during snowmelt (Stottlemyer and Toczydlowski, 1999a), either of which could provide a C source for denitrifying bacteria. Discharge varies seasonally corresponding to rainy seasons or snowmelt, which leads to predictable patterns of disturbance that can move, scour and mobilize riverbed substrates, disturbing microbial communities, and the microenvironments of these communities, which control nutrient cycling processes (O'Connor et al., 2012). Seasonal changes in discharge can also alter nutrient concentrations both by altering patterns of nutrient delivery and via dilution (Meyer and Likens, 1979; Horner et al., 1990; Stottlemyer and Toczydlowski, 1999a). Denitrification and N_2 fixing microbial communities can both be sensitive to seasonal temperature changes, with N_2 - fixing microbes often preferring warmer temperatures (Scott and Marcarelli, 2012), and denitrification activity decreasing during periods with colder temperatures such as in the winter (Christensen et al., 1990; Kim et al., 2006). Light also varies seasonally, especially in forested streams where canopy cover will shade streams differently depending on the season and vegetation type (Roberts and Mulholland, 2007; Bowes et al., 2012). Increased light availability in spring and fall due to low canopy cover can lead to higher rates of in-stream nutrient uptake due to increases in primary productivity (Roberts and Mulholland, 2007).

Environmental factors in temperate streams can also shift much more rapidly over hours to days, which could instead affect rates through enzymatic regulation (Grimm, 1987; Marcarelli and Wurtsbaugh, 2006; Marcarelli et al., 2008). N cycling rates can both increase and decrease over the course of hours with factors such as temperature or substrate availability affecting enzyme activity (Grimm, 1987; Marcarelli and Wurtsbaugh, 2006). Temperature variation can occur on an hourly scale, and the enzymes and microbes facilitating N cycling processes may have reduced activity at cold temperatures (Yvon-Durocher et al., 2010; Boulêtreau et al., 2012). Also, changes in light can alter N_2 fixation rates by cyanobacteria in some ecosystems on a diel cycle because it is tightly coupled with photosynthesis (Howarth et al., 1988; Grimm and Petrone, 1997). Nutrient concentrations can rapidly change due to storm runoff, which could also alter N process rates. Because N_2 fixation is so energy intensive, enzymes can be rapidly deregulated to save energy following an influx of DIN (Howarth et al., 1988; Grimm and Fisher, 1989; Grimm and Petrone, 1997). But, the distinction between community or enzymatic shifts can be complicated because some environmental factors such as temperature and light vary at multiple time scales. For example, hydrological disturbances such as flooding caused by storms are overlain on seasonal patterns created by snowmelt and rainfall, and may therefore alter biogeochemical process rates through a mix of shorter-term enzymatic changes and longer-term community shifts (Grimm and Fisher, 1989). Stormdriven surface runoff can carry nutrients (Meyer and Likens, 1979; Horner et al., 1990; Stottlemyer and Toczydlowski, 1999a), change water temperature,

and create short-term increases in water velocity and benthic disturbance (O'Connor et al., 2012), all of which may act independently and in concert to create variation in process rates.

Understanding relationships between N cycling rates and different environmental drivers that vary across time scales may provide insight into how climate change will affect these processes in the future. Climate change can lead to temperature, stream flow, and growing season changes (Barnett et al., 2005; Backlund et al., 2008; Christiansen et al., 2011), all environmental characteristics which may be influential to denitrification and N_2 fixation rates. Storms are becoming separated by longer periods of drought, but overall higher rainfall can occur driven by more intense storm events (Mathbout et al., 2018). It is also not wellunderstood how resistant and resilient rates of N transformations are to hydrologic disturbances, which is important in the face of ongoing climate change. Resistance and recovery (or resilience) are concepts used throughout ecological studies, including studies on the impact of hydrological disturbances, anthropogenic disturbances, and climate change (Holling, 1973; Bahadur et al., 2013; Pope et al., 2014; Reisinger et al., 2017). Resistance describes how much the structure and processes of an ecosystem change in response to disturbances of different magnitudes, while recovery describes how quickly the ecosystem returns to the same structure and/or function after a disturbance event (Hershkovitz and Gasith, 2013; Reisinger et al., 2017). Initial declines in process rates following a hydrologic disturbance may be rapid; however, recovery may be slower if the microbial community was altered or reduced in biomass rather than enzymatic downregulation. Understanding how large hydrological disturbances affect denitrification and N_2 fixation can provide insight into how they will be affected by changing climactic conditions.

The complexity of seasonal vs. shorter-term environmental changes coupled with different mechanisms of microbial responses that lead to changes in rates makes it difficult to decipher how and why N cycle processes vary

References

Backlund, P., Schimel, D., Janetos, A., Hatfield, J., Ryan, M. G., Archer, S. R., et al. (2008). Introduction. The effects of climate change on agriculture, land resources, water resources, and biodiversity in the United States, in *United States Climate Change Science Program Synthesis and Assessment Product*, Vol. 4, 11–20..

Bahadur, A. V., Ibrahim, M., and Tanner, T. (2013). Characterising resilience: unpacking the concept for tackling climate change and development. *Clim. Dev.* 5, 55–65. doi: 10.1080/17565529.2012.762334.

Bandyopadhyay, A., Elvitigala, T., Liberton, M., and Pakrasi, H. B. (2013). Variations in the rhythms of respiration and nitrogen fixation in members of the unicellular diazotrophic cyanobacterial genus Cyanothece. *Plant Physiol.* 161, 1334–1346. doi: 10.1104/pp.112.208231.

Barnett, T. P., Adam, J. C., and Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* 438, 303–309. doi: 10.1038/nature04141.

Boulêtreau, S., Salvo, E., Lyautey, E., Mastrorillo, S., and Garabetian, F. (2012). Temperature dependence of denitrification in phototrophic river biofilms. *Sci. Total Environ.* 416, 323–328. doi: 10.1016/j.scitotenv.2011.11.066.

Bowes, M. J., Ings, N. L., McCall, S. J., Warwick, A., Barrett, C., Wickham, H. D., et al. (2012). Nutrient and light limitation of periphyton in the River Thames: implications for catchment management. *Sci. Total Environ.* 434, 201–212. doi: 10.1016/j.scitotenv.2011.09.082

Burris, R. H., and Roberts, G. P. (1993). Biological nitrogen fixation. *Annu. Rev. Nutr.* 13, 317–335. doi: 10.1146/annurev.nu.13.070193.001533.

Christensen, P. B., Nielsen, L. P., Sørensen, J., and Revsbech, N. P. (1990). Denitrification in nitrate-rich streams: Diurnal and seasonal variation related to benthic oxygen metabolism. *Limnol. Oceanogr.* 35, 640–651. doi: 10.4319/lo.1990.35.3.0640.

Coble, A. A., Marcarelli, A. M., and Kane, E. S. (2019). Year-round measurements reveal seasonal drivers of nutrient uptake in a snowmelt-driven headwater stream. *Freshw. Sci.* 38, 156–169. doi: 10.1086/701733.

Fox, R. J., Fisher, T. R., Gustafson, A. B., Jordan, T. E., Kana, T. M., and Lang, M. W. (2014). Searching for the missing nitrogen: biogenic nitrogen gasses in groundwater and streams. *J. Agric. Sci.* 152, S96–S106. doi: 10.1017/S0021859614000070.

Grimm, N. B. (1987). Nitrogen dynamics during succession in a desert stream. *Ecology* 68, 1157–1170. doi: 10.2307/1939200.

Grimm, N., and Petrone, K. (1997). Nitrogen fixation in a desert stream ecosystem. *Biogeochemistry* 37, 33–61. doi: 10.1023/A:1005798410819.

Grimm, N. B., and Fisher, S. G. (1989). Stability of periphyton and macroinvertebrates to disturbance by flash floods in a desert stream. *J. North Am. Benthol. Soc.* 8, 293–307. doi: 10.2307/1467493.

Hershkovitz, Y, and Gasith, A. (2013). Resistance, resilience, and community dynamics in mediterranean-climate streams. *Hydrobiologia* 719, 59–75. doi: 10.1007/s10750-012-1387-3.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Ann. Rev. Ecol. Syst.* 4, 1–23. doi: 10.1146/annurev.es.04.110173.000245.

Horner, R. R., Welch, E. B., Seele, M. R., and Jacoby, J. M. (1990). Response of periphyton to changes in current velocity, suspended sediment, and phosphorus concentration. *Freshw. Biol.* 24, 215–232. doi: 10.1111/j.1365-2427.1990.tb00704.x.

Howarth, R. W., Marino, R., and Cole, J. J. (1988). Nitrogen fixation in freshwater, estuarine, and marine ecosystems. 2. Biogeochemical control. *Limnol. Oceanogr.* 33(4, Part 2), 688–701. doi: 10.4319/lo.1988.33.4part2.0688.

Kim, D. J., Lee, D. I., and Keller, J. (2006). Effect of temperature and free ammonia on nitrification and nitrite accumulation in landfill leachate and analysis of its nitrifying bacterial community by FISH. *Bioresour. Technol.* 97, 459–468. doi: 10.1016/j.biortech.2005.03.032.

Marcarelli, A. M., and Wurtsbaugh, W. A. (2006). Temperature and nutrient supply interact to control nitrogen fixation in oligotrophic streams: an experimental examination. *Limnol. Oceanogr.* 51, 2278–2289. doi: 10.4319/lo.2006.51.5.2278.

Mathbout, S., Lopez-Bustins, J. A., Roy,é, D., Martin-Vide, J., Bech, J., and Rodrigo, F. S. (2018). Observed changes in daily precipitation extremes at annual timescale over the eastern Mediterranean during 1961–2012. *Pure Appl. Geophys.* 175, 3875–3890. doi: 10.1007/s00024-017-1695-7.

Meyer, J. L., and Likens, G. E. (1979). Transport and transformation of phosphorus in a forest stream ecosystem. *Ecography* 60, 1255–1269. doi: 10.2307/1936971

Nifong, R. L., Taylor, J. M., Adams, G., Moore, M. T., and Farris, J. L. (2020). Recognizing both denitrification and nitrogen consumption improves performance of stream diel N2 flux models. *Limnol. Oceanogr.: Methods* 18, 169–182. doi: 10.1002/lom3.10361.

O'Connor, B. L., Harvey, J. W., and McPhillips, L. E. (2012). Thresholds of flow-induced bed disturbances and their effects on stream metabolism in an agricultural river. *Water Resour. Res.* 48:W08504. doi: 10.1029/2011WR011488.

Pope, K. L., Allen, C. R., and Angeler, D. G. (2014). Fishing for resilience. *Trans. Am. Fish. Soc.* 143, 467–478. doi: 10.1080/00028487.2014.880735.

Reisinger, A. J., Rosi, E. J., Bechtold, H. A., Doody, T. R., Kaushal, S. S., and Groffman, P. M. (2017). Recovery and resilience of urban stream metabolism following Superstorm Sandy and other floods. *Ecosphere* 8, e01776. doi: 10.1002/ecs2.1776.

Roberts, B. J., and Mulholland, P. J. (2007). In-stream biotic control on nutrient biogeochemistry in a forested stream, West Fork of Walker Branch. *J. Geophys. Res.: Biogeosci.* 112, 1–11. doi: 10.1029/2007JG000422.

Scott, J. T., and Marcarelli, A. M. (2012). Cyanobacteria in freshwater benthic environments, in *Ecology of the Cyanobacteria II: Their Diversity in Time and Space*, 271–289. doi: 10.1007/978-94-007-3855-3_9.

Seitzinger, S., Harrison, J., Bohlke, J., Bouwman, A., Lowrance, R., Peterson, B., et al. (2006). Denitrification across landscapes and waterscapes: a synthesis. *Ecol. Appl.* 16, 2064–2090. doi: 10.1890/1051-0761(2006)016[2064:DALAWA]2.0.CO;2.

Stottlemyer, R., and Toczydlowski, D. (1999a). Seasonal change in precipitation, snowpack, snowmelt, soil water and stream water chemistry, northern Michigan. *Hydrogeol. J.* 2231, 2215–2231. doi: 10.1002/(SICI)1099-1085(199910)13:14/15<2215::AID-HYP882>3.0.CO;2-V.

Stottlemyer, R., and Toczydlowski, D. (1999b). Seasonal relationships between precipitation, forest floor, and stream water nitrogen, Isle Royale, Michigan. *Soil Sci. Soc. Am. J.* 63, 389–398. doi: 10.2136/sssaj1999.03615995006300020018x.

Vitousek, P. M., Cassman, K., Cleveland, C., Crews, T., Field, C. B., Grimm, N. B., et al. (2002). Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry* 57–58, 1–45. doi: 10.1007/978-94-017-3405-9_1.

Wall, L. G., Tank, J. L., Royer, T. V., and Bernot, M. J. (2005). Spatial and temporal variability in sediment denitrification within an agriculturally influenced reservoir. *Biogeochemistry* 76, 85–111. doi: 10.1007/s10533-005-2199-6.

Yvon-Durocher, G., Jones, J. I., Trimmer, M., Woodward, G., and Montoya, J. M. (2010). Warming alters the metabolic balance of ecosystems. *Philos. Trans. R. Soc. B: Biol. Sci.* 365, 2117–2126. doi: 10.1098/rstb.2010.0038

Excerpted from:

Nevorski, K. C., & Marcarelli, A. M. (2022). High Daily and Year-Round Variability in Denitrification and Nitrogen Fixation in a Northern Temperate River. *Frontiers in Water*, *4*, 894554. Accessed December 2023 at https://www.frontiersin.org/journals/water/articles/10.3389/frwa.2022.894554/full#B65. CC-BY-4.0

12.3: Temporal variations in the global nitrogen cycle is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

SECTION OVERVIEW

12.4: Nitrous Oxide (N2O)

12.4.1: Terrestrial

12.4.2: Marine

12.4.3: Human Sources

12.4: Nitrous Oxide (N2O) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

12.4.1: Terrestrial

From the major processes involved in the nitrogen cycle in soil and sediments, four microbially mediated reaction pathways can contribute to N_2O emissions (Baggs, 2011; Quick et al., 2019): nitrification and nitrifier denitrification in oxic environments and denitrification and dissimilatory nitrate reduction to ammonium (DNRA) in suboxic and anoxic environments. The reactive nitrogen species are oxidized or reduced through a sequence of electron transfer steps, promoted by enzymatic reaction pathways. In all of these pathways, N_2O is produced as an intermediate reaction product (Fig. 1). In aquatic ecosystems, denitrification is regarded as the predominant source of N_2O, and nitrifier-denitrification is likely more significant than nitrification (Quick et al., 2019). These processes can co-occur over a broad range of oxygen (O_2)/redox and moisture content (MC) conditions, within oxic/anoxic microsites in sediments (Jørgensen and Revsbech, 1985; Seitzinger et al., 2006).

Denitrification, a facultative anaerobic process, is the reduction of nitrate (NO_3^-) or nitrite (NO_2^-) to N_2O and di-nitrogen (N_2) performed by heterotrophic bacteria (denitrifiers). Denitrifying microorganisms also include ammonia-oxidizing chemolithotrophic bacteria, which reduce NO_2^- to N_2O aerobically, archaea, fungi and other eukaryotes (Baggs, 2011). Part of the denitrifying bacteria and archaea are missing the genes encoding the enzymes involved in the reduction of nitric oxide (NO) and N_2O to N_2, which can lead to incomplete pathways and N_2O release (Stein & Klotz, 2016). Denitrification enzymes are inhibited by O_2, particularly N_2O reductase, which catalyzes the reduction of N_2O to N_2. Thus, under suboxic conditions, N_2O may be the end product of denitrification (Knowles, 1982). Apart from O_2 conditions, several other environmental factors control N_2O production from denitrification, specifically the N_2O yield (N_2O/(N_2O+ N_2)), including water content, NO_3^- availability, C quality and availability and C:NO_3^- (Quick et al., 2019).

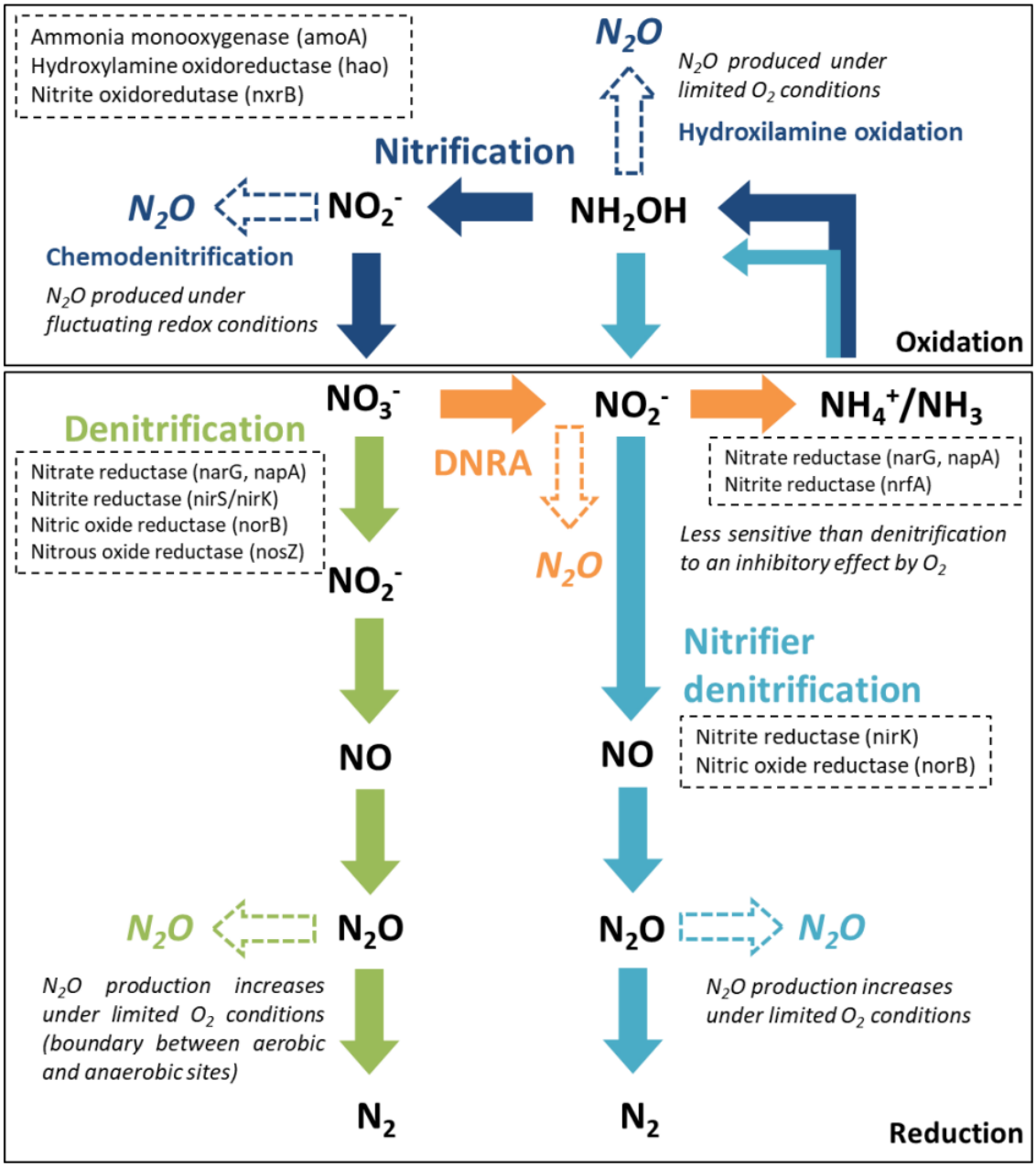


Figure 12.4.1.1: Major nitrogen cycle processes that produce N_2O. Dashed arrows indicate the steps were N_2O emissions may occur, along each reaction pathway (adapted from Giles et al., 2012, CC BY 3.0 SA). The enzymes and encoding genes of each process are indicated in the dashed boxes. A reference is made to the O_2 conditions promoting N_2O production.

Nitrification and nitrifier denitrification occur under different environmental conditions and both oxidize ammonia. Nitrification is the oxidation of ammonia (NH_3) or ammonium (NH_4^+) to NO_2^- by ammonia oxidizers (cohort I; primary nitrifiers) and to NO_3^- by nitrite oxidizers (cohort II; secondary nitrifiers). Ammonia/um can be directly oxidized to nitrate by complete ammonia oxidizers (comammox, cohort III). Cohorts II and III only include chemolithotrophic microbes (Stein & Klotz, 2016). Under certain conditions, ammonia oxidizers can significantly contribute to N_2O emissions by two reactions along this pathway, hydroxylamine oxidation (biotic and abiotic) and chemodenitrification (abiotic). The main factors influencing hydroxylamine oxidation are aerobic conditions and NH_3availability, whilst chemodenitrification is limited by NO_2^- availability and may occur under fluctuating redox conditions (Quick et al., 2019). Nitrifier denitrification, strictly carried out by ammonia oxidizers, converts $NH_3$to N_2 gas. It is supported by different O_2 conditions, having both oxidation and reduction steps. The first steps are oxidative (ammonia is oxidized to (NO_2^-) and the final steps are reductive (NO_2^- is sequentially reduced to NO, N_2O and N_2). Factors influencing N_2O production

from nitrifier denitrification include O_2 conditions, $NH_4^+$and C availability (Quick et al., 2019). It differs from nitrification and coupled nitrification-denitrification as there is no (NO_3^- involved.

DNRA is performed by both bacteria and fungi, using C as an electron donor. During nitrate ammonification, nitrate is reduced to NO_2^- and NH_4^+, and N_2O is produced as a by-product during the NO_2^- reduction stage. Reducing conditions are an important factor controlling this process, which is mostly anaerobic but can also occur under relatively oxic conditions, being less sensitive to O_2 than denitrifiers (Giles et al., 2012). The C:(NO_3^- ratio is also considered an important controlling factor in the process (Quick et al., 2019).

References

Baggs EM. 2011. Soil microbial sources of nitrous oxide: recent advances in knowledge, emerging challenges and future direction. Current Opinion in Environmental Sustainability 3(5):321–327 DOI 10.1016/j.cosust.2011.08.011

Giles M, Morley N, Baggs E, Daniell T. 2012. Soil nitrate reducing processes—drivers, mechanisms for spatial variation, and significance for nitrous oxide production. Frontiers in Microbiology 3:407 DOI 10.3389/fmicb.2012.00407.

Jørgensen BB, Revsbech NP. 1985. Diffusive boundary-layers and the oxygenuptake of sediments and detritus. Limnology and Oceanography 30:111–122 DOI 10.4319/lo.1985.30.1.0111.

Knowles R. 1982. Denitrification. Microbiological Reviews 46(1):43–70.

Quick AM, Reeder WJ, Farrell TB, Tonina D, Feris KP, Benner SG. 2019. Nitrous oxide from streams and rivers: A review of primary biogeochemical pathways and environmental variables. Earth-Science Reviews 191:224–262 DOI 10.1016/j.earscirev.2019.02.021.

Seitzinger S, Harrison JA, Bohlke JK, Bouwman AF, Lowrance R, Peterson B, Tobias C, Van Drecht G. 2006. Denitrification across landscapes and waterscapes: a synthesis. Ecological Applications 16:2064–2090 DOI 10.1890/1051-0761(2006)016%5b2064:DALAWA.

Stein LY, Klotz MG. 2016. Primer: the nitrogen cycle. Current Biology 26:R94–R98 DOI 10.1016/j.cub.2015.12.021.

Excerpted from

Pinto R, Weigelhofer G, Brito AóG, Hein T. 2021. Effects of dry-wet cycles on nitrous oxide emissions in freshwater sediments: a synthesis. PeerJ 9:e10767 http://doi.org/10.7717/peerj.10767 CC-BY-4.0

12.4.1: Terrestrial is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

12.4.2: Marine

The marine nitrogen cycle has been widely investigated, as nitrogen is one of the main limiting factors of primary production in the upper sunlit layers of the oceans (Arrigo, 2005; Codispoti, 1997) and the ocean accounts for about half of the global net primary production (Field et al., 1998; Gruber and Galloway, 2008). In the traditional view, the marine nitrogen cycle includes nitrogen fixation as the main input of nitrogen in the ocean and dinitrogen gas formed by denitrification as the main output, so that these two pathways are mainly responsible for the marine nitrogen budget status (Karl et al., 1997). Codispoti et al. (2001) suggested that, in the present-day ocean, the nitrogen budget is not in a steady state but rather out of balance, with denitrification fluxes being underestimated . Nitrogen fixation is mediated by few microorganisms, including cyanobacteria, while denitrification is performed by a wide range of microorganisms with different metabolic features, able to switch from aerobic to anaerobic nitrate (NO_3^-)-dependent respiration modes (Lam and Kuypers, 2011). In this classical view, nitrification, representing the major oxidative part of the cycle, connecting organic nitrogen to NO_3^-(Codispoti et al., 2001; Lam and Kuypers, 2011), was seen exclusively as an aerobic process carried out by ammonia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria, members of the β - and γ proteobacteria. The nitrification reaction is divided into two steps, performed by distinct bacterial groups. In the first part, ammonium $(NH_4^+)$is oxidized to nitrite (NO_2^-), whereas in the second $NO_2^-$is oxidized to nitrate (NO_3^-). In both cases, oxygen serves as the electron acceptor, although AOB have been reported to perform nitrification in suboxic conditions (Lam et al., 2007; Schmidt and Bock, 1997).

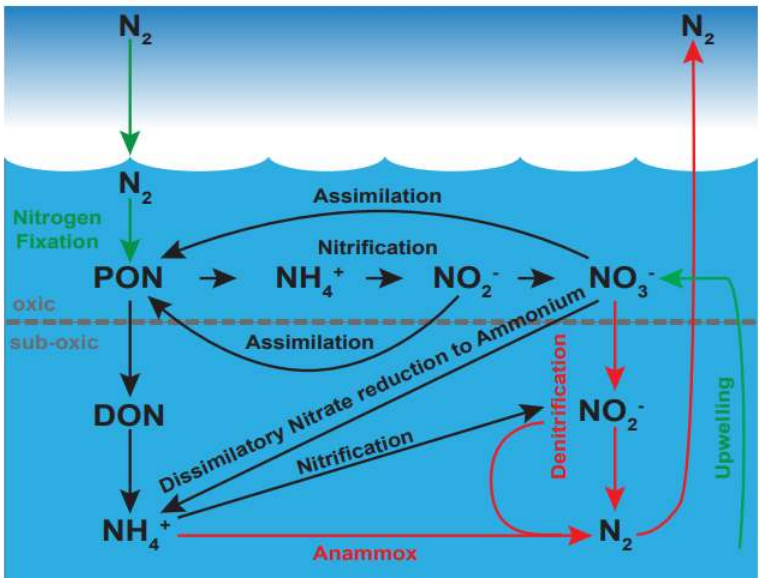


Figure 12.4.2.1: The marine biogeochemical nitrogen cycle. The main redox reactions involved are included: in green are the processes responsible for gains of nitrogen to the ocean, and in red are the losses. Modified from Arrigo (2005).

The overall understanding of the marine nitrogen cycle has substantially changed in the last decade (Fig. 1). Specific archaea were discovered to be important players in the marine nitrogen cycle (Venter et al., 2004) as some of them perform nitrification in the marine water column and sediment (Francis et al., 2005; Könneke et al., 2005; Wuchter et al., 2006). The group of archaea capable of nitrification has recently been relocated in a separate phylum named Thaumarchaeota (Brochier-Armanet et al., 2008; Spang et al., 2010). Compared to their bacterial counterpart, ammonia-oxidizing archaea (AOA) are often more abundant in the ocean (Karner et al., 2001; Lam et al., 2007; Wuchter et al., 2006), accounting for 20% of picoplankton and 40% of the estimated total number of cells (Karner et al., 2001). These microorganisms are able to cope with low-oxygen conditions (Coolen et al., 2007; Lam et al., 2007; Park et al., 2010; Pitcher et al., 2011b; Sinninghe Damsté et al., 2002a), have low substrate requirements (Martens-Habbena et al., 2009) and are able to utilize a highly energy-efficient CO_2-fixation pathway (Könneke et al., 2014); new coastal marine AOA isolates show obligate mixotrophy and vary in their adaptive ability to different environmental parameters (Qin et al., 2014). All these features have been suggested to provide a reason for the observed dominance of AOA over AOB as ammonia oxidizers in the open oceans (Könneke et al., 2014; Pester et al., 2011). Moreover, a "novel" process in the nitrogen cycle, named anammox, was discovered. Anaerobic ammonia oxidizing (anammox) bacteria are a unique group of microorganisms member of the order Planctomycetales (Strous et al., 1999). They are able to oxidize ammonium $(NH_4^+)$to molecular nitrogen (N_2) under

anoxic conditions, using nitrite (NO_2^-)as the electron acceptor (van de Graaf et al., 1995). Anammox bacterial activity has been detected in marine anoxic sediments and waters (Dalsgaard et al., 2003; Kuypers et al., 2003; Thamdrup and Dalsgaard, 2002) and has been recognized to contribute, along with denitrifying bacteria, to the loss of N_2 from the ocean (Galán et al., 2009; Hamersley et al., 2007; Kuypers et al., 2005; Lam et al., 2009; Thamdrup et al., 2006). Despite different oxygen tolerances, anammox bacteria and Thaumarchaeota have been observed to coexist in different settings, particularly in oxygen-deficient zones (ODZs) and anoxic waters (Coolen et al., 2007; Francis et al., 2005; Lam et al., 2007; Pitcher et al., 2011b; Woebken et al., 2007). These two microbial groups can potentially benefit from each other, because the thaumarchaeotal nitrification might be coupled with the anammox process by providing the NO_2^-anammox bacteria need and, at the same time, consume oxygen, to which anammox bacteria are sensitive. Alternatively, when nitrite is provided to anammox by other sources, the two groups might compete for NH_4^+(Yan et al., 2012).

References

Arrigo, K. R.: Marine microorganisms and global nutrient cycles, Nature, 437, 349–356, 2005.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P.: Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota, Nat. Rev. Microbiol., 6, 245–52, 2008.

Codispoti, L. A.: The limits to growth, Nature, 387, 237–238, 1997.

Codispoti, L. A., Brandes, J. A., Christensen, J. P., Devol, A. H., Naqvi, S. W. A., Paerl, H. W., and Yoshinari, T.: The oceanic fixed nitrogen and nitrous oxide budgets: moving targets as we enter the anthropocene?, Sci. Mar., 65, 85–105, 2001.

Coolen, M. J. L., Abbas, B., van Bleijswijk, J., Hopmans, E. C., Kuypers, M. M. M., Wakeham, S. G., and Sinninghe Damsté, J. S.: Putative ammonia-oxidizing Crenarchaeota in suboxic waters of the Black Sea: a basin-wide ecological study using 16S ribosomal and functional genes and membrane lipids., Environ. Microbiol., 9, 1001–16, 2007.

Dalsgaard, T., Canfield, D. E., Petersen, J., Thamdrup, B., and Acuña-González, J.: N2 production by the anammox reaction in the anoxic water column of Golfo Dulce, Costa Rica., Nature, 422, 606–8, 2003.

Field, C. B., Behernefeld, M. J., Randerson, J. T., and Falkowski, P. G.: Primary production of the biosphere: integrating terrestrial and oceanic components, Science, 281, 237–240, 1998.

Francis, C. A., Roberts, K. J., Beman, J. M., Santoro, A. E., and Oakley, B. B.: Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean, Proc. Natl. Acad. Sci. USA, 102, 14683–14688, 2005.

Galán, A., Molina, V., Thamdrup, B., Woebken, D., Lavik, G., Kuypers, M. M. M., and Ulloa, O.: Anammox bacteria and the anaerobic oxidation of ammonium in the oxygen minimum zone off northern Chile, Deep Sea Res. Part II Top. Stud. Oceanogr., 56, 1021–1031, 2009

Gruber, N. and Galloway, J. N.: An Earth-system perspective of the global nitrogen cycle, Nature, 451, 293–6, 2008.

Karl, D., Letelier, R., Tupas, L., Dore, J., Christian, J., and Hebel, D.: The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean, Nature, 388, 533–538, 1997

Karner, M. B., DeLong, E. F., and Karl, D. M.: Archaeal dominance in the mesopelagic zone of the Pacific Ocean, Nature, 409, 507– 10, 2001.

Könneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., and Stahl, D. A.: Isolation of an autotrophic ammonia-oxidizing marine archaeon, Nature, 437, 543–6, 2005.

Kuypers, M. M. M., Lavik, G., Woebken, D., Schmid, M., Fuchs, B. M., Amann, R., Jørgensen, B. B., and Jetten, M. S. M.: Massive nitrogen loss from the Benguela upwelling system through anaerobic ammonium oxidation, Proc. Natl. Acad. Sci. USA, 102, 6478–83, 2005.

Lam, P. and Kuypers, M. M. M.: Microbial nitrogen cycling processes in oxygen minimum zones, Ann. Rev. Mar. Sci., 3, 317–345, 2011.

Lam, P., Lavik, G., Jensen, M. M., van de Vossenberg, J., Schmid, M., Woebken, D., Gutiérrez, D., Amann, R., Jetten, M. S. M., and Kuypers, M. M. M.: Revising the nitrogen cycle in the Peruvian oxygen minimum zone, Proc. Natl. Acad. Sci. USA, 106,

4752– 4757, 2009.

Lam, P., Jensen, M. M., Lavik, G., McGinnis, D. F., Müller, B., Schubert, C. J., Amann, R., Thamdrup, B., and Kuypers, M. M. M.: Linking crenarchaeal and bacterial nitrification to anammox in the Black Sea, Proc. Natl. Acad. Sci. USA, 104, 7104–7109, 2007.

Martens-Habbena, W., Berube, P. M., Urakawa, H., de la Torre, J. R., and Stahl, D. A.: Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria., Nature, 461, 976–979, 2009.

Park, B.-J., Park, S.-J., Yoon, D.-N., Schouten, S., Sinninghe Damsté, J. S., and Rhee, S.-K.: Cultivation of autotrophic ammonia-oxidizing archaea from marine sediments in coculture with sulfur-oxidizing bacteria, Appl. Environ. Microbiol., 76, 7575–87, 2010.

Pester, M., Schleper, C., and Wagner, M.: The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology, Curr. Opin. Microbiol., 14, 300–306, 2011.

Pitcher, A., Villanueva, L., Hopmans, E. C., Schouten, S., Reichart, G.-J., and Sinninghe Damsté, J. S.: Niche segregation of ammonia-oxidizing archaea and anammox bacteria in the Arabian Sea oxygen minimum zone, ISME J., 5, 1896–904, 2011b.

Qin, W., Amin, S. A., Martens-Habbena, W., Walker, C. B., Urakawa, H., Devol, A. H., Ingalls, A. E., Moffett, J. W., Armbrust, E. V., and Stahl, D. A.: Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation, Proc. Natl. Acad. Sci., 111, 12504–12509, 2014.

Schmidt, I. and Bock, E.: Anaerobic ammonia oxidation with nitrogen dioxide by Nitrosomonas eutropha, Arch. Microbiol., 167, 106–111, 1997.

Sinninghe Damsté, J. S., Strous, M., Rijpstra, W. I. C., Hopmans, E. C., Geenevasen, J. A. J., van Duin, A. C. T., van Niftrik, L., and Jetten, M. S. M.: Linearly concatenated cyclobutane lipids form a dense bacterial membrane, Nature, 419, 708–712, 2002c.

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D. A., Wagner, M., and Schleper, C.: Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota, Trends Microbiol., 18, 331–40, 2010.

Strous, M., Fuerst, J., Kramer, E. H. M., Logemann, S., Muyzer, G., van de Pas-Schoonen, K., Webb, R., Kuenen, J. G., and Jetten, M. S. M.: Missing lithotroph identified as new planctomycete, Nature, 400, 446–449, 1999.

Thamdrup, B., Dalsgaard, T., Jensen, M. M., Ulloa, O., Farías, L., and Escribano, R.: Anaerobic ammonium oxidation in the oxygen-deficient waters off northern Chile, Limnol. Oceanogr., 51, 2145–2156, 2006.

Thamdrup, B. and Dalsgaard, T.: Production of N2 through Anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments, Appl. Environ. Microbiol., 68, 1312–1318, 2002.

Van de Graaf, A. A., Mulder, A., de Bruijn, P., Jetten, M. S. M., Robertson, L. A., and Kuenen, J. G.: Anaerobic oxidation of ammonium is a biologically mediated process, Appl. Environ. Microbiol., 61, 1246–1251, 1995.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H., and Smith, H. O.: Environmental genome shotgun sequencing of the Sargasso Sea., Science, 304, 66–74, 2004.

Woebken, D., Fuchs, B. M., Kuypers, M. M. M., and Amann, R.: Potential interactions of particle-associated anammox bacteria with bacterial and archaeal partners in the Namibian upwelling system, Appl. Environ. Microbiol., 73, 4648–4657, 2007.

Wuchter, C., Abbas, B., Coolen, M. J. L., Herfort, L., van Bleijswijk, J., Timmers, P., Strous, M., Teira, E., Herndl, G. J., Middelburg, J. J., Schouten, S., and Sinninghe Damsté, J. S.: Archaeal nitrification in the ocean, Proc. Natl. Acad. Sci. USA, 103, 12317–12322, 2006.

Yan, J., Haaijer, S. C. M., Op den Camp, H. J. M., van Niftrik, L., Stahl, D. A., Könneke, M., Rush, D., Sinninghe Damsté, J. S., Hu, Y. Y., and Jetten, M. S. M.: Mimicking the oxygen minimum zones: stimulating interaction of aerobic archaeal and anaerobic bacterial ammonia oxidizers in a laboratory-scale model system, Environ. Microbiol., 14, 3146–3158, 2012.

LibreTexts

Excerpted from:

Sollai, M., Hopmans, E. C., Schouten, S., Keil, R. G., and Sinninghe Damsté, J. S.: Intact polar lipids of Thaumarchaeota and anammox bacteria as indicators of N cycling in the eastern tropical North Pacific oxygen-deficient zone, Biogeosciences, 12, 4725–4737, Accessed December 2023 at https://doi.org/10.5194/bg-12-4725-2015, 2015. CC-BY-4.0

12.4.2: Marine is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

12.4.3: Human Sources

Properties

- Lifetime in Atmosphere: 114 years
- **Global Warming Potential (100-year)**: 265[1]

In 2021, nitrous oxide (N_2O) accounted for 6% of all U.S. greenhouse gas emissions from human activities. Human activities such as agriculture, fuel combustion, wastewater management, and industrial processes are increasing the amount of N_2O in the atmosphere. Nitrous oxide is also naturally present in the atmosphere as part of the Earth's nitrogen cycle and has a variety of natural sources. Nitrous oxide molecules stay in the atmosphere for an average of 121 years before being removed by a sink or destroyed through chemical reactions. The impact of 1 pound of N_2O on warming the atmosphere is 265 times that of 1 pound of carbon dioxide.[1]

Globally, 40% of total N_2O emissions come from human activities.[2] Nitrous oxide is emitted from agriculture, land use, transportation, industry, and other activities, described below.

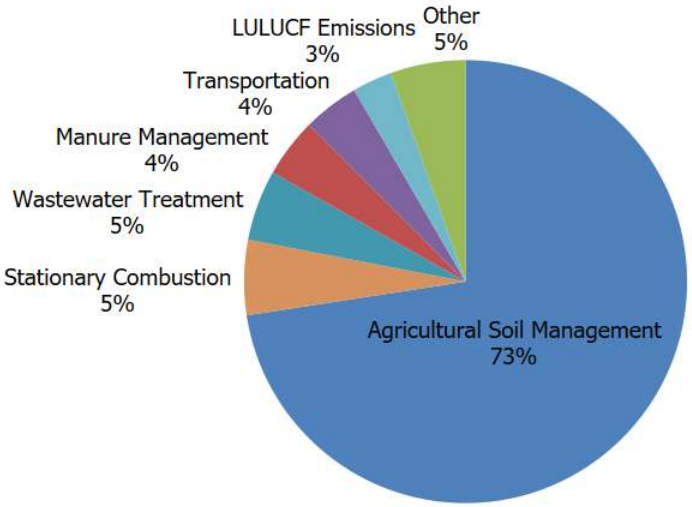


Figure 12.4.3.1: Note: All emission estimates from the *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2021*.

- **Agriculture.** Nitrous oxide can result from various agricultural soil management activities, such as application of synthetic and organic fertilizers and other cropping practices, the management of manure, or burning of agricultural residues. Agricultural soil management is the largest source of N_2O emissions in the United States, accounting for 75% of total U.S. N_2O emissions in 2021. While not shown in the figure and less significant, emissions of N_2O also occur as a result of land use and land management activities in the Land Use, Land-Use Change, and Forestry sector (e.g. forest and grassland fires, application of synthetic nitrogen fertilizers to urban soils (e.g., lawns, golf courses) and forest lands, etc.).
- **Fuel Combustion.** Nitrous oxide is emitted when fuels are burned. The amount of N_2O emitted from burning fuels depends on the type of fuel and combustion technology, maintenance, and operating practices.
- **Industry.** Nitrous oxide is generated as a byproduct during the production of chemicals such as nitric acid, which is used to make synthetic commercial fertilizer, and in the production of adipic acid, which is used to make fibers, like nylon, and other synthetic products. Nitrous oxide is also emitted from use in other applications such as anesthesia and semiconductor manufacturing.
- **Waste.** Nitrous oxide is also generated from treatment of domestic wastewater during nitrification and denitrification of the nitrogen present, usually in the form of urea, ammonia, and proteins.

Nitrous oxide emissions occur naturally through many sources associated with the nitrogen cycle, which is the natural circulation of nitrogen among the atmosphere, plants, animals, and microorganisms that live in soil and water. Nitrogen takes on a variety of chemical forms throughout the nitrogen cycle, including N_2O. Natural emissions of N_2O are mainly from bacteria breaking down nitrogen in soils and the oceans. Nitrous oxide is removed from the atmosphere when it is absorbed by certain types of bacteria or destroyed by ultraviolet radiation or chemical reactions.

Trends

Nitrous oxide emissions in the United States decreased by 3% between 1990 and 2021. During this time, nitrous oxide emissions from mobile combustion decreased by 56% as a result of criteria pollutant emission standards for on-road vehicles. Nitrous oxide emissions from agricultural soils have varied during this period and were about the same in 2021 as in 1990.

Reducing Nitrous Oxide Emissions

There are a number of ways to reduce emissions of N_2O, discussed below.

Examples of Reduction Opportunities for Nitrous Oxide Emissions

Emissions Source	Examples of How Emissions Can be Reduced
Agriculture	The application of nitrogen fertilizers accounts for the majority of N_2O emissions in the United States. Emissions can be reduced by reducing nitrogen-based fertilizer applications and applying these fertilizers more efficiently,[3] as well as modifying a farm's manure management practices.
Fuel Combustion	• Nitrous oxide is a byproduct of fuel combustion, so reducing fuel consumption in motor vehicles and secondary sources can reduce emissions. • Additionally, the introduction of pollution control technologies (e.g., catalytic converters to reduce exhaust pollutants from passenger cars) can also reduce emissions of N_2O.
Industry	• Nitrous oxide is generally emitted from industry through fossil fuel combustion, so technological upgrades and fuel switching are effective ways to reduce industry emissions of N_2O. • Production of nitric acid and adipic acid result in N_2O emissions that can be reduced through technological upgrades and use of abatement equipment.

References

1. IPCC (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. [Stocker, T.F., D. Qin, G.K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.
2. IPCC (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2391 pp.
3. EPA (2005). *Greenhouse Gas Mitigation Potential in U.S. Forestry and Agriculture*. U.S. Environmental Protection Agency, Washington, DC, USA.

Excerpted from:

Overview of Greenhouse Gases, United States Environmental Protection Agency. Accessed December 2023, https://www.epa.gov/ghgemissions/overview-greenhouse-gases, Public Domain

12.4.3: Human Sources is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

12.5: The global phosphorus cycle

Key Points

- Phosphorus, a major component of nucleic acid and phospholipids, also makes up the supportive components of our bones; it is often necessary for growth in aquatic ecosystems.
- Phosphates (PO_4^{3-}) are sent into rivers, lakes, and the ocean by leaching and natural surface runoff.
- Phosphate-containing ocean sediments slowly move to land by the uplifting of areas of the earth's surface.
- Excess phosphorus and nitrogen in the ecosystem leads to the death of many organisms, causing dead zones.
- Dead zones are caused by by eutrophication, oil spills, dumping of toxic chemicals, and other human activities.

The Phosphorus Cycle

Phosphorus is an essential nutrient for living processes; it is a major component of nucleic acid s, both DNA and RNA; of phospholipids, the major component of cell membranes; and, as calcium phosphate, makes up the supportive components of our bones. Phosphorus is often the limiting nutrient (necessary for growth) in aquatic, particularly freshwater, ecosystems. Several forms of nitrogen (nitrogen gas, ammnoium, nitrates, etc.) were involved in the nitrogen cycle, but phosphorus remains primarily in the form of the phosphate ion (PO_4^{3-}). Also in contrast to the nitrogen cycle, there is no form of phosphorus in the atmosphere. In addition to phosphate runoff as a result of human activity, natural surface runoff occurs when it is leached from phosphate-containing rock by weathering, thus sending phosphates into rivers, lakes, and the ocean.

Rocks are a reservoir for phosphorus, and these rocks have their origins in the ocean. Phosphate-containing ocean sediments form primarily from the bodies of ocean organisms and from their excretions. However, volcanic ash, aerosols, and mineral dust may also be significant phosphate sources. This sediment then is moved to land over geologic time by the uplifting of Earth's surface (Figure 12.5.1). Phosphorus is also reciprocally exchanged between phosphate dissolved in the ocean and marine ecosystems. The movement of phosphate from the ocean to the land and through the soil is extremely slow, with the average phosphate ion having an oceanic residence time between 20,000 and 100,000 years.

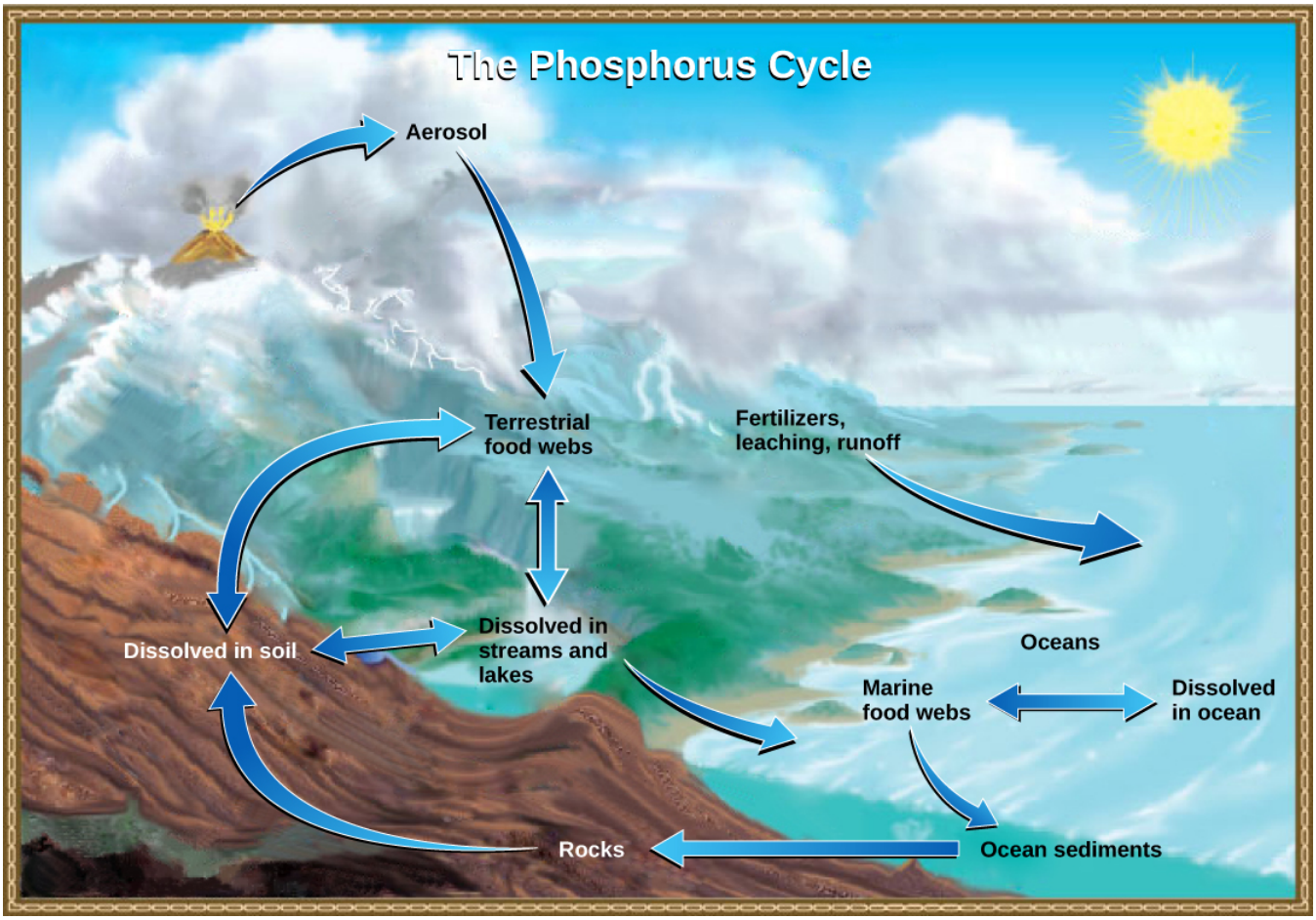


Figure 12.5.5: In nature, phosphorus exists as the phosphate ion (PO_4^{3-}). Weathering of rocks and volcanic activity releases phosphate into the soil, water, and air, where it becomes available to terrestrial food webs. Phosphate enters the oceans in surface runoff, groundwater flow, and river flow. Phosphate dissolved in ocean water cycles into marine food webs. Some phosphate from the marine food webs falls to the ocean floor, where it forms sediment. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Marine birds play a unique role in the phosphorous cycle. These birds take up phosphorous from ocean fish. Their droppings on land (**guano**) contain high levels of phosphorous and are sometimes mined for commercial use. A 2020 study estimated that the **ecosystem services** (natural processes and products that benefit humans) provided by guano are worth $470 million per year.

Weathering of rocks releases phosphates into the soil and bodies of water. Plants can assimilate phosphates in the soil and incorporate it into organic molecules, making phosphorus available to consumers in terrestrial food webs. Waste and dead organisms are decomposed by fungi and bacteria, releasing phosphates back into the soil. Some phosphate is leached from the soil, entering into rivers, lakes, and the ocean. Primary producers in aquatic food webs, such as algae and photosynthetic bacteria, assimilate phosphate, and organic phosphate is thus available to consumers in aquatic food webs. Similar to terrestrial food webs, phosphorus is reciprocally exchanged between phosphate dissolved in the ocean and organic phosphorus in marine organisms.

The movement of phosphorus from rock to living organisms is normally a very slow process, but some human activities speed up the process. Phosphate-bearing rock is often mined for use in the manufacture of fertilizers and detergents. This commercial production greatly accelerates the phosphorous cycle. In addition, runoff from agricultural land and the release of sewage into water systems can cause a local overload of phosphate. The increased availability of phosphate can cause overgrowth of algae. This reduces the oxygen level, causing eutrophication and the destruction of other aquatic species.

Contributors and Attributions

Modified by Kyle Whittinghill (University of Pittsburgh) and Melissa Ha from the following sources:

- Biogeochemical Cycles, Energy, and Energy Enters Ecosystems Through Photosynthesis from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Energy Flow through Ecosystems from *General Biology* by OpenStax (licensed under CC-BY)
- Soil and Sustainability and Biogeochemical Cycles and the Flow of Energy in the Earth System from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (licensed under CC-BY). Download for free at CNX.
- Cycling of Matter from *AP Environmental Science* by University of California College Prep, University of California (licensed under CC-BY). Download for free at CNX.
- Nutrient Cycles from *Life Sciences Grade 10* by Siyavula (CC-BY)
- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd).

  20.2: Biogeochemical Cycles by OpenStax, is licensed CC BY
- General Microbiology **Provided by**: Boundless.com. **License**: ***CC BY-SA: Attribution-ShareAlike***

12.5: The global phosphorus cycle is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **20.2: Biogeochemical Cycles** by OpenStax is licensed CC BY 4.0.

12.6: Potassium- An overlooked limiting element?

Nitrogen (N) is one of the most abundant elements on the earth and plays an important role in modern agriculture (Chen et al., 2018). N has an irreplaceable role in organ construction, material metabolism, fruit yield, and the quality formation of fruit trees (Warner et al., 2004; Liu et al., 2010; Bai et al., 2016). Sufficient N application not only improves the photosynthetic efficiency of the leaves but also promotes flower bud differentiation (Li et al., 2012; Wei et al., 2016; Liu et al., 2018), enhances fruit setting rate, and increases yield (Raese et al., 2007). Due to the one-sided pursuit of high yield and large fruits by fruit farmers, the excessive application of N fertilizer in apple orchards has become common in China. The applied N fertilizer dose is are more above the demand of the tree (Ge et al., 2018). Furthermore, the over application of N increases the input cost of farmers, decreases fruit yield and quality (Chen et al., 2017), and directly leads to a decreased N utilization rate. Excessive N also indirectly results in adverse ecological effects, including the increased N deposition, the intensification of the greenhouse effect, soil acidification, and the eutrophication of water bodies (Liu et al., 2013; Zhu and Zhang, 2017). Therefore, it is a main issue for Chinese apple production to promote the efficient utilization of N to improve the quality and yield of apple.

Potassium (K) is the most abundant inorganic cation, and it is important for ensuring optimal plant growth (White and Karley, 2010). K is an activator of dozens of important enzymes, such as protein synthesis, sugar transport, N and C metabolism, and photosynthesis. It plays an important role in the formation of yield and quality improvement (Marschner, 2012; Oosterhuis et al., 2014). K is also very important for cell growth, which is an important process for the function and development of plants (Hepler et al., 2001). In terms of the growth-promoting mechanism of K, it is generally agreed that K stimulates and controls ATPase in the plasma membrane to generate acid stimulation, which then triggers cell wall loosening and hydrolase activation (Oosterhuis et al., 2014), thus promoting cell growth. K has strong mobility in plants and plays an important role in regulating cell osmotic pressure and balancing the cations and anions in the cytoplasm (Kaiser, 1982; Hu et al., 2016a). Through these processes, K is involved in the regulation of stomatal opening and closing, cell elongation, and other important physiological processes. There are many studies on the effect of K level on plant growth. Jin et al. (2007) found that the highest yield and fruit quality were obtained in Red Fuji apple under treatment with $600\ \mathrm{kg/ha}$ K; Wang et al. (2017) determined that $6\mathrm{mMK}$ treatment promoted pear growth and improved photosynthetic efficiency; and Lu et al. (2001) also reported increased production with better fruit quality parameters in navel orange supplied under $500\ \mathrm{kg/haK}$. There is an interaction between K and other nutrient ions. High K concentrations in the soil solution inhibit Mg uptake and may induce Mg deficiency in plants (Tränkner et al., 2018). However, K deficiency could promote the absorption of $\mathrm{Na}^{+}$and Ca^{2+} in maize (Du et al., 2017), and could inhibit N absorption in cotton and significantly reduce the content of NO_3^-in the leaves (Hu et al., 2017). It is evident that K affects significantly the absorption and utilization of other nutrients by plants, and the appropriate K level differs in different crops.

Among the interactions between K and other nutrients, the interaction with N is the most important. Some studies evaluated the relationship between K and N metabolism. In contrast to the antagonistic relationship between $\mathrm{K}^{+}$and NH_4^{+}nutrition, the acquisition rates of $\mathrm{K}^{+}$and $\mathrm{NO}_3^{-}$are often found to be positively correlated (Rufty et al., 1981; Coskun et al., 2016), and sufficient K supply can promote N metabolism and enhance the synthesis of amino acids and proteins (Ruan et al., 1998; Ruiz and Romero, 2002). Hu et al. (2016b) found that K deficiency could reduce Nitrate reductase (NR), Glutamine synthetase (GS), and Glutamate synthase (GOGAT) activities and inhibit nitrate absorption in cotton, whereas Armengaud et al. (2009) found that K deficiency could up-regulate the activities of GS and Glu dehydrogenase (GDH) in Arabidopsis. Metabolism of N affected by K appears to vary in different types of plants. Meanwhile, the level of K has a significant impact on C metabolism, and also a strong interaction exists between C metabolism and N metabolism in the metabolic process and energy level (Hu et al., 2017).

References

Armengaud, P., Sulpice, R., Miller, A. J., Stitt, M., Amtmann, A., and Gibon, Y. (2009). Multilevel analysis of primary metabolism provides new insights into the role of potassium nutrition for glycolysis and nitrogen assimilation in *Arabidopsis* roots. *Plant Physiol.* 150, 772–785. doi: 10.1104/pp.108.133629.

Bai, L. Q., Deng, H. H., Zhang, X. C., Yu, X. C., and Li, Y. S. (2016). Gibberellin is involved in inhibition of Cucumber growth and nitrogen uptake at Suboptimal root-Zone temperatures. *PLoS One* 11:e0156188. doi: 10.1371/journal.pone.0156188.

Chen, Q., Ding, N., Peng, L., Ge, S. F., and Jiang, Y. M. (2017). Effects of different nitrogen application rates on ^{15} N-urea absorption, utilization, loss and fruit yield and quality of dwarf apple. *Chin. J. Appl. Ecol.* 28, 2247–2253..

Chen, G. D., Wang, L., Fabrice, M. R., Tian, Y. N., Qi, K. J., Tao, S. T., et al. (2018). Physiological and nutritional responses of pear seedlings to nitrate concentrations. *Front. Plant Sci.* 9:1679. doi: 10.3389/fpls.2018.01679.

Coskun, D., Britto, D. T., and Kronzucker, H. J. (2016). The nitrogen–potassium intersection: membranes, metabolism, and mechanism. *Plant Cell Environ.* 10, 2029–2041. doi: 10.1111/pce.12671.

Du, Q., Zhao, X. H., Jiang, C. J., Wang, X. G., Han, Y., Wang, J., et al. (2017). Effect of potassium deficiency on root growth and nutrient uptake in maize (*Zea mays L.*). *Agric. Sci.* 8, 1263–1277. doi: 10.4236/as.2017.811091.

Ge, S. F., Zhu, Z. L., Peng, L., Chen, Q., and Jiang, Y. M. (2018). Soil nutrient status and leaf nutrient diagnosis in the main apple producing regions in China. *Hortic. Plant J.* 4, 89–93. doi: 10.1016/j.hpj.2018.03.009.

Hepler, P. K., Vidali, L., and Cheung, A. Y. (2001). Polarized cell growth in higher plants. *Annu.Rev. Cell. Dev. Biol.* 17, 159–187. doi: 10.1146/annurev.cellbio.17.1.159.

Hu, W., Jiang, N., Yang, J., Meng, Y., Wang, Y., Chen, B., et al. (2016a). Potassium (K) supply affects K accumulation and photosynthetic physiology in two cotton (*Gossypium hirsutum* L.) cultivars with different K sensitivities. *Field Crop. Res.* 196, 51–63. doi: 10.1016/j.fcr.2016.06.005.

Hu, W., Zhao, W., Yang, J., Oosterhuis, D. M., Loka, D. A., and Zhou, Z. (2016b). Relationship between potassium fertilization and nitrogen metabolism in the leaf subtending the cotton (*Gossypium hirsutum* L.) boll during the boll development stage. *Plant Physiol. Biochem.* 101, 113–123. doi: 10.1016/j.plaphy.2016.01.019.

Hu, W., Coomer, T. D., Loka, D. A., Oosterhuis, D. M., and Zhou, Z. (2017). Potassium deficiency affects the carbon-nitrogen balance in cotton leaves. *Plant Physiol. Biochem.* 115, 408–417. doi: 10.1016/j.plaphy.2017.04.005.

Jin, H. C., Zhang, L. S., Li, B. Z., Han, M. Y., and Liu, X. G. (2007). Effect of potassium on the leaf nutrition and quality of Red Fuji apple. *Acta Agric. Bor Occid. Sin.* 16, 100–104. doi: 10.3969/j.issn.1004-1389.2007.03.026.

Kaiser, W. M. (1982). Correlation between changes in photosynthetic activity and changes in total protoplast volume in leaf tissue from hygro-, meso-and xerophytes under osmotic stress. *Planta* 154, 538–545. doi: 10.1007/BF00402997.

Li, G., Zhang, Z. S., Gao, H. Y., Liu, P., Dong, S. T., Zhang, J. W., et al. (2012). Effects of nitrogen on photosynthetic characteristics of leaves from two different stay-green corn (*Zea mays* L.) varieties at the grain-filling stage. *Can. J. Plant Sci.* 92, 671–680. doi: 10.4141/cjps2012-039.

Liu, L., Peng, F., and Wang, X. (2010). Effects of bag-controlled release fertilizer on nitrogen utilization rate, growth and fruiting of the ‘Fuji’ apple. *J. Plant Nutr.* 33, 1904–1913. doi: 10.1080/01904167.2010.512050.

Liu, X. J., Zhang, Y., Han, W. X., Tang, A. H., Shen, J. L., Zhang, F. S., et al. (2013). Enhanced nitrogen deposition over China. *Nature* 494, 459–462. doi: 10.1038/nature11917

Liu, Z., Gao, J., Gao, F., Liu, P., Zhao, B., and Zhang, J. W. (2018). Photosynthetic characteristics and chloroplast ultrastructure of summer Maize response to different nitrogen supplies. *Front. Plant Sci.* 9:576. doi: 10.3389/fpls.2018.00576.

Lu, J. W., Chen, F., Wan, Y. F., Liu, D. B., Yu, C. B., Wang, Y. Q., et al. (2001). Effect of application of potassium on the yield and quality of Navel Orang. *J. Fruit Sci.* 18, 272–275..

Marschner, H. (2012). *Marschner’s Mineral Nutrition of Higher Plants.* Cambridge, MA: Academic press..

Oosterhuis, D., Loka, D., Kawakami, E., and Pettigrew, W. (2014). The physiology of potassium in crop production. *Adv. Agron.* 126, 203–234. doi: 10.1016/B978-0-12-800132-5.00003-1.

Raese, J. T., Drake, R. S., and Curry, A. E. (2007). Nitrogen fertilizer influences fruit quality, soil nutrients and cover crops, leaf color and nitrogen content, biennial bearing and cold hardiness of ‘Golden Delicious’. *J. Plant Nutr.* 30, 1585–1604. doi: 10.1080/01904160701615483.

Ruan, J., Wu, X., Ye, Y., and Hardter, R. (1998). Effect of potassium, magnesium and sulphur applied in different forms of fertilisers on free amino acid content in leaves of tea (*Camellia sinensis* L). *J. Sci. Food Agric.* 76, 389–396. doi: 10.1002/(SICI)1097-0010(199803)76:3<389::AID-JSFA963<3.0.CO;2-X.

Rufty, T. W., Jackson, W. A., and Raper, C. D. (1981). Nitrate reduction in roots as affected by the presence of potassium and by flux of nitrate through the roots. *Plant Physiol.* 68, 605–609. doi: 10.2307/4266953

Ruiz, J., and Romero, L. (2002). Relationship between potassium fertilisation and nitrate assimilation in leaves and fruits of cucumber (*Cucumis sativus*) plants. *Ann. Appl. Biol.* 140, 241–245. doi: 10.1111/j.1744-7348.2002.tb00177.x.

Tränkner, M., Tavakol, E., and Jákli, B. (2018). Functioning of potassium and magnesium in photosynthesis, photosynthate translocation and photoprotection. *Physiol. Plant.* 163, 414–431. doi: 10.1111/ppl.12747.

Wang, Y. Z., Zhang, H. P., Huang, X. S., Wang, J. Z., Cheng, R., Chen, G. D., et al. (2017). Effect of potassium supply on plant potassium distribution and growth and leaf photosynthetic capacity of *Pyrus pyrifolia*. *J. Nanjing Agric. Univ.* 40, 60–67. doi: 10.7685/jnau.201603054.

Wei, S. S., Wang, X. Y., Shi, D. Y., Li, Y. H., Zhang, J. W., Dong, S. T., et al. (2016). The mechanisms of low nitrogen induced weakened photosynthesis in summer maize (*Zea mays* L.) under field conditions. *Plant Physiol. Biochem.* 105, 118–128. doi: 10.1016/j.plaphy.2016.04.007.

White, P. J., and Karley, A. J. (2010). *Potassium Cell Biology of Metals and Nutrients.* Berlin: Springer, 199–224..

Zhu, L. X., and Zhang, W. (2017). Effects of controlled-release urea combined with conventional urea on nitrogen uptake, root yield, and quality of *Platycodon grandiflorum*. *J. Plant Nutr.* 40, 662–672. doi: 10.1080/01904167.2016.1249799

Excerpted from:

Xu, X., Du, X., Wang, F., Sha, J., Chen, Q., Tian, G., ... & Jiang, Y. (2020). Effects of potassium levels on plant growth, accumulation and distribution of carbon, and nitrate metabolism in apple dwarf rootstock seedlings. *Frontiers in Plant Science*, *11*, 904. Accessed December 2023 at https://www.frontiersin.org/articles/10.3389/fpls.2020.00904/full. CC-BY-4.0

12.6: Potassium- An overlooked limiting element? is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by X. Xu, et al..

12.7: Linking global biogeochemical cycles

Eutrophication and Dead Zones

Nutrients in soil and water are generally beneficial when they exist at naturally occurring levels. Nitrogen fertilizers have been applied to farm fields for decades in order to maximize production of agricultural lands. However, an unintended consequence is that the same nutrients can be detrimental to aquatic ecosystems when introduced excessively for agricultural or other purposes. Nitrogen (N) and Phosphorus (P) are introduced by fertilizers that are used intensively in agriculture, as well as golf courses and some lawns and gardens. Farm animal waste and sewage also provide large amounts of reactive N and P. Phosphorus was formerly used heavily as an additive in laundry and dishwater detergents, but since the 1970's it has been phased out in both through a combination of state and federal regulations. Overall, our modern society has altered the global N and P cycles such that there is an overabundance in many settings.

Excessive nutrients (not utilized) are often washed into drainage ways, streams, and rivers during rainfall and storm events. **Eutrophication** occurs when excess phosphorus and nitrogen from fertilizer runoff or sewage causes excessive growth of algae. Algal blooms that block light and therefore kill aquatic plants in rivers, lakes, and seas. The subsequent death and decay of these organisms depletes dissolved oxygen, which leads to the death of aquatic organisms such as shellfish and fish. This process is responsible for **dead zones**, large areas in lakes and oceans near the mouths of rivers that are periodically depleted of their normal flora and fauna, and for massive fish kills, which often occur during the summer months (figure 12.7.1). There are more than 500 dead zones worldwide. Phosphate and nitrate runoff from fertilizers also negatively affect several lake and bay ecosystems including the Chesapeake Bay in the eastern United States.

One of the worst dead zones is off the coast of the United States in the Gulf of Mexico (figure 12.7.1). Fertilizer runoff from the Mississippi River basin created a dead zone, which reached its peak size of 8,776 square miles in 2017. The Mississippi River dumps high-nutrient runoff from its drainage basin that includes vast agricultural lands in the American Midwest. Increased algal growth produced by these nutrients has affected important shrimp fishing grounds in the Gulf. The primary source of the nutrients is the heavily tile-drained areas of farmland in the Midwest corn and soybean belt (SW Minnesota, N Iowa, NE Illinois, N Indiana and NW Ohio). Improved soil drainage systems over the past century or more have allowed for effective transport of nitrate compounds as stormwater runoff into drainage basins (Ohio River, Wabash River, Illinois River, Missouri River, etc.) that feed into the Mississippi River. In other words, the same drainage tiles that allow for the agricultural benefit of having rich bottomland/wetland soils in production, have the disadvantage of increased and more rapid movements of nitrate solutes to the Gulf of Mexico. Such large-scale problems, across state governmental boundaries, can only be fully addressed in the future with a national system of incentives, regulations, or laws.

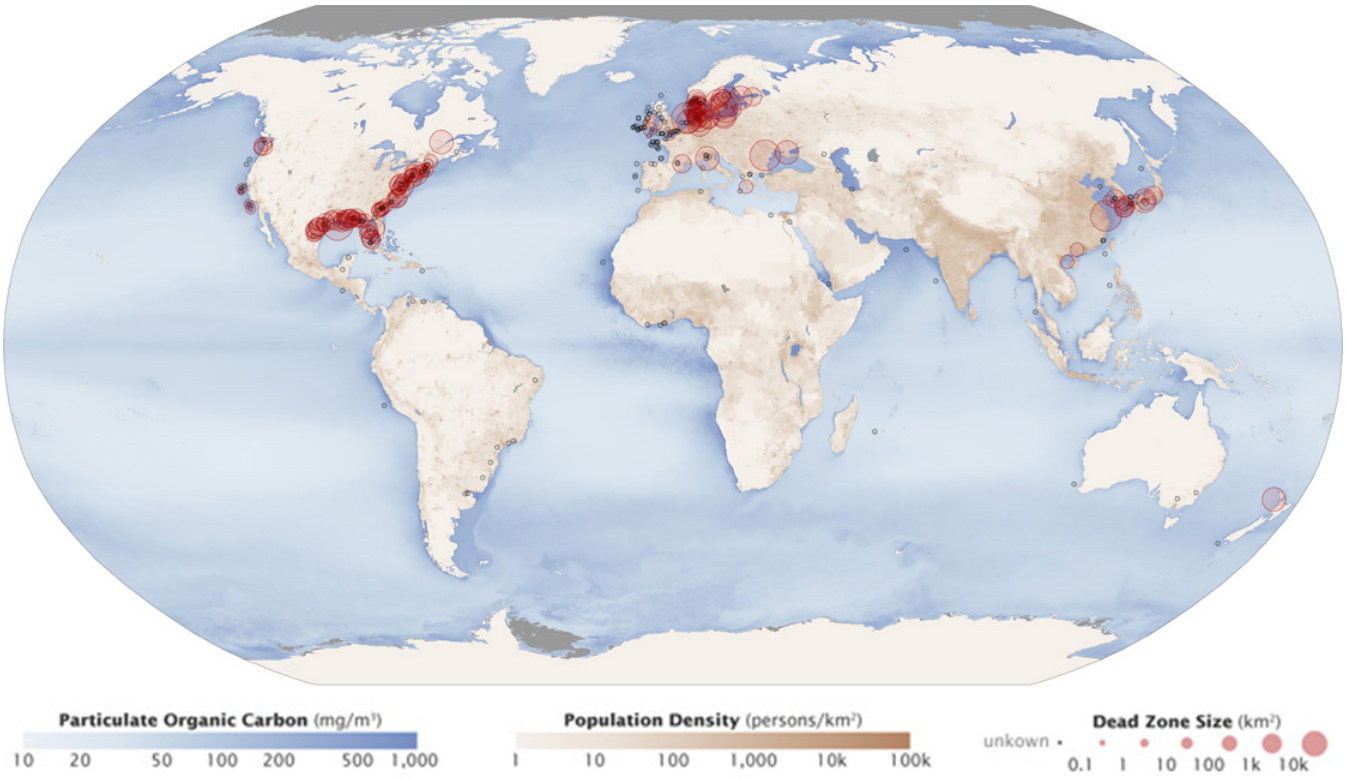


Figure 12.7.1: Dead zones occur when phosphorus and nitrogen from fertilizers cause excessive growth of microorganisms, which depletes oxygen and kills fauna. This map shows dead zones around the world in 2008. Worldwide, large dead zones are found in coastal areas of high population density. (credit: NASA Earth Observatory)

In addition to fertilizers, Nitrogen inputs to watersheds can also include atmospheric deposition, livestock waste, and sewage, but nitrogen fertilizers comprise a significant majority of the input to monitored streams, particularly in springtime when much fertilizer is applied. Possible solutions to this problem include encouraging farmers to apply a more limited quantity of fertilizer in the spring (only as much as necessary), rather than in the fall, to allow for considerably less time for stormwater or meltwater runoff. Other solutions include maintaining cover crops, or restoring wetlands in key locations to contain nitrate losses. An overall strategy that limits the excess capacity of nutrients can simultaneously benefit farmers (by limiting cost), the ecology of stream watersheds and coastal ecosystems (also locally stressed by oil spills and other pollution). Over the long term, more efforts will need to be made in the Mississippi River Basin, and globally in similarly stressed agricultural or urban watersheds (figure 12.7.1), to improve the health and sustainability of our soil, land, and aquatic ecosystems.

Everyday Connection: Chesapeake Bay

The Chesapeake Bay has long been valued as one of the most scenic areas on Earth; it is now in distress and is recognized as a declining ecosystem. In the 1970s, the Chesapeake Bay was one of the first ecosystems to have identified dead zones, which continue to kill many fish and bottom-dwelling species, such as clams, oysters, and worms (figure 12.7.2). Several species have declined in the Chesapeake Bay due to surface water runoff containing excess nutrients from artificial fertilizer used on land. The source of the fertilizers (with high nitrogen and phosphate content) is not limited to agricultural practices. There are many nearby urban areas and more than 150 rivers and streams empty into the bay that are carrying fertilizer runoff from lawns and gardens. Thus, the decline of the Chesapeake Bay is a complex issue and requires the cooperation of industry, agriculture, and everyday homeowners.



(a) (b)

Figure 12.7.2: This (a) satellite image shows the Chesapeake Bay, an ecosystem affected by phosphate and nitrate runoff. A (b) member of the Army Corps of Engineers holds a clump of oysters being used as a part of the oyster restoration effort in the bay. (credit a: modification of work by NASA/MODIS; credit b: modification of work by U.S. Army)

Of particular interest to conservationists is the oyster population; it is estimated that more than 200,000 acres of oyster reefs existed in the bay in the 1700s, but that number has now declined to only 36,000 acres. Oyster harvesting was once a major industry for Chesapeake Bay, but it declined 88 percent between 1982 and 2007. This decline was due not only to fertilizer runoff and dead zones but also to overexploitation. Oysters require a certain minimum population density because they must be in close proximity to reproduce. Human activity has altered the oyster population and locations, greatly disrupting the ecosystem.

The restoration of the oyster population in the Chesapeake Bay has been ongoing for several years with mixed success. Not only do many people find oysters good to eat, but they also clean up the bay. Oysters are filter feeders, and as they eat, they clean the water around them. In the 1700s, it was estimated that it took only a few days for the oyster population to filter the entire volume of the bay. Today, with changed water conditions, it is estimated that the present population would take nearly a year to do the same job.

Restoration efforts have been ongoing for several years by non-profit organizations, such as the Chesapeake Bay Foundation. The restoration goal is to find a way to increase population density so the oysters can reproduce more efficiently. Many disease-resistant varieties (developed at the Virginia Institute of Marine Science for the College of William and Mary) are now available and have been used in the construction of experimental oyster reefs. Efforts to clean and restore the bay by Virginia and Delaware have been hampered because much of the pollution entering the bay comes from other states, which stresses the need for inter-state cooperation to gain successful restoration.

The new, hearty oyster strains have also spawned a new and economically viable industry—oyster aquaculture—which not only supplies oysters for food and profit, but also has the added benefit of cleaning the bay.

Suggested Supplementary Reading

Bruckner, M. 2018. The Gulf of Mexico Dead Zone. [Website]

References

Cell Press. (2020, August 6). Researchers hope to save seabirds by calculating the value of their excrement. Retrieved August 7, 2020 from *ScienceDaily*.

Attributions

Modified by Kyle Whittinghill (University of Pittsburgh) and Melissa Ha from the following sources:

- Biogeochemical Cycles, Energy, and Energy Enters Ecosystems Through Photosynthesis from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Carbon Cycle and Nitrogen Cycle from *Biology* by John W. Kimball (licensed under CC-BY)
- Energy Flow through Ecosystems from *General Biology* by OpenStax (licensed under CC-BY)
- Soil and Sustainability and Biogeochemical Cycles and the Flow of Energy in the Earth System from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (licensed under CC-BY). Download for free at CNX.
- Cycling of Matter from *AP Environmental Science* by University of California College Prep, University of California (licensed under CC-BY). Download for free at CNX.
- Nutrient Cycles from *Life Sciences Grade 10* by Siyavula (CC-BY)
- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd).

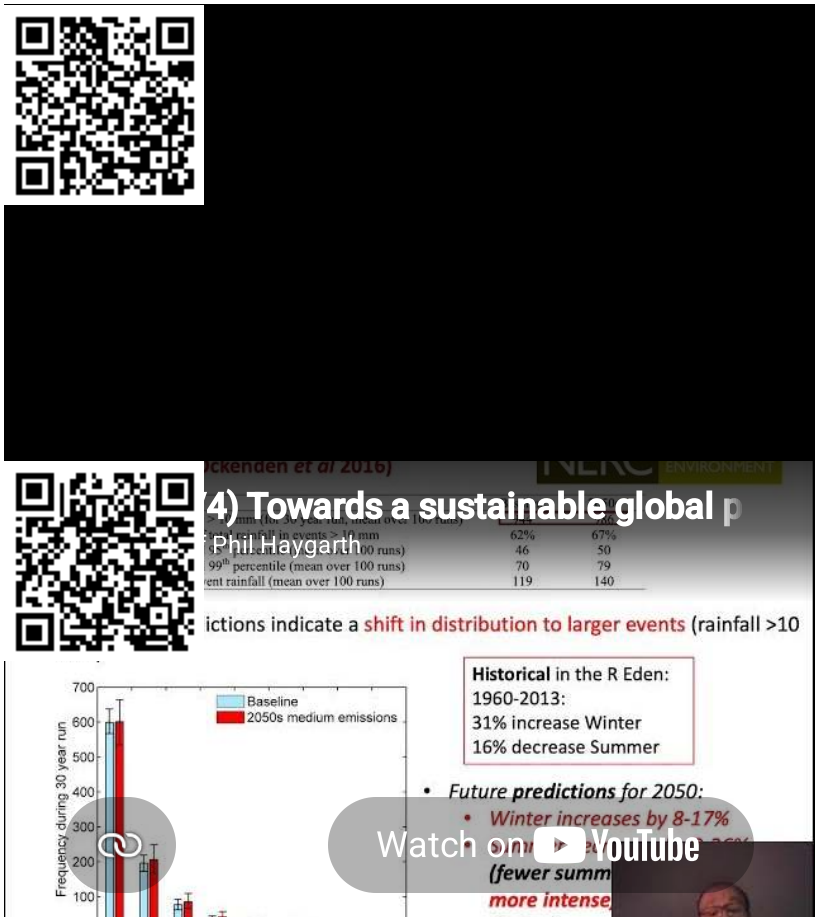
  20.2: Biogeochemical Cycles by OpenStax, is licensed CC BY

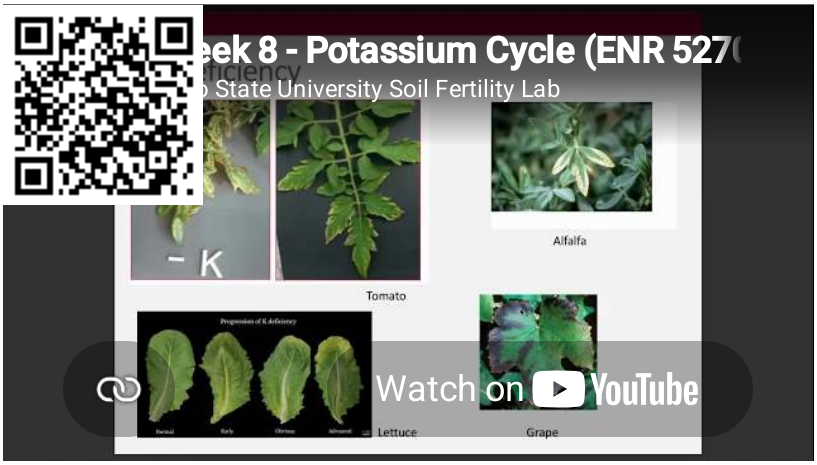
12.7: Linking global biogeochemical cycles is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

12.8: Summary



Under Development





12.8: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

CHAPTER OVERVIEW

13: The Global Cycles of Sulfur and Mercury

13.1: Introduction
13.2: The global sulfur cycle
13.3: Interactions of the sulfur and oxygen cycles
13.4: Temporal perspectives on the global sulfur cycle
13.5: The global mercury cycle
13.6: Summary

13: The Global Cycles of Sulfur and Mercury is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

13.1: Introduction

13.1: Introduction is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

13.2: The global sulfur cycle

Key Points

- Sulfur is an essential element for the macromolecules of living things since it determines the 3-D folding patterns of proteins.
- On land, sulfur enters the atmosphere via acid rain, fallout, the weathering of rocks, decompostion of organic materials, and geothermal vents.
- Sulfur enters the ocean via runoff, fallout, and underwater geothermal vents; some marine ecosystems also rely on chemoautotrophs as a sulfur source.
- The burning of fossil fuels increases the amount of sulfide in the atmosphere and causes acid rain.
- Acid rain is corrosive rain that causes damage to aquatic ecosystems by lowering the pH of lakes, killing many of the resident fauna; it also degrades buildings and human-made structures.

The Sulfur Cycle

Sulfur is an essential element for the macromolecules of living things. As a part of the amino acid cysteine, it is involved in the formation of disulfide bonds within proteins, which help to determine their 3-D folding patterns and, hence, their functions. As shown in Figure 13.2.1, sulfur cycles between the oceans, land, and atmosphere. Atmospheric sulfur is found in the form of sulfur dioxide (SO_2), which enters the atmosphere in three ways: first, from the decomposition of organic molecules; second, from volcanic activity and geothermal vents; and, third, from the burning of fossil fuels by humans.

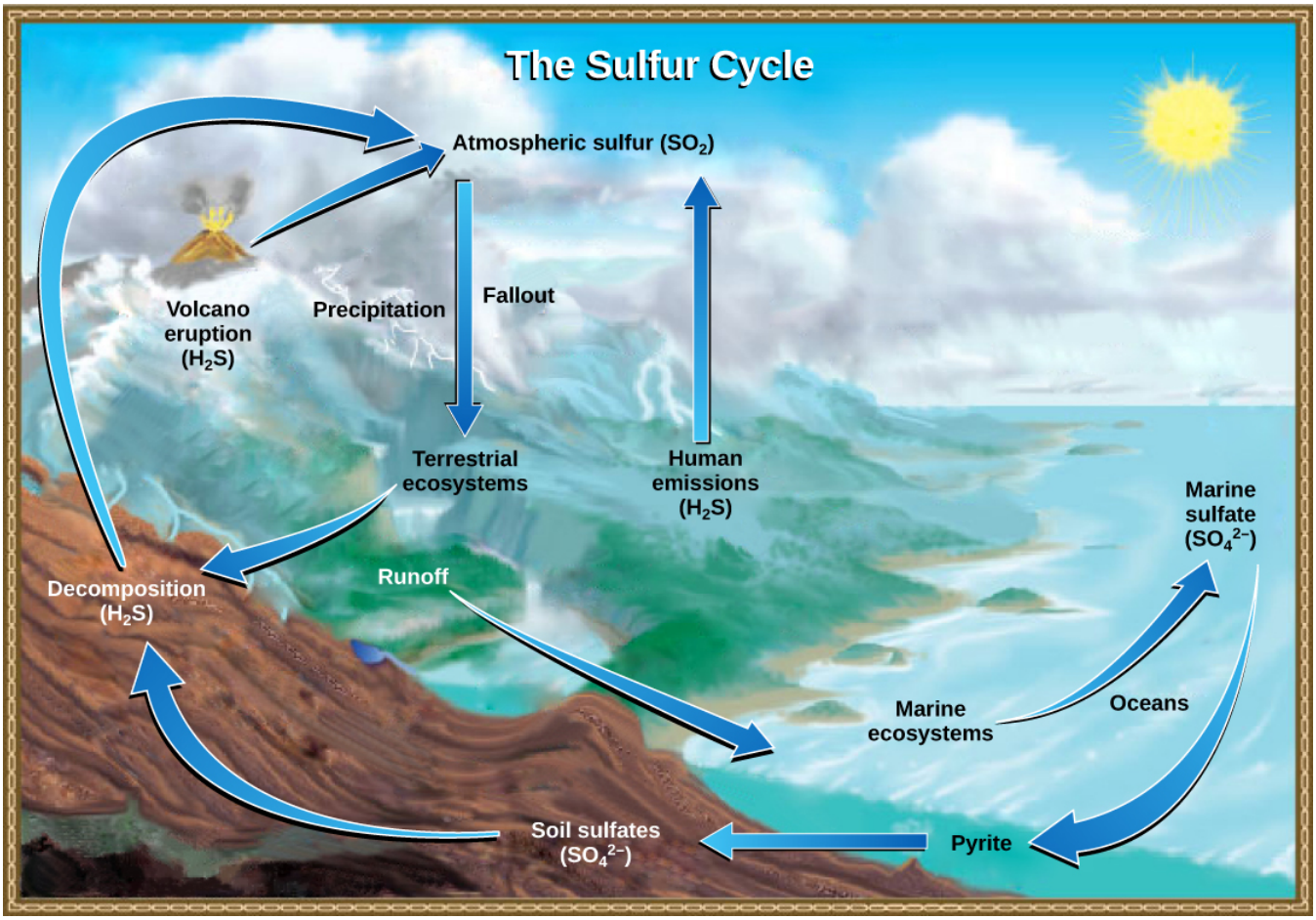


Figure 13.2.1: The sulfur cycle. Sulfur dioxide (SO_2) from the atmosphere is dissolved in precipitation as weak sulfuric acid or falls directly to Earth as fallout. This releases sulfates (SO_4^{2-}) into the soil and water. Soil sulfates can be carried as runoff into the water. Marine sulfate can form pyrite, and this can break down to release soil sulfates. Organisms in terrestrial and marine ecosystems assimilate sulfate, adding sulfur to organic molecules, such as proteins (not shown). Decomposition of these organisms returns sulfates to the soil. Microorganisms can convert sulfates to hydrogen sulfide (H_2S) and vice versa. Decomposition, volcanic

eruptions, and human activities (including burning fossil fuels) can release hydrogen sulfide (H_2S) or sulfur dioxide into the atmosphere. (credit: Melissa Ha modification of work by John M. Evans and Howard Perlman, USGS)

On land, sulfur is deposited in five major ways: precipitation, direct fallout from the atmosphere, rock weathering, decomposition of organic material and geothermal vents (Figure 13.2.2). As rain falls through the atmosphere, sulfur dioxide (SO_2) in the atmosphere is dissolved in the form of weak sulfuric acid (H_2SO_4), creating acid rain. Sulfur can also fall directly from the atmosphere in a process called fallout. The weathering of sulfur-containing rocks also releases sulfur into the soil. These rocks originate from ocean sediments that are moved to land by the geologic uplift. Terrestrial ecosystems can then make use of these soil sulfates (SO_4^{2-}), which enter the food web by being taken up by plant roots. When these plants decompose and die, sulfur is released back into the atmosphere as hydrogen sulfide (H_2S) gas.



Figure 13.2.2: At this sulfur vent in Lassen Volcanic National Park in northeastern California, the yellowish sulfur deposits are visible near the mouth of the vent. (credit: “Calbear22”/Wikimedia Commons)

Sulfur enters the ocean in runoff from land, from atmospheric fallout, and from hydrothermal vents. Some ecosystems rely on microorganisms using sulfur as a biological energy source (chemoautotrophs) in contrast to ecosystems with photosynthetic producers. This sulfur then supports marine ecosystems in the form of sulfates.

Human activities have played a major role in altering the balance of the global sulfur cycle. The burning of large quantities of fossil fuels, especially from coal, releases sulfur dioxide, which reacts with the atmosphere to form sulfuric acid. Like nitric acid, sulfuric acid contributes to acid deposition. Acid deposition (sometimes referred to simply as acid rain) damages the natural environment by lowering the pH of lakes, thus killing many of the resident plants and animals. Acid deposition also affects the man-made environment through the chemical degradation of buildings. For example, many marble monuments, such as the Lincoln Memorial in Washington, DC, have suffered significant damage from acid rain over the years. These examples show the wide-ranging effects of human activities on our environment and the challenges that remain for our future.

Chemosynthesis

Why do bacteria that live deep below the ocean’s surface rely on chemical compounds instead of sunlight for energy to make food?

Most autotrophs make food by photosynthesis, but this isn’t the only way that autotrophs produce food. Some bacteria make food by another process, which uses chemical energy instead of light energy. This process is called **chemosynthesis**. In chemosynthesis, one or more carbon molecules (usually carbon dioxide or methane, CH_4) and nutrients is converted into organic matter, using the oxidation of inorganic molecules (such as hydrogen gas, hydrogen sulfide (H_2S) or ammonia (NH_3)) or methane as a source of energy, rather than sunlight. In hydrogen sulfide chemosynthesis, in the presence of carbon dioxide and oxygen,carbohydrates (CH_2O) can be produced:

$$CO_2 + O_2 + 4H_2S \rightarrow CH_2O + 4S + 3H_2O$$

Many organisms that use chemosynthesis are **extremophiles,** living in harsh conditions, such as in the absence of sunlight and a wide range of water temperatures, some approaching the boiling point. Some chemosynthetic bacteria live around deep-ocean vents known as “black smokers.” Compounds such as hydrogen sulfide, which flow out of the vents from Earth’s interior, are used by the

bacteria for energy to make food. These organisms are known as **chemoautotrophs.** Many chemosynthetic microorganisms are consumed by other organisms in the ocean, and symbiotic associations between these organisms and respiring heterotrophs are quite common. Consumers that depend on these bacteria to produce food for them include giant tubeworms (Figure 13.2.3), and Pompeii worms ((Figure 13.2.4).

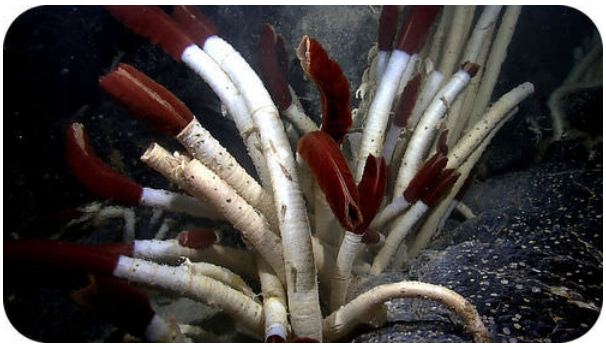


Figure 13.2.3**:** Tubeworms deep in the Galapagos Rift get their energy from chemosynthetic bacteria. Tubeworms have no mouth, eyes or stomach. Their survival depends on a symbiotic relationship with the billions of bacteria that live inside them. These bacteria convert the chemicals that shoot out of the hydrothermal vents into food for the worm.

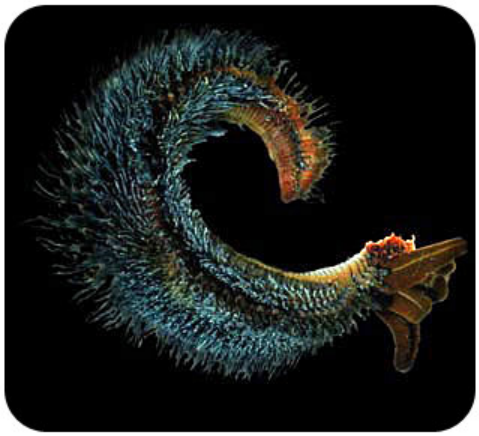


Figure 13.2.3**: Is it possible to live in temperatures over 175°F?** It is if you're a Pompeii worm. The Pompeii worm, the most heat-tolerant animal on Earth, lives in the deep ocean at super-heated hydrothermal vents. Covering this deep-sea worm's back is a fleece of bacteria. These microbes contain all the genes necessary for life in extreme environments.

Contributors and Attributions

This section was modified from the original by Kyle Whittinghill

- Samantha Fowler (Clayton State University), Rebecca Roush (Sandhills Community College), James Wise (Hampton University). Original content by OpenStax (CC BY 4.0; Access for free at https://cnx.org/contents/b3c1e1d2-83...4-e119a8aafbdd).
- **20.2: Biogeochemical Cycles** by OpenStax, is licensed CC BY
- **2.24: Chemosynthesis** by CK-12: Biology Concepts, is licensed CC BY-NC
- General Microbiology **Provided by**: Boundless.com. **License**: ***CC BY-SA: Attribution-ShareAlike***
- Atmosphere-Biosphere-Hydrosphere-Lithosphere. **Provided by**: Wikimedia. **Located at**: **commons.wikimedia.org/wiki/Fi...ithosphere.png**. **License**: ***CC BY: Attribution***

13.2: The global sulfur cycle is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

- **20.2: Biogeochemical Cycles** by OpenStax is licensed CC BY 4.0.
- **2.24: Chemosynthesis** by CK12 is licensed CK-12. Original source: http://www.ck12.org/book/CK-12-Biology-Concepts.

13.3: Interactions of the sulfur and oxygen cycles

The sulfur cycle of marine sediments is primarily driven by the dissimilatory sulfate reduction (DSR) to sulfide by anaerobic microorganisms (e.g., Jørgensen and Kasten, 2006). This process links the complex food web of organic matter degradation to the terminal organic carbon oxidation to CO_2. Most of the sulfide is ultimately reoxidized back to sulfate, via diverse sulfur intermediates, by geochemical or microbial reactions that involve oxygen, nitrate, manganese [Mn(IV)], iron [Fe(III)], and other potential oxidants (e.g., Rickard, 2012). A fraction of the sulfide precipitates with iron and other metals or reacts with organic matter and is buried deeply into the seabed. The microbial sulfur transformations affect the isotopic composition of sulfate and sulfides and the resulting isotope fractionation is thereby diagnostic for both process rates and pathways of the sulfur cycle (e.g., Canfield, 2001).

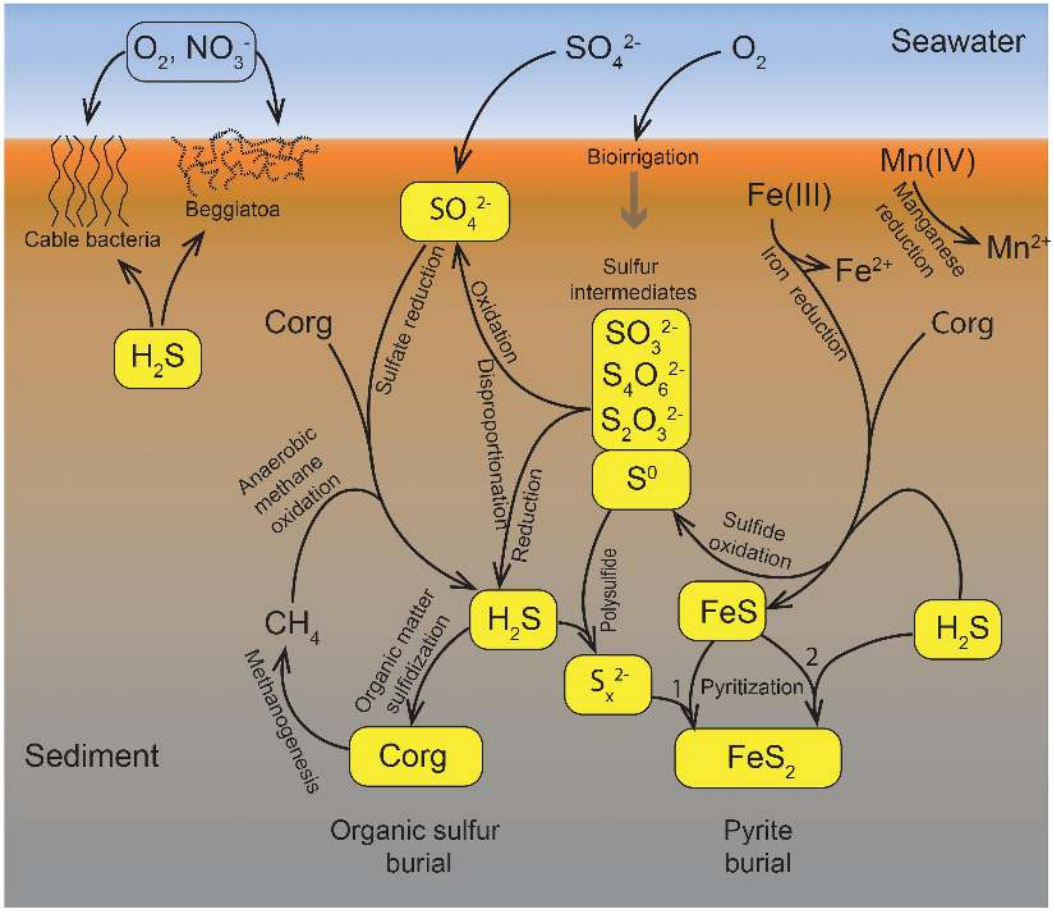


Figure 13.3.1: The biogeochemical sulfur cycle of marine sediments. Arrows indicate fluxes and pathways of biological or chemical processes. For further explanation, see text. (Copyright; Barker Jørgensen, Alyssa J. Findlay and André Pellerin via Frontiers in Microbiology)

Figure 13.3.1 presents the sulfur cycle of marine sediments. The processes include chemical reactions, microbially catalyzed pathways, and a combination of both. Sulfate (SO_4^{2-}) reduction to sulfide ($H_2S + HS^- + S^{2-}$) is driven by the oxidation of buried organic carbon (C_{org}), supplemented by the anaerobic oxidation of methane (CH_4) at the subsurface sulfate-methane transition (SMT). Manganese and iron reduction are focused toward the surface sediment, but Fe(III) is also buried and acts as an oxidant for sulfide in the deeper sediment layers where it partly binds the produced sulfide as iron sulfide (FeS) and pyrite (FeS_2). Pyrite is the end product of iron-sulfide mineral formation and provides a deep sink for sulfur. Two pathways of pyrite formation are the "polysulfide pathway" (1) and the "H_2S pathway" (2) (Figure 13.3.1). The sulfidization of buried organic matter provides an additional deep sink for sulfur. Intermediate sulfur species, such as elemental sulfur (S^0), thiosulfate ($S_2O_3^{2-}$), tetrathionate ($S_4O_6^{2-}$), and sulfite (SO_3^{2-}), are formed during the oxidation of sulfide by, for example, buried Fe(III). These intermediates may be reduced back to sulfide, oxidized further to sulfate, or disproportionated to form both sulfide and sulfate. In very sulfidic sediments, a part of the sulfide diffuses up to the surface sediment where it may be oxidized by cable bacteria, by large sulfur bacteria such as

Beggiatoa spp., or by other, less conspicuous sulfide oxidizers. The different pathways of sulfide oxidation ultimately depend on oxygen (and less on nitrate) as the ultimate oxidant, and thereby consume a considerable part of the total oxygen uptake of the seabed (Jørgensen, 1982b). The oxygen flux into the sediment is enhanced by bioirrigation (ventillation of burrows) by the benthic macrofauna (e.g., Kristensen et al., 2013).

Sulfate Reduction

Organic Matter Degradation

Organic matter deposited on the seafloor provides food for the benthic communities, either at the sediment surface or upon burial into the sediment layers below. Oxygen is available for respiration and chemical reactions near the surface and through faunal burrows. Beneath this mixed surface zone, marine sediments constitute an anoxic world inhabited by anaerobic microorganisms. These subsurface organisms become increasingly sparse with depth, yet they account for half of all microbial cells in the ocean (Kallmeyer et al., 2012). Their energy source in most of the seabed is the buried organic matter, which they oxidize to CO_2 and inorganic nutrients. Due to the high concentration of sulfate in seawater (28 mM at an ocean salinity of 35), sulfate generally penetrates meters down into the seabed and supplies the sulfate reducing microorganisms (SRM) with an electron acceptor for their respiration. As the sediment ages with increasing burial depth beneath the seafloor, the remaining organic matter becomes steadily more refractory to microbial degradation. The time-course of organic matter degradation in the sediment, and thus of sulfate reduction rates (SRR), can be described by the sum of several exponential decay functions relating to different organic matter components, each of which is being degraded by first-order kinetics (Westrich and Berner, 1984).

The anaerobic degradation of organic matter involves complex microbial food chains, starting with the hydrolysis of macromolecular structures by extracellular enzymes and the formation of organic molecules small enough (generally < ca. 600 dalton, but for polysaccharides possibly larger) to be taken up by bacteria or archaea (Arnosti, 2011; Reintjes et al., 2017). It is this initial hydrolysis of the complex organic material that is rate-limiting for the overall degradation rate of organic matter (Kristensen and Holmer, 2001; Arnosti, 2004; Beulig et al., 2018). Microbial cells, which take up the small organic molecules such as sugars, amino acids, lipids, organic acids etc., conserve energy and grow by multistep fermentation processes that produce a range of volatile fatty acids (VFAs), such as formate, acetate, propionate and butyrate, plus H_2 and CO_2.

Sulfide Oxidation

Mass balance estimates and diffusion gradients of sulfide indicate that a significant fraction of the sulfide produced by sulfate reduction in marine sediments is reoxidized (Jørgensen, 1982b; Canfield et al., 1992; Pellerin et al., 2015b). This reoxidation occurs through diverse biological and geochemical pathways, forming a variety of reactive intermediates (Figure 13.3.2). The extent of sulfide reoxidation depends upon the quantity and type of available oxidant as well as the presence of microorganisms (e.g., Luther et al., 2011).

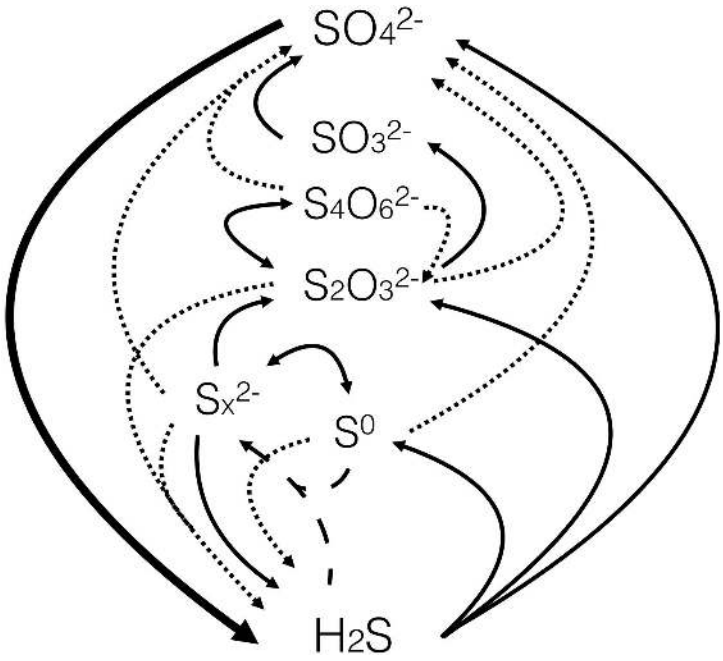


Figure 13.3.2: Overview of the processes and major inorganic species of the sulfur cycle. The large black arrow represents sulfate reduction, thin black lines represent oxidation and dotted lines represent disproportionation reactions. The dashed lines represent equilibration to form polysulfides. Schematic adapted from Zopfi et al. (2004). (Copyright; Copyright; Barker Jørgensen, Alyssa J. Findlay and André Pellerin via Frontiers in Microbiologyauthor via source)

Formation of Pyrite

The formation of pyrite (FeS_2) represents the main burial of sulfur, and thereby of reducing potential, in marine sediments, as pyrite is stable over geological timescales under anoxic conditions (Bottrell and Newton, 2006; Fike et al., 2015). Very generally, pyrite forms from the reaction of sulfide with buried ferric iron minerals, initially forming a mixture of elemental sulfur, polysulfides and ferrous iron minerals. Different overall reactions leading to pyrite formation in marine sediments have been proposed over the years, depending upon the initial reacting iron, and sulfur species. However, it has been argued that despite this potential variety only two reaction mechanisms are important: the reaction between FeS and H_2S ("H_2S pathway"; Equation 3) (Rickard and Luther, 1997; Thiel et al., 2019) and the reaction between FeS and polysulfide ("polysulfide pathway"; Equation 4) (Rickard and Luther, 2007).

$$\mathrm{FeS} + \mathrm{H_2\ S} \rightarrow \mathrm{FeS_2} + \mathrm{H_2} \tag{13.3.1}$$

$$\mathrm{FeS} + \mathrm{S_x^{2-}} \rightarrow \mathrm{FeS_2} + \mathrm{S_{x-1}}^{2-} \tag{13.3.2}$$

References

Arnosti, C. (2004). Speed bumps and barricades in the carbon cycle: substrate structural effects on carbon cycling. *Mar. Chem.* 92, 263–273. doi: 10.1016/j.marchem.2004.06.030

Arnosti, C. (2011). Microbial extracellular enzymes and the marine carbon cycle. *Ann. Rev. Mar. Sci.* 3, 401–425. doi: 10.1146/annurev-marine-120709-142731

Beulig, F., Røy, H., Glombitza, C., and Jørgensen, B. B. (2018). Control on rate and pathway of anaerobic organic carbon degradation in the seabed. *Proc. Natl. Acad. Sci. U.S.A.* 115, 367–372. doi: 10.1073/pnas.1715789115

Bottrell, S. H., and Newton, R. J. (2006). Reconstruction of changes in global sulfur cycling from marine sulfate isotopes. *Earth Sci. Rev.* 75, 59–83. doi: 10.1016/j.earscirev.2005.10.004

Canfield, D. E. (2001). Biogeochemistry of sulfur isotopes. *Rev. Mineral. Geochem.* 43, 607–636. doi: 10.2138/gsrmg.43.1.607

Canfield, D. E., Raiswell, R., and Bottrell, S. (1992). The reactivity of sedimentary iron minerals toward sulfide. *Am. J. Sci.* 292, 659–683. doi: 10.2475/ajs.292.9.659

Fike, D. A., Bradley, A. S., and Rose, C. V. (2015). Rethinking the ancient sulfur cycle. *Annu. Rev. Earth Planet. Sci.* 43, 593–622. doi: 10.1146/annurev-earth-060313-054802

Jørgensen, B. B. (1982b). Mineralization of organic matter in the sea bed - the role of sulfate reduction. *Nature* 296, 643–645. doi: 10.1038/296643a0:

Jørgensen, B. B., and Kasten, S. (2006). "Sulfur cycling and methane oxidation," in *Marine Geochemistry*, eds H. D. Schulz and M. Zabel (Berlin: Springer), 271–309. doi: 10.1007/3-540-32144-6_8:

Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C., and D'Hondt, S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16213–16216. doi: 10.1073/pnas.1203849109

Kristensen, E., Haese, R. R., and Kostka, J. E. (2013). *Interactions between Macro and Microorganisms in Marine Sediments, Coastal and Estuarine Studies*, Vol. 60. Washington, DC: American Geophysical Union. doi: 10.1007/bf00413135

Luther, G. W. III, Findlay, A. J., MacDonald, D. J., Owings, S. M., Hanson, T. E., Beinart, R. A., et al. (2011). Thermodynamics and kinetics of sulfide oxidation by oxygen: a look at inorganically controlled reactions and biologically mediated processes in the environment. *Front. Microbiol.* 2:62. doi: 10.3389/fmicb.2011.00062

Pellerin, A., Bui, T. H., Rough, M., Mucci, A., Canfield, D. E., and Wing, B. A. (2015b). Mass-dependent sulfur isotope fractionation during reoxidative sulfur cycling: a case study from Mangrove Lake, Bermuda. *Geochim. Cosmochim. Acta* 149, 152–164. doi: 10.1016/j.gca.2014.11.007

Reintjes, G., Arnosti, C., Fuchs, B. M., and Amann, R. (2017). An alternative polysaccharide uptake mechanism of marine bacteria. *ISME J.* 11, 1640–1650. doi: 10.1038/ismej.2017.26

Rickard, D. (2012). *Sulfidic Sediments and Sedimentary Rocks. Developments in Sedimentology*, Vol. 65. Amsterdam: Elsevier. doi: 10.1126/science.207.4437.1355:

Rickard, D., and Luther, G. W. III (1997). Kinetics of pyrite formation by the H2S oxidation of iron (II) monosulfide in aqueous solutions between 25 and 125°C: The mechanism. *Geochim. Cosmochim. Acta* 61, 135–147. doi: 10.1016/s0016-7037(96)00322-5

Rickard, D., and Luther, G. W. III (2007). Chemistry of iron sulfides. *Chem. Rev.* 107, 514–562. doi: 10.1021/cr0503658

Thiel, J., Byrne, J. M., Kappler, A., Schink, B., and Pester, M. (2019). Pyrite formation from FeS and H_2S is mediated through microbial redox activity. *Proc. Natl. Acad. Sci. U.S.A* 116, 6897–6902. doi: 10.1073/pnas.1814412116

Westrich, J. T., and Berner, R. A. (1984). The role of sedimentary organic matter in bacterial sulfate reduction - the G model tested. *Limnol. Oceanogr.* 29, 236–249. doi: 10.4319/lo.1984.29.2.0236

Zopfi, J., Ferdelman, T. G., and Fossing, H. (2004). "Distribution and fate of sulfur intermediates - sulfite, tetrathionate, thiosulfate, and elemental sulfur - in marine sediments," in *Sulfur Biogeochemistry—Past and Present*, Vol. 379, eds J. P. Amend, K. J. Edwards, and T. W. Lyons (Boulder, CO: Geological Society of America), 97–116. doi: 10.1130/0-8137-2379-5.97

Excerpted from:

Jørgensen, B. B., Findlay, A. J., & Pellerin, A. (2019). The biogeochemical sulfur cycle of marine sediments. *Frontiers in microbiology*, *10*, 849. Accessed December 2023 at https://www.frontiersin.org/articles/10.3389/fmicb.2019.00849/full. CC-BY-4.0

13.3: Interactions of the sulfur and oxygen cycles is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

13.4: Temporal perspectives on the global sulfur cycle

13.4: Temporal perspectives on the global sulfur cycle is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

13.5: The global mercury cycle

Learning Objectives

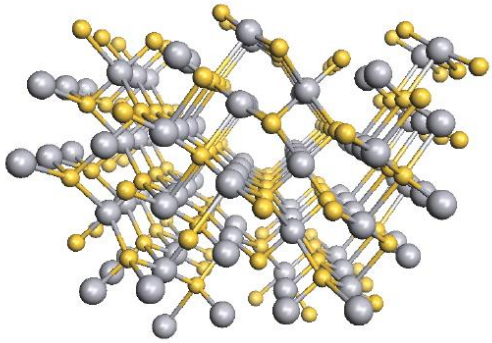
In this lecture you will learn the following

- Mercury cycle
- Mercury toxicity.
- Mercury poisoning.

mercury

The Mercury Cycle

Mercury is 62nd in terms of natural abundance and is found everywhere, usually as the mineral cinnabar, HgS, although 30 minerals containing Hg are known. The oxidation states are Hg(0), Hg(I) and Hg(II), where Hg(I) has been shown to exist as Hg_2^{2+}.



HgS - cinnabar

The three Hg species are related by the disproportionation:

$Hg_2^{2+} \rightarrow Hg^0 + Hg^{2+}$ $E° = -0.131$ V or $K = 6 \times 10^{-3}$

In addition:

Hg_2^{2+}/Hg $E° = 0.789$ V
Hg^{2+}/Hg $E° = 0.854$ V

This means that to oxidise Hg to Hg_2^{2+} an oxidising agent with potential > 0.789V is required, but very importantly < 0.854 V, otherwise oxidation to Hg^{2+} will occur. There are no common oxidants that fit this arrangement so if any reaction occurs the product will be Hg^{2+}. The equilibrium constant of 6×10^{-3} shows that when Hg(I) is formed it is moderately stable, however any agent that reduces the Hg(II) concentration automatically drives the reaction from Left → Right. Given that many Hg(II) derivatives are insoluble then this clearly restricts the range of Hg(I) compounds.

Minamata Disease

There have been several serious outbreaks of mercury poisoning. The most famous was between 1953 and 1965 at Minamata Bay in Japan when 46 people died and 120 suffered severe symptoms. As of March 2001, 2,265 victims had been officially recognised (1,784 of whom had died) and over 10,000 had received financial compensation from Chisso. By 2004, Chisso Corporation had paid $86 million in compensation, and in the same year was ordered to clean up its contamination. On March 29th, 2010, a settlement was reached to compensate as-yet uncertified victims.

The disease was first noticed in cats (who were seen throwing themselves into the sea) and was quickly traced to mercury poisoning acquired as a result of eating contaminated fish (5-10 ppm Hg). The investigations that followed showed that the fish had

acquired the high mercury due to the dumping of inorganic mercury salts and methylmercury from the Chisso Co. plastics factory upstream.

Analysis of fish exhibits from museums, some over 90 years old, has shown that mercury levels for ocean fish are similar but that river fish levels have risen as a result of man-made contamination. The forms of mercury occuring in the environment are Hg^{2+} and methylmercury, either $MeHg^{+}$ or Me_2Hg. Interconversion can be affected by microorganisms.

Aerobes can solubilize Hg^{2+} from cinnabar (K_{sp} ~10^{-53}) which in sediments was considered safe since the solubility product was so small. The conversion of $S^{2-} \rightarrow SO_3^{2-} \rightarrow SO_4^{2-}$ allows the insoluble sulfide to breakdown and in the process other Hg(II) salts are formed or the mercury may get reduced to Hg(0) enzymatically.

$$Hg^{2+} + NADH + H^{+} \longrightarrow Hg^{0} + NAD^{+} + 2\,H^{+} \tag{13.5.1}$$

where NADH = reduced form of nicotinamideadeninedinucleotide. This conversion can be considered as a detoxification process since Hg^0 is more easily eliminated.

In the environment, sulfate-reducing bacteria take up mercury in its inorganic form and through metabolic processes convert it to methylmercury. Sulfate-reducing bacteria are found in anaerobic conditions, typical of the well-buried muddy sediments of rivers, lakes, and oceans where methylmercury concentrations tend to be highest. Sulfate-reducing bacteria use sulfur rather than oxygen as their cellular energy-driving system. One hypothesis is that the uptake of inorganic mercury by sulfate-reducing bacteria occurs via passive diffusion of the dissolved complex HgS. Once the bacterium has taken up this complex, it utilizes detoxification enzymes to strip the sulfur group from the complex and replaces it with a methyl group:

$$HgS \longrightarrow CH_3Hg(II)X + H_2S \tag{13.5.2}$$

Upon methylation, the sulfate-reducing bacteria transport the new mercury complex back to the aquatic environment, where it is taken up by other microorganisms. Bacteria eliminate Hg by methylating it first to $MeHg^{+}$ and Me_2Hg. The detoxification process for them is the reverse for us unfortunately! The conversion probably involves vitamin B12 a methyl-cobalt organometallic compound so this is another example of synthesis involving transmetallation.

The major source of methylmercury exposure in humans is consumption of fish, marine mammals, and crustaceans. Once inside the human body, roughly 95% of the fish-derived methylmercury is absorbed from the gastrointestinal tract and distributed throughout the body. Uptake and accumulation of methylmercury is rapid due to the formation of methylmercury-cysteine complexes. Methylmercury is believed to cause toxicity by binding the sulfhydryl groups at the active centers of critical enzymes and structural proteins. Binding of methylmercury to these moieties constitutively alters the structure of the protein, inactivating or significantly lowering its functional capabilities.

Once the Me_2Hg is formed it is volatile and when released into the atmosphere it is readily photolysed by UV light

$$Me_2Hg \longrightarrow Hg^{0} + 2\,CH_3^{\cdot} \longrightarrow CH_4 \;\; \text{text or} \;\; C_2H_6 \tag{13.5.3}$$

Other microorganisms can convert $MeHg^{+}$ to Hg^0 + CH_4 that is make the mercury considerably less toxic to humans.

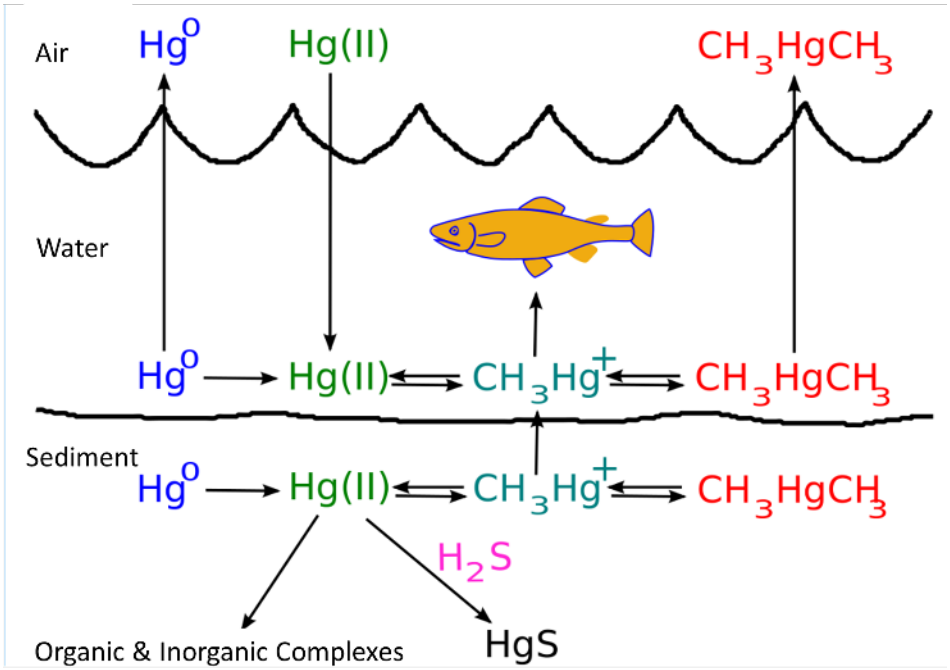


Figure 13.5.1: Chemistry of the mercury cycle. (GNU FDL; after Xhmikos Wkimedia Commons)

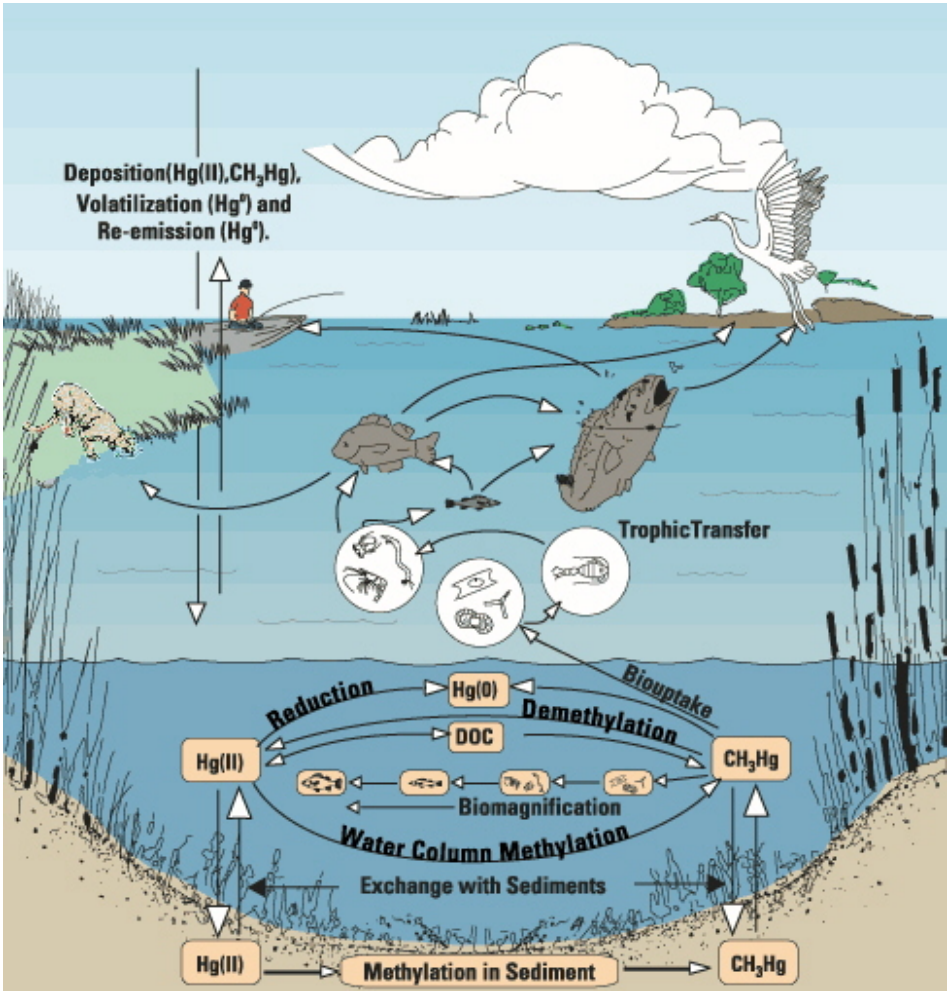


Figure 13.5.2: The mercury cycle showing biogeochemical interactions. (Copyright; author via source)

Mercury Toxicity

The toxicity of mercury arises from the very high affinity of the soft Hg atom for sulfhydryl (—SH) groups in enzymes. Simple mercury-sulfur compounds have been studied as potential analogs of natural systems. The Hg atoms are most commonly four-coordinated, as in $[Hg_2(SMe)_6]^{2-}$.

Mercury poisoning was a serious concern even from early days. Issac Newton, Alfred Stock worked in the early 20th century. Later in 60s awareness came following the incidence of brain damage and death it caused among the inhabitants in Minamata, Japan. Mercury from a plastic company was allowed to escape into a bay where it found its way into fish that were later eaten. Research has shown that bacteria found in sediments are capable of methylating mercury, and that species such $HgMe_2$ and $[HgCH_3]^+$ enter the food chain because they readily penetrate cell walls. The bacteria appear to produce $HgMe_2$ as a means of eliminating toxic mercury ions through their cell walls and into the environment.

References

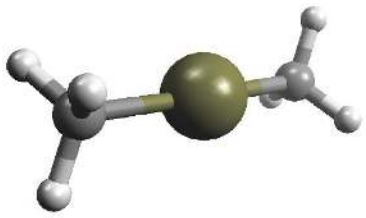
- Inorganic Chemistry, Principles of structure and reactivity, 4th edition; 1993, J. E. Huheey, E. A. Keiter, R. L. Keiter, Addison-Wesley Publishing Co, New York.
- Advanced Inorganic Chemistry, 6th edition, 1999, F. A. Cotton, G. Wilkinson, C. A. Murillo, M. Bochmann, John Wiley and Sons, New York.
- Organometallics, A Concise Introduction, 2nd edition (revised), 1992, Ch. Elschenbroich, A. Salzer, Weinheim, Germany.
- Inorganic Chemistry, 3rd Edition, 1999, D. F. Shriver, P. W. Atkins, Oxford University Press, Oxford.
- Inorganic Chemistry, 2nd Edition, 2005, C.C. Housecroft and A. G. Sharpe, Pearson, Prentice Hall, England.

Methylmercury

The usual method of ingestion of a metal into the body is:
a) Orally - the Gastrointestinal Tract
b) via the lungs - the Respiratory Tract

a) Orally - (mouth, stomach, small intestine): Food digestion begins in the mouth where saliva containing the enzyme amylase beaks down starch to lower sugars. Most digestion occurs in the stomach in the presence of HCl (pH ~1.6 ie. 0.17M HCl). In the case of mercury it is quite readily absorbed through the stomach since reaction of mercury salts with HCl produces $HgCl_2$. This neutral covalent molecule (solubility in H_2O 0.5 g/100 mL but 8g/100 mL in ethanol) is absorbed far better than most inorganic ions and this no doubt contributes to its high toxicity.

b) Respiratory tract (nose, lungs): This is usually unimportant for most metals but in some cases it can be more efficient that via the gastrointestinal tract. For example, lead where 50% of Pb in air can be absorbed but only 5-10% via the gastrointestinal tract. Volatile dimethyl mercury is another case for concern.



dimethyl

Summary

Organic mercury tends to increase up the food chain, particularly in lakes. The mud at the bottom of a lake may have 100 or 1000 times the amount of mercury than is in the water. Bacteria, worms and insects in the mud extract and concentrate the organic mercury. Small fish that eat them further concentrate the mercury in their bodies. This concentration process, known as "bioaccumulation", continues as larger fish eat smaller fish until the top predator fish in the lake may have methylmercury levels in their tissues that are up to 1,000,000 times the level in the water in which they live. We then eat the fish....

> *To consume a human being would be extremely unhealthy for any animal. Humans carry the highest concentration of toxic chemicals of all creatures on the planet. Their livers, hearts, kidneys and brains are so heavily contaminated with hundreds of different synthetic chemicals that if humans were slaughtered as a meat source, they'd never pass USDA food safety standards.*

Table 13.5.1: Properties of Me_2Hg

Molecular formula	C_2H_6Hg
Molar mass	230.66 g mol^{-1}
Appearance	Colourless liquid
Density	2.96 g/mL
Melting point	-43 °C
Boiling point	87-97 °C
Solubility in water	Insoluble

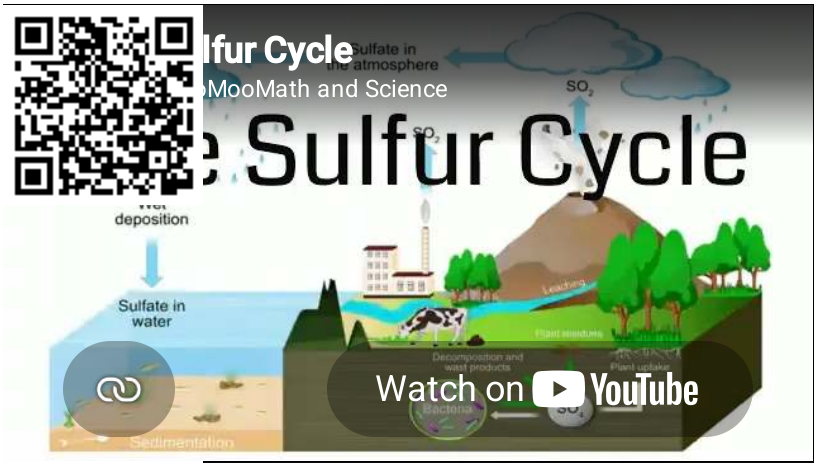
13.5: The global mercury cycle is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

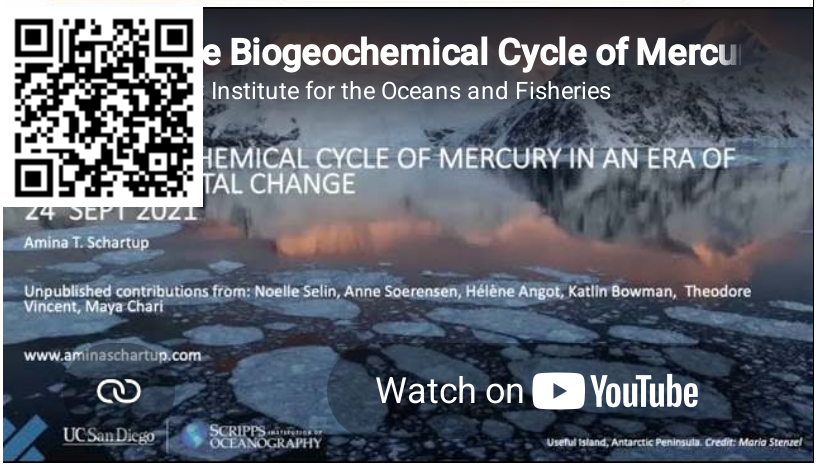
- **24.16.2: Organomercury Chemistry** is licensed CC BY-NC-SA 4.0.

13.6: Summary



Under Development





13.6: Summary is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by LibreTexts.

Index

A

absorb
5.8.1: Introduction to and Components of Food Webs
acid rain
3.7.1: Acid Deposition
Antarctic Ozone Hole
3.5.4: Stratospheric Ozone Formation
apex predator
5.8.1: Introduction to and Components of Food Webs
5.8.3: Trophic Cascades
aquatic
5.8.3: Trophic Cascades
Atmospheric aerosol
3.5.6: Where do Cloud Condensation Nuclei (CCN) come from?
ATP
5.8.1: Introduction to and Components of Food Webs
autotroph
5.8.1: Introduction to and Components of Food Webs

B

beaches
8.8: Waves and Beaches
bedrock
4.4: Bedrock
biodiversity
5.8.3: Trophic Cascades
biogenous sediments
9.5.4: Biogenous Sediments
biomass
5.8.2: Quantifying Food Webs
bottom feeder
5.8.1: Introduction to and Components of Food Webs
boulder lifting
4.5: Weathering

C

carbon
5.8.3: Trophic Cascades
carnivore
5.8.1: Introduction to and Components of Food Webs
chemoautotroph
5.8.1: Introduction to and Components of Food Webs
chemosynthesis
5.8.1: Introduction to and Components of Food Webs
cloud condensation nuclei (CCN)
3.5.1: Atmospheric Composition
3.5.6: Where do Cloud Condensation Nuclei (CCN) come from?
colloid
4.9: Soils
complexity
5.8.2: Quantifying Food Webs
connectance
5.8.2: Quantifying Food Webs
coprophagy
5.8.1: Introduction to and Components of Food Webs
cumulative distribution curve
4.8: Sediment

D

Daniel Pauly
5.8.1: Introduction to and Components of Food Webs
decomposer
5.8.1: Introduction to and Components of Food Webs
decomposition
5.8.1: Introduction to and Components of Food Webs
detritivore
5.8.1: Introduction to and Components of Food Webs
detritus
5.8.1: Introduction to and Components of Food Webs
detritus eater
5.8.1: Introduction to and Components of Food Webs
detritus feeder
5.8.1: Introduction to and Components of Food Webs
detrivore
5.8.1: Introduction to and Components of Food Webs
digest
5.8.1: Introduction to and Components of Food Webs
dissolved organic carbon
5.8.1: Introduction to and Components of Food Webs
DOC
5.8.1: Introduction to and Components of Food Webs

E

ecological pyramid
5.8.1: Introduction to and Components of Food Webs
5.8.2: Quantifying Food Webs
ecosystem
5.8.2: Quantifying Food Webs
ecosystem dynamics
5.8.1: Introduction to and Components of Food Webs
ecosystem structure
5.8.1: Introduction to and Components of Food Webs
electromagnetic spectrum
5.2: Photosynthesis
energy
5.8.1: Introduction to and Components of Food Webs
5.8.2: Quantifying Food Webs
erosion
4.13: Weathering, Erosion, and Mass Movement
exposures
4.4: Bedrock

F

fatty acid
5.8.2: Quantifying Food Webs
food chain
5.8.1: Introduction to and Components of Food Webs
food chain length
5.8.2: Quantifying Food Webs
food web
5.8.1: Introduction to and Components of Food Webs
5.8.2: Quantifying Food Webs
food web dynamics
5.8.3: Trophic Cascades
food web linkage
5.8.2: Quantifying Food Webs
frost wedging
4.5: Weathering
functional group
5.8.1: Introduction to and Components of Food Webs

G

Gaia hypothesis
1.4: Gaia - Bioregulation of the Environment
Geology of Mars
2.6.4: The Geology of Mars
gray wolf
5.8.3: Trophic Cascades
grus
4.5: Weathering
gut content analysis
5.8.2: Quantifying Food Webs

H

herbivore
5.8.1: Introduction to and Components of Food Webs
5.8.3: Trophic Cascades
heterotroph
5.8.1: Introduction to and Components of Food Webs
holistic model
5.8.1: Introduction to and Components of Food Webs
hydrogen sulfide
5.8.1: Introduction to and Components of Food Webs
hydrothermal vent
5.8.1: Introduction to and Components of Food Webs

I

ingest
5.8.1: Introduction to and Components of Food Webs
inverted pyramid
5.8.1: Introduction to and Components of Food Webs

K

kelp
5.8.3: Trophic Cascades
keystone species
5.8.3: Trophic Cascades

M

macromolecule
5.8.1: Introduction to and Components of Food Webs
magma
4.3: Rocks
marine snow
5.8.1: Introduction to and Components of Food Webs
mass movement
4.13: Weathering, Erosion, and Mass Movement
mass wasting
4.13: Weathering, Erosion, and Mass Movement
meteorite
2.6.4: The Geology of Mars
microbial loop
5.8.1: Introduction to and Components of Food Webs
mineral soils
4.9: Soils
mixotroph
5.8.1: Introduction to and Components of Food Webs
molecular tool
5.8.2: Quantifying Food Webs

N

nucleosynthesis
- 2.2: Origin of the Elements

O

omnivore
- 5.8.1: Introduction to and Components of Food Webs
- 5.8.3: Trophic Cascades

outcrops
- 4.4: Bedrock

oxidative phosphorylation
- 5.3: Respiration

Ozone
- 3.5.4: Stratospheric Ozone Formation

P

parameter
- 5.8.2: Quantifying Food Webs

pedology
- 4.9: Soils

pedosphere
- 4.9: Soils

photoautotroph
- 5.8.1: Introduction to and Components of Food Webs

photosynthesis
- 5.2: Photosynthesis
- 5.8.1: Introduction to and Components of Food Webs

phytoplankton
- 5.8.1: Introduction to and Components of Food Webs

plant defense
- 5.8.3: Trophic Cascades

Prebiotic Atmosphere
- 2.5.2: Origin and Evolution of the Atmosphere
- 3.3: Origin and Evolution of the Atmosphere

predator
- 5.8.3: Trophic Cascades

primary consumer
- 5.8.1: Introduction to and Components of Food Webs
- 5.8.3: Trophic Cascades

primary producer
- 5.8.1: Introduction to and Components of Food Webs
- 5.8.3: Trophic Cascades

producer
- 5.8.1: Introduction to and Components of Food Webs

productivity
- 5.8.2: Quantifying Food Webs

R

Raymond L. Lindemann
- 5.8.1: Introduction to and Components of Food Webs
- 5.8.2: Quantifying Food Webs

regolith
- 4.7: Regolith

relief
- 4.4: Bedrock

remineralization
- 5.8.1: Introduction to and Components of Food Webs

remotely operated vehicle
- 5.8.2: Quantifying Food Webs

respiration
- 5.3: Respiration

ROV
- 5.8.2: Quantifying Food Webs

runaway greenhouse effect
- 2.6.3: The Massive Atmosphere of Venus

S

saprotroph
- 5.8.1: Introduction to and Components of Food Webs

scat content analysis
- 5.8.2: Quantifying Food Webs

scavenger
- 5.8.1: Introduction to and Components of Food Webs

sea otter
- 5.8.3: Trophic Cascades

sea urchin
- 5.8.3: Trophic Cascades

secondary consumer
- 5.8.1: Introduction to and Components of Food Webs

soil organic matter
- 5.8.1: Introduction to and Components of Food Webs

species composition
- 5.8.2: Quantifying Food Webs

species richness
- 5.8.2: Quantifying Food Webs

stability
- 5.8.2: Quantifying Food Webs

stable isotopes
- 5.8.2: Quantifying Food Webs

Stratospheric Ozone
- 3.5.4: Stratospheric Ozone Formation

surface area
- 4.9: Soils

surface charge
- 4.9: Soils

T

tectonic
- 2.6.2: The Geology of Venus

ten percent law
- 5.8.2: Quantifying Food Webs

terrestrial
- 5.8.3: Trophic Cascades

tertiary consumer
- 5.8.1: Introduction to and Components of Food Webs

top consumer
- 5.8.3: Trophic Cascades

transfer efficiency
- 5.8.2: Quantifying Food Webs

trophic cascade
- 5.8.3: Trophic Cascades

trophic group
- 5.8.1: Introduction to and Components of Food Webs

trophic interaction
- 5.8.1: Introduction to and Components of Food Webs
- 5.8.3: Trophic Cascades

trophic level
- 5.8.1: Introduction to and Components of Food Webs
- 5.8.2: Quantifying Food Webs
- 5.8.3: Trophic Cascades

trophic link
- 5.8.2: Quantifying Food Webs

trophic pyramid
- 5.8.1: Introduction to and Components of Food Webs
- 5.8.2: Quantifying Food Webs

trophic relationship
- 5.8.1: Introduction to and Components of Food Webs

trophic trickle
- 5.8.3: Trophic Cascades

Twin Mars Exploration Rovers (MER)
- 2.6.4: The Geology of Mars

V

volatile organic compounds (VOCs)
- 3.5.3: Other Trace Gases

W

waves
- 8.8: Waves and Beaches

weathering
- 4.5: Weathering
- 4.13: Weathering, Erosion, and Mass Movement

Y

Yellowstone
- 5.8.3: Trophic Cascades

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: BioGeoChemistry (LibreTexts)

Webpages: 228

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-SA 4.0: 51.3% (117 pages)
- CC BY 4.0: 16.7% (38 pages)
- CC BY-NC-SA 4.0: 14% (32 pages)
- Undeclared: 9.6% (22 pages)
- Public Domain: 6.6% (15 pages)
- CC BY-NC 4.0: 1.8% (4 pages)

By Page

- BioGeoChemistry (LibreTexts) - *CC BY-SA 4.0*
  - Front Matter - *CC BY-SA 4.0*
    - TitlePage - *CC BY-SA 4.0*
    - InfoPage - *CC BY-SA 4.0*
    - Table of Contents - *Undeclared*
    - Licensing - *CC BY-SA 4.0*
  - 1: Introduction - *CC BY-SA 4.0*
    - 1.1: What is biogeochemistry? - *CC BY-SA 4.0*
    - 1.2: Chemistry and Energetics of the Life Process - *CC BY-SA 4.0*
    - 1.3: Systems, Cycles, Reservoirs, and Fluxes - *CC BY-NC-SA 4.0*
    - 1.4: Gaia - Bioregulation of the Environment - *CC BY-SA 4.0*
  - 2: Origins - *CC BY-SA 4.0*
    - 2.1: Introduction - *CC BY 4.0*
    - 2.2: Origin of the Elements - *CC BY-SA 4.0*
    - 2.3: Formation and evolution of the Earth - *CC BY-SA 4.0*
    - 2.4: The Earth's crust - *CC BY-SA 4.0*
    - 2.5: Origin of the atmosphere and the oceans - *CC BY-SA 4.0*
      - 2.5.1: Origin of the Oceans - *CC BY 4.0*
      - 2.5.2: Origin and Evolution of the Atmosphere - *CC BY-SA 4.0*
    - 2.6: Earthlike Planets - Venus and Mars - *CC BY 4.0*
      - 2.6.1: The Nearest Planets - An Overview - *CC BY 4.0*
      - 2.6.2: The Geology of Venus - *CC BY 4.0*
      - 2.6.3: The Massive Atmosphere of Venus - *CC BY 4.0*
      - 2.6.4: The Geology of Mars - *CC BY 4.0*
      - 2.6.5: Water and Life on Mars - *CC BY 4.0*
      - 2.6.6: Divergent Planetary Evolution - *CC BY 4.0*
      - 2.6.E: Earthlike Planets - Venus and Mars (Exercises) - *CC BY 4.0*
    - 2.7: Summary - *CC BY-SA 4.0*
  - 3: The Atmosphere - *CC BY-SA 4.0*
    - 3.1: Introduction - *CC BY 4.0*
    - 3.2: Structure and Composition of the Atmosphere - *CC BY-SA 4.0*
    - 3.3: Origin and Evolution of the Atmosphere - *CC BY-SA 4.0*
    - 3.4: General Circulation - *CC BY-NC-SA 4.0*
      - 3.4.1: A Simplified Description of the Global Circulation - *CC BY-NC-SA 4.0*
      - 3.4.2: Radiative Differential Heating - *CC BY-NC-SA 4.0*
      - 3.4.3: Pressure Profiles - *CC BY-NC-SA 4.0*
      - 3.4.4: Geostrophic Wind and Geostrophic Adjustment - *CC BY-NC-SA 4.0*
      - 3.4.5: Thermal Wind Effect - *CC BY-NC-SA 4.0*
      - 3.4.6: Horizontal Circulation - *CC BY-NC-SA 4.0*
      - 3.4.7: Extratropical Ridges and Troughs (Rossby Waves) - *CC BY-NC-SA 4.0*
      - 3.4.8: Three-band Global Circulation - *CC BY-NC-SA 4.0*
      - 3.4.9: Ekman Spiral of Ocean Currents - *CC BY-NC-SA 4.0*
      - 3.4.10: Explaining the General Circulation - *CC BY-NC-SA 4.0*

- 3.4.11: Jet Streams - *CC BY-NC-SA 4.0*
- 3.4.12: Types of Vorticity - *CC BY-NC-SA 4.0*
- 3.4.13: Key Terms - *CC BY-NC-SA 4.0*
- 3.4.14: Review - *CC BY-NC-SA 4.0*
- 3.4.15: Homework Exercises - *CC BY-NC-SA 4.0*

- 3.5: Atmospheric Composition - *CC BY-NC-SA 4.0*
  - 3.5.1: Atmospheric Composition - *CC BY-NC-SA 4.0*
  - 3.5.2: Changes in Atmospheric Composition - *CC BY-NC-SA 4.0*
  - 3.5.3: Other Trace Gases - *CC BY-NC-SA 4.0*
  - 3.5.4: Stratospheric Ozone Formation - *CC BY-NC-SA 4.0*
  - 3.5.5: The Story of the Atmosphere's PAC-MAN - *CC BY-NC-SA 4.0*
  - 3.5.6: Where do Cloud Condensation Nuclei (CCN) come from? - *CC BY-NC-SA 4.0*
  - 3.5.7: Summary and Final Tasks - *CC BY-NC-SA 4.0*
- 3.6: Biogeochemical reactions in the troposphere - *Public Domain*
- 3.7: Atmospheric deposition - *Public Domain*
  - 3.7.1: Acid Deposition - *CC BY-NC 4.0*
- 3.8: Biogeochemical reactions in the stratosphere - *Public Domain*
- 3.9: Models - *CC BY-SA 4.0*
- 3.10: Impacts - *CC BY-SA 4.0*
- 3.11: Summary - *CC BY-SA 4.0*

- 4: The Lithosphere - *CC BY-SA 4.0*
  - 4.1: The Earth - The Ground Beneath Your Feet - *CC BY-SA 4.0*
  - 4.2: Minerals - *CC BY-SA 4.0*
  - 4.3: Rocks - *CC BY-SA 4.0*
  - 4.4: Bedrock - *CC BY-SA 4.0*
  - 4.5: Weathering - *CC BY-SA 4.0*
  - 4.6: The Mineral Particles of the Earth's Surface Materials - *CC BY-SA 4.0*
  - 4.7: Regolith - *CC BY-SA 4.0*
  - 4.8: Sediment - *CC BY-SA 4.0*
  - 4.9: Soils - *CC BY-SA 4.0*
  - 4.10: Soil Chemistry - *CC BY-SA 4.0*
  - 4.11: Soil Biodiversity and Ecology - *CC BY-NC 4.0*
  - 4.12: Soil Nutrient Cycling - *CC BY-NC 4.0*
  - 4.13: Weathering, Erosion, and Mass Movement - *CC BY-SA 4.0*
    - 4.13.1: Weathering - *CC BY-SA 4.0*
    - 4.13.2: Mass Movement - *CC BY-SA 4.0*
    - 4.13.3: Water Erosion - *CC BY-SA 4.0*
    - 4.13.4: Review and Additional Resources - *CC BY-SA 4.0*
  - 4.14: Summary - *CC BY-SA 4.0*
- 5: The Carbon Cycle of Terrestrial Ecosystems - *CC BY-SA 4.0*
  - 5.1: Introduction - *CC BY-SA 4.0*
  - 5.2: Photosynthesis - *CC BY-SA 4.0*
  - 5.3: Respiration - *CC BY-SA 4.0*
  - 5.4: Net Primary Production - *Undeclared*
    - 5.4.1: Marine Primary Production - *Undeclared*
      - 5.4.1.1: Primary production - *CC BY-SA 4.0*
      - 5.4.1.2: The Producers - *CC BY 4.0*
      - 5.4.1.3: Factors Influencing Production - *CC BY 4.0*
      - 5.4.1.4: Patterns of Primary Production - *CC BY 4.0*
    - 5.4.2: Surface Primary Production - *Public Domain*
  - 5.5: Net ecosystem production and Eddy-covariance studies - *CC BY-SA 4.0*
  - 5.6: Remote sensing of primary production and biomass - *CC BY-SA 4.0*
  - 5.7: Global estimates of net primary production and biomass - *Public Domain*
  - 5.8: The fate of net primary production - *CC BY-SA 4.0*
    - 5.8.1: Introduction to and Components of Food Webs - *CC BY-NC-SA 4.0*
    - 5.8.2: Quantifying Food Webs - *CC BY-NC-SA 4.0*
    - 5.8.3: Trophic Cascades - *CC BY-NC-SA 4.0*
    - 5.8.4: Sources and Attributions - *CC BY-NC-SA 4.0*
  - 5.9: Net primary production and global change - *Public Domain*
  - 5.10: Detritus - *CC BY-SA 4.0*
  - 5.11: Soil organic matter and global change - *CC BY 4.0*
  - 5.12: Summary - *CC BY-SA 4.0*
- 6: Biogeochemical Cycling on Land - *CC BY-SA 4.0*
  - 6.1: Biogeochemical cycling in land plants - *CC BY-SA 4.0*
  - 6.2: Nutrient allocations and cycling in land vegetation - *CC BY 4.0*
  - 6.3: Biogeochemical cycling in the soil - *CC BY-SA 4.0*
    - 6.3.1: Soils and the carbon cycle - *Undeclared*
    - 6.3.2: Soils and the Water Cycle - *Undeclared*

- 6.4: Calculating landscape mass balance - *CC BY-SA 4.0*
- 6.5: Ecosystem models and remote sensing - *CC BY-SA 4.0*
- 6.6: Ecosystem Monitoring - *CC BY-SA 4.0*
- 6.7: Human impacts on terrestrial biogeochemistry - *Public Domain*
  - 6.7.1: Introduction - *Public Domain*
  - 6.7.2: Human-Induced Changes - *Public Domain*
  - 6.7.3: Sinks and Cycles - *Public Domain*
  - 6.7.4: Impacts and Options - *Public Domain*
  - 6.7.5: References - *Undeclared*
- 6.8: Summary - *CC BY-SA 4.0*

7: Wetland Ecosystems - *CC BY-SA 4.0*
- 7.1: Introduction - *CC BY-SA 4.0*
- 7.2: Types of wetlands - *CC BY-SA 4.0*
- 7.3: Productivity in wetland ecosystems - *CC BY-SA 4.0*
- 7.4: Organic matter storage in wetlands - *CC BY 4.0*
- 7.5: Microbial metabolism in saturated sediments - *CC BY 4.0*
- 7.6: Anaerobic and aerobic metabolic pathways - *CC BY 4.0*
- 7.7: The redox ladder - *CC BY-SA 4.0*
- 7.8: Wetlands and water quality - *CC BY-SA 4.0*
- 7.9: Wetlands and global change - *CC BY 4.0*
  - 7.9.1: Introduction - *Undeclared*
  - 7.9.2: Carbon Accumulation and GHG Emissions from Freshwater Wetlands (Including Permafrost) In a Changing Climate - *CC BY 4.0*
  - 7.9.3: Ecological Consequences for Freshwater Wetlands in a Changing Climate - *CC BY 4.0*
  - 7.9.4: Salt Marsh and Mangrove Response to a Changing Climate and Associated Sea Level Rise - *CC BY 4.0*
- 7.10: Summary - *CC BY-SA 4.0*

8: Inland waters - *CC BY-SA 4.0*
- 8.1: Introduction - *CC BY-SA 4.0*
- 8.2: The structure of aquatic ecosystems - *CC BY-SA 4.0*
- 8.3: The special properties of water - *CC BY-NC-SA 4.0*
- 8.4: Inland waters and their watersheds - *CC BY-SA 4.0*
  - 8.4.1: The Physical Template - *Undeclared*
  - 8.4.2: Watershed Structure - *Undeclared*
  - 8.4.3: Watershed Functions - *Undeclared*
  - 8.4.4: Summary - *Undeclared*
- 8.5: Lakes - *CC BY-SA 4.0*
  - 8.5.1: Freshwater Lakes - *Undeclared*
  - 8.5.2: Saline Lakes - *Undeclared*
- 8.6: Streams and Rivers - *CC BY-SA 4.0*
- 8.7: Estuaries- Where the turf meets the surf - *CC BY-SA 4.0*
- 8.8: Waves and Beaches - *Undeclared*
- 8.9: Human impacts on inland waters - *CC BY-SA 4.0*
  - 8.9.1: Fresh Water Supply and the Water Cycle - *CC BY-SA 4.0*
  - 8.9.2: Water Usage - *CC BY-SA 4.0*
  - 8.9.3: Water Scarcity and Solutions - *CC BY-SA 4.0*
  - 8.9.4: Data Dive- Aqueducts of Rome - *CC BY-SA 4.0*
  - 8.9.5: Review - *CC BY-SA 4.0*
- 8.10: Summary - *CC BY-SA 4.0*

9: The Oceans - *CC BY-SA 4.0*
- 9.1: Introduction - *CC BY-NC 4.0*
- 9.2: Properties of Seawater - *CC BY-SA 4.0*
  - 9.2.1: Properties of Seawater - *CC BY-SA 4.0*
  - 9.2.2: Specific Heat and Latent Heat Capacity of Water - *CC BY-SA 4.0*
  - 9.2.3: Energy Associated With Evaporation and Condensation of Water In the Air - *CC BY-SA 4.0*
  - 9.2.4: Salinity - *CC BY-SA 4.0*
  - 9.2.5: Formation of Sea Ice - *CC BY-SA 4.0*
  - 9.2.6: Relationship of Salinity, Density and Temperature - *CC BY-SA 4.0*
  - 9.2.7: Salinity and Latitude - *CC BY-SA 4.0*
  - 9.2.8: "Cline Curves" and Mixing (Surface) Zone - *CC BY-SA 4.0*
  - 9.2.9: Gases Dissolved In Seawater - *CC BY-SA 4.0*
  - 9.2.10: Local Conditions in the San Diego Region - *CC BY-SA 4.0*
  - 9.2.11: Quiz Questions Properties of Seawater - *CC BY-SA 4.0*
- 9.3: Biogeochemistry of the surface ocean - *CC BY-SA 4.0*
  - 9.3.1: Introduction - *CC BY 4.0*
  - 9.3.2: Biological Pump - *Undeclared*
  - 9.3.3: Export from the Upper Ocean to the Depths and Sediments - *Undeclared*
  - 9.3.4: Summary - *Undeclared*
- 9.4: 12.4 Biogeochemistry of the Deep Ocean - *CC BY-NC-SA 4.0*

- 9.5: Biogeochemistry in ocean sediments - *CC BY-SA 4.0*
  - 9.5.1: Introduction - *Undeclared*
  - 9.5.2: Classifying Sediments - *CC BY 4.0*
  - 9.5.3: Lithogenous Sediments - *CC BY 4.0*
  - 9.5.4: Biogenous Sediments - *CC BY 4.0*
  - 9.5.5: Hydrogenous Sediments - *CC BY 4.0*
  - 9.5.6: Cosmogenous Sediments - *CC BY 4.0*
  - 9.5.7: Sediment Distribution - *Undeclared*
- 9.6: Marine element cycles - *CC BY-SA 4.0*
- 9.7: The Oceans and Global Climate Change - *Undeclared*
- 9.8: Summary - *CC BY-SA 4.0*

10: The Global Water Cycle - *CC BY-SA 4.0*
- 10.1: Introduction - *Public Domain*
- 10.2: Models of the hydrologic cycle - *CC BY-SA 4.0*
- 10.3: The history of the water cycle - *CC BY 4.0*
- 10.4: The water cycle and climate change - *CC BY-SA 4.0*
- 10.5: Summary - *CC BY-SA 4.0*

11: The Global Carbon and Oxygen Cycles - *CC BY-SA 4.0*
- 11.1: Introduction - *CC BY-SA 4.0*
- 11.2: What might stimulate net primary production? - *Public Domain*
- 11.3: Temporal perspectives of the carbon cycle - *Public Domain*
- 11.4: Atmospheric methane - *Public Domain*
- 11.5: Carbon monoxide - *CC BY 4.0*
- 11.6: Synthesis- Linking the carbon and oxygen cycles - *CC BY-NC-SA 4.0*
- 11.7: Summary - *CC BY-SA 4.0*

12: The Global Cycles of Nitrogen, Phosphorus and Potassium - *CC BY-SA 4.0*
- 12.1: Introduction - *CC BY-SA 4.0*
- 12.2: The global nitrogen cycle - *CC BY-SA 4.0*
  - 12.2.1: Chemistry of the nitrogen cycle - *CC BY 4.0*
- 12.3: Temporal variations in the global nitrogen cycle - *CC BY 4.0*
- 12.4: Nitrous Oxide (N2O) - *CC BY 4.0*
  - 12.4.1: Terrestrial - *CC BY 4.0*
  - 12.4.2: Marine - *CC BY 4.0*
  - 12.4.3: Human Sources - *CC BY 4.0*
- 12.5: The global phosphorus cycle - *CC BY-SA 4.0*
- 12.6: Potassium- An overlooked limiting element? - *CC BY 4.0*
- 12.7: Linking global biogeochemical cycles - *CC BY-SA 4.0*
- 12.8: Summary - *CC BY-SA 4.0*

13: The Global Cycles of Sulfur and Mercury - *CC BY-SA 4.0*
- 13.1: Introduction - *CC BY-SA 4.0*
- 13.2: The global sulfur cycle - *CC BY-SA 4.0*
- 13.3: Interactions of the sulfur and oxygen cycles - *CC BY-SA 4.0*
- 13.4: Temporal perspectives on the global sulfur cycle - *CC BY-SA 4.0*
- 13.5: The global mercury cycle - *CC BY-SA 4.0*
- 13.6: Summary - *CC BY-SA 4.0*

Back Matter - *CC BY-SA 4.0*
- Index - *CC BY-SA 4.0*
- Glossary - *Undeclared*
- Detailed Licensing - *CC BY-SA 4.0*
- Detailed Licensing - *Undeclared*

Detailed Licensing

Overview

Title: BioGeoChemistry (LibreTexts)

Webpages: 228

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-SA 4.0: 51.3% (117 pages)
- CC BY 4.0: 16.7% (38 pages)
- CC BY-NC-SA 4.0: 14% (32 pages)
- Undeclared: 9.6% (22 pages)
- Public Domain: 6.6% (15 pages)
- CC BY-NC 4.0: 1.8% (4 pages)

By Page

- BioGeoChemistry (LibreTexts) - *CC BY-SA 4.0*
  - Front Matter - *CC BY-SA 4.0*
    - TitlePage - *CC BY-SA 4.0*
    - InfoPage - *CC BY-SA 4.0*
    - Table of Contents - *Undeclared*
    - Licensing - *CC BY-SA 4.0*
  - 1: Introduction - *CC BY-SA 4.0*
    - 1.1: What is biogeochemistry? - *CC BY-SA 4.0*
    - 1.2: Chemistry and Energetics of the Life Process - *CC BY-SA 4.0*
    - 1.3: Systems, Cycles, Reservoirs, and Fluxes - *CC BY-NC-SA 4.0*
    - 1.4: Gaia - Bioregulation of the Environment - *CC BY-SA 4.0*
  - 2: Origins - *CC BY-SA 4.0*
    - 2.1: Introduction - *CC BY 4.0*
    - 2.2: Origin of the Elements - *CC BY-SA 4.0*
    - 2.3: Formation and evolution of the Earth - *CC BY-SA 4.0*
    - 2.4: The Earth's crust - *CC BY-SA 4.0*
    - 2.5: Origin of the atmosphere and the oceans - *CC BY-SA 4.0*
      - 2.5.1: Origin of the Oceans - *CC BY 4.0*
      - 2.5.2: Origin and Evolution of the Atmosphere - *CC BY-SA 4.0*
    - 2.6: Earthlike Planets - Venus and Mars - *CC BY 4.0*
      - 2.6.1: The Nearest Planets - An Overview - *CC BY 4.0*
      - 2.6.2: The Geology of Venus - *CC BY 4.0*
      - 2.6.3: The Massive Atmosphere of Venus - *CC BY 4.0*
      - 2.6.4: The Geology of Mars - *CC BY 4.0*
      - 2.6.5: Water and Life on Mars - *CC BY 4.0*
      - 2.6.6: Divergent Planetary Evolution - *CC BY 4.0*
      - 2.6.E: Earthlike Planets - Venus and Mars (Exercises) - *CC BY 4.0*
    - 2.7: Summary - *CC BY-SA 4.0*
  - 3: The Atmosphere - *CC BY-SA 4.0*
    - 3.1: Introduction - *CC BY 4.0*
    - 3.2: Structure and Composition of the Atmosphere - *CC BY-SA 4.0*
    - 3.3: Origin and Evolution of the Atmosphere - *CC BY-SA 4.0*
    - 3.4: General Circulation - *CC BY-NC-SA 4.0*
      - 3.4.1: A Simplified Description of the Global Circulation - *CC BY-NC-SA 4.0*
      - 3.4.2: Radiative Differential Heating - *CC BY-NC-SA 4.0*
      - 3.4.3: Pressure Profiles - *CC BY-NC-SA 4.0*
      - 3.4.4: Geostrophic Wind and Geostrophic Adjustment - *CC BY-NC-SA 4.0*
      - 3.4.5: Thermal Wind Effect - *CC BY-NC-SA 4.0*
      - 3.4.6: Horizontal Circulation - *CC BY-NC-SA 4.0*
      - 3.4.7: Extratropical Ridges and Troughs (Rossby Waves) - *CC BY-NC-SA 4.0*
      - 3.4.8: Three-band Global Circulation - *CC BY-NC-SA 4.0*
      - 3.4.9: Ekman Spiral of Ocean Currents - *CC BY-NC-SA 4.0*
      - 3.4.10: Explaining the General Circulation - *CC BY-NC-SA 4.0*

- 3.4.11: Jet Streams - *CC BY-NC-SA 4.0*
- 3.4.12: Types of Vorticity - *CC BY-NC-SA 4.0*
- 3.4.13: Key Terms - *CC BY-NC-SA 4.0*
- 3.4.14: Review - *CC BY-NC-SA 4.0*
- 3.4.15: Homework Exercises - *CC BY-NC-SA 4.0*
- 3.5: Atmospheric Composition - *CC BY-NC-SA 4.0*
  - 3.5.1: Atmospheric Composition - *CC BY-NC-SA 4.0*
  - 3.5.2: Changes in Atmospheric Composition - *CC BY-NC-SA 4.0*
  - 3.5.3: Other Trace Gases - *CC BY-NC-SA 4.0*
  - 3.5.4: Stratospheric Ozone Formation - *CC BY-NC-SA 4.0*
  - 3.5.5: The Story of the Atmosphere's PAC-MAN - *CC BY-NC-SA 4.0*
  - 3.5.6: Where do Cloud Condensation Nuclei (CCN) come from? - *CC BY-NC-SA 4.0*
  - 3.5.7: Summary and Final Tasks - *CC BY-NC-SA 4.0*
- 3.6: Biogeochemical reactions in the troposphere - *Public Domain*
- 3.7: Atmospheric deposition - *Public Domain*
  - 3.7.1: Acid Deposition - *CC BY-NC 4.0*
- 3.8: Biogeochemical reactions in the stratosphere - *Public Domain*
- 3.9: Models - *CC BY-SA 4.0*
- 3.10: Impacts - *CC BY-SA 4.0*
- 3.11: Summary - *CC BY-SA 4.0*

4: The Lithosphere - *CC BY-SA 4.0*

- 4.1: The Earth - The Ground Beneath Your Feet - *CC BY-SA 4.0*
- 4.2: Minerals - *CC BY-SA 4.0*
- 4.3: Rocks - *CC BY-SA 4.0*
- 4.4: Bedrock - *CC BY-SA 4.0*
- 4.5: Weathering - *CC BY-SA 4.0*
- 4.6: The Mineral Particles of the Earth's Surface Materials - *CC BY-SA 4.0*
- 4.7: Regolith - *CC BY-SA 4.0*
- 4.8: Sediment - *CC BY-SA 4.0*
- 4.9: Soils - *CC BY-SA 4.0*
- 4.10: Soil Chemistry - *CC BY-SA 4.0*
- 4.11: Soil Biodiversity and Ecology - *CC BY-NC 4.0*
- 4.12: Soil Nutrient Cycling - *CC BY-NC 4.0*
- 4.13: Weathering, Erosion, and Mass Movement - *CC BY-SA 4.0*
  - 4.13.1: Weathering - *CC BY-SA 4.0*
  - 4.13.2: Mass Movement - *CC BY-SA 4.0*
  - 4.13.3: Water Erosion - *CC BY-SA 4.0*
  - 4.13.4: Review and Additional Resources - *CC BY-SA 4.0*
- 4.14: Summary - *CC BY-SA 4.0*

5: The Carbon Cycle of Terrestrial Ecosystems - *CC BY-SA 4.0*

- 5.1: Introduction - *CC BY-SA 4.0*
- 5.2: Photosynthesis - *CC BY-SA 4.0*
- 5.3: Respiration - *CC BY-SA 4.0*
- 5.4: Net Primary Production - *Undeclared*
  - 5.4.1: Marine Primary Production - *Undeclared*
    - 5.4.1.1: Primary production - *CC BY-SA 4.0*
    - 5.4.1.2: The Producers - *CC BY 4.0*
    - 5.4.1.3: Factors Influencing Production - *CC BY 4.0*
    - 5.4.1.4: Patterns of Primary Production - *CC BY 4.0*
  - 5.4.2: Surface Primary Production - *Public Domain*
- 5.5: Net ecosystem production and Eddy-covariance studies - *CC BY-SA 4.0*
- 5.6: Remote sensing of primary production and biomass - *CC BY-SA 4.0*
- 5.7: Global estimates of net primary production and biomass - *Public Domain*
- 5.8: The fate of net primary production - *CC BY-SA 4.0*
  - 5.8.1: Introduction to and Components of Food Webs - *CC BY-NC-SA 4.0*
  - 5.8.2: Quantifying Food Webs - *CC BY-NC-SA 4.0*
  - 5.8.3: Trophic Cascades - *CC BY-NC-SA 4.0*
  - 5.8.4: Sources and Attributions - *CC BY-NC-SA 4.0*
- 5.9: Net primary production and global change - *Public Domain*
- 5.10: Detritus - *CC BY-SA 4.0*
- 5.11: Soil organic matter and global change - *CC BY 4.0*
- 5.12: Summary - *CC BY-SA 4.0*

6: Biogeochemical Cycling on Land - *CC BY-SA 4.0*

- 6.1: Biogeochemical cycling in land plants - *CC BY-SA 4.0*
- 6.2: Nutrient allocations and cycling in land vegetation - *CC BY 4.0*
- 6.3: Biogeochemical cycling in the soil - *CC BY-SA 4.0*
  - 6.3.1: Soils and the carbon cycle - *Undeclared*
  - 6.3.2: Soils and the Water Cycle - *Undeclared*

- 6.4: Calculating landscape mass balance - *CC BY-SA 4.0*
- 6.5: Ecosystem models and remote sensing - *CC BY-SA 4.0*
- 6.6: Ecosystem Monitoring - *CC BY-SA 4.0*
- 6.7: Human impacts on terrestrial biogeochemistry - *Public Domain*
  - 6.7.1: Introduction - *Public Domain*
  - 6.7.2: Human-Induced Changes - *Public Domain*
  - 6.7.3: Sinks and Cycles - *Public Domain*
  - 6.7.4: Impacts and Options - *Public Domain*
  - 6.7.5: References - *Undeclared*
- 6.8: Summary - *CC BY-SA 4.0*

7: Wetland Ecosystems - *CC BY-SA 4.0*
- 7.1: Introduction - *CC BY-SA 4.0*
- 7.2: Types of wetlands - *CC BY-SA 4.0*
- 7.3: Productivity in wetland ecosystems - *CC BY-SA 4.0*
- 7.4: Organic matter storage in wetlands - *CC BY 4.0*
- 7.5: Microbial metabolism in saturated sediments - *CC BY 4.0*
- 7.6: Anaerobic and aerobic metabolic pathways - *CC BY 4.0*
- 7.7: The redox ladder - *CC BY-SA 4.0*
- 7.8: Wetlands and water quality - *CC BY-SA 4.0*
- 7.9: Wetlands and global change - *CC BY 4.0*
  - 7.9.1: Introduction - *Undeclared*
  - 7.9.2: Carbon Accumulation and GHG Emissions from Freshwater Wetlands (Including Permafrost) In a Changing Climate - *CC BY 4.0*
  - 7.9.3: Ecological Consequences for Freshwater Wetlands in a Changing Climate - *CC BY 4.0*
  - 7.9.4: Salt Marsh and Mangrove Response to a Changing Climate and Associated Sea Level Rise - *CC BY 4.0*
- 7.10: Summary - *CC BY-SA 4.0*

8: Inland waters - *CC BY-SA 4.0*
- 8.1: Introduction - *CC BY-SA 4.0*
- 8.2: The structure of aquatic ecosystems - *CC BY-SA 4.0*
- 8.3: The special properties of water - *CC BY-NC-SA 4.0*
- 8.4: Inland waters and their watersheds - *CC BY-SA 4.0*
  - 8.4.1: The Physical Template - *Undeclared*
  - 8.4.2: Watershed Structure - *Undeclared*
  - 8.4.3: Watershed Functions - *Undeclared*
  - 8.4.4: Summary - *Undeclared*
- 8.5: Lakes - *CC BY-SA 4.0*
  - 8.5.1: Freshwater Lakes - *Undeclared*
  - 8.5.2: Saline Lakes - *Undeclared*
- 8.6: Streams and Rivers - *CC BY-SA 4.0*
- 8.7: Estuaries- Where the turf meets the surf - *CC BY-SA 4.0*
- 8.8: Waves and Beaches - *Undeclared*
- 8.9: Human impacts on inland waters - *CC BY-SA 4.0*
  - 8.9.1: Fresh Water Supply and the Water Cycle - *CC BY-SA 4.0*
  - 8.9.2: Water Usage - *CC BY-SA 4.0*
  - 8.9.3: Water Scarcity and Solutions - *CC BY-SA 4.0*
  - 8.9.4: Data Dive- Aqueducts of Rome - *CC BY-SA 4.0*
  - 8.9.5: Review - *CC BY-SA 4.0*
- 8.10: Summary - *CC BY-SA 4.0*

9: The Oceans - *CC BY-SA 4.0*
- 9.1: Introduction - *CC BY-NC 4.0*
- 9.2: Properties of Seawater - *CC BY-SA 4.0*
  - 9.2.1: Properties of Seawater - *CC BY-SA 4.0*
  - 9.2.2: Specific Heat and Latent Heat Capacity of Water - *CC BY-SA 4.0*
  - 9.2.3: Energy Associated With Evaporation and Condensation of Water In the Air - *CC BY-SA 4.0*
  - 9.2.4: Salinity - *CC BY-SA 4.0*
  - 9.2.5: Formation of Sea Ice - *CC BY-SA 4.0*
  - 9.2.6: Relationship of Salinity, Density and Temperature - *CC BY-SA 4.0*
  - 9.2.7: Salinity and Latitude - *CC BY-SA 4.0*
  - 9.2.8: "Cline Curves" and Mixing (Surface) Zone - *CC BY-SA 4.0*
  - 9.2.9: Gases Dissolved In Seawater - *CC BY-SA 4.0*
  - 9.2.10: Local Conditions in the San Diego Region - *CC BY-SA 4.0*
  - 9.2.11: Quiz Questions Properties of Seawater - *CC BY-SA 4.0*
- 9.3: Biogeochemistry of the surface ocean - *CC BY-SA 4.0*
  - 9.3.1: Introduction - *CC BY 4.0*
  - 9.3.2: Biological Pump - *Undeclared*
  - 9.3.3: Export from the Upper Ocean to the Depths and Sediments - *Undeclared*
  - 9.3.4: Summary - *Undeclared*
- 9.4: 12.4 Biogeochemistry of the Deep Ocean - *CC BY-NC-SA 4.0*

- 9.5: Biogeochemistry in ocean sediments - *CC BY-SA 4.0*
  - 9.5.1: Introduction - *Undeclared*
  - 9.5.2: Classifying Sediments - *CC BY 4.0*
  - 9.5.3: Lithogenous Sediments - *CC BY 4.0*
  - 9.5.4: Biogenous Sediments - *CC BY 4.0*
  - 9.5.5: Hydrogenous Sediments - *CC BY 4.0*
  - 9.5.6: Cosmogenous Sediments - *CC BY 4.0*
  - 9.5.7: Sediment Distribution - *Undeclared*
- 9.6: Marine element cycles - *CC BY-SA 4.0*
- 9.7: The Oceans and Global Climate Change - *Undeclared*
- 9.8: Summary - *CC BY-SA 4.0*

10: The Global Water Cycle - *CC BY-SA 4.0*
- 10.1: Introduction - *Public Domain*
- 10.2: Models of the hydrologic cycle - *CC BY-SA 4.0*
- 10.3: The history of the water cycle - *CC BY 4.0*
- 10.4: The water cycle and climate change - *CC BY-SA 4.0*
- 10.5: Summary - *CC BY-SA 4.0*

11: The Global Carbon and Oxygen Cycles - *CC BY-SA 4.0*
- 11.1: Introduction - *CC BY-SA 4.0*
- 11.2: What might stimulate net primary production? - *Public Domain*
- 11.3: Temporal perspectives of the carbon cycle - *Public Domain*
- 11.4: Atmospheric methane - *Public Domain*
- 11.5: Carbon monoxide - *CC BY 4.0*
- 11.6: Synthesis- Linking the carbon and oxygen cycles - *CC BY-NC-SA 4.0*
- 11.7: Summary - *CC BY-SA 4.0*

12: The Global Cycles of Nitrogen, Phosphorus and Potassium - *CC BY-SA 4.0*
- 12.1: Introduction - *CC BY-SA 4.0*
- 12.2: The global nitrogen cycle - *CC BY-SA 4.0*
  - 12.2.1: Chemistry of the nitrogen cycle - *CC BY 4.0*
- 12.3: Temporal variations in the global nitrogen cycle - *CC BY 4.0*
- 12.4: Nitrous Oxide (N2O) - *CC BY 4.0*
  - 12.4.1: Terrestrial - *CC BY 4.0*
  - 12.4.2: Marine - *CC BY 4.0*
  - 12.4.3: Human Sources - *CC BY 4.0*
- 12.5: The global phosphorus cycle - *CC BY-SA 4.0*
- 12.6: Potassium- An overlooked limiting element? - *CC BY 4.0*
- 12.7: Linking global biogeochemical cycles - *CC BY-SA 4.0*
- 12.8: Summary - *CC BY-SA 4.0*

13: The Global Cycles of Sulfur and Mercury - *CC BY-SA 4.0*
- 13.1: Introduction - *CC BY-SA 4.0*
- 13.2: The global sulfur cycle - *CC BY-SA 4.0*
- 13.3: Interactions of the sulfur and oxygen cycles - *CC BY-SA 4.0*
- 13.4: Temporal perspectives on the global sulfur cycle - *CC BY-SA 4.0*
- 13.5: The global mercury cycle - *CC BY-SA 4.0*
- 13.6: Summary - *CC BY-SA 4.0*

Back Matter - *CC BY-SA 4.0*
- Index - *CC BY-SA 4.0*
- Glossary - *Undeclared*
- Detailed Licensing - *CC BY-SA 4.0*
- Detailed Licensing - *Undeclared*