# ENVIRONMENTAL SCIENCE



*Kyle Whittinghill*
University of Pittsburgh

LibreTexts™

# CHAPTER OVERVIEW

## Front Matter


https://bio.libretexts.org/@go/page/69342

# Environmental Science (Whittinghill) 3

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 03/04/2026

# TABLE OF CONTENTS


Front Matter

Licensing

## 1: Introduction to Environmental Science

- 1.1: What is Environmental Science?
- 1.2: Introduction to Sustainability
- 1.3: Environmental Ethics and Environmental Justice
- 1.4: Environmental Policy

## 2: The Process of Science

## 3: Matter, Energy, Life

- 3.1: Energy
- 3.2: The Building Blocks of Molecules
- 3.3: A Cell is the Smallest Unit of Life
- 3.4: Water
- 3.5: Themes and Concepts of Biology

## 4: Evolution and the Origin of Species

- 4.1: Understanding Evolution
- 4.2: Formation of New Species

## 5: Cycling of Matter in the Earth System

- 5.1: Introduction - Systems, Cycles, Reservoirs, and Fluxes
- 5.2: Water (Hydrologic) Cycle
- 5.3: Carbon Cycle
- 5.4: Photosynthesis
- 5.5: Nitrogen Cycle
- 5.6: Phosphorus Cycle
- 5.7: Eutrophication and Dead Zones
- 5.8: Sulfur Cycle

## 6: Geology

- 6.1: Introduction to Geology
- 6.2: Fundamentals of Plate Tectonics
- 6.3: The Rock Cycle
- 6.4: Geologic Hazards
  - 6.4.1: Earthquakes
  - 6.4.2: Volcanoes
  - 6.4.3: Mass Wasting
  - 6.4.4: Floods
- 6.5: Geologic Resources
  - 6.5.1: Mineral Resources

- 6.5.2: Mining
- 6.5.3: Fossil Fuels- Formation and Mining
- 6.6: Soils

## 7: Energy

- 7.1: Energy from Fossil Fuels
- 7.2: Nuclear Energy
- 7.3: Solar Energy
- 7.4: Hydropower
- 7.5: Geothermal Energy
- 7.6: Wind Energy
- 7.7: Biofuels (Biomass Energy)
- 7.8: The Electrical Grid, Battery Challenges, and Energy Conservation

## 8: Atmosphere and Air Pollution

- 8.1: The Atmosphere, Climate, and Weather
- 8.2: Earth's Energy Balance
- 8.3: Atmospheric and Ocean Circulation
- 8.4: Air Pollution
- 8.5: Acid Deposition
- 8.6: Stratospheric Ozone Depletion

## 9: Climate Change

- 9.1: What Makes the Climate Change?
- 9.2: Prehistoric Climate Change and Proxys for Studying Past Climates
- 9.3: Anthropogenic Climate Change
- 9.4: Implications of Climate Change

## 10: Urbanization

- 10.1: Urbanization
- 10.2: Environmental Challenges of Urbanization
- 10.3: Sustainable Cities
- 10.4: Case Study- Sustainable Stormwater Management

## 11: Solid Waste Management

- 11.1: Waste Generation
- 11.2: Waste Reduction
- 11.3: Life Cycle Assessment
- 11.4: Waste Disposal
- 11.5: Waste Regulations
- 11.6: Solid Waste and Marine Life
- 11.7: Case Study- Electronic Waste and Extended Producer Responsibility

## 12: Water Supply and Water Pollution

- 12.1: Fresh Water Supply
- 12.2: Case Study - The Aral Sea - Going, Going, Gone
- 12.3: Water Pollutants and Their Sources

- 12.4: Water and Wastewater Treatment
- 12.5: Mitigating Water Pollution

## 13: Agriculture

- 13.1: Food Security
- 13.2: Industrial Agriculture
- 13.3: Pests and Pesticides
- 13.4: Soil Degradation
- 13.5: Selective Breeding and Genetic Engineering
- 13.6: Pollination and Seed Dispersal
- 13.7: Sustainable Agriculture

## 14: Population Ecology

- 14.1: Population Demography
- 14.2: Population Growth and Regulation
- 14.3: Human Population Growth

## 15: Community and Ecosystem Ecology

- 15.1: Community Ecology- Species Interactions
- 15.2: Community Structure
- 15.3: Community Dynamics
- 15.4: Ecology of Ecosystems
- 15.5: Energy Flow Through Ecosystems

## 16: Conservation Biology

- 16.1: Introduction to Biodiversity
- 16.2: Measures of Biodiversity Loss
- 16.3: Threats to Biodiversity
- 16.4: Importance of Biodiversity
- 16.5: Preserving Biodiversity

## Index

## Glossary

## Detailed Licensing

# Licensing

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

# CHAPTER OVERVIEW

## 1: Introduction to Environmental Science


1.1: What is Environmental Science?

1.2: Introduction to Sustainability

1.3: Environmental Ethics and Environmental Justice

1.4: Environmental Policy


---


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

# 1.1: What is Environmental Science?

## What is Environmental Science?

**Environmental science** is the dynamic, interdisciplinary study of the interaction of living and non-living parts of the environment, with special focus on the impact of humans on the environment. The study of environmental science includes circumstances, objects, or conditions by which an organism or community is surrounded and the complex ways in which they interact.

### Environmental Science is Interdisciplinary

Environmental science includes disciplines in the physical sciences (like geology, soil science, physical geography, chemistry, and atmospheric sciences), life sciences (like ecology, conservation biology, restoration biology, and population biology), social sciences (like human geography, economics, law, political science, and anthropology), and humanities (philosophy, ethics). As such, environmental scientists are a diverse bunch. However, they all come together for the focus of studying and identifying past, current, and future environmental issues while also exploring solutions to the environmental issues and also considering practical needs.

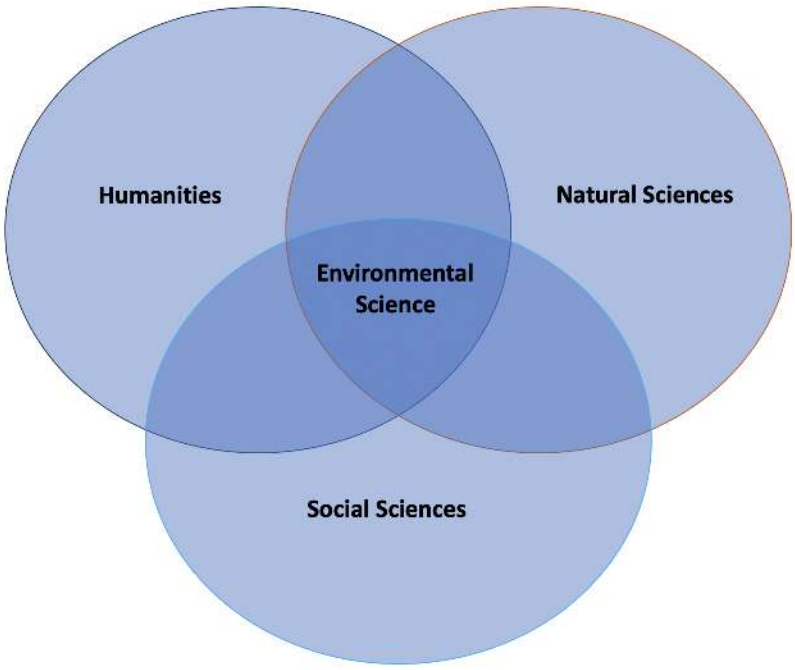



Figure **1.1.*a***: Venn diagram illustrating how environmental science is interdisciplinary between humanities, natural sciences (which includes physical and life sciences) and social sciences. Image by Rachel Schleiger (CC-BY-NC).

### Why Study Environmental Science?

The need for equitable, ethical, and sustainable use of Earth's resources by a global population that nears the carrying capacity of the planet requires us not only to understand how human behaviors affect the environment, but also the scientific principles that govern interactions between the living and non-living. Our future depends on our ability to understand and evaluate evidence-based arguments about the environmental consequences of human actions and technologies, and to make informed decisions based on those arguments.

From global climate change to habitat loss driven by human population growth and development, Earth is becoming a different planet—right before our eyes. The global scale and rate of environmental change are beyond anything in recorded human history. Our challenge is to acquire an improved understanding of Earth's complex environmental systems; systems characterized by

interactions within and among their natural and human components that link local to global and short-term to long-term phenomena, and individual behavior to collective action. The complexity of environmental challenges demands that we all participate in finding and implementing solutions leading to long-term environmental sustainability.

### The Tragedy of the Commons

In his essay, The Tragedy of the Commons, Garrett Hardin (1968) looked at what happens when humans do not limit their actions by including the land as part of their ethic. The **tragedy of the commons** develops in the following way: Imagine a pasture open to all (the 'commons'). It is to be expected that each herder will try to keep as many cattle as possible on the commons. As rational beings, each herdsman seeks to maximize their gain. Adding more cattle increases their profit, and they do not suffer any immediate negative consequence because the commons are shared by all. The rational herder concludes that the only sensible course is to add another animal to their herd, and then another, and so forth. However, this same conclusion is reached by each and every rational herder sharing the commons. Therein lies the tragedy: each person is locked into a system that compels them to increase their herd, without limit, in a world that is limited. Eventually this leads to the degradation of the commons. In a society that believes in the freedom of the commons, freedom brings ruin to all because each person acts selfishly.

Hardin went on to apply the situation to modern commons: overgrazing of public lands, overuse of public forests and parks, depletion of fish populations in the ocean, use of rivers as a common dumping ground for sewage, and releasing air pollutants.

The "Tragedy of the Commons" is applicable to what is arguably the most consequential environmental problem: global climate change. The atmosphere is a commons into which countries are dumping carbon dioxide from the burning of fossil fuels. Although we know that the generation of greenhouse gases will have damaging effects upon the entire globe, we continue to burn fossil fuels. As a country, the immediate benefit from the continued use of fossil fuels is seen as a positive component (because of economic growth). All countries, however, will share the negative long-term effects.

## Some Indicators of Global Environmental Stress

- **Forests** — Deforestation remains a main issue. 1 million hectares of forest were lost every year in the decade 1980-1990. The largest losses of forest area are taking place in the tropical moist deciduous forests, the zone best suited to human settlement and agriculture. Recent estimates suggest that nearly two-thirds of tropical deforestation is due to farmers clearing land for agriculture. There is increasing concern about the decline in forest quality associated with intensive use of forests and unregulated access.
- **Soil** — As much as 10% of the earth's vegetated surface is now at least moderately degraded. Trends in soil quality and management of irrigated land raise serious questions about longer-term sustainability. It is estimated that about 20% of the world's 250 million hectares of irrigated land are already degraded to the point where crop production is seriously reduced.
- **Fresh water** — Some 20% of the world's population lacks access to safe water and 50% lacks access to safe sanitation. If current trends in water use persist, two-thirds of the world's population could be living in countries experiencing moderate or high water stress by 2025.
- **Marine fisheries** — 25% of the world's marine fisheries are being fished at their maximum level of productivity and 35% are overfished (yields are declining). In order to maintain current per capita consumption of fish, global fish harvests must be increased; much of the increase might come through aquaculture which is a known source of water pollution, wetland loss and mangrove swamp destruction.
- **Biodiversity** — Biodiversity is increasingly coming under threat from development, which destroys or degrades natural habitats, and from pollution from a variety of sources. The first comprehensive global assessment of biodiversity put the total number of species at close to 14 million and found that between 1% and 11% of the world's species may be threatened by extinction every decade. Coastal ecosystems, which host a very large proportion of marine species, are at great risk with perhaps one-third of the world's coasts at high potential risk of degradation and another 17% at moderate risk.
- **Atmosphere** — The Intergovernmental Panel on Climate Change has established that human activities are having a discernible influence on global climate. $CO_2$ emissions in most industrialized countries have risen during the past few years and countries generally failed to stabilize their greenhouse gas emissions at 1990 levels by 2000 as required by the Climate Change convention.

- **Toxic chemicals** — About 100,000 chemicals are now in commercial use and their potential impacts on human health and ecological function represent largely unknown risks. Persistent organic pollutants are now so widely distributed by air and ocean currents that they are found in the tissues of people and wildlife everywhere; they are of particular concern because of their high levels of toxicity and persistence in the environment.
- **Hazardous wastes** — Pollution from heavy metals, especially from their use in industry and mining, is also creating serious health consequences in many parts of the world. Incidents and accidents involving uncontrolled radioactive sources continue to increase, and particular risks are posed by the legacy of contaminated areas left from military activities involving nuclear materials.
- **Waste** — Domestic and industrial waste production continues to increase in both absolute and per capita terms, worldwide. In the developed world, per capita waste generation has increased threefold over the past 20 years; in developing countries, it is highly likely that waste generation will double during the next decade. The level of awareness regarding the health and environmental impacts of inadequate waste disposal remains rather poor; poor sanitation and waste management infrastructure is still one of the principal causes of death and disability for the urban poor.

## Attributions

Modified by Melissa Ha and Rachel Schleiger from The Earth, Humans, & the Environment, Environment and Sustainability, and Environmental Ethics from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)

## Contributors and Attributions

- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0. Modified from the original by Matthew R. Fisher.

---

This page titled 1.1: What is Environmental Science? is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **1: Environmental Science Preface** by Melissa Ha and Rachel Schleiger is licensed CC BY-NC 4.0.
- **1.3: Environment and Sustainability** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# 1.2: Introduction to Sustainability

Cities in today's modern world are struggling to balance a sustainable society, economy, and environment. According to a 2016 report that rated sustainability across 32 indicators, Zurich, Switzerland, was ranked number one. Zurich invests in renewable energies, sustainable public transport, public green space, and public education. One of the most notable ways Zurich is leading global sustainability efforts is their dedication to keeping their carbon dioxide emissions low. Figure **1.2.1** below shows the difference in emissions for the United States and Switzerland:

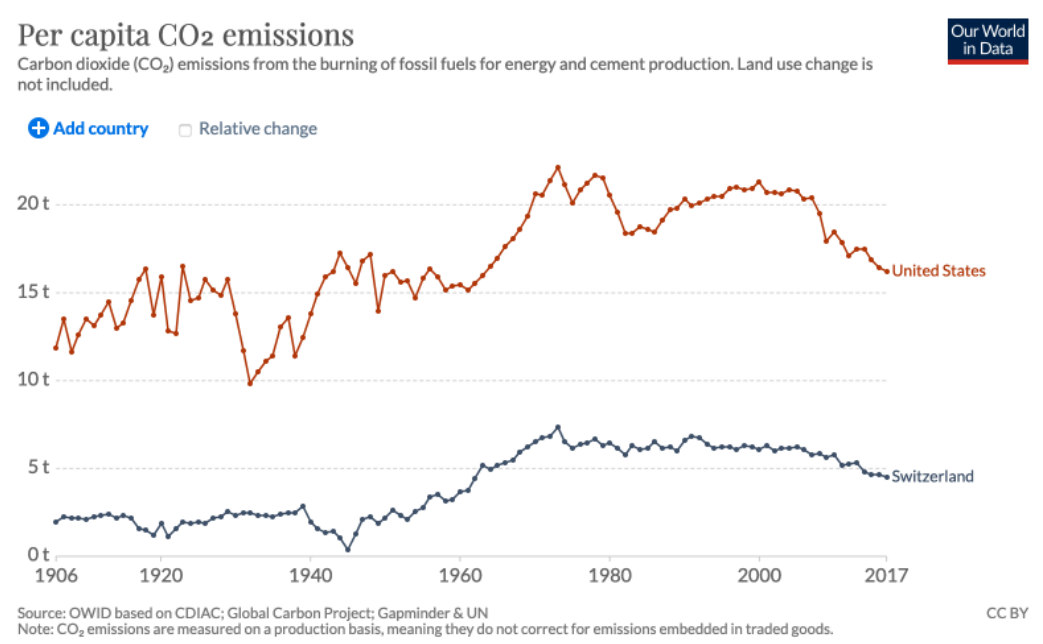



Figure **1.2.1**: Per capita emissions of carbon dioxide between the United States and Switzerland. Graph by Our World in Data (CC-BY4.0)

## Introduction

**Sustainability** is derived from two Latin words: *sus* which means up, and *tenere* which means to hold. Thus, sustainability is essentially about holding up human existence by meeting the needs of the present without compromising the ability of future generations to meet their needs.

There are three dimensions that sustainability seeks to integrate: economic, environmental, and social (including sociopolitical).

- Economic interests define the framework for making decisions, the flow of financial capital, and the facilitation of commerce, including the knowledge, skills, competences and other attributes embodied in individuals that are relevant to economic activity.
- Environmental aspects recognize the diversity and interdependence within living systems, the goods and services produced by the world's ecosystems, and the impacts of human wastes.
- Social/Socio-political refers to interactions between institutions/firms and people, functions expressive of human values, aspirations and well-being, ethical issues, and decision-making that depends upon collective action.

The intersection of social and economic elements can form the basis of social "equitability". In the sense of enlightened management, "viability" is formed through consideration of economic and environmental interests. Between environment and social elements lies "bearability," the recognition that the functioning of societies is dependent on environmental resources and services. At the intersection of all three of these lies sustainability (figure **1.2.2**).

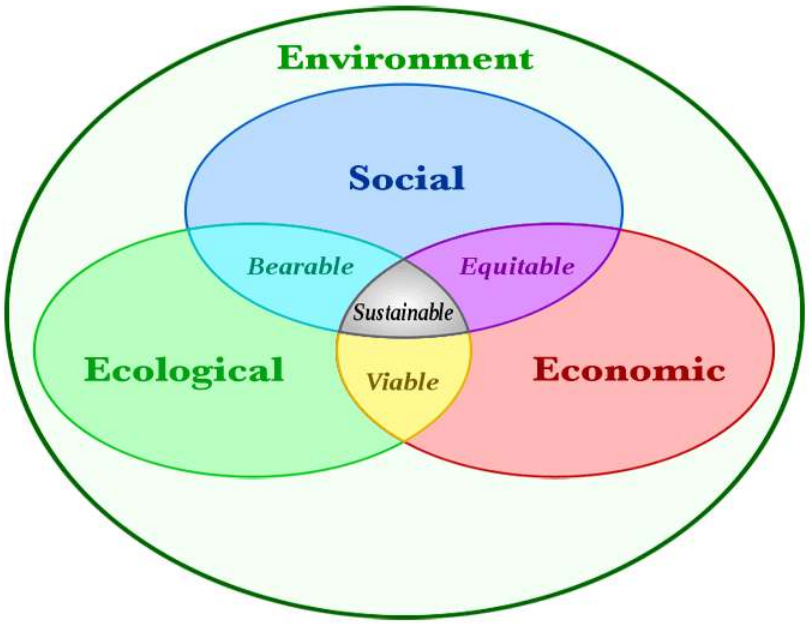



Figure 1.2.2: Venn diagram of defining the components of sustainability. Image by Andrew, Sunray, based on "File:Sustainable development.svg" by Johann Dréo in Wikimedia Commons (CC-BY-SA4.0)

The three main elements of the sustainability paradigm are thought of as equally important, however, tradeoffs occur depending on the local/global objective. For example, in some instances it may be deemed necessary to degrade a particular ecosystem in order to facilitate commerce, or food production, or housing. In reality, the extent to which tradeoffs can be made before irreversible damage results is not always known, and in any case there are definite limits on how much substitution among the three elements is wise (to date, humans have treated economic development as the dominant one of the three). This has led to the notion of strong sustainability, where tradeoffs among natural, human, and social capital are not allowed or are very restricted, and weak sustainability, where tradeoffs are unrestricted or have few limits. Whether or not one follows the strong or weak form of sustainability, it is important to understand that while economic and social systems are human creations, the environment is not. Rather, a functioning environment underpins both society and the economy.

## The Evolution of Sustainability

*Our Common Future* (1987), the report of the World Commission on Environment and Development, is widely credited with having popularized the concept of **sustainable development.** It defines sustainable development in the following ways…

- …development that meets the needs of the present without compromising the ability of future generations to meet their own needs.
- … sustainable development is not a fixed state of harmony, but rather a process of change in which the exploitation of resources, the orientation of the technological development, and institutional change are made consistent with future as well as present needs.

The concept of sustainability, however, can be traced back much farther to the oral histories of indigenous cultures. For example, the principle of inter-generational equity is captured in the Inuit saying, ‘we do not inherit the Earth from our parents, we borrow it from our children’. The Native American ‘Law of the Seventh Generation’ is another illustration. According to this, before any major action was to be undertaken its potential consequences on the seventh generation had to be considered. For a species that at present is only 6,000 generations old and whose current political decision-makers operate on time scales of months or few years at most, the thought that other human cultures have based their decision-making systems on time scales of many decades seems wise but unfortunately inconceivable in the current political climate.

## Environmental Equity

While much progress is being made to improve resource efficiency, far less progress has been made to improve resource distribution. Currently, just one-fifth of the global population is consuming three-quarters of the earth’s resources. If the remaining

four-fifths were to exercise their right to grow to the level of the rich minority it would result in ecological devastation. So far, global income inequalities and lack of purchasing power have prevented poorer countries from reaching the standard of living (and also resource consumption/waste emission) of the industrialized countries.

Countries such as China, Brazil, India, and Malaysia are, however, catching up fast. In such a situation, global consumption of resources and energy needs to be drastically reduced to a point where it can be repeated by future generations. But who will do the reducing? Poorer nations want to produce and consume more. Yet so do richer countries: their economies demand ever greater consumption-based expansion. Such stalemates have prevented any meaningful progress towards equitable and sustainable resource distribution at the international level. These issue of fairness and distributional justice remain unresolved.

## Ecological Footprint

The **ecological footprint** (EF), developed by Canadian ecologist and planner William Rees, is basically an accounting tool that uses land as the unit of measurement to assess per capita consumption, production, and discharge needs. It starts from the assumption that every category of energy and material consumption and waste discharge requires the productive or absorptive capacity of a finite area of land or water. If we (add up) all the land requirements for all categories of consumption and waste discharge by a defined population, the total area represents the Ecological Footprint of that population on Earth whether or not this area coincides with the population's home region.

Land is used as the unit of measurement for the simple reason that, according to Rees, "Land area not only captures planet Earth's finiteness, it can also be seen as a proxy for numerous essential life support functions from gas exchange to nutrient recycling … land supports photosynthesis, the energy conduit for the web of life. Photosynthesis sustains all important food chains and maintains the structural integrity of ecosystems."

**WHAT IS YOUR CARBON FOOTPRINT?** There are many personal calculators available on the internet. Here are a few to try:

- EPA Household Emissions Calculator
- Cool Climate Network (UC Berkeley)
- Carbon Footprint

What does the ecological footprint tell us? Ecological footprint analysis can tell us in a vivid, ready-to-grasp manner how much of the Earth's environmental functions are needed to support human activities. It also makes visible the extent to which consumer lifestyles and behaviors are ecologically sustainable calculated that the ecological footprint of the average American is – conservatively – 5.1 hectares per capita of productive land. With roughly 7.4 billion hectares of the planet's total surface area of 51 billion hectares available for human consumption, if the current global population were to adopt American consumer lifestyles we would need two additional planets to produce the resources, absorb the wastes, and provide general life-support functions.

For more information on ecological footprints see the Living Planet Report by the World Wildlife Fund, Zoological Society of London, and The Global Footprint Network. You can find the 2020 Living Planet Report as well as summaries on the Global Footprint Network website.

## The IPAT Equation

As attractive as the concept of sustainability may be as a means of framing our thoughts and goals, its definition is rather broad and difficult to work with when confronted with choices among specific courses of action. Here we introduce one general way to begin to apply sustainability concepts: the IPAT equation. As is the case for any equation, IPAT expresses a balance among interacting factors. It can be stated as:

$$I=P\times A\times T$$

where $I$ represents the impacts of a given course of action on the environment, $P$ is the relevant human population for the problem at hand, $A$ is the level of consumption per person, and $T$ is impact per unit of consumption. Impact per unit of consumption is a general term for technology, interpreted in its broadest sense as any human-created invention, system, or organization that serves to either worsen or uncouple consumption from impact. The equation is not meant to be mathematically rigorous; rather it provides a way of organizing information for a "first-order" analysis.

## The Precautionary Principle

The **precautionary principle** is an important concept in environmental sustainability. A 1998 consensus statement characterized the precautionary principle this way: "when an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically". For example, if a new pesticide chemical is created, the precautionary principle would dictate that we presume, for the sake of safety, that the chemical may have potential negative consequences for the environment and/or human health, even if such consequences have not been proven yet. In other words, it is best to proceed cautiously in the face of incomplete knowledge about something's potential harm.

## Natural Resources

All natural resources (also known as natural capital) can be divided into two categories: non-renewable and renewable.

### Non-Renewables

Non-renewable resources are present in a finite quantity and do not regenerate after they are harvested and used. Consequently, as non-renewable resources are used, their remaining stocks in the environment are depleted. This means that non-renewable resources can never be used in a sustainable fashion – they can only be "mined" (figure **1.2.3**). Examples of non-renewable resources include metal ores, petroleum, coal, and natural gas. Although continuing exploration may discover additional stocks of non-renewable resources that can be exploited, this does not change the fact that there is a finite quantity of these resources present on Earth. For example, the discovery of a large amount of metal ore in a remote place may substantially increase the known, exploitable reserves of those non-renewable materials. That discovery does not, however, affect the amounts of the metal present on Earth.

Metals are often used to manufacture parts of buildings and machinery. To some degree, the metals can be recovered after these uses and recycled back into the economy, effectively extending the lifespan of their reserves. However, due to the growth and increasing industrialization of the economy, the demand for metals is accelerating. Because recycling cannot keep up with the increasing demands for metals, large additional quantities must be mined from their known reserves in the environment. For valuable metals, such as gold and platinum, there is a high efficiency of recycling, but it is much less so for iron and other less-costly metals.

Fossil fuels are the other major category of non-renewable resources. They are mostly combusted to provide energy for transportation and heating, which converts their organic compounds into carbon dioxide and water, which are released into the environment. Some of that $CO_2$ and $H_2O$ may be absorbed by plants and other photosynthetic organisms and be converted back into organic materials, a process that might be interpreted as being a kind of recycling. However, the rate at which this happens is insignificantly small compared with the release of the $CO_2$ and $H_2O$ by the combustion of fossil fuels, so these materials should be viewed as being as non-renewable as metals are. A more minor use of fossil fuels is to manufacture various kinds of plastics. These synthetic materials can be recycled after initial uses, which does help to extend the lifespan of the reserves of fossil fuels. Nevertheless, because the dominant use of fossil fuels is as sources of energy, they essentially flow through an industrial economy, with little new recycling.

> *Figure* **1.2.3***: Non-renewable resources can only be mined. This is a view of the Etaki open-pit diamond mine in the Northwest Territories. Three open pits can be seen as a cluster, plus another at the top-left of the image, along with an extensive tailings-disposal area and other infrastructure. Source: Jason Pineau, Wikimedia Commons;* http://commons.wikimedia.org/wiki/File:Ekati_mine_640px.jpg .



## Renewables

Renewable resources are capable of regenerating after harvesting, so potentially their stocks can be utilized forever. Most renewable resources are biological (Figure 1.2.4, although some are non-biological. **Biological Renewable Resources** Renewable resources that are biological in nature (bio-resources) include the following:

- wild animals that are hunted as food or for bio-materials, such as deer, moose, hare, ducks, fish, lobster, and seals
- forest biomass that is harvested for lumber, fiber, or energy
- wild plants that are gathered as sources of food
- plants cultivated as sources of food, medicine, materials, or energy
- the organic-based capability of soil to sustain the productivity of agricultural crops

*Figure 1.2.4: Renewable resources, such as timber and fish, are capable of regenerating after they are harvested. Provided they are not over-harvested or managed inappropriately, renewable resources can be harvested in a sustainable fashion. This photo shows a load of timber that was harvested on Vancouver Island. Source: B. Freedman.*



**Non-Biological Renewable Resources** The following are renewable resources that are non-biological:

- sunlight, of which there is a continuous input to Earth
- surface water and groundwater, which are renewed through the hydrologic cycle
- winds, which are renewed through the heat-distribution system of the atmosphere
- water currents and waves, which are renewed through the heat-distribution system of the oceans, as well as the tidal influence of the Moon

Many renewable resources can be managed to increase their rates of recruitment and productivity and to decrease mortality. In the following section we explain how management practices can be used to increase the productivity of biological resources. Although a renewable resource can regenerate after harvesting, it can also be badly degraded by excessive use or by inappropriate management. These practices can damage the ability to regenerate and may ultimately cause a collapse of the stock. If this happens, the renewable resource is being “mined”, or used as if it were a non-renewable resource. As such, it becomes depleted by excessive use. For this reason, ecologists commonly use the qualified term: potentially renewable resources.

## Sustainable Living

**Sustainable living** describes a lifestyle that attempts to reduce/eliminate an individuals (or societies) use of resources so as to be as close as possible to "net zero living". As such, an individual (or society) focuses on reducing their footprint (ecologically, carbon, socially, etc) through their choices/methods of:

- Use of resources (energy/diet/transportation/water/etc)
- Support of people/companies (voting/economic/social/etc)
- Practices of reducing, reusing, and recycling
- Consciousness of priority and focus for local concerns/needs versus global concerns/needs
- Sharing knowledge with their community (all ages)

The thought of getting started with sustainable living can be overwhelming! However, it is important to note that although we can acknowledge all the things that need to be accomplished, that that entire burden does not lie with just one individuals control (Figure 1.2.5). It takes a multi-layerd approach to appropriately take action at the individual, societal, and political levels (Figure 1.2.6).

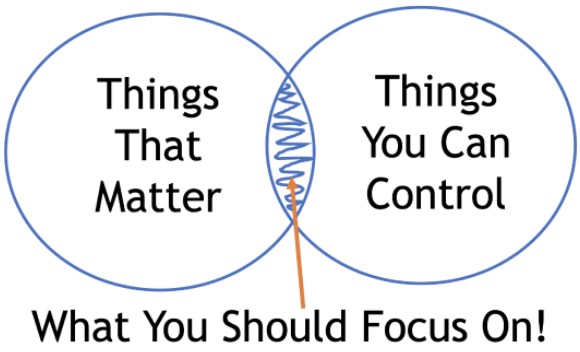



Figure 1.2.5: Focus for individual sustainability lifestyle goals. Image by Rachel Schleiger (CC-BY-NC)




Figure 1.2.6: Considerations for sustainable living goals. Image by Sustainability Week Switzerland in Wikimedia Commons (CC-BY-SA)

## References Cited and Further Reading

Begon, M., R.W. Howorth, and C.R. Townsend. 2014. Essentials of Ecology. 4th ed. Wiley, Cambridge, UK.

Brown, L.R. 2001. Eco-Economy: Building an Economy for the Earth. W.W. Norton and Company, New York, NY.

Brown, L.R. 2003. Plan C: Rescuing a Planet Under Stress and a Civilization in Trouble. W.W. Norton and Company, New York, NY.

Chambers, N., C. Simmons, and N. Wackernagel. 2001. Sharing Nature's Interest: Ecological Footprints as an Indicator of Sustainability. Earthscan Publications, London, UK.

Chiras, D.D. and J.P. Reganold. 2009. Natural Resource Conservation: Management for a Sustainable Future. 10th ed. Prentice Hall, Upper Saddle River, NJ.

Clark, W.C. 1989. Clear-cut economies: Should we harvest everything now? The Sciences, Jan./Feb.: 16-19.

Clark, W.C. and R.E. Munn (eds.). 1986. Sustainable Development of the Biosphere. Cambridge University Press, New York, NY.

Costanza, R. 1991. Ecological Economics: The Science and Management of Sustainability. Columbia University Press, New York, NY.

Costanza, R. and H.E. Daly. 1992. Natural capital and sustainable development. Conservation Biology, 6: 37-46.

Daly, H.E. 1997. Beyond Growth: The Economics of Sustainable Development. Beacon Press, Boston, MA.

Diamond, J.M. 1982. Man the exterminator. Nature, 298: 787-789.

Diamond, J. 2004. Collapse: How Societies Choose to Fail or Succeed. Viking, East Rutherford, NJ.

Ehrlich, P.R. 1989. Facing the habitat crisis. BioScience, 39: 480-482.

Faber, M.M., R. Manstetter, and J. Proops. 1996. Ecological Economics: Concepts and Methods. Edward Elgar Publishers, New York, NY.

Farley, J. and H.E. Daly. 2003. Ecological Economics: Principles and Applications. Island Press, Washington, DC.

Freedman, B. 1995. Environmental Ecology. 2nd ed. Academic Press, San Diego, CA.

Freedman, B. 2010. Environmental Science. A Canadian Perspective. 5th ed. Pearson Education Canada, Toronto, ON.

Freedman, B., J. Hutchings, D. Gwynne, J. Smol, R. Suffling, R. Turkington, R. Walker, and D. Bazeley. 2014. Ecology: A Canadian Context. 2nd ed. Nelson Canada, Toronto, ON.

Hardin, G. 1968. The tragedy of the commons. Science, 162: 1243-1248.

Holechek, J.L., R.A. Cole, J.T. Fisher, and R. Valdez. 2003. Natural Resources: Ecology, Economics, and Policy. 2nd ed. Prentice Hall, East Rutherford, NJ.

Martin, P.S. 1984. Catastrophic extinctions and late Pleistocene blitzkrieg: two radiocarbon tests. pp. 153-589 In: Extinctions (M.H. Nitecki, ed.). University of Chicago Press, Chicago, IL.

Martin, P.S. and H.E. Wright (eds.). 1967. Pleistocene Extinctions: The Search for a Cause. Yale University Press, New Haven, CT.

Meadows, D.H., D.L. Meadows, and J. Randers. 1992. Beyond the Limits: Global Collapse or a Sustainable Future. Earthscan, London, UK.

Meadows, D.H., J. Randers, and D.L. Meadows. 2003. Limits to Growth: The 30-Year Update. Chelsea Green Publishing Co., White River Junction, VT.

Mowat, F. 1984. Sea of Slaughter. McClelland & Stewart, Toronto, ON.

Ponting, C. 1991. A Green History of the World. Penguin, Middlesex, UK.

Rees, W.E. 1990. The ecology of sustainable development. Ecologist, 20(1): 18-23.

Statistics Canada. 2014. Gross domestic product, expenditure-based, provincial and territorial. Cansim Table 384-00384. http://www5.statcan.gc.ca/cansim/a05?lang=eng&id=3840038&pattern=3840038&searchTypeByValue=1&p2=35

Thirgood, J.V. 1981. Man and the Mediterranean Forest: A History of Resource Depletion. Academic Press, New York, NY.

Tietenberg, T. 2005. Environmental and Natural Resource Economics. 9th ed. Addison Wesley Boston, MA.

Tristram, H.B. 1873. The Natural History of the Bible. Society for Promoting Christian Knowledge, London, UK.

White, L. 1967. The historical roots of our ecological crisis. Science, 155: 1203-1207.

World Commission on Environment and Development (WCED). 1987. Our Common Future. Oxford University Press, Oxford, UK.

## Contributors and Attributions

- Modified by Kyle Whittinghill (University of Pittsburgh)
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0. Modified from the original by Matthew R. Fisher.
- Modified by Rachel Schleiger from Sustainability: A Comprehensive Foundation by Openstax (licensed under CC-BY)
- Modified by Melissa Ha and Rachel Schleiger from Environment and Sustainability from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY) and The Myths of Restoration Ecology by Robert H. Hilderbrand, Adam C. Watts and April M. Randle 2005 (CC-BY-NC)
- *Environmental Science: A Canadian perspective* by Bill Freedman Chapter 12: Resources and Sustainable Development
- *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin Section 3.3 The IPAT Equation
- *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin Section 11.3.2 Footprinting: Carbon, Ecological and Water

This page titled 1.2: Introduction to Sustainability is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **1.3: Environment and Sustainability** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# 1.3: Environmental Ethics and Environmental Justice

## Environmental Ethics

The concept of **ethics** involves standards of conduct. These standards help to distinguish between behavior that is considered right and that which is considered wrong. As we all know, it is not always easy to distinguish between right and wrong, as there is no universal code of ethics. For example, a poor farmer clears an area of rainforest in order to grow crops. Some would not oppose this action, because the act allows the farmer to provide a livelihood for his family. Others would oppose the action, claiming that the deforestation will contribute to soil erosion and global warming. Right and wrong are usually determined by an individual's morals, and to change the ethics of an entire society, it is necessary to change the individual ethics of a majority of the people in that society.

**Moral extensionism** defines how far a persons values extend outside of themselves (Figure **1.3.1**). There are many variables that can influence where the limit lies for each individual person. For example: religion, culture, education, and personal interests (just to name a few).




Figure **1.3.1**: Scope of consideration for moral extensionism. Image by Rachel Schleiger (CC-BY-NC).

### Frontier Ethic

The ways in which humans interact with the land and its natural resources are determined by ethical attitudes and behaviors. Early European settlers in North America rapidly consumed the natural resources of the land. After they depleted one area, they moved westward to new frontiers. Their attitude towards the land was that of a **frontier ethic**. A frontier ethic assumes that the earth has an unlimited supply of resources. If resources run out in one area, more can be found elsewhere or alternatively human ingenuity will find substitutes. This attitude sees humans as masters who manage the planet. The frontier ethic is completely anthropocentric (human-centered), for only the needs of humans are considered.

Most industrialized societies experience population and economic growth that are based upon this frontier ethic, assuming that infinite resources exist to support continued growth indefinitely. In fact, economic growth is considered a measure of how well a society is doing. The late economist Julian Simon pointed out that life on earth has never been better, and that population growth means more creative minds to solve future problems and give us an even better standard of living. However, now that the human population has passed seven billion and few frontiers are left, many are beginning to question the frontier ethic. Such people are moving toward an environmental ethic, which includes humans as part of the natural community rather than managers of it. Such an ethic places limits on human activities (e.g., uncontrolled resource use), that may adversely affect the natural community.

Some of those still subscribing to the frontier ethic suggest that outer space may be the new frontier. If we run out of resources (or space) on earth, they argue, we can simply populate other planets. This seems an unlikely solution, as even the most aggressive colonization plan would be incapable of transferring people to extraterrestrial colonies at a significant rate. Natural population growth on earth would outpace the colonization effort. A more likely scenario would be that space could provide the resources (e.g. from asteroid mining) that might help to sustain human existence on earth.

### Sustainable Ethic

A **sustainable ethic** is an environmental ethic by which people treat the earth as if its resources are limited. This ethic assumes that the earth's resources are not unlimited and that humans must use and conserve resources in a manner that allows their continued use in the future. A sustainable ethic also assumes that humans are a part of the natural environment and that we suffer when the health of a natural ecosystem is impaired. A sustainable ethic includes the following tenets:

- The earth has a limited supply of resources.
- Humans must conserve resources.

- Humans share the earth's resources with other living things.
- Growth is not sustainable.
- Humans are a part of nature.
- Humans are affected by natural laws.
- Humans succeed best when they maintain the integrity of natural processes sand cooperate with nature.

For example, if a fuel shortage occurs, how can the problem be solved in a way that is consistent with a sustainable ethic? The solutions might include finding new ways to conserve oil or developing renewable energy alternatives. A sustainable ethic attitude in the face of such a problem would be that if drilling for oil damages the ecosystem, then that damage will affect the human population as well. A sustainable ethic can be either anthropocentric or biocentric (life-centered). An advocate for conserving oil resources may consider all oil resources as the property of humans. Using oil resources wisely so that future generations have access to them is an attitude consistent with an anthropocentric ethic. Using resources wisely to prevent ecological damage is in accord with a biocentric ethic.

## Land Ethic

Aldo Leopold, an American wildlife natural historian and philosopher, advocated a biocentric ethic in his book, *A Sand County Almanac*. He suggested that humans had always considered land as property, just as ancient Greeks considered slaves as property. He believed that mistreatment of land (or of slaves) makes little economic or moral sense, much as today the concept of slavery is considered immoral. All humans are merely one component of an ethical framework. Leopold suggested that land be included in an ethical framework, calling this the land ethic.

"The **land ethic** simply enlarges the boundary of the community to include soils, waters, plants and animals; or collectively, the land. In short, a land ethic changes the role of *Homo sapiens* from conqueror of the land-community to plain member and citizen of it. It implies respect for his fellow members, and also respect for the community as such." (Aldo Leopold, 1949)

Leopold divided conservationists into two groups: one group that regards the soil as a commodity and the other that regards the land as biota, with a broad interpretation of its function. If we apply this idea to the field of forestry, the first group of conservationists would grow trees like cabbages, while the second group would strive to maintain a natural ecosystem. Leopold maintained that the conservation movement must be based upon more than just economic necessity. Species with no discernible economic value to humans may be an integral part of a functioning ecosystem. The land ethic respects all parts of the natural world regardless of their utility, and decisions based upon that ethic result in more stable biological communities.

"Anything is right when it tends to preserve the integrity, stability and beauty of the biotic community. It is wrong when it tends to do otherwise." (Aldo Leopold, 1949)

## Hetch Hetchy Valley

In 1913, the Hetch Hetchy Valley – located in Yosemite National Park in California – was the site of a conflict between two factions, one with an anthropocentric ethic and the other, a biocentric ethic. As the last American frontiers were settled, the rate of forest destruction started to concern the public.



Figure **1.3.2**. Yosemite valley, California, USA.

The conservation movement gained momentum, but quickly broke into two factions. One faction, led by Gifford Pinchot, Chief Forester under Teddy Roosevelt, advocated utilitarian conservation (i.e., conservation of resources for the good of the public). The

other faction, led by John Muir, advocated preservation of forests and other wilderness for their inherent value. Both groups rejected the first tenet of frontier ethics, the assumption that resources are limitless. However, the conservationists agreed with the rest of the tenets of frontier ethics, while the preservationists agreed with the tenets of the sustainable ethic.

The Hetch Hetchy Valley was part of a protected National Park, but after the devastating fires of the 1906 San Francisco earthquake, residents of San Francisco wanted to dam the valley to provide their city with a stable supply of water. Gifford Pinchot favored the dam.

> *"As to my attitude regarding the proposed use of Hetch Hetchy by the city of San Francisco…I am fully persuaded that… the injury…by substituting a lake for the present swampy floor of the valley…is altogether unimportant compared with the benefits to be derived from it's use as a reservoir.*
>
> *"The fundamental principle of the whole conservation policy is that of use, to take every part of the land and its resources and put it to that use in which it will serve the most people." (Gifford Pinchot, 1913)*

John Muir, the founder of the Sierra Club and a great lover of wilderness, led the fight against the dam. He saw wilderness as having an intrinsic value, separate from its utilitarian value to people. He advocated preservation of wild places for their inherent beauty and for the sake of the creatures that live there. The issue aroused the American public, who were becoming increasingly alarmed at the growth of cities and the destruction of the landscape for the sake of commercial enterprises. Key senators received thousands of letters of protest.

> *"These temple destroyers, devotees of ravaging commercialism, seem to have a perfect contempt for Nature, and instead of lifting their eyes to the God of the Mountains, lift them to the Almighty Dollar." (John Muir, 1912)*

Despite public protest, Congress voted to dam the valley. The preservationists lost the fight for the Hetch Hetchy Valley, but their questioning of traditional American values had some lasting effects. In 1916, Congress passed the "National Park System Organic Act," which declared that parks were to be maintained in a manner that left them unimpaired for future generations. As we use our public lands, we continue to debate whether we should be guided by preservationism or conservationism.

## Environmental Justice

**Environmental Justice** is defined as the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies. It will be achieved when everyone enjoys the same degree of protection from environmental and health hazards and equal access to the decision-making process to have a healthy environment in which to live, learn, and work.

Whenever a community is faced with the potential of an environmentally undesirable facility, such as the placement of a hazardous waste dump in its midst, the usual response from residents is: "Not in my back yard!" Such a response is known as the NIMBY principle. Such reactions are usually reactions to visions of previous environmental irresponsibility: uncontrolled dumping of noxious industrial wastes and rusty steel drums oozing hazardous chemicals into the environment. Such occurrences were all too real in the past and some are still taking place. It is now possible -- and much more common -- to build environmentally sound, state-of-the-art disposal facilities. However, the NIMBY principle usually prevents the construction of such new facilities. Instead, hazardous waste facilities tend to be built upon pre-existing, already contaminated sites, even though the geology of such locations may be less favorable for containment than potential new sites.

During the 1980's minority groups protested that hazardous waste sites were preferentially sited in minority neighborhoods. In 1987, Benjamin Chavis of the United Church of Christ Commission for Racism and Justice coined the term **environmental racism** to describe such a practice. The charges generally failed to consider whether the facility or the demography of the area came first. Most hazardous waste sites are located on property that was used as disposal sites long before modern facilities and disposal methods were available. Areas around such sites are typically depressed economically, often as a result of past disposal activities.

Persons with low incomes are often constrained to live in such undesirable, but affordable, areas. The problem more likely resulted from one of insensitivity rather than racism. Indeed, the ethnic makeup of potential disposal facilities was most likely not considered when the sites were chosen.

Decisions in citing hazardous waste facilities are generally made on the basis of economics, geological suitability and the political climate. For example, a site must have a soil type and geological profile that prevents hazardous materials from moving into local aquifers. The cost of land is also an important consideration. The high cost of buying land would make it economically unfeasible to build a hazardous waste site in Beverly Hills. Some communities have seen a hazardous waste facility as a way of improving their local economy and quality of life. Emelle County, Alabama had illiteracy and infant mortality rates that were among the highest in the nation. A landfill constructed there provided jobs and revenue that ultimately helped to reduce both figures.

In an ideal world, there would be no hazardous waste facilities, but we do not live in an ideal world. Unfortunately, we live in a world plagued by rampant pollution and dumping of hazardous waste. Our industrialized society has necessarily produced wastes during the manufacture of products for our basic needs. Until technology can find a way to manage (or eliminate) hazardous waste, disposal facilities will be necessary to protect both humans and the environment. By the same token, this problem must be addressed. Industry and society must become more socially sensitive in the selection of future hazardous waste sites. All humans who help produce hazardous wastes must share the burden of dealing with those wastes, not just the poor and minorities.

## Indigenous People

Since the end of the 15th century, most of the world's frontiers have been claimed and colonized by established nations. Invariably, these conquered frontiers were home to people indigenous to those regions. Some were wiped out or assimilated by the invaders, while others survived while trying to maintain their unique cultures and way of life. The United Nations officially classifies **indigenous people** as those "having an historical continuity with pre-invasion and pre-colonial societies," and "consider themselves distinct from other sectors of the societies now prevailing in those territories or parts of them." Furthermore, indigenous people are "determined to preserve, develop and transmit to future generations, their ancestral territories, and their ethnic identity, as the basis of their continued existence as peoples in accordance with their own cultural patterns, social institutions and legal systems." A few of the many groups of indigenous people around the world are: the many tribes of Native Americans (i.e., Navajo, Sioux) in the contiguous 48 states, the Inuit of the arctic region from Siberia to Canada, the rainforest tribes in Brazil, and the Ainu of northern Japan.

Many problems face indigenous people including the lack of human rights, exploitation of their traditional lands and themselves, and degradation of their culture. In response to the problems faced by these people, the United Nations proclaimed an "International Decade of the World's Indigenous People" beginning in 1994. The main objective of this proclamation, according to the United Nations, is "the strengthening of international cooperation for the solution of problems faced by indigenous people in such areas as human rights, the environment, development, health, culture and education." Its major goal is to protect the rights of indigenous people. Such protection would enable them to retain their cultural identity, such as their language and social customs, while participating in the political, economic and social activities of the region in which they reside.

Despite the lofty U.N. goals, the rights and feelings of indigenous people are often ignored or minimized, even by supposedly culturally sensitive developed countries. In the United States many of those in the federal government are pushing to exploit oil resources in the Arctic National Wildlife Refuge on the northern coast of Alaska. The "Gwich'in," an indigenous people who rely culturally and spiritually on the herds of caribou that live in the region, claim that drilling in the region would devastate their way of life. Thousands of years of culture would be destroyed for a few months' supply of oil. Drilling efforts have been stymied in the past, but mostly out of concern for environmental factors and not necessarily the needs of the indigenous people. Curiously, another group of indigenous people, the "Inupiat Eskimo," favor oil drilling in the Arctic National Wildlife Refuge. Because they own considerable amounts of land adjacent to the refuge, they would potentially reap economic benefits from the development of the region.



Figure **1.3.3**. An Inupiaq woman, Nome, Alaska, c. 1907. Credit: This work is in the Public Domain, CC0

The heart of most environmental conflicts faced by governments usually involves what constitutes proper and sustainable levels of development. For many indigenous peoples, sustainable development constitutes an integrated wholeness, where no single action is separate from others. They believe that sustainable development requires the maintenance and continuity of life, from generation to generation and that humans are not isolated entities, but are part of larger communities, which include the seas, rivers, mountains, trees, fish, animals and ancestral spirits. These, along with the sun, moon and cosmos, constitute a whole. From the point of view of indigenous people, sustainable development is a process that must integrate spiritual, cultural, economic, social, political, territorial and philosophical ideals.

## Contributors and Attributions

- Kyle Whittinghill (University of Pittsburgh)
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0. Modified from the original by Matthew R. Fisher.
- Modified by Rachel Schleiger from Sustainability: A Comprehensive Foundation by Openstax (licensed under CC-BY)

---

This page titled 1.3: Environmental Ethics and Environmental Justice is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Matthew R. Fisher (OpenOregon) via source content that was edited to the style and standards of the LibreTexts platform.

- **1.4: Environmental Ethics** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.
- **1.5: Environmental Justice and Indigenous Struggles** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# 1.4: Environmental Policy

## Introduction

It is not uncommon to think of the sustainability paradigm as being a recent interpretation of environmental policy, one that was given credence by the United Nations report "Our Common Future" (the Brundtland Report) when it was first presented in 1987. However, it would be a mistake to conclude that sustainability as a mental construct and policy framework for envisioning the relationship of humans and nature came into being suddenly and at a single moment in time. Most environmental historians who have studied U.S. policy have discerned at least three distinct periods during which new concepts and ideas, scientific understandings, technological advances, political institutions, and laws and regulations came or were brought into being in order to understand and manage human impacts on the environment.

1. **The American conservation movement:** This 19th century movement had the main objective for conservation of American land and resources, it grew from opposite philosophies. On one side you had people like Gifford Pinchot (first head of the US National Forest Service) that were very pragmatic about their desire for conservation. For example, these people observed that policies for harvesting some resource (like lumber) were not sustainable, so they began conservation/sustainability research and new policies so the resource would not run out in the foreseeable future. On the other side were people like John Muir and Theodore Roosevelt that saw land/habitats/wildlife as intrinsic entities, meaning that had value in their own right (meaning nothing to do with money).
2. **The rise of environmental risk management as a basis for policy:** The beginnings of environmental risk management can be traced to the fields of public health, industrial hygiene, and sanitary engineering, which came into prominence in the latter decades of the 19$^{th}$ century and beginning of the 20$^{th}$. The spread of disease was a particularly troublesome problem as the country continued to urbanize. In addition, environmental scientists of the day were alarmed by the extent and degree of damage that they were documenting. The publication of *Silent Spring* in 1962 by Rachel Carson (1907-1964), about the impact of the widespread and indiscriminate use of pesticides, was a watershed moment, bringing environmental concerns before a large portion of the American, and global, public. Carson, collected scientifically documented evidence on the effects of pesticides, particularly DDT, heptachlor, and dieldrin, on humans and mammals, and the systemic disruption they caused to ecosystems. *Silent Spring* is credited with bringing about a ban on the use of DDT in the United States, and setting in motion a chain of events that would ultimately result in the transformation of environmental public policy from one based on the problems and attitudes that brought about nineteenth century conservation, to one based on the management of risks from chemical toxins. The U.S. Environmental Protection Agency was established in 1970, just eight years after the publication of *Silent Spring*. The same year Earth Day was created.
3. **The integration of social and economic factors to create what we now refer to as the sustainability paradigm:** Eventually, we (humanity) realized that in order to solve environmental issues (like ones identified by Rachel Carson), that the "human factors" (both social and economical) had to be accounted for. This is what defines the sustainability paradigm: environmental, social, and economic disciplines. Only Band-Aid solutions can be applied if we only take a one-minded approach.

## History of Environmental Policy

Environmental policies are typically driven by problems of the day, real and perceived, that require systemic solutions. For example:

- Early conservationists were alarmed at the inefficiencies of human resource management and the encroachment of humans on unspoiled lands... Leading to conservation legislation.
- During the 20$^{th}$ century many groups (scientists, economists, politicians, and ordinary citizens) became alarmed and fearful of the consequences of toxic pollutant loads to the environment that included localized effects on human health and well-being... Leading to clean air and water legislation.
- As we proceed into the 21$^{st}$ century, an array of complex problems that have the potential to alter substantially the structure and well-being of large segments of human societies, calls for a renewal and reassessment of our approach to environmental policy.

This has, thus far, proven to be a difficult transition. Many of these complex problems have multiple causes and impacts, affect some groups of people more than others, are economically demanding, and are often not as visibly apparent to casual observers as previous impacts, nor are the benefits perceived to be commensurate with costs. Devising a regulatory strategy for such problems requires an adaptive and flexible approach that current laws do not foster.

**Table 1.4.*a*:** Table illustrating evolution of environmental policy. Table by Sustainability: A Comprehensive Foundation in Openstax (CC-BY).

| | 1850-1920 | 1960-1990 | 1990-Present |
|---|---|---|---|
| **Focus** | Conservation/sanitation | Media/site/problem specific | Complex regional/ global problems |
| **Outcome** | Land preservation/efficiency/control of disease | Manage anthropocentric and ecological risk | Global sustainable development |
| **Principal Activity** | Resource management reform/simple contaminant controls | Compliance/remediation/technological emphasis on problem solving | Integration of social, economic, and technological information for holistic problem solving |
| **Economic Focus** | Profit maximization/public health | Cost minimization | Strategic investments/long-term societal well-being |
| **Regulatory Activity** | Low | Heavy | Adaptive and flexible |
| **Conceptual Model** | Expansion vs preservation | Command and control | Systems/life cycle approach |
| **Disciplinary Approach** | Disciplinary and insular | Multidisciplinary | Interdisciplinary/integrative |

**Table 1.4.*b*:** Table illustrating major environmental legislation, society, and organization conceptions. Table by Rachel Schleiger (CC-BY-NC).

| Year | Policy/Society/Organization Conception |
|---|---|
| **1899** | Refuse Act |
| **1905** | National Forest Service |
| **1916** | National Park Service |
| **1918** | Migratory Bird Treaty Act |
| **1930's** | State Parks |
| **1948** | Federal Water Pollution Control Act |
| **1950** | The Nature Conservancy |
| **1955** | Air Pollution Control Act |
| **1961** | World Wildlife Federation |
| **1963** | Clean Air Act (1970, 1977, 1990 updates) |
| **1965** | Solid Waste Disposal Act (1976 update) |
| **1965** | Water Quality Act (1987 update) |
| **1967** | Air Quality Act |
| **1969** | National Environmental Policy Act |
| **1970** | Occupational Safety and Health Act |
| **1970** | Environmental Protection Agency |
| **1971** | Occupational Safety and Health Administration |
| **1972** | Consumer Product Safety Act |
| **1972** | Federal Insecticide, Fungicide, and Rodenticide Act |
| **1972** | Clean Water Act (1977 update) |
| **1972** | Noise Control Act |
| **1973** | Endangered Species Act |
| **1974** | Safe Drinking Water Act (1986, 1996 updates) |
| **1975** | Hazardous Materials Transportation Act |
| **1976** | Resource Conservation and Recovery Act |
| **1976** | Toxic Substances Control Act |
| **1980** | CERCLA (Superfund) |
| **1984** | Hazardous and Solid Waste Amendments |
| **1986** | Superfund Amendments and Reauthorization Act |
| **1986** | Emergency Wetlands Resources Act |
| **1990** | Oil Pollution Act |
| **1993** | North American Free Trade Agreement |
| **2003** | Healthy Forest Initiative |

## Attribution

Modified by Rachel Schleiger from Sustainability: A Comprehensive Foundation by Openstax (CC-BY).

---

This page titled 1.4: Environmental Policy is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **22.4: Environmental Policy** by Melissa Ha and Rachel Schleiger has no license indicated.

# 2: The Process of Science

Like other natural sciences, environmental science gathers knowledge about the natural world. Science is more than just a body of knowledge, science provides a means to evaluate and create new knowledge. The discoveries of environmental science are made by a community of researchers who work individually and together using agreed-on methods. In this sense, science is a social enterprise like politics or the arts.




Figure **2.1**: Formerly called blue-green algae, the (a) cyanobacteria seen through a light microscope are some of Earth's oldest life forms. These (b) stromatolites along the shores of Lake Thetis in Western Australia are ancient structures formed by the layering of cyanobacteria in shallow waters. (credit a: modification of work by NASA; scale-bar data from Matt Russell; credit b: modification of work by Ruth Ellison)

The methods of science include careful observation, record keeping, logical and mathematical reasoning, experimentation, and submitting conclusions to the scrutiny of others. Science also requires considerable imagination and creativity; a well-designed experiment is commonly described as elegant, or beautiful. Science has considerable practical implications and some science is dedicated to practical applications, such as the prevention of disease (Figure **2.2**). Other science proceeds largely motivated by curiosity. Whatever its goal, there is no doubt that science, including biology, has transformed human existence and will continue to do so.

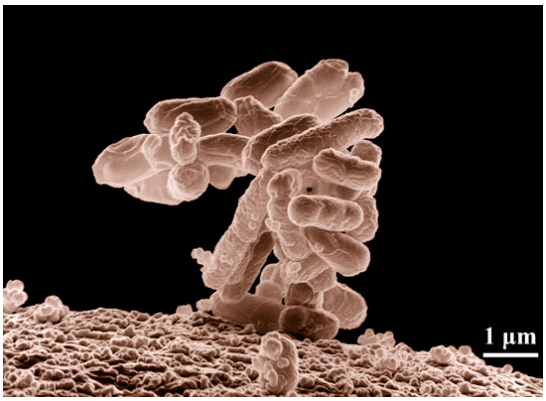



Figure **2.2**: Biologists may choose to study *Escherichia coli* (*E. coli*), a bacterium that is a normal resident of our digestive tracts but which is also sometimes responsible for disease outbreaks. In this micrograph, the bacterium is visualized using a scanning electron microscope and digital colorization. (credit: Eric Erbe; digital colorization by Christopher Pooley, USDA-ARS)

Science is a very specific way of learning, or knowing, about the world. The history of the past 500 years demonstrates that science is a very powerful way of knowing about the world; it is largely responsible for the technological revolutions that have taken place during this time. There are however, areas of knowledge and human experience that the methods of science cannot be applied to. These include such things as answering purely moral questions, aesthetic questions, or what can be generally categorized as spiritual questions. Science has cannot investigate these areas because they are outside the realm of material phenomena, the phenomena of matter and energy, and cannot be observed and measured.

The scientific method is a method of research with defined steps that include experiments and careful observation. The steps of the scientific method will be examined in detail later, but one of the most important aspects of this method is the testing of hypotheses. A hypothesis is a suggested explanation for an event, which can be tested. Hypotheses, or tentative explanations, are generally

produced within the context of a scientific theory. A scientific theory is a generally accepted, thoroughly tested and confirmed explanation for a set of observations or phenomena. Scientific theory is the foundation of scientific knowledge. In addition, in many scientific disciplines (less so in biology) there are scientific laws, often expressed in mathematical formulas, which describe how elements of nature will behave under certain specific conditions. There is not an evolution of hypotheses through theories to laws as if they represented some increase in certainty about the world. Hypotheses are the day-to-day material that scientists work with and they are developed within the context of theories. Laws are concise descriptions of parts of the world that are amenable to formulaic or mathematical description.

### Natural Sciences

What would you expect to see in a museum of natural sciences? Frogs? Plants? Dinosaur skeletons? Exhibits about how the brain functions? A planetarium? Gems and minerals? Or maybe all of the above? Science includes such diverse fields as astronomy, biology, computer sciences, geology, logic, physics, chemistry, and mathematics (Figure **2.3**). However, those fields of science related to the physical world and its phenomena and processes are considered natural sciences. Thus, a museum of natural sciences might contain any of the items listed above.

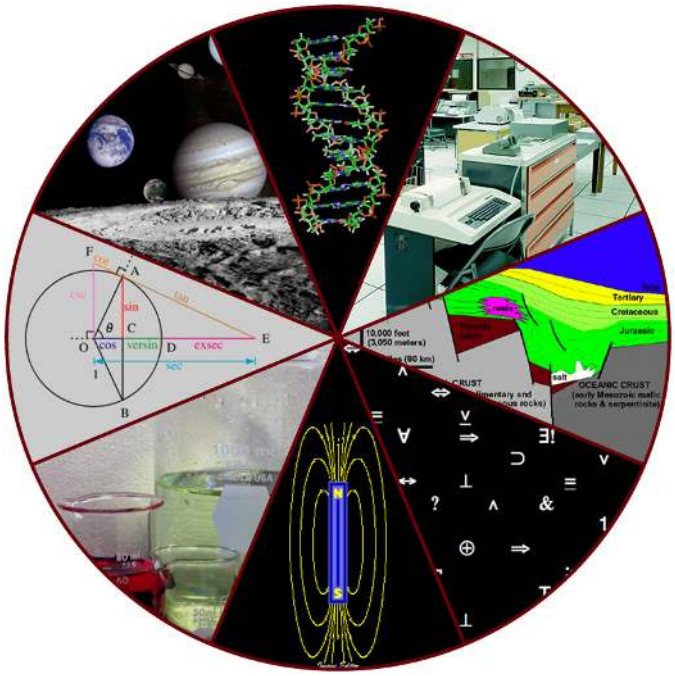



Figure **2.3**: Some fields of science include astronomy, biology, computer science, geology, logic, physics, chemistry, and mathematics. (credit: "Image Editor"/Flickr)

There is no complete agreement when it comes to defining what the natural sciences include. For some experts, the natural sciences are astronomy, biology, chemistry, earth science, and physics. Other scholars choose to divide natural sciences into life sciences, which study living things and include biology, and physical sciences, which study nonliving matter and include astronomy, physics, and chemistry. Some disciplines such as biophysics and biochemistry build on two sciences and are interdisciplinary.

## Evidence, Measurements, and Observations

Scientists use objective evidence over subjective evidence, to reach sound and logical conclusions. An **objective** observation is without personal bias and the same by all individuals. **Bias** refers to favoring one thing over another, and it can lead to inaccurate results. Humans are biased by nature, so they cannot be completely objective; the goal is to be as unbiased as possible. A **subjective** observation is based on a person's feelings and beliefs and is unique to that individual (figure **2.4**).



Figure 2.4: This is Grand Canyon of the Yellowstone in Yellowstone National Park. An objective statement about this would be, “The picture is of a waterfall.” A subjective statement would be, “The picture is beautiful.”

Another way scientists avoid bias is by using quantitative over qualitative measurements whenever possible. A **quantitative measurement** is expressed with a specific numerical value. **Qualitative observations** are general or relative descriptions. For example, describing a rock as red or heavy is a qualitative observation. Determining a rock’s color by measuring wavelengths of reflected light or its density by measuring the proportions of minerals it contains is quantitative. Numerical values are more precise than general descriptions, and they can be analyzed using statistical calculations. This is why quantitative measurements are much more useful to scientists than qualitative observations.

## Inductive and Deductive Reasoning

One thing is common to all forms of science: an ultimate goal “to know.” Curiosity and inquiry are the driving forces for the development of science. Scientists seek to understand the world and the way it operates. Two methods of logical thinking are used: **inductive reasoning** and **deductive reasoning**.

**Inductive reasoning** is a form of logical thinking that uses related observations to arrive at a general conclusion. This type of reasoning is common in **descriptive science.** A life scientist such as a biologist makes observations and records them. These data can be qualitative (descriptive) or quantitative (consisting of numbers), and the raw data can be supplemented with drawings, pictures, photos, or videos. From many observations, the scientist can infer conclusions (inductions) based on evidence. Inductive reasoning involves formulating generalizations inferred from careful observation and the analysis of a large amount of data. Surveying land use (which areas are forested, agricultural, urban, etc.) across the United States and then concluding that forested areas are concentrated in the West is an example of descriptive science.

**Deductive reasoning** or deduction is the type of logic used in hypothesis-based science. In deductive reasoning, the pattern of thinking moves in the opposite direction as compared to inductive reasoning. Deductive reasoning is a form of logical thinking that uses a general principle or law to forecast specific results. From those general principles, a scientist can extrapolate and predict the specific results that would be valid as long as the general principles are valid. For example, a prediction would be that if the climate is becoming warmer in a region, the distribution of plants and animals should change. Comparisons have been made between distributions in the past and the present, and the many changes that have been found are consistent with a warming climate. Finding the change in distribution is evidence that the climate change conclusion is a valid one. **Deductive reasoning**, or deduction, is the type of logic used in hypothesis-based science (see below).

To summarize, inductive reasoning moves from the specific (an observation) to general (conclusion), and deductive reasoning moves from the general (a hypothesis or principle) to the specific (results).

Both types of logical thinking are related to the two main pathways of scientific study: **descriptive science** and **hypothesis-based science**. **Descriptive** (or discovery) **science** aims to observe, explore, and discover, while **hypothesis-based science** begins with a specific question or problem and a potential answer or solution that can be tested. The boundary between these two forms of study is often blurred, because most scientific endeavors combine both approaches. Observations lead to questions, questions lead to forming a hypothesis as a possible answer to those questions, and then the hypothesis is tested. Thus, descriptive science and hypothesis-based science are in continuous dialogue.

## Hypothesis Testing and The scientific Method

The **scientific method** is a process of research with defined steps that include data collection and careful observation. The scientific method was used even in ancient times, but it was first documented by England's Sir Francis Bacon (1561–1626) (Figure **2.5**), who set up inductive methods for scientific inquiry.



Figure **2.5**: Sir Francis Bacon is credited with being the first to document the scientific method.

### Observation

Scientific advances begin with **observations**. This involves noticing a pattern, either directly or indirectly from the literature. An example of a direct observation is noticing that there have been a lot of toads in your yard ever since you turned on the sprinklers, where as an indirect observation would be reading a scientific study reporting high densities of toads in urban areas with watered lawns.

During the Vietnam War (figure **2.6**), press reports from North Vietnam documented an increasing rate of birth defects. While this credibility of this information was initially questioned by the U.S., it evoked questions about what could be causing these birth defects. Furthermore, increased incidence of certain cancers and other diseases later emerged in Vietnam veterans who had returned to the U.S. This leads us to the next step of the scientific method, the question.

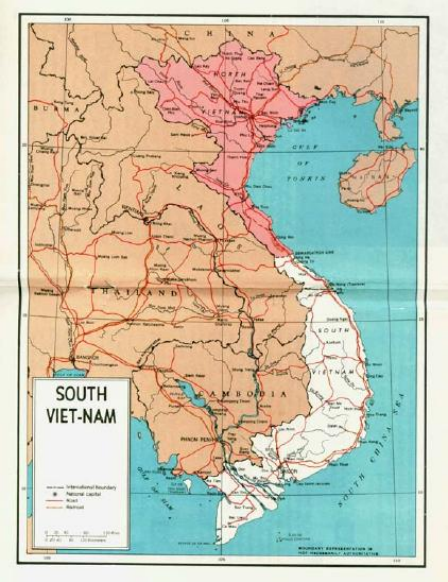



Figure 2.6: A map of Vietnam 1954-1975. Image from Bureau of Public Affairs U.S. Government Printing Office (public domain).

## Question

The **question** step of the scientific method is simply asking, what explains the observed pattern? Multiple questions can stem from a single observation. Scientists and the public began to ask, what is causing the birth defects in Vietnam and diseases in Vietnam veterans? Could it be associated with the widespread military use of the herbicide Agent Orange to clear the forests (figure 2.7 – 8), which helped identify enemies more easily?



Figure 2.7: Agent Orange drums in Vietnam. Image by U.S. Government (public domain).



Figure 2.8: A healthy mangrove forest (top), and another forest after application of Agent Orange. Image by unknown author (public domain).

## Hypothesis and Prediction

The **hypothesis** is the expected answer to the question. The best hypotheses state the proposed direction of the effect (increases, decreases, etc.) and explain why the hypothesis could be true.

- OK hypothesis: Agent Orange influences rates of birth defects and disease.
- Better hypothesis: Agent Orange increases the incidence of birth defects and disease.
- Best hypothesis: Agent Orange increases the incidence of birth defects and disease because these health problems have been frequently reported by individuals exposed to this herbicide.

If two or more hypotheses meet this standard, the simpler one is preferred.

**Predictions** stem from the hypothesis. The prediction explains what results would support hypothesis. The prediction is more specific than the hypothesis because it references the details of the experiment. For example, "If Agent Orange causes health problems, then mice experimentally exposed to TCDD, a contaminant of Agent Orange, during development will have more frequent birth defects than control mice" (figure 2.9).

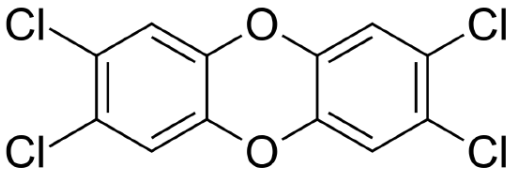



Figure 2.9: The chemical structure of TCDD (2,3,7,8-tetrachlorodibenzo-p-dioxin), which is produced when synthesizing the chemicals in Agent Orange. It contaminates Agent Orange at low but harmful concentrations. Image by Emeldir (public domain).

Hypotheses and predictions must be testable to ensure that it is valid. For example, a hypothesis that depends on what a bear thinks is not testable, because it can never be known what a bear thinks. It should also be **falsifiable**, meaning that they have the capacity to be tested and demonstrated to be untrue. An example of an unfalsifiable hypothesis is “Botticelli’s *Birth of Venus* is beautiful.” There is no experiment that might show this statement to be false. To test a hypothesis, a researcher will conduct one or more experiments designed to eliminate one or more of the hypotheses. This is important. A hypothesis can be disproven, or eliminated, but it can never be proven. Science does not deal in proofs like mathematics. If an experiment fails to disprove a hypothesis, then we find support for that explanation, but this is not to say that down the road a better explanation will not be found, or a more carefully designed experiment will be found to falsify the hypothesis.

Hypotheses are tentative explanations and are different from scientific theories. A **scientific theory** is a widely-accepted, thoroughly tested and confirmed explanation for a set of observations or phenomena. Scientific theory is the foundation of scientific knowledge. In addition, in many scientific disciplines (less so in biology) there are **scientific laws**, often expressed in mathematical formulas, which describe how elements of nature will behave under certain specific conditions, but they do not offer explanations for why they occur.

## Design an Experiment

Next, a scientific study (experiment) is planned to test the hypothesis and determine whether the results match the predictions. Each experiment will have one or more variables. The **explanatory variable** is what scientists hypothesize might be causing something else. In a manipulative experiment (see below), the explanatory variable is manipulated by the scientist. The **response variable** is the response, the variable ultimately measured in the study. **Controlled variables** (confounding factors) might affect the response variable, but they are not the focus of the study. Scientist attempt to standardize the controlled variables so that they do not influence the results. In our previous example, exposure to Agent Orange is the explanatory variable. It is hypothesized to cause a change in health (likelihood of having children with birth defects or developing a disease), the response variable. Many other things could affect health, including diet, exercise, and family history. These are the controlled variables.

There are two main types of scientific studies: experimental studies (manipulative experiments) and observational studies.

In a **manipulative experiment**, the explanatory variable is altered by the scientists, who then observe the response. In other words, the scientists apply a **treatment**. An example would be exposing developing mice to TCDD and comparing the rate of birth defects to a control group. The **control group** is group of test subjects that are as similar as possible to all other test subjects, with the exception that they don't receive the experimental treatment (those that do receive it are known as the experimental, treatment, or **test group**). The purpose of the control group is to establish what the dependent variable would be under normal conditions, in the absence of the experimental treatment. It serves as a baseline to which the test group can be compared. In this example, the control group would contain mice that were not exposed to TCDD but were otherwise handled the same way as the other mice (figure 2.10)



Figure 2.10: Laboratory mice. In a proper scientific study, the treatment would be applied to multiple mice. Another group of mice would not receive the treatment (the control group). Image by Aaron Logan (CC-BY).

In an **observational study**, scientists examine multiple samples with and without the presumed cause. An example would be monitoring the health of veterans who had varying levels of exposure to Agent Orange.

Scientific studies contain many **replicates.** Multiple samples ensure that any observed pattern is due to the treatment rather than naturally occurring differences between individuals. A scientific study should also be **repeatable**, meaning that if it is conducted again, following the same procedure, it should reproduce the same general results. Additionally, multiple studies will ultimately test the same hypothesis.

### Results

Finally, the data are collected and the results are analyzed. As described in the Math Blast chapter, statistics can be used to describe the data and summarize data. They also provide a criterion for deciding whether the pattern in the data is strong enough to support the hypothesis.

The manipulative experiment in our example found that mice exposed to high levels of 2,4,5-T (a component of Agent Orange) or TCDD (a contaminant found in Agent Orange) during development had a cleft palate birth defect more frequently than control mice (figure 2.11). Mice embryos were also more likely to die when exposed to TCDD compared to controls.



**Figure 2.11:** Cleft lip and palate, a birth defect in which these structures are split. Image by James Heilman, MD (CC-BY-SA).

An observational study found that self-reported exposure to Agent Orange was positively correlated with incidence of multiple diseases in Korean veterans of the Vietnam War, including various cancers, diseases of the cardiovascular and nervous systems, skin diseases, and psychological disorders. Note that a **positive correlation** simply means that the independent and dependent variables both increase or decrease together, but further data, such as the evidence provided by manipulative experiments is needed to document a **cause-and-effect relationship**. (A **negative correlation** occurs when one variable increases as the other decreases.)

### Conclusion

Lastly, scientists make a conclusion regarding whether the data support the hypothesis. In the case of Agent Orange, the data, that mice exposed to TCDD and 2,4,5-T had higher frequencies of cleft palate, matches the prediction. Additionally, veterans exposed to Agent Orange had higher rates of certain diseases, further supporting the hypothesis. We can thus accept the hypothesis that Agent Orange increases the incidence of birth defects and disease.

### Scientific Method in Practice

In practice, the scientific method is not as rigid and structured as it might first appear. Sometimes an experiment leads to conclusions that favor a change in approach; often, an experiment brings entirely new scientific questions to the puzzle. Many times, science does not operate in a linear fashion; instead, scientists continually draw inferences and make generalizations, finding patterns as their research proceeds (figure **2.12**). Even if the hypothesis was supported, scientists may still continue to test it in different ways. For example, scientists explore the impacts of Agent Orange, examining long-term health impacts as Vietnam veterans age.

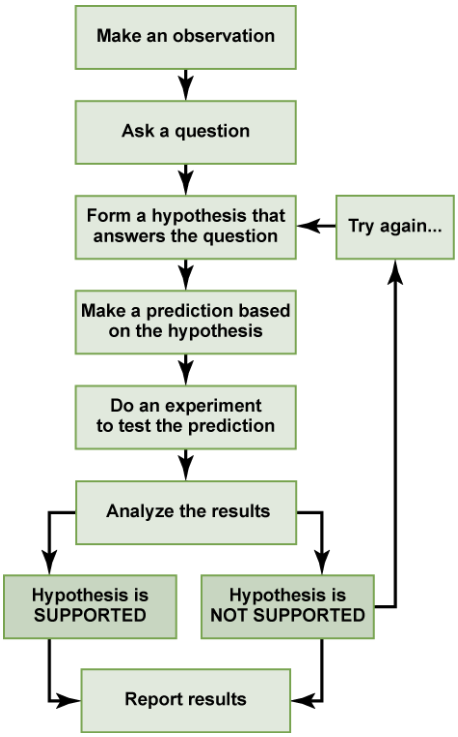



Figure 2.12: The scientific method is a series of defined steps that include experiments and careful observation. The steps are as follows: make an observation; ask a question; form a hypothesis that answers the question; make a prediction based on the hypothesis; do an experiment to test the prediction; analyze the results; and report the results. Whether the hypothesis is supported or not supported, the results are still reported. If a hypothesis is not supported by data, a new hypothesis can be proposed.

Scientific findings can influence decision making. In response to evidence regarding the effect of Agent Orange on human health, compensation is now available for Vietnam veterans who were exposed to Agent Orange and develop certain diseases. The use of Agent Orange is also banned in the U.S. Finally, the U.S. has began cleaning sites in Vietnam that are still contaminated with TCDD.

As another simple example, an experiment might be conducted to test the hypothesis that phosphate limits the growth of algae in freshwater ponds. A series of artificial ponds are filled with water and half of them are treated by adding phosphate each week, while the other half are treated by adding a salt that is known not to be used by algae. The variable here is the phosphate (or lack of phosphate), the experimental or treatment cases are the ponds with added phosphate and the control ponds are those with something inert added, such as the salt. Just adding something is also a control against the possibility that adding extra matter to the pond has an effect. If the treated ponds show lesser growth of algae, then we have found support for our hypothesis. If they do not, then we reject our hypothesis. Be aware that rejecting one hypothesis does not determine whether or not the other hypotheses can be accepted; it simply eliminates one hypothesis that is not valid (Figure 2.12). Using the scientific method, the hypotheses that are inconsistent with experimental data are rejected.

## Science is also a Social Process

Scientists must share their findings for other researchers to expand and build upon their discoveries. Communication and collaboration within and between sub disciplines of science are key to the advancement of knowledge in science. For this reason, an important aspect of a scientist's work is disseminating results and communicating with peers. Scientists share their ideas with peers at conferences, seeking guidance and feedback (figure 2.13), but this approach can reach only the limited few who are present.



Figure 2.13: Scientists share information by publishing and attending conferences. The conference shown here focuses on peanut and mycotoxin (harmful chemicals produced by fungi) research. Image by Sharon Dowdy (CC-BY-NC).

## Peer Review

Most scientists present their results in peer-reviewed articles that are published in scientific journals (figure 2.14). **Peer-reviewed articles** are scientific papers that are reviewed, usually anonymously by a scientist's colleagues, or peers. These colleagues are qualified individuals, often experts in the same research area, who judge whether or not the scientist's work is suitable for publication. The process of peer review helps to ensure that the research described in a scientific paper or grant proposal is original, significant, logical, and thorough. Grant proposals, which are requests for research funding, are also subject to peer review. Scientists publish their work so other scientists can reproduce their experiments under similar or different conditions to expand on the findings. The experimental results must be consistent with the findings of other scientists.




Figure 2.14: The findings of scientific studies are published in peer-reviewed scientific journals. Image by free svg (public domain).

There are many journals and the popular press that do not use a peer-review system. A large number of online open-access journals, journals with articles available without cost, are now available many of which use rigorous peer-review systems, but some of which do not. Results of any studies published in these forums without peer review are not reliable and should not form the basis for other scientific work. In one exception, journals may allow a researcher to cite a personal communication from another researcher about unpublished results with the cited author's permission.

The scientific review process aims to weed out misinformation, invalid research results, and wild speculation. Thus, it is slow, cautious, and conservative. Scientists tend to wait until a hypothesis is supported by an overwhelming amount of evidence from many independent researchers before accepting it as a **scientific theory**.

As you review scientific information, whether in an academic setting or as part of your day-to-day life, it is important to think about the credibility of that information. You might ask yourself: has this scientific information been through the rigorous process

of peer review? Are the conclusions based on available data and accepted by the larger scientific community? Scientists are inherently skeptical, especially if conclusions are not supported by evidence (and you should be too).

### Building on the Work of Others

Only rarely does a scientific discovery spring full-blown on the scene. When it does, it is likely to create a revolution in the way scientists perceive the world around them and to open up new areas of scientific investigation. Darwin's theory of evolution and Mendel's rules of inheritance are examples of such revolutionary developments. Most science, however, consists of adding another brick to an edifice that has been slowly and painstakingly constructed by prior work.

The development of a new technique often lays the foundation for rapid advances along many different scientific avenues. Just consider the advances in biology that discovery of the light microscope and, later, the electron microscope have made possible. Throughout these pages, there are many examples of experimental procedures. Each was developed to solve a particular problem. However, each was then taken up by workers in other laboratories and applied to their problems.

In a similar way, the creation of a new explanation (hypothesis) in a scientific field often stimulates workers in related fields to reexamine their own field in the light of the new ideas. Darwin's theory of evolution, for example, has had an enormous impact on virtually every subspecialty in biology as well as environmental science. To this very day, scientists in specialties as different as biochemistry and conservation biology are guided in their work by evolutionary theory (figure 2.15).



Figure 2.15: Understanding the endangered mountain yellow-legged frog through the lens of evolution and genetics has aided its conservation. Image by Iasaac Chellman/NPS (CC-BY).

## Structure of Scientific Papers

Scientific papers are usually divided into several sections (not necessarily in this order).

**Summary or Abstract:** This section includes only the essence of the other sections. It should be as brief as possible, telling the reader what the goal of the experiment was, what was found, and the significance of the findings. The abstract is often placed at the beginning of the paper rather than at its end.

**Introduction:** This section of the paper describes the scientific question or problem that was the subject of the investigation. The introduction also includes references to earlier reports of these and other scientists that have served as the foundation for the present work. Finally, the introduction states the hypothesis.

**Materials and Methods:** Here are precisely described the materials used (e.g., strains of organism, source of the reagents) and all the methods followed. The goal of this section is to give all the details necessary for workers in other laboratories to be able to **repeat the experiments exactly**. When many complex procedures are involved, it is acceptable to refer to earlier papers describing these methods in greater detail.

**Results:** Here the authors report what happened in their experiments. This report is usually supplemented with graphs, tables, and photographs.

**Discussion:** Here the authors point out what they think is the significance of their findings. This is the place to show that the results are compatible with certain hypotheses and less compatible, or even incompatible, with others. If the results contradict the results of similar experiments in other laboratories, the discrepancies are noted here, and an attempt may be made to reconcile the differences.

**Acknowledgments:** In this brief but important section, the authors give credit to those who have assisted them in the work. These usually include technicians (who may have actually performed most of the experiments!) and other scientists who donated materials for the experiments and/or gave advice about them.

**References:** This section gives a careful listing of all earlier scientific work referred to in the main body of the paper. Most of the references are to other scientific papers. Each reference should provide enough information so that another person can locate the document. This means that each reference should include the name(s) of the author(s), the journal or book in which the report appears, and the year of publication. In the case of scientific journals, the volume number in which the paper appears and the page number on which the paper begins should be included. A digital object identifier (DOI) or url is often included. The full title of the article is often included as well, although some citation styles omit the title from the reference.

## Basic and Applied Science

The scientific community has been debating for the last few decades about the value of different types of science. Is it valuable to pursue science for the sake of simply gaining knowledge, or does scientific knowledge only have worth if we can apply it to solving a specific problem or bettering our lives? This question focuses on the differences between two types of science: **basic science** and **applied science**.

**Basic science** or "pure" science seeks to expand knowledge regardless of the short-term application of that knowledge. It is not focused on developing a product or a service of immediate public or commercial value. The immediate goal of basic science is knowledge for knowledge's sake, though this does not mean that in the end it may not result in an application.

In contrast, **applied science** or "technology," aims to use science to solve real-world problems, making it possible, for example, to improve a crop yield, find a cure for a particular disease, or save animals threatened by a natural disaster. In applied science, the problem is usually defined for the researcher.

Some individuals may perceive applied science as "useful" and basic science as "useless." A question these people might pose to a scientist advocating knowledge acquisition would be, "What for?" A careful look at the history of science, however, reveals that basic knowledge has resulted in many remarkable applications of great value. Many scientists think that a basic understanding of science is necessary before an application is developed; therefore, applied science relies on the results generated through basic science. Other scientists think that it is time to move on from basic science and instead to find solutions to actual problems. Both approaches are valid. It is true that there are problems that demand immediate attention; however, few solutions would be found without the help of the knowledge generated through basic science.

One example of how basic and applied science can work together to solve practical problems occurred after the discovery of DNA structure led to an understanding of the molecular mechanisms governing DNA replication. Strands of DNA, unique in every human, are found in our cells, where they provide the instructions necessary for life. During DNA replication, new copies of DNA are made, shortly before a cell divides to form new cells. Understanding the mechanisms of DNA replication enabled scientists to develop laboratory techniques that are now used to identify genetic diseases, pinpoint individuals who were at a crime scene, and determine paternity. Without basic science, it is unlikely that applied science would exist.

Another example of the link between basic and applied research is the Human Genome Project, a study in which each human chromosome was analyzed and mapped to determine the precise sequence of DNA subunits and the exact location of each gene. (The gene is the basic unit of heredity; an individual's complete collection of genes is his or her genome.) Other organisms have also been studied as part of this project to gain a better understanding of human chromosomes. The Human Genome Project (Figure **2.16**) relied on basic research carried out with non-human organisms and, later, with the human genome. An important end goal eventually became using the data for applied research seeking cures for genetically related diseases.

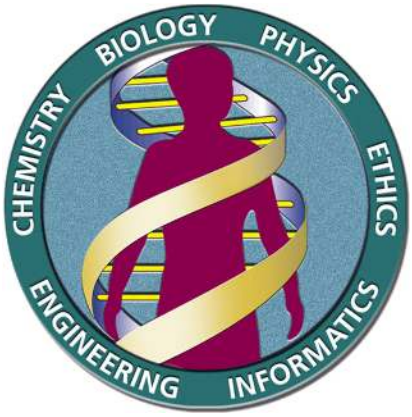



Figure **2.16**: The Human Genome Project was a 13-year collaborative effort among researchers working in several different fields of science. The project was completed in 2003. (credit: the U.S. Department of Energy Genome Programs)

While research efforts in both basic science and applied science are usually carefully planned, it is important to note that some discoveries are made by serendipity, that is, by means of a fortunate accident or a lucky surprise. Penicillin was discovered when biologist Alexander Fleming accidentally left a petri dish of *Staphylococcus* bacteria open. An unwanted mold grew, killing the bacteria. The mold turned out to be *Penicillium*, and a new antibiotic was discovered. Even in the highly organized world of science, luck—when combined with an observant, curious mind—can lead to unexpected breakthroughs.

## Predictability

A few phenomena are considered to have a universal character and are consistent wherever and whenever they are accurately measured. One of the best examples of such a universal constant is the speed of light, which always has a value of $2.998 \times 10^8$ meters per second, regardless of where it is measured or of the speed of the body from which the light is emitted. Similarly, certain relationships describing transformations of energy and matter, known as the laws of thermodynamics, always give reliable predictions.

However, most natural phenomena are not so consistent—depending on circumstances, there are exceptions to general predictions about them. This circumstance is particularly true of biology and ecology, related fields of science in which almost all general predictions have exceptions. In fact, laws or unifying principles of biology or ecology have not yet been discovered, in contrast to the several esteemed laws and 11 universal constants of physics. For this reason, biologists and ecologists have great difficulties making accurate predictions about the responses of organisms and ecosystems to environmental change. This is why biologists and ecologists are sometimes said to have “physics envy.”

In large part, the inaccuracies of biology and ecology occur because key functions are controlled by complexes of poorly understood, and sometimes unidentified, environmental influences. Consequently, predictions about future values of biological and ecological variables or the causes of changes are seldom accurate. For example, even though ecologists in eastern Canada have been monitoring the population size of spruce budworm (an important pest of conifer forests) for some years, they cannot accurately predict its future abundance in particular stands of forest or in larger regions. This is because the abundance of this moth is influenced by a complex of environmental factors, including tree-species composition, age of the forest, abundance of its predators and parasites, quantities of its preferred foods, weather at critical times of year, and insecticide use to reduce its populations. Biologists and ecologists do not fully understand this complexity, and perhaps they never will.

## Variability

Many natural phenomena are highly variable in space and time. This is true of physical and chemical variables as well as of biological and ecological ones. Within a forest, for example, the amount of sunlight reaching the ground varies greatly with time, depending on the hour of the day and the season of the year. It also varies spatially, depending on the density of foliage over any place where sunlight is being measured. Similarly, the density of a particular species of fish within a river typically varies in response to changes in habitat conditions and other influences. Most fish populations also vary over time, especially migratory species such as salmon. In environmental science, replicated (or independently repeated) measurements and statistical analyses are used to measure and account for these kinds of temporal and spatial variations.

## Accuracy and Precision

Accuracy refers to the degree to which a measurement or observation reflects the actual, or true, value of the subject. For example, the insecticide DDT and the metal mercury are potentially toxic chemicals that occur in trace concentrations in all organisms, but their small residues are difficult to analyze chemically. Some of the analytical methods used to determine the concentrations of DDT and mercury are more accurate than others and therefore provide relatively useful and reliable data compared with less accurate methods. In fact, analytical data are usually approximations of the real values – rigorous accuracy is rarely attainable.

Precision is related to the degree of repeatability of a measurement or observation. For example, suppose that the actual number of caribou in a migrating herd is 10,246 animals. A wildlife ecologist might estimate that there were about 10,000 animals in that herd, which for practical purposes is a reasonably accurate reckoning of the actual number of caribou. If other ecologists also independently estimate the size of the herd at about 10,000 caribou, there is a good degree of precision among the values. If, however, some systematic bias existed in the methodology used to count the herd, giving consistent estimates of 15,000 animals (remember, the actual population is 10 246 caribou), these estimates would be considered precise, but not particularly accurate.

Precision is also related to the number of digits with which data are reported. If you were using a flexible tape to measure the lengths of 10 large, wriggly snakes, you would probably measure the reptiles only to the nearest centimetre. The strength and squirminess of the animals make more precise measurements impossible. The reported average length of the 10 snakes should reflect the original measurements and might be given as 204 cm and not a value such as 203.8759 cm. The latter number might be displayed as a digital average by a calculator or computer, but it is unrealistically precise.

Significant figures are related to accuracy and precision and can be defined as the number of digits used to report data from analyses or calculations (see also Appendix A). Significant figures are most easily understood by examples. The number 179 has three significant figures, as does the number 0.0849 and also 0.000794 (the zeros preceding the significant integers do not count). However, the number 195,000,000 has nine significant figures (the zeros following are meaningful), although the number $195 \times 10^6$ has only three significant figures.

It is rarely useful to report environmental or ecological data to more than 2-4 significant figures. This is because any more would generally exceed the accuracy and precision of the methodology used in the estimation and would therefore be unrealistic. For example, the approximate population of Canada in 2015 was 35.1 million people (or $35.1 \times 10^6$; both of these notations have three significant figures). However, the population should not be reported as 33,100,000, which implies an unrealistic accuracy and precision of eight significant figures.

## References


Institute of Medicine (US) Committee to Review the Health Effects in Vietnam Veterans of Exposure to Herbicides. Veterans and Agent Orange: Health Effects of Herbicides Used in Vietnam. Washington (DC): National Academies Press (US); 1994. 2, History of the Controversy Over the Use of Herbicides.

Neubert, D., Dillmann, I. Embryotoxic effects in mice treated with 2,4,5-trichlorophenoxyacetic acid and 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **272,** 243–264 (1972).

Stellman, J. M., & Stellman, S. D. (2018). Agent Orange During the Vietnam War: The Lingering Issue of Its Civilian and Military Health Impact. *American journal of public health*, *108*(6), 726–728.

Yi, S. W., Ohrr, H., Hong, J. S., & Yi, J. J. (2013). Agent Orange exposure and prevalence of self-reported diseases in Korean Vietnam veterans. *Journal of preventive medicine and public health = Yebang Uihakhoe chi*, *46*(5), 213–225.

American Association for the Advancement of Science (AAAS). 1990. Science for All Americans. AAAS, Washington, DC.

Barnes, B. 1985. About Science. Blackwell Ltd ,London, UK.

Giere, R.N. 2005. Understanding Scientific Reasoning. 5th ed. Wadsworth Publishing, New York, NY.

Kuhn, T.S. 1996. The Structure of Scientific Revolutions. 3rd ed. University of Chicago Press, Chicago, IL.

McCain, G. and E.M. Siegal. 1982. The Game of Science. Holbrook Press Inc., Boston, MA.

Moore, J.A. 1999. Science as a Way of Knowing. Harvard University Press, Boston, MA.

Popper, K. 1979. Objective Knowledge: An Evolutionary Approach. Clarendon Press, Oxford, UK.

Raven, P.H., G.B. Johnson, K.A. Mason, and J. Losos. 2013. Biology. 10th ed. McGraw-Hill, Columbus, OH.

Silver, B.L. 2000. The Ascent of Science. Oxford University Press, Oxford, UK.

## Contributors and Attributions

- Modified by Kyle Whittinghill (University of Pittsburgh)
- 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).
- Modified by Melissa Ha
- **1.2: The Process of Science** by OpenStax, is licensed CC BY
- What is Science? from *An Introduction to Geology* by Chris Johnson et al. (licensed under CC-BY-NC-SA)
- The Process of Science from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Scientific Methods from *Biology* by John W. Kimball (licensed under CC-BY)
- Scientific Papers from *Biology* by John W. Kimball (CC-BY)
- *Environmental Science: A Canadian perspective* by Bill Freedman Chapter 2: Science as a Way of Understanding the Natural World

---

This page titled 2: The Process of Science is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **1.2: The Process of Science** by OpenStax is licensed CC BY 4.0.

# CHAPTER OVERVIEW

## 3: Matter, Energy, Life


3.1: Energy

3.2: The Building Blocks of Molecules

3.3: A Cell is the Smallest Unit of Life

3.4: Water

3.5: Themes and Concepts of Biology


---


3: Matter, Energy, Life is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 3.1: Energy

Virtually every task performed by living organisms requires energy. Nutrients and other molecules are imported into the cell to meet these energy demands. For example, energy is required for the synthesis and breakdown of molecules, as well as the transport of molecules into and out of cells. In addition, processes such as ingesting and breaking down food, exporting wastes and toxins, and movement of the cell all require energy.

Scientists use the term **bioenergetics** to describe the concept of energy flow through living systems, such as cells. Cellular processes such as the building and breaking down of complex molecules occur through step-wise chemical reactions. Some of these chemical reactions are spontaneous and release energy, whereas others require energy to proceed. Together, all of the chemical reactions that take place inside cells, including those that consume or generate energy, are referred to as the cell's **metabolism**.

From where, and in what form, does this energy come? How do living cells obtain energy, and how do they use it? This section will discuss different forms of energy and the physical laws that govern energy transfer.

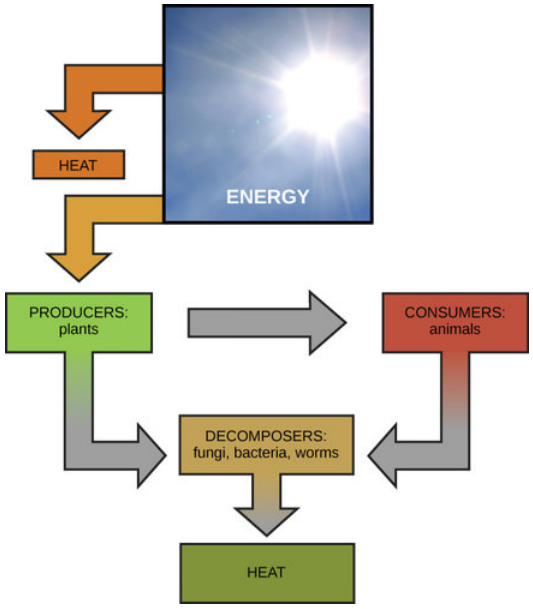



Figure **3.1.1**. Ultimately, most life forms get their energy from the sun. Plants use photosynthesis to capture sunlight, and herbivores eat the plants to obtain energy. Carnivores eat the herbivores, and eventual decomposition of plant and animal material contributes to the nutrient pool.

## Energy

**Thermodynamics** refers to the study of energy and energy transfer involving physical matter. The matter relevant to a particular case of energy transfer is called a system, and everything outside of that matter is called the surroundings. For instance, when heating a pot of water on the stove, the system includes the stove, the pot, and the water. Energy is transferred within the system (between the stove, pot, and water). There are two types of systems: open and closed. In an **open system**, energy can be exchanged with its surroundings. The stovetop system is open because heat can be lost to the air. A **closed system** cannot exchange energy with its surroundings.

Biological organisms are open systems. Energy is exchanged between them and their surroundings as they use energy from the sun to perform photosynthesis or consume energy-storing molecules and release energy to the environment by doing work and releasing heat. Like all things in the physical world, energy is subject to physical laws. The laws of thermodynamics govern the transfer of energy in and among all systems in the universe. In general, **energy** is defined as the ability to do work, or to create some kind of change. Energy exists in different forms: electrical energy, light energy, mechanical energy, and heat energy are all different types of energy. To appreciate the way energy flows into and out of biological systems, it is important to understand two of the physical laws that govern energy.

The **first law of thermodynamics** states that the total amount of energy in the universe is constant and conserved. In other words, there has always been, and always will be, exactly the same amount of energy in the universe. Energy exists in many different forms. According to the first law of thermodynamics, energy may be transferred from place to place or transformed into different forms, but it cannot be created or destroyed. The transfers and transformations of energy take place around us all the time. Light bulbs transform electrical energy into light and heat energy. Gas stoves transform chemical energy from natural gas into heat energy. Plants perform one of the most biologically useful energy transformations on earth: that of converting the energy of sunlight to chemical energy stored within organic molecules (Figure 3.1.2 below).

The challenge for all living organisms is to obtain energy from their surroundings in forms that are usable to perform cellular work. Cells have evolved to meet this challenge. Chemical energy stored within organic molecules such as sugars and fats is transferred and transformed through a series of cellular chemical reactions into energy within molecules of ATP (adenosine triphosphate). Energy in ATP molecules is easily accessible to do work. Examples of the types of work that cells need to do include building complex molecules, transporting materials, powering the motion of cilia or flagella, and contracting muscles to create movement.

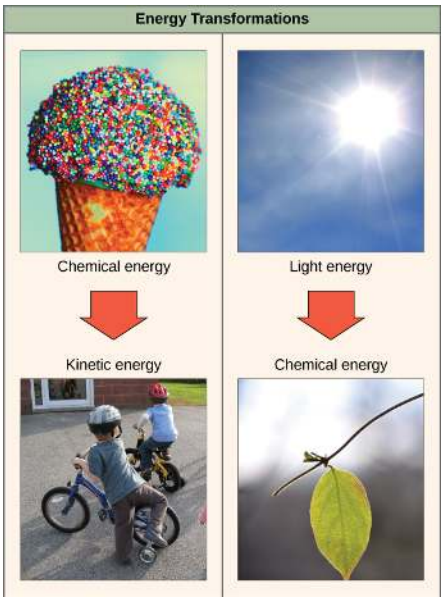



Figure 3.1.2. Shown are some examples of energy transferred and transformed from one system to another and from one form to another. The food we consume provides our cells with the energy required to carry out bodily functions, just as light energy provides plants with the means to create the chemical energy they need. (credit “ice cream”: modification of work by D. Sharon Pruitt; credit “kids”: modification of work by Max from Providence; credit “leaf”: modification of work by Cory Zanker)

A living cell’s primary tasks of obtaining, transforming, and using energy to do work may seem simple. However, the **second law of thermodynamics** explains why these tasks are harder than they appear. All energy transfers and transformations are never completely efficient. In every energy transfer, some amount of energy is lost in a form that is unusable. In most cases, this form is heat energy.

Thermodynamically, **heat energy** is defined as the energy transferred from one system to another that is not work. For example, when a light bulb is turned on, some of the energy being converted from electrical energy into light energy is lost as heat energy. Likewise, some energy is lost as heat energy during cellular metabolic reactions.

An important concept in physical systems is that of order and disorder. The more energy that is lost by a system to its surroundings, the less ordered and more random the system is. Scientists refer to the measure of randomness or disorder within a system as **entropy**. High entropy means high disorder and low energy. Molecules and chemical reactions have varying entropy as well. For example, entropy increases as molecules at a high concentration in one place diffuse and spread out. The second law of thermodynamics says that energy will always be lost as heat in energy transfers or transformations. Living things are highly ordered, requiring constant energy input to be maintained in a state of low entropy.

## Potential and Kinetic Energy

When an object is in motion, there is energy associated with that object. Think of a wrecking ball. Even a slow-moving wrecking ball can do a great deal of damage to other objects. Energy associated with objects in motion is called **kinetic energy**. A speeding bullet, a walking person, and the rapid movement of molecules in the air all have kinetic energy. Now what if that same motionless wrecking ball is lifted two stories above ground with a crane? If the suspended wrecking ball is not moving, is there energy associated with it? The answer is yes. The energy that was required to lift the wrecking ball did not disappear, but is now stored in the wrecking ball by virtue of its position and the force of gravity acting on it. This type of energy is called **potential energy** (Figure **3.1.3** below). If the ball were to fall, the potential energy would be transformed into kinetic energy until all of the potential energy was exhausted when the ball rested on the ground. Wrecking balls also swing like a pendulum; through the swing, there is a constant change of potential energy (highest at the top of the swing) to kinetic energy (highest at the bottom of the swing). Other examples of potential energy include the energy of water held behind a dam or a person about to skydive out of an airplane.



Figure **3.1.3**. Still water has potential energy; moving water, such as in a waterfall or a rapidly flowing river, has kinetic energy. (credit “dam”: modification of work by “Pascal”/Flickr; credit “waterfall”: modification of work by Frank Gualtieri)

Potential energy is not only associated with the location of matter, but also with the structure of matter. Even a spring on the ground has potential energy if it is compressed; so does a rubber band that is pulled taut. On a molecular level, the bonds that hold the atoms of molecules together exist in a particular structure that has potential energy. The fact that energy can be released by the breakdown of certain chemical bonds implies that those bonds have potential energy. In fact, there is potential energy stored within the bonds of all the food molecules we eat, which is harnessed for use. The type of potential energy that exists within chemical bonds, and is released when those bonds are broken, is called **chemical energy**. Chemical energy is responsible for providing living cells with energy from food. The release of energy occurs when the molecular bonds within food molecules are broken.

## What Is Light Energy?

The sun emits an enormous amount of electromagnetic radiation (solar energy). Humans can see only a fraction of this energy, which portion is therefore referred to as “visible light.” The manner in which solar energy travels is described as waves. Scientists can determine the amount of energy of a wave by measuring its wavelength, the distance between consecutive points of a wave. A single wave is measured from two consecutive points, such as from crest to crest or from trough to trough (Figure **3.1.4**).

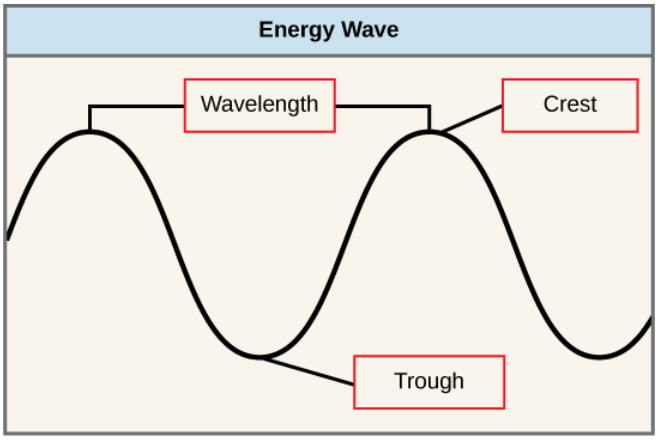



Figure **3.1.4**: The wavelength of a single wave is the distance between two consecutive points of similar position (two crests or two troughs) along the wave.

Visible light constitutes only one of many types of electromagnetic radiation emitted from the sun and other stars. Scientists differentiate the various types of radiant energy from the sun within the electromagnetic spectrum. The electromagnetic spectrum is the range of all possible frequencies of radiation (Figure **3.1.5**). The difference between wavelengths relates to the amount of energy carried by them.

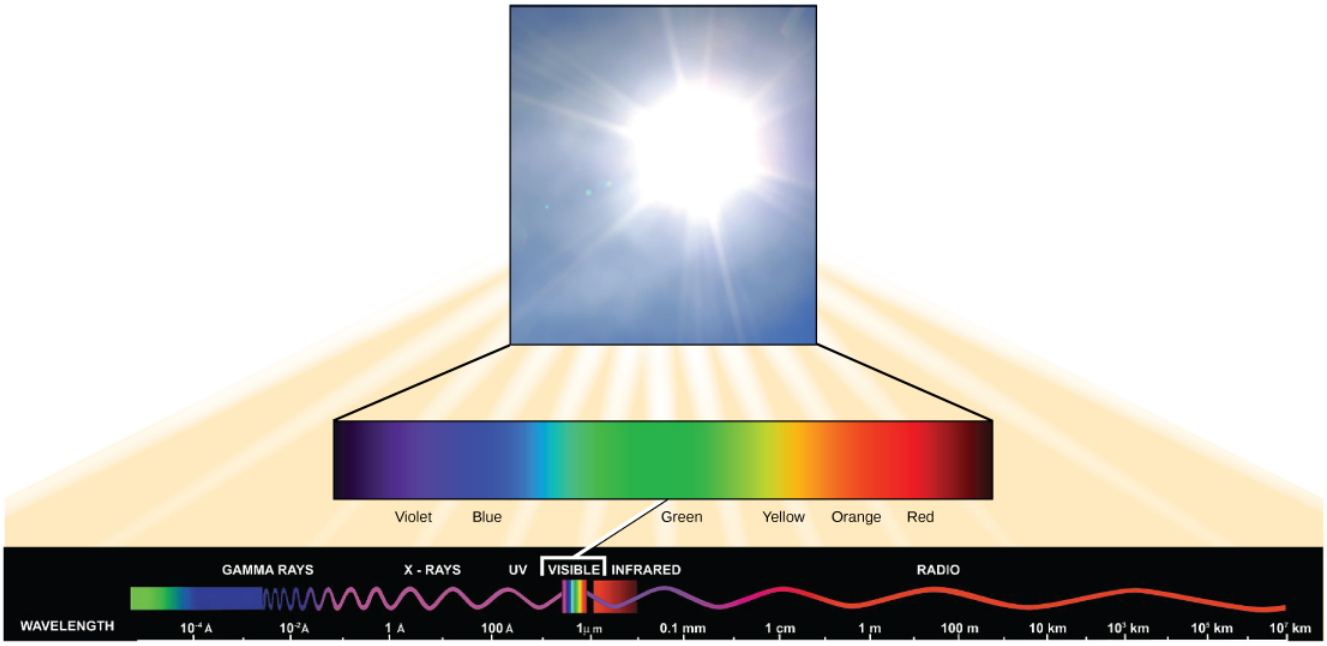



Figure **3.1.5**: The sun emits energy in the form of electromagnetic radiation. This radiation exists at different wavelengths, each of which has its own characteristic energy. All electromagnetic radiation, including visible light, is characterized by its wavelength.

Each type of electromagnetic radiation travels at a particular wavelength. The longer the wavelength (or the more stretched out it appears in the diagram), the less energy is carried. Short, tight waves carry the most energy. This may seem illogical, but think of it in terms of a piece of moving a heavy rope. It takes little effort by a person to move a rope in long, wide waves. To make a rope move in short, tight waves, a person would need to apply significantly more energy.

The electromagnetic spectrum (Figure **3.1.5**) shows several types of electromagnetic radiation originating from the sun, including X-rays and ultraviolet (UV) rays. The higher-energy waves can penetrate tissues and damage cells and DNA, explaining why both X-rays and UV rays can be harmful to living organisms. The visible light seen by humans as white light actually exists in a rainbow of colors. Certain objects, such as a prism or a drop of water, disperse white light to reveal the colors to the human eye. The visible light portion of the electromagnetic spectrum shows the rainbow of colors, with violet and blue having shorter wavelengths, and therefore higher energy. At the other end of the spectrum toward red, the wavelengths are longer and have lower energy (Figure **3.1.5**).

## Contributors and Attributions

Modified by Kyle Whittinghill (University of Pittsburgh) from the following sources

- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0. Modified from the original by Matthew R. Fisher.
- 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).

  **8.2: The Light-Dependent Reactions of Photosynthesis** by OpenStax, is licensed CC BY

This page titled 3.1: Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Matthew R. Fisher (OpenOregon) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.2: Energy** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.
- **8.2: The Light-Dependent Reactions of Photosynthesis** by OpenStax is licensed CC BY 4.0.

# 3.2: The Building Blocks of Molecules

At its most fundamental level, life is made up of matter. Matter occupies space and has mass. All matter is composed of elements, substances that cannot be broken down or transformed chemically into other substances. Each element is made of atoms, each with a constant number of protons and unique properties. A total of 118 elements have been defined; however, only 92 occur naturally, and fewer than 30 are found in living cells. The remaining 26 elements are unstable and, therefore, do not exist for very long or are theoretical and have yet to be detected.

Each element is designated by its chemical symbol (such as H, N, O, C, and Na), and possesses unique properties. These unique properties allow elements to combine and to bond with each other in specific ways.

## Atoms

An atom is the smallest component of an element that retains all of the chemical properties of that element. For example, one hydrogen atom has all of the properties of the element hydrogen, such as it exists as a gas at room temperature, and it bonds with oxygen to create a water molecule. Hydrogen atoms cannot be broken down into anything smaller while still retaining the properties of hydrogen. If a hydrogen atom were broken down into subatomic particles, it would no longer have the properties of hydrogen.

At the most basic level, all organisms are made of a combination of elements. They contain atoms that combine together to form molecules. In multicellular organisms, such as animals, molecules can interact to form cells that combine to form tissues, which make up organs. These combinations continue until entire multicellular organisms are formed.

All atoms contain protons, electrons, and neutrons (Figure 3.2.1). The only exception is hydrogen (H), which is made of one proton and one electron. A proton is a positively charged particle that resides in the nucleus (the core of the atom) of an atom and has a mass of 1 and a charge of +1. An electron is a negatively charged particle that travels in the space around the nucleus. In other words, it resides outside of the nucleus. It has a negligible mass and has a charge of –1.

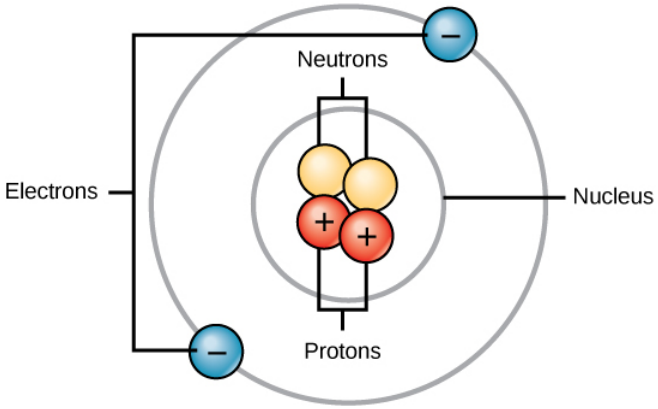



**Figure 3.2.1:** Atoms are made up of protons and neutrons located within the nucleus, and electrons surrounding the nucleus.

Neutrons, like protons, reside in the nucleus of an atom. They have a mass of 1 and no charge. The positive (protons) and negative (electrons) charges balance each other in a neutral atom, which has a net zero charge.

Because protons and neutrons each have a mass of 1, the mass of an atom is equal to the number of protons and neutrons of that atom. The number of electrons does not factor into the overall mass, because their mass is so small.

As stated earlier, each element has its own unique properties. Each contains a different number of protons and neutrons, giving it its own atomic number and mass number. The atomic number of an element is equal to the number of protons that element contains. The mass number, or atomic mass, is the number of protons plus the number of neutrons of that element. Therefore, it is possible to determine the number of neutrons by subtracting the atomic number from the mass number.

These numbers provide information about the elements and how they will react when combined. Different elements have different melting and boiling points, and are in different states (liquid, solid, or gas) at room temperature. They also combine in different ways. Some form specific types of bonds, whereas others do not. How they combine is based on the number of electrons present.

Because of these characteristics, the elements are arranged into the periodic table of elements, a chart of the elements that includes the atomic number and relative atomic mass of each element. The periodic table also provides key information about the properties of elements (Figure **3.2.2**)—often indicated by color-coding. The arrangement of the table also shows how the electrons in each element are organized and provides important details about how atoms will react with each other to form molecules.

Isotopes are different forms of the same element that have the same number of protons, but a different number of neutrons. Some elements, such as carbon, potassium, and uranium, have naturally occurring isotopes. Carbon-12, the most common isotope of carbon, contains six protons and six neutrons. Therefore, it has a mass number of 12 (six protons and six neutrons) and an atomic number of 6 (which makes it carbon). Carbon-14 contains six protons and eight neutrons. Therefore, it has a mass number of 14 (six protons and eight neutrons) and an atomic number of 6, meaning it is still the element carbon. These two alternate forms of carbon are isotopes. Some isotopes are unstable and will lose protons, other subatomic particles, or energy to form more stable elements. These are called radioactive isotopes or radioisotopes.

ART CONNECTION

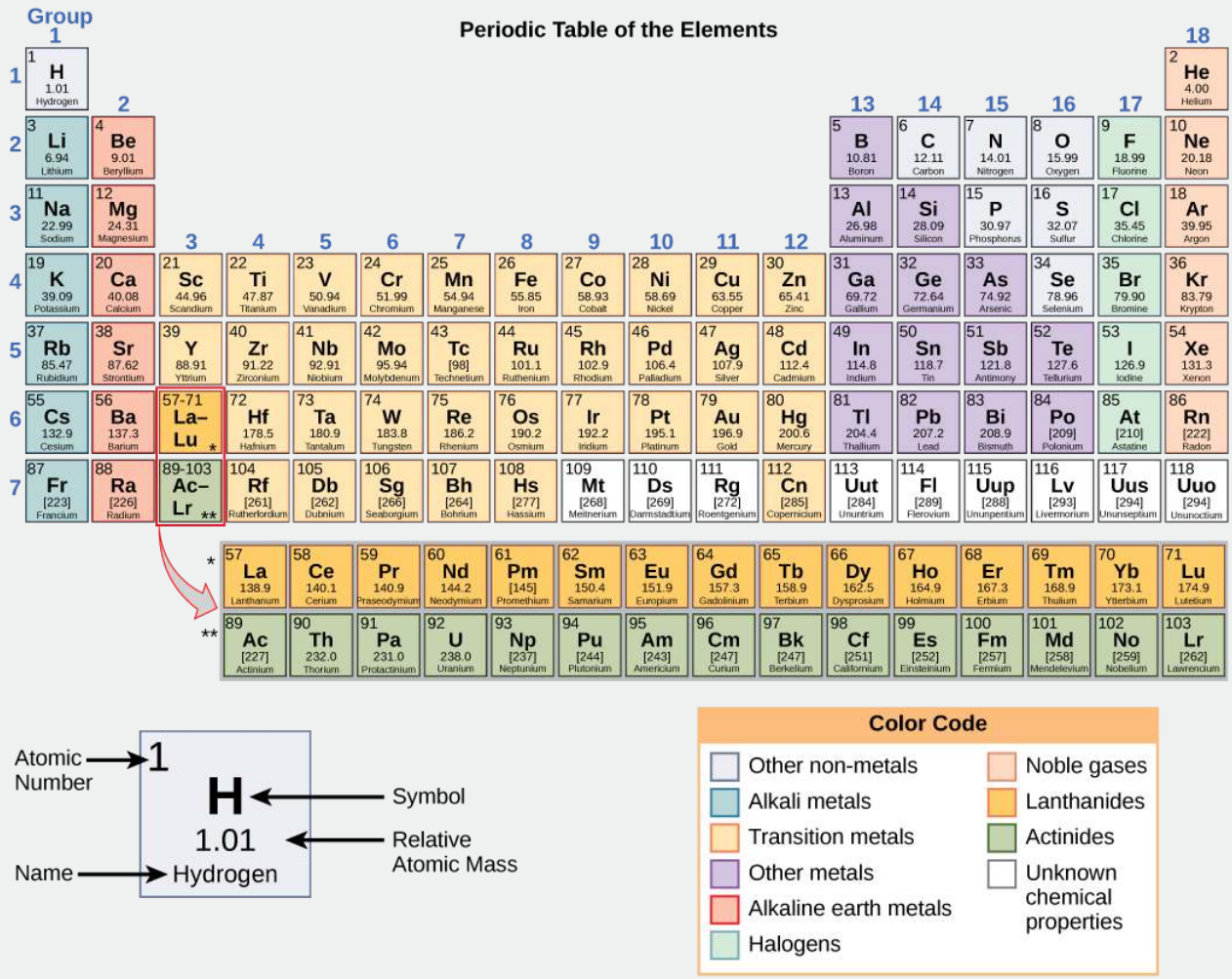



**Figure 3.2.2:** Arranged in columns and rows based on the characteristics of the elements, the periodic table provides key information about the elements and how they might interact with each other to form molecules. Most periodic tables provide a key or legend to the information they contain.

How many neutrons do (K) potassium-39 and potassium-40 have, respectively?

EVOLUTION IN ACTION: Carbon Dating

Carbon-14 ($^{14}C$) is a naturally occurring radioisotope that is created in the atmosphere by cosmic rays. This is a continuous process, so more $^{14}C$ is always being created. As a living organism develops, the relative level of $^{14}C$ in its body is equal to the concentration of $^{14}C$ in the atmosphere. When an organism dies, it is no longer ingesting $^{14}C$, so the ratio will decline. $^{14}C$ decays to $^{14}N$ by a process called beta decay; it gives off energy in this slow process.

After approximately 5,730 years, only one-half of the starting concentration of $^{14}C$ will have been converted to $^{14}N$. The time it takes for half of the original concentration of an isotope to decay to its more stable form is called its half-life. Because the half-life of $^{14}C$ is long, it is used to age formerly living objects, such as fossils. Using the ratio of the $^{14}C$ concentration found in an object to the amount of $^{14}C$ detected in the atmosphere, the amount of the isotope that has not yet decayed can be determined. Based on this amount, the age of the fossil can be calculated to about 50,000 years (Figure **3.2.3**). Isotopes with longer half-lives, such as potassium-40, are used to calculate the ages of older fossils. Through the use of carbon dating, scientists can reconstruct the ecology and biogeography of organisms living within the past 50,000 years.



**Figure 3.2.3:** The age of remains that contain carbon and are less than about 50,000 years old, such as this pygmy mammoth, can be determined using carbon dating. (credit: Bill Faulkner/NPS)

CONCEPT IN ACTION

To learn more about atoms and isotopes, and how you can tell one isotope from another, visit this site and run the simulation.

## Chemical Bonds

How elements interact with one another depends on how their electrons are arranged and how many openings for electrons exist at the outermost region where electrons are present in an atom. Electrons exist at energy levels that form shells around the nucleus. The closest shell can hold up to two electrons. The closest shell to the nucleus is always filled first, before any other shell can be filled. Hydrogen has one electron; therefore, it has only one spot occupied within the lowest shell. Helium has two electrons; therefore, it can completely fill the lowest shell with its two electrons. If you look at the periodic table, you will see that hydrogen and helium are the only two elements in the first row. This is because they only have electrons in their first shell. Hydrogen and helium are the only two elements that have the lowest shell and no other shells.

The second and third energy levels can hold up to eight electrons. The eight electrons are arranged in four pairs and one position in each pair is filled with an electron before any pairs are completed.

Looking at the periodic table again (Figure **3.2.2**), you will notice that there are seven rows. These rows correspond to the number of shells that the elements within that row have. The elements within a particular row have increasing numbers of electrons as the columns proceed from left to right. Although each element has the same number of shells, not all of the shells are completely filled with electrons. If you look at the second row of the periodic table, you will find lithium (Li), beryllium (Be), boron (B), carbon (C), nitrogen (N), oxygen (O), fluorine (F), and neon (Ne). These all have electrons that occupy only the first and second shells. Lithium

has only one electron in its outermost shell, beryllium has two electrons, boron has three, and so on, until the entire shell is filled with eight electrons, as is the case with neon.

Not all elements have enough electrons to fill their outermost shells, but an atom is at its most stable when all of the electron positions in the outermost shell are filled. Because of these vacancies in the outermost shells, we see the formation of chemical bonds, or interactions between two or more of the same or different elements that result in the formation of molecules. To achieve greater stability, atoms will tend to completely fill their outer shells and will bond with other elements to accomplish this goal by sharing electrons, accepting electrons from another atom, or donating electrons to another atom. Because the outermost shells of the elements with low atomic numbers (up to calcium, with atomic number 20) can hold eight electrons, this is referred to as the octet rule. An element can donate, accept, or share electrons with other elements to fill its outer shell and satisfy the octet rule.

When an atom does not contain equal numbers of protons and electrons, it is called an ion. Because the number of electrons does not equal the number of protons, each ion has a net charge. Positive ions are formed by losing electrons and are called cations. Negative ions are formed by gaining electrons and are called anions.

For example, sodium only has one electron in its outermost shell. It takes less energy for sodium to donate that one electron than it does to accept seven more electrons to fill the outer shell. If sodium loses an electron, it now has 11 protons and only 10 electrons, leaving it with an overall charge of +1. It is now called a sodium ion.

The chlorine atom has seven electrons in its outer shell. Again, it is more energy-efficient for chlorine to gain one electron than to lose seven. Therefore, it tends to gain an electron to create an ion with 17 protons and 18 electrons, giving it a net negative (–1) charge. It is now called a chloride ion. This movement of electrons from one element to another is referred to as electron transfer. As Figure **3.2.4** illustrates, a sodium atom (Na) only has one electron in its outermost shell, whereas a chlorine atom (Cl) has seven electrons in its outermost shell. A sodium atom will donate its one electron to empty its shell, and a chlorine atom will accept that electron to fill its shell, becoming chloride. Both ions now satisfy the octet rule and have complete outermost shells. Because the number of electrons is no longer equal to the number of protons, each is now an ion and has a +1 (sodium) or –1 (chloride) charge.

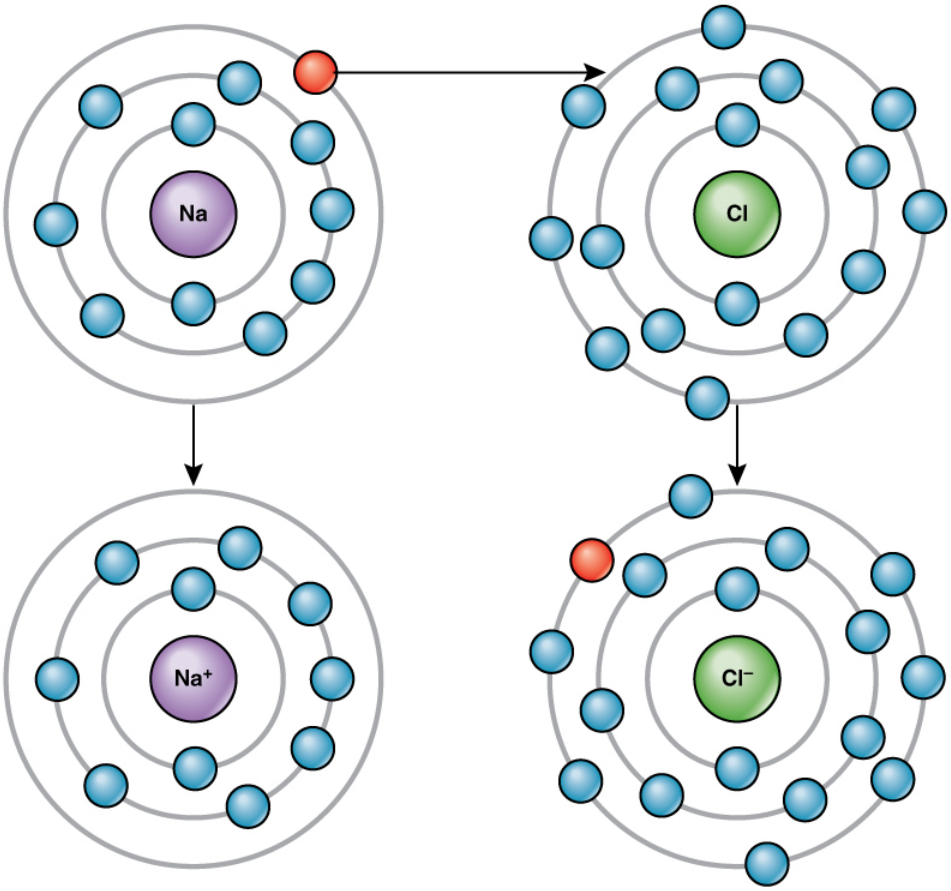



**Figure 3.2.4:** Elements tend to fill their outermost shells with electrons. To do this, they can either donate or accept electrons from other elements.

## Ionic Bonds

There are four types of bonds or interactions: ionic, covalent, hydrogen bonds, and van der Waals interactions. Ionic and covalent bonds are strong interactions that require a larger energy input to break apart. When an element donates an electron from its outer shell, as in the sodium atom example above, a positive ion is formed. The element accepting the electron is now negatively charged. Because positive and negative charges attract, these ions stay together and form an ionic bond, or a bond between ions. The elements bond together with the electron from one element staying predominantly with the other element. When $Na^+$ and $Cl^-$ ions combine to produce NaCl, an electron from a sodium atom stays with the other seven from the chlorine atom, and the sodium and chloride ions attract each other in a lattice of ions with a net zero charge.

## Covalent Bonds

Another type of strong chemical bond between two or more atoms is a covalent bond. These bonds form when an electron is shared between two elements and are the strongest and most common form of chemical bond in living organisms. Covalent bonds form between the elements that make up the biological molecules in our cells. Unlike ionic bonds, covalent bonds do not dissociate in water.

The hydrogen and oxygen atoms that combine to form water molecules are bound together by covalent bonds. The electron from the hydrogen atom divides its time between the outer shell of the hydrogen atom and the incomplete outer shell of the oxygen atom. To completely fill the outer shell of an oxygen atom, two electrons from two hydrogen atoms are needed, hence the subscript “2” in $H_2O$. The electrons are shared between the atoms, dividing their time between them to “fill” the outer shell of each. This sharing is a lower energy state for all of the atoms involved than if they existed without their outer shells filled.

There are two types of covalent bonds: polar and nonpolar. Nonpolar covalent bonds form between two atoms of the same element or between different elements that share the electrons equally. For example, an oxygen atom can bond with another oxygen atom to fill their outer shells. This association is nonpolar because the electrons will be equally distributed between each oxygen atom. Two

covalent bonds form between the two oxygen atoms because oxygen requires two shared electrons to fill its outermost shell. Nitrogen atoms will form three covalent bonds (also called triple covalent) between two atoms of nitrogen because each nitrogen atom needs three electrons to fill its outermost shell. Another example of a nonpolar covalent bond is found in the methane ($CH_4$) molecule. The carbon atom has four electrons in its outermost shell and needs four more to fill it. It gets these four from four hydrogen atoms, each atom providing one. These elements all share the electrons equally, creating four nonpolar covalent bonds (Figure **3.2.5**).

In a polar covalent bond, the electrons shared by the atoms spend more time closer to one nucleus than to the other nucleus. Because of the unequal distribution of electrons between the different nuclei, a slightly positive (δ+) or slightly negative (δ–) charge develops. The covalent bonds between hydrogen and oxygen atoms in water are polar covalent bonds. The shared electrons spend more time near the oxygen nucleus, giving it a small negative charge, than they spend near the hydrogen nuclei, giving these molecules a small positive charge.

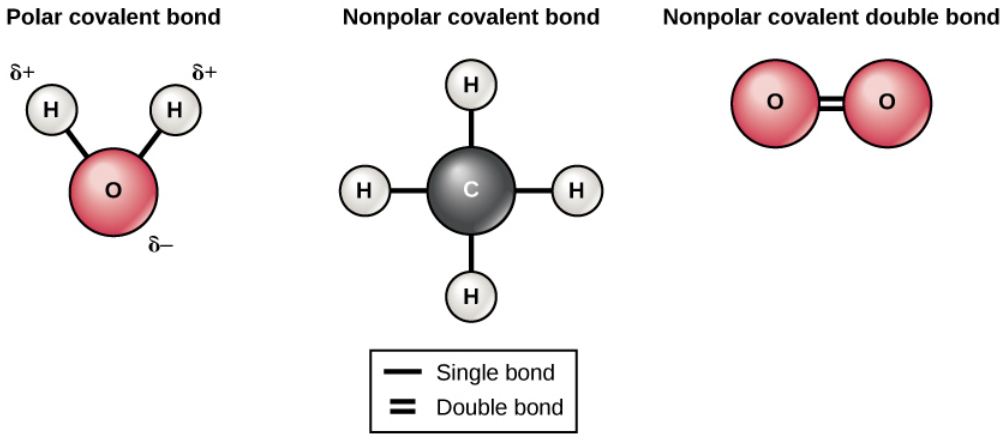



**Figure 3.2.5:** The water molecule (left) depicts a polar bond with a slightly positive charge on the hydrogen atoms and a slightly negative charge on the oxygen. Examples of nonpolar bonds include methane (middle) and oxygen (right).

## Hydrogen Bonds

Ionic and covalent bonds are strong bonds that require considerable energy to break. However, not all bonds between elements are ionic or covalent bonds. Weaker bonds can also form. These are attractions that occur between positive and negative charges that do not require much energy to break. Two weak bonds that occur frequently are hydrogen bonds and van der Waals interactions. These bonds give rise to the unique properties of water and the unique structures of DNA and proteins.

When polar covalent bonds containing a hydrogen atom form, the hydrogen atom in that bond has a slightly positive charge. This is because the shared electron is pulled more strongly toward the other element and away from the hydrogen nucleus. Because the hydrogen atom is slightly positive (δ+), it will be attracted to neighboring negative partial charges (δ–). When this happens, a weak interaction occurs between the δ+ charge of the hydrogen atom of one molecule and the δ– charge of the other molecule. This interaction is called a hydrogen bond. This type of bond is common; for example, the liquid nature of water is caused by the hydrogen bonds between water molecules (Figure **3.2.6**). Hydrogen bonds give water the unique properties that sustain life. If it were not for hydrogen bonding, water would be a gas rather than a liquid at room temperature.

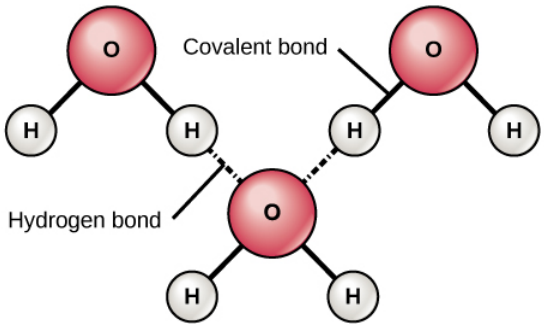



**Figure 3.2.6:** Hydrogen bonds form between slightly positive (δ+) and slightly negative (δ–) charges of polar covalent molecules, such as water.

Hydrogen bonds can form between different molecules and they do not always have to include a water molecule. Hydrogen atoms in polar bonds within any molecule can form bonds with other adjacent molecules. For example, hydrogen bonds hold together two long strands of DNA to give the DNA molecule its characteristic double-stranded structure. Hydrogen bonds are also responsible for some of the three-dimensional structure of proteins.

### van der Waals Interactions

Like hydrogen bonds, van der Waals interactions are weak attractions or interactions between molecules. They occur between polar, covalently bound, atoms in different molecules. Some of these weak attractions are caused by temporary partial charges formed when electrons move around a nucleus. These weak interactions between molecules are important in biological systems.

#### CAREERS IN ACTION: Radiography Technician

Have you or anyone you know ever had a magnetic resonance imaging (MRI) scan, a mammogram, or an X-ray? These tests produce images of your soft tissues and organs (as with an MRI or mammogram) or your bones (as happens in an X-ray) by using either radiowaves or special isotopes (radiolabeled or fluorescently labeled) that are ingested or injected into the body. These tests provide data for disease diagnoses by creating images of your organs or skeletal system.

MRI imaging works by subjecting hydrogen nuclei, which are abundant in the water in soft tissues, to fluctuating magnetic fields, which cause them to emit their own magnetic field. This signal is then read by sensors in the machine and interpreted by a computer to form a detailed image.

Some radiography technologists and technicians specialize in computed tomography, MRI, and mammography. They produce films or images of the body that help medical professionals examine and diagnose. Radiologists work directly with patients, explaining machinery, preparing them for exams, and ensuring that their body or body parts are positioned correctly to produce the needed images. Physicians or radiologists then analyze the test results.

Radiography technicians can work in hospitals, doctors' offices, or specialized imaging centers. Training to become a radiography technician happens at hospitals, colleges, and universities that offer certificates, associate's degrees, or bachelor's degrees in radiography.

## Summary

Matter is anything that occupies space and has mass. It is made up of atoms of different elements. All of the 92 elements that occur naturally have unique qualities that allow them to combine in various ways to create compounds or molecules. Atoms, which consist of protons, neutrons, and electrons, are the smallest units of an element that retain all of the properties of that element. Electrons can be donated or shared between atoms to create bonds, including ionic, covalent, and hydrogen bonds, as well as van der Waals interactions.

## Art Connections

Figure 3.2.2: How many neutrons do (K) potassium-39 and potassium-40 have, respectively?

**Answer**

Potassium-39 has twenty neutrons. Potassium-40 has twenty one neutrons.

## Glossary

**anion**
a negative ion formed by gaining electrons

**atomic number**
the number of protons in an atom

**cation**
a positive ion formed by losing electrons

**chemical bond**
an interaction between two or more of the same or different elements that results in the formation of molecules

**covalent bond**
a type of strong bond between two or more of the same or different elements; forms when electrons are shared between elements

**electron**
a negatively charged particle that resides outside of the nucleus in the electron orbital; lacks functional mass and has a charge of –1

**electron transfer**
the movement of electrons from one element to another

**element**
one of 118 unique substances that cannot be broken down into smaller substances and retain the characteristic of that substance; each element has a specified number of protons and unique properties

**hydrogen bond**
a weak bond between partially positively charged hydrogen atoms and partially negatively charged elements or molecules

**ion**
an atom or compound that does not contain equal numbers of protons and electrons, and therefore has a net charge

**ionic bond**
a chemical bond that forms between ions of opposite charges

**isotope**
one or more forms of an element that have different numbers of neutrons

**mass number**
the number of protons plus neutrons in an atom

**matter**
anything that has mass and occupies space

**neutron**
a particle with no charge that resides in the nucleus of an atom; has a mass of 1

**nonpolar covalent bond**
a type of covalent bond that forms between atoms when electrons are shared equally between atoms, resulting in no regions with partial charges as in polar covalent bonds

**nucleus**
(chemistry) the dense center of an atom made up of protons and (except in the case of a hydrogen atom) neutrons

**octet rule**
states that the outermost shell of an element with a low atomic number can hold eight electrons

**periodic table of elements**
an organizational chart of elements, indicating the atomic number and mass number of each element; also provides key information about the properties of elements

**polar covalent bond**

a type of covalent bond in which electrons are pulled toward one atom and away from another, resulting in slightly positive and slightly negative charged regions of the molecule

**proton**

a positively charged particle that resides in the nucleus of an atom; has a mass of 1 and a charge of +1

**radioactive isotope**

an isotope that spontaneously emits particles or energy to form a more stable element

**van der Waals interaction**

a weak attraction or interaction between molecules caused by slightly positively charged or slightly negatively charged atoms

## Contributors and Attributions

- 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).

---

This page titled 3.2: The Building Blocks of Molecules is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **2.1: The Building Blocks of Molecules** by OpenStax is licensed CC BY 4.0.

# 3.3: A Cell is the Smallest Unit of Life

## Levels of Organization of Living Things

Living things are highly organized and structured, following a hierarchy of scale from small to large (Figure 3.3.1). The **atom** is the smallest and most fundamental unit of matter. It consists of a **nucleus** surrounded by **electrons**. Atoms combine to form **molecules**, which are chemical structures consisting of at least two atoms held together by a chemical bond. Many molecules that are biologically important are **macromolecules**, large molecules that are typically formed by polymerization (a polymer is a large molecule that is made by combining smaller units called monomers, which are simpler than macromolecules). An example of a macromolecule is deoxyribonucleic acid (**DNA**) (Figure 3.3.3), which contains the instructions for the structure and functioning of all living organisms.

In plants, animals, and many other types of organisms, molecules come together in specific ways to create structures called **organelles** (Figure 3.3.1). Organelles are small structures that exist within cells and perform specialized functions. Examples of organelles include mitochondria and chloroplasts (Figure 3.3.5 – 8), which carry out indispensable functions: mitochondria produce energy to power the cell, while chloroplasts enable green plants to utilize the energy in sunlight to make sugars.

All living things are made of **cells** (Figure 3.3.1); the cell itself is the smallest fundamental unit of structure and function in living organisms. (This requirement is why viruses are not considered living: they are not made of cells. To make new viruses, they have to invade and hijack the reproductive mechanism of a living cell; only then can they obtain the materials they need to reproduce.) Some organisms consist of a single cell and others are multicellular. Cells are classified as prokaryotic or eukaryotic. **Prokaryotes** are single-celled or colonial organisms that do not have membrane-bound nuclei; in contrast, the cells of **eukaryotes** do have membrane-bound organelles and a membrane-bound nucleus.

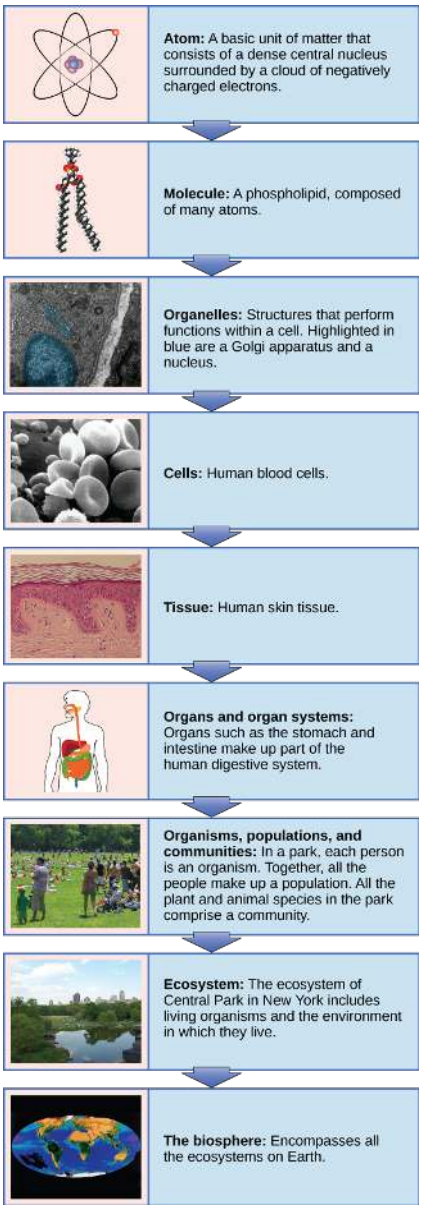



Figure **3.3.1**. From an atom to the entire Earth, biology examines all aspects of life. (credit "molecule": modification of work by Jane Whitney; credit "organelles": modification of work by Louisa Howard; credit "cells": modification of work by Bruce Wetzel, Harry Schaefer, National Cancer Institute; credit "tissue": modification of work by "Kilbad"/Wikimedia Commons; credit "organs": modification of work by Mariana Ruiz Villareal, Joaquim Alves Gaspar; credit "organisms": modification of work by Peter Dutton; credit "ecosystem": modification of work by "gigi4791″/Flickr; credit "biosphere": modification of work by NASA)

In most multicellular organisms, cells combine to make **tissues** (Figure **3.3.1**), which are groups of similar cells carrying out similar or related functions. **Organs** are collections of tissues grouped together performing a common function (Figure **3.3.1**). Organs are present not only in animals but also in plants. An organ system is a higher level of organization that consists of functionally related organs. For example vertebrate animals have many organ systems, such as the circulatory system that transports blood throughout the body and to and from the lungs; it includes organs such as the heart and blood vessels.

**Organisms** are individual living entities (Figure **3.3.1**). For example, each tree in a forest is an organism. Single-celled prokaryotes and single-celled eukaryotes are also considered organisms and are typically referred to as microorganisms.

All the individuals of a species living within a specific area are collectively called a **population** (Figure **3.3.1**). For example, a forest may include many pine trees. All of these pine trees represent the population of pine trees in this forest. Different populations may live in the same specific area. For example, the forest with the pine trees includes populations of flowering plants and also insects and microbial populations.

A **community** is the set of different populations inhabiting a common area (Figure **3.3.1**). For instance, all of the trees, flowers, insects, and other populations in a forest form the forest's community.

The forest itself is an **ecosystem** (Figure **3.3.1**). An ecosystem consists of all the living things in a particular area together with the abiotic, or non-living, parts of that environment such as nitrogen in the soil or rainwater.

At the highest level of organization, the **biosphere** (Figure **3.3.1**) is the collection of all ecosystems, and it represents the zones of life on Earth. It includes land, water, and portions of the atmosphere.

## Biological Molecules

Besides water, the molecules necessary for life are organic. **Organic molecules** are those that contain carbon covalently bonded to hydrogen. In addition, they may contain oxygen, nitrogen, phosphorus, sulfur, and additional elements.There are four major classes of organic molecules: **carbohydrates**, **lipids**, **proteins**, and **nucleic acids.** Each is an important component of the cell and performs a wide array of functions.

It is often said that life is "carbon-based." This means that carbon atoms, bonded to other carbon atoms or other elements, form the fundamental components of many of the molecules found uniquely in living things. Other elements play important roles in biological molecules, but carbon certainly qualifies as the "foundation" element for molecules in living things. It is the bonding properties of carbon atoms that are responsible for its important role.

Carbon can form four covalent bonds with other atoms or molecules. The simplest organic carbon molecule is methane (CH4), in which four hydrogen atoms bind to a carbon atom (Figure **3.3.2**). **Carbohydrates** include what are commonly referred to as simple sugars, like glucose, and complex carbohydrates such as starch. While many types of carbohydrates are used for energy, some are used for structure by most organisms, including plants and animals. For example, cellulose is a complex carbohydrate that adds rigidity and strength to the cell walls of plants. The suffix "-ose" denotes a carbohydrate, but note that not all carbohydrates were given that suffix when names (e.g., starch).

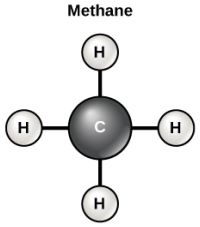



Figure **3.3.2**. Carbon can form four covalent bonds to create an organic molecule. The simplest carbon molecule is methane (CH4), depicted here.

**Lipids** include a diverse group of compounds that are united by a common feature. Lipids are hydrophobic ("water-fearing"), or insoluble in water, because they are non-polar molecules (molecules that contain non-polar covalent bonds) . Lipids perform many different functions in a cell. Cells store energy for long-term use in the form of lipids called fats. Lipids also provide insulation from the environment for plants and animals. For example, they help keep aquatic birds and mammals dry because of their water-repelling nature. Lipids are also the building blocks of many hormones and are an important constituent of cellular membranes. Lipids include fats, oils, waxes, phospholipids, and steroids.

**Proteins** are one of the most abundant organic molecules in living systems and have the most diverse range of functions of all macromolecules. They are all polymers of amino acids. The functions of proteins are very diverse because there are 20 different chemically distinct amino acids that form long chains, and the amino acids can be in any order. Proteins can function as enzymes, hormones, contractile fibers, cytoskeleton rods, and much more. **Enzymes** are vital to life because they act as catalyst in biochemical reactions (like digestion). Each enzyme is specific for the substrate (a reactant that binds to an enzyme) upon which it acts. Enzymes can function to break molecular bonds, to rearrange bonds, or to form new bonds.

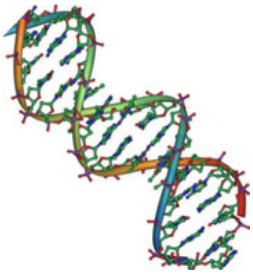


Figure **3.3.3**. The double-helix model shows DNA as two parallel strands of intertwining molecules.

Link to Learning

Video: Watch this video that animates the three-dimensional structure of the DNA molecule shown in Figure **3.3.3**.

**Nucleic acids** are very large molecules that are important to the continuity of life. They carry the genetic blueprint of a cell and thus the instructions for its functionality. The two main types of nucleic acids are deoxyribonucleic acid (**DNA**) and ribonucleic acid (RNA). DNA is the genetic material found in all organisms, ranging from single-celled bacteria to multicellular mammals. The other type of nucleic acid, **RNA**, is mostly involved in protein synthesis. The DNA molecules never leave the nucleus, but instead use an RNA intermediary to communicate with the rest of the cell. Other types of RNA are also involved in protein synthesis and its regulation. DNA and RNA are made up of small building blocks known as **nucleotides**. The nucleotides combine with each other to form a polynucleotide: DNA or RNA. Each nucleotide is made up of three components: a nitrogenous base, a pentose (five-carbon) sugar, and a phosphate. DNA has a beautiful double-helical structure (Figure **3.3.3**).

## Cell Theory

Close your eyes and picture a brick wall. What is the basic building block of that wall? It is a single brick, of course. Like a brick wall, your body is composed of basic building blocks and the building blocks of your body are cells. Your body has many kinds of cells, each specialized for a specific purpose. Just as a home is made from a variety of building materials, the human body is constructed from many cell types. For example, bone cells help to support and protect the body. Cells of the immune system fight invading bacteria. And red blood cells carry oxygen throughout the body. Each of these cell types plays a vital role during the growth, development, and day-to-day maintenance of the body. In spite of their enormous variety, however, all cells share certain fundamental characteristics.

The microscopes we use today are far more complex than those used in the 1600s by Antony van Leeuwenhoek, a Dutch shopkeeper who had great skill in crafting lenses. Despite the limitations of his now-ancient lenses, van Leeuwenhoek observed the movements of single-celled organism and sperm, which he collectively termed “animalcules.” In a 1665 publication called *Micrographia*, experimental scientist Robert Hooke coined the term “cell” (from the Latin *cella*, meaning “small room”) for the box-like structures he observed when viewing cork tissue through a lens. In the 1670s, van Leeuwenhoek discovered bacteria and protozoa. Later advances in lenses and microscope construction enabled other scientists to see different components inside cells.

By the late 1830s, botanist Matthias Schleiden and zoologist Theodor Schwann were studying tissues and proposed the unified **cell theory**, which states that all living things are composed of one or more cells, that the cell is the basic unit of life, and that all new cells arise from existing cells. These principles still stand today. There are many types of cells, and all are grouped into one of two broad categories: prokaryotic and eukaryotic. Animal, plant, fungal, and protist cells are classified as eukaryotic, whereas bacteria and archaea cells are classified as prokaryotic.

All cells share four common components: 1) a plasma membrane, an outer covering that separates the cell’s interior from its surrounding environment; 2) cytoplasm, consisting of a jelly-like region within the cell in which other cellular components are

found; 3) DNA, the genetic material of the cell; and 4) ribosomes, particles that synthesize proteins. However, prokaryotes differ from eukaryotic cells in several ways.

## Components of Prokaryotic Cells

A **prokaryotic cell** is a simple, single-celled (unicellular) organism that lacks a nucleus, or any other membrane-bound organelle. We will shortly come to see that this is significantly different in eukaryotes. Prokaryotic DNA is found in the central part of the cell: a darkened region called the nucleoid (Figure **3.3.4**).

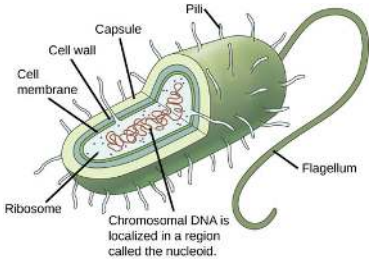



Figure **3.3.4**. This figure shows the generalized structure of a prokaryotic cell.

Unlike Archaea and eukaryotes, bacteria have a cell wall made of peptidoglycan (molecules comprised of sugars and amino acids) and many have a polysaccharide capsule. The cell wall acts as an extra layer of protection, helps the cell maintain its shape, and prevents dehydration. The capsule enables the cell to attach to surfaces in its environment. Some prokaryotes have flagella, pili, or fimbriae. Flagella are used for locomotion. Pili are used to exchange genetic material during a type of reproduction called conjugation. Fimbriae are protein appendages used by bacteria to attach to other cells.

## Eukaryotic Cells

A **eukaryotic cell** is a cell that has a membrane-bound nucleus and other membrane-bound compartments called **organelles.** There are many different types of organelles, each with a highly specialized function (see Figure **3.3.5 – 7**). The word eukaryotic means “true kernel” or “true nucleus,” alluding to the presence of the membrane-bound nucleus in these cells. The word “organelle” means “little organ,” and, as already mentioned, organelles have specialized cellular functions, just as the organs of your body have specialized functions.

## Cell Size

At 0.1–5.0 µm in diameter, most prokaryotic cells are significantly smaller than eukaryotic cells, which have diameters ranging from 10–100 µm (Figure **3.3.5**). The small size of prokaryotes allows ions and organic molecules that enter them to quickly spread to other parts of the cell. Similarly, any wastes produced within a prokaryotic cell can quickly move out. However, larger eukaryotic cells have evolved different structural adaptations to enhance cellular transport. Indeed, the large size of these cells would not be possible without these adaptations. In general, cell size is limited because volume increases much more quickly than does cell surface area. As a cell becomes larger, it becomes more and more difficult for the cell to acquire sufficient materials to support the processes inside the cell, because the relative size of the surface area through which materials must be transported declines.

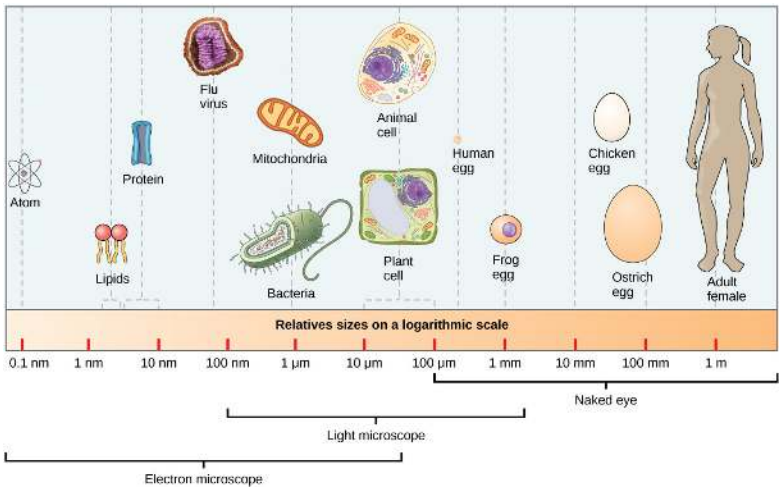



Figure **3.3.5**. This figure shows the relative sizes of different kinds of cells and cellular components. An adult human is shown for comparison.

## Animal Cells versus Plant Cells

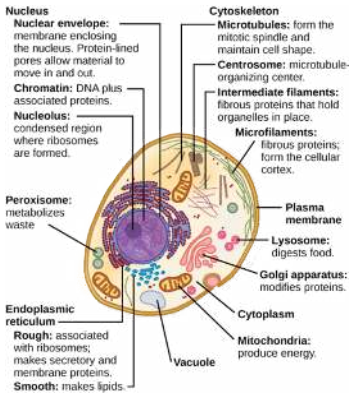



Figure **3.3.6**. An example of a typical animal cell.

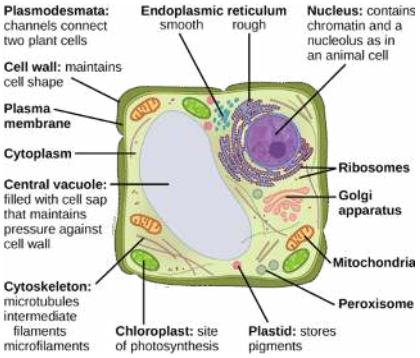



Figure **3.3.7**. An example of a typical plant cell.

Despite their fundamental similarities, there are some striking differences between animal and plant cells (Figure **3.3.6 – 7**). Animal cells have centrioles, centrosomes, and lysosomes, whereas plant cells do not. Plant cells have a rigid cell wall that is external to the plasma membrane, chloroplasts, plasmodesmata, and plastids used for storage, and a large central vacuole, whereas animal cells do not.

## Chloroplasts

From an ecological perspective, **chloroplasts** are a particularly important type of organelle because they perform photosynthesis (Figure **3.3.8**). Photosynthesis forms the foundation of food chains in most ecosystems. Chloroplasts are only found in eukaryotic cells such as plants and algae. During photosynthesis, carbon dioxide, water, and light energy are used to make glucose and

molecular oxygen. One major difference between algae/plants and animals is that plants/algae are able to make their own food, like glucose, whereas animals must obtain food by consuming other organisms.

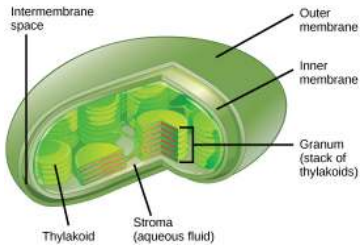



Figure 3.3.8. This simplified diagram of a chloroplast shows its structure.

Chloroplasts have outer and inner membranes, but within the space enclosed by a chloroplast's inner membrane is a set of interconnected and stacked, fluid-filled membrane sacs called thylakoids (Figure 3.3.8 below). Each stack of thylakoids is called a granum (plural = grana). The fluid enclosed by the inner membrane and surrounding the grana is called the stroma. Each structure within the chloroplast has an important function, which is enabled by its particular shape. A common theme in biology is that form and function are interrelated. For example, the membrane-rich stacks of the thylakoids provide ample surface area to embed the proteins and pigments that are vital to photosynthesis.

## Contributors and Attributions

- Kyle Whittinghill (University of Pittsburgh)
- 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).
- "Essentials of Environmental Science" by Kamala Doršner is licensed under CC BY 4.0. "Levels of Organization of Living Things" by Open Stax is licensed under CC BY 4.0. Modified from the originals by Matthew R. Fisher.
- Concepts of Biology by OpenStax is licensed under CC BY 4.0. Modified from original by Matthew R. Fisher.
- **1.2: Themes and Concepts of Biology** by OpenStax, is licensed CC BY

---

This page titled 3.3: A Cell is the Smallest Unit of Life is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Matthew R. Fisher (OpenOregon) via source content that was edited to the style and standards of the LibreTexts platform.

- **2.3: A Cell is the Smallest Unit of Life** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.
- **2.1: Matter** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# 3.4: Water

**Skills to Develop**

- Describe the properties of water that are critical to maintaining life
- Explain why water is an excellent solvent
- Provide examples of water's cohesive and adhesive properties
- Discuss the role of acids, bases, and buffers in homeostasis

Why do scientists spend time looking for water on other planets? Why is water so important? It is because water is essential to life as we know it. Water is one of the more abundant molecules and the one most critical to life on Earth. Approximately 60–70 percent of the human body is made up of water. Without it, life as we know it simply would not exist.

The polarity of the water molecule and its resulting hydrogen bonding make water a unique substance with special properties that are intimately tied to the processes of life. Life originally evolved in a watery environment, and most of an organism's cellular chemistry and metabolism occur inside the watery contents of the cell's cytoplasm. Special properties of water are its high heat capacity and heat of vaporization, its ability to dissolve polar molecules, its cohesive and adhesive properties, and its dissociation into ions that leads to the generation of pH. Understanding these characteristics of water helps to elucidate its importance in maintaining life.

## Water's Polarity

One of water's important properties is that it is composed of polar molecules: the hydrogen and oxygen within water molecules ($H_2O$) form polar covalent bonds. While there is no net charge to a water molecule, the polarity of water creates a slightly positive charge on hydrogen and a slightly negative charge on oxygen, contributing to water's properties of attraction. 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.

As a result of water's polarity, each water molecule attracts other water molecules because of the opposite charges between water molecules, forming hydrogen bonds. Water also attracts or is attracted to other polar molecules and ions. A polar substance that interacts readily with or dissolves in water is referred to as hydrophilic (hydro- = "water"; -philic = "loving"). In contrast, non-polar molecules such as oils and fats do not interact well with water, as shown in Figure **3.4.1** and separate from it rather than dissolve in it, as we see in salad dressings containing oil and vinegar (an acidic water solution). These nonpolar compounds are called hydrophobic (hydro- = "water"; -phobic = "fearing").

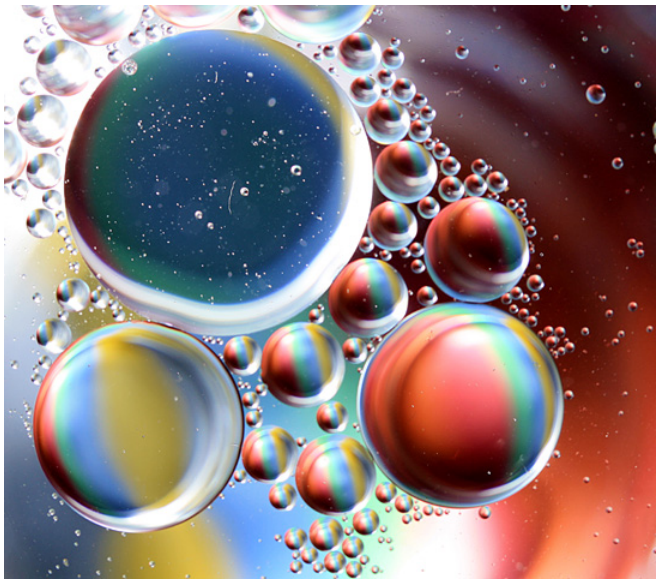


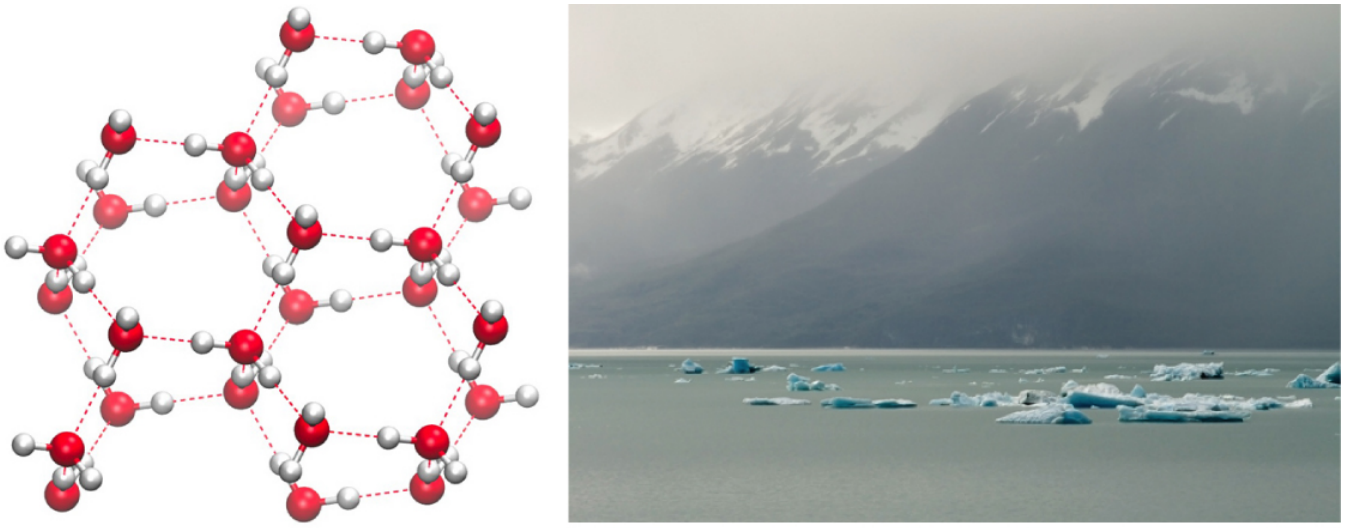
Figure **3.4.1**: Oil and water do not mix. As this macro image of oil and water shows, oil does not dissolve in water but forms droplets instead. This is due to it being a nonpolar compound. (credit: Gautam Dogra).

## Water's States: Gas, Liquid, and Solid

The formation of hydrogen bonds is an important quality of the liquid water that is crucial to life as we know it. As water molecules make hydrogen bonds with each other, water takes on some unique chemical characteristics compared to other liquids and, since living things have a high water content, understanding these chemical features is key to understanding life. In liquid water, hydrogen bonds are constantly formed and broken as the water molecules slide past each other. The breaking of these bonds is caused by the motion (kinetic energy) of the water molecules due to the heat contained in the system. When the heat is raised as water is boiled, the higher kinetic energy of the water molecules causes the hydrogen bonds to break completely and allows water molecules to escape into the air as gas (steam or 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 (there is not enough energy to break the hydrogen bonds) that makes ice less dense than liquid water, a phenomenon not seen in the solidification of other liquids.

Water's lower density in its solid form is due to the way hydrogen bonds are oriented as it freezes: the water molecules are pushed farther apart compared to liquid water. With most other liquids, solidification when the temperature drops includes the lowering of kinetic energy between molecules, allowing them to pack even more tightly than in liquid form and giving the solid a greater density than the liquid.

The lower density of ice, illustrated and pictured in Figure **3.4.2**, an anomaly, causes it to float at the surface of liquid water, such as in an iceberg or in the ice cubes in a glass of ice water. In lakes and ponds, ice will form on the surface of the water creating an insulating barrier that protects the animals and plant life in the pond from freezing. Without this layer of insulating ice, plants and animals living in the pond would freeze in the solid block of ice and could not survive. The detrimental effect of freezing on living organisms is caused by the expansion of ice relative to liquid water. The ice crystals that form upon freezing rupture the delicate membranes essential for the function of living cells, irreversibly damaging them. Cells can only survive freezing if the water in them is temporarily replaced by another liquid like glycerol.



(a) (b)

Figure **3.4.2**: Hydrogen bonding makes ice less dense than liquid water. The (a) lattice structure of ice makes it less dense than the freely flowing molecules of liquid water, enabling it to (b) float on water. (credit a: modification of work by Jane Whitney, image created using Visual Molecular Dynamics (VMD) software[1]; credit b: modification of work by Carlos Ponte)

Link to Learning



Video: Click here to see a 3-D animation of the structure of an ice lattice. (Image credit: Jane Whitney. Image created using Visual Molecular Dynamics VMD software.[2])

## Water's High Heat Capacity

Water's high heat capacity is a property caused by hydrogen bonding among water molecules. Water has the highest specific heat capacity of any liquids. 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 therefore takes water a long time to heat and long time to cool. In fact, the specific heat capacity of water is about five times more than that of sand. This explains why the land cools faster than the sea. Due to its high heat capacity, water is used by warm blooded animals to more evenly disperse heat 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.

## Water's 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 cal) 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 its gaseous phase (steam). As a result, water acts as a heat sink or heat reservoir and requires much more heat to boil than does a liquid such as ethanol (grain alcohol), whose hydrogen bonding with other ethanol molecules is weaker than water's hydrogen bonding. Eventually, as water reaches its boiling point of 100° Celsius (212° Fahrenheit), the heat is able to break the hydrogen bonds between the water molecules, and the kinetic energy (motion) 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.

The fact that hydrogen bonds need to be broken for water to evaporate means that a substantial amount of energy is used in the 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's Solvent Properties

Since water is a polar molecule with slightly positive and slightly negative charges, ions and polar molecules can readily dissolve in it. Therefore, water is referred to as a solvent, a substance capable of dissolving other polar molecules and ionic compounds. The charges associated with these molecules will form hydrogen bonds with water, surrounding the particle with water molecules. This is referred to as a sphere of hydration, or a hydration shell, as illustrated in Figure **3.4.3** and serves to keep the particles separated or dispersed in the water.

When ionic compounds are added to water, the individual ions react with the polar regions of the water molecules and their ionic bonds are disrupted in the process of dissociation. Dissociation occurs when atoms or groups of atoms break off from molecules and form ions. Consider table salt (NaCl, or sodium chloride): when NaCl crystals are added to water, the molecules of NaCl dissociate into $Na^+$ and $Cl^-$ ions, and spheres of hydration form around the ions, illustrated in Figure **3.4.3**. The positively charged sodium ion is surrounded by the partially negative charge of the water molecule's oxygen. The negatively charged chloride ion is surrounded by the partially positive charge of the hydrogen on the water molecule.

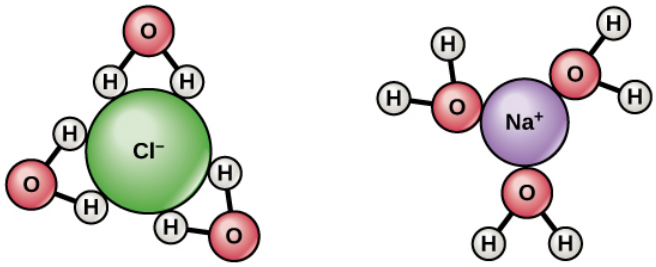



Figure **3.4.3**: When table salt (NaCl) is mixed in water, spheres of hydration are formed around the ions.

## 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. This water can stay above the glass because of the property of cohesion. 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 withstand being ruptured 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. When a small scrap of paper is placed onto the droplet of water, the paper floats on top of the water droplet even though paper 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's even possible to "float" a needle on top of a glass of water if it is placed gently without breaking the surface tension, as shown in Figure **3.4.4**.

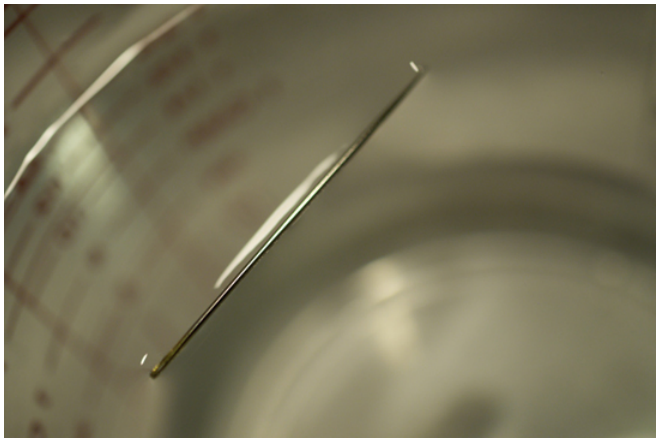


Figure **3.4.4**: The weight of the needle is pulling the surface downward; at the same time, the surface tension is pulling it up, suspending it on the surface of the water and keeping it from sinking. Notice the indentation in the water around the needle. (credit: Cory Zanker)

These cohesive forces are related to water's property of adhesion, or the attraction between water molecules and other molecules. This attraction is sometimes stronger than water's cohesive forces, especially when the water is exposed to charged surfaces such as those found 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 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 **3.4.5**.

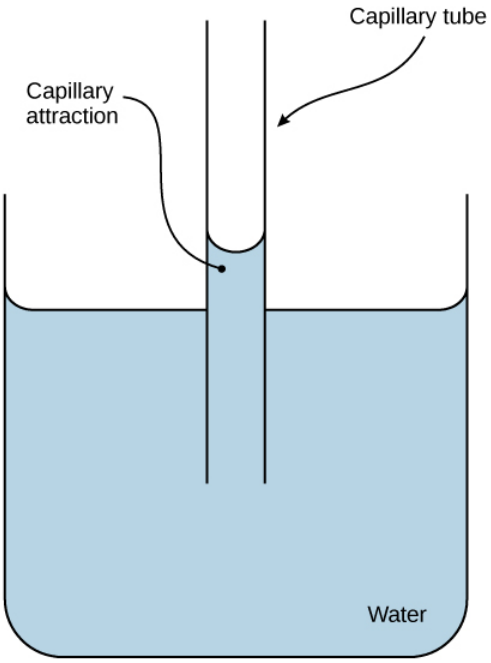



Figure 3.4.5: Capillary action in a glass tube is caused by the adhesive forces exerted by the internal surface of the glass exceeding the cohesive forces between the water molecules themselves. (credit: modification of work by Pearson-Scott Foresman, donated to the Wikimedia Foundation)

Why are cohesive and adhesive forces important for life? Cohesive and adhesive forces are important for the transport of water from the roots to the leaves in plants. These forces create a "pull" on the water column. This pull results from the tendency of water molecules being evaporated on the surface of the plant to stay connected to water molecules below them, and so they are pulled along. Plants use this natural phenomenon to help transport water from their roots to their leaves. Without these properties of water, plants would be unable to receive the water and the dissolved minerals they require. In another example, insects such as the water strider, shown in Figure 3.4.6, use the surface tension of water to stay afloat on the surface layer of water and even mate there.

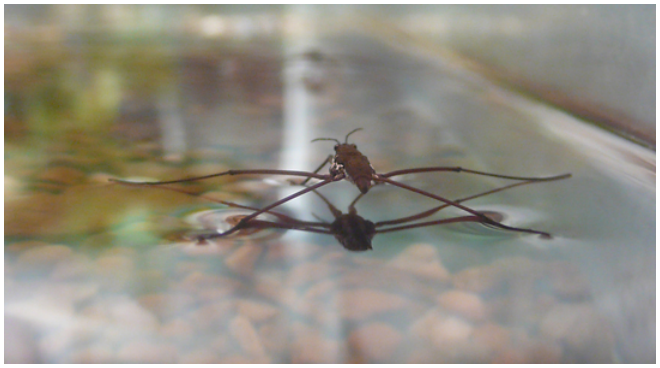


Figure 3.4.6: Water's cohesive and adhesive properties allow this water strider (*Gerris* sp.) to stay afloat. (credit: Tim Vickers)

## pH, Buffers, Acids, and Bases

The pH of a solution indicates its acidity or alkalinity.

$$H_2O(l) \rightleftharpoons H^+(aq) + O^-(aq)$$

litmus or pH paper, filter paper that has been treated with a natural water-soluble dye so it can be used as a pH indicator, to test how much acid (acidity) or base (alkalinity) exists in a solution. You might have even used some to test whether the water in a swimming pool is properly treated. In both cases, the pH test measures the concentration of hydrogen ions in a given solution.

Hydrogen ions are spontaneously generated in pure water by the dissociation (ionization) of a small percentage of water molecules into equal numbers of hydrogen ($H^+$) ions and hydroxide ($OH^-$) ions. While the hydroxide ions are kept in solution by their

hydrogen bonding with other water molecules, the hydrogen ions, consisting of naked protons, are immediately attracted to un-ionized water molecules, forming hydronium ions ($H_30^+$). Still, by convention, scientists refer to hydrogen ions and their concentration as if they were free in this state in liquid water.

The concentration of hydrogen ions dissociating from pure water is $1 \times 10^{-7}$ moles $H^+$ ions per liter of water. Moles (mol) are a way to express the amount of a substance (which can be atoms, molecules, ions, etc), with one mole being equal to $6.02 \times 10^{23}$ particles of the substance. Therefore, 1 mole of water is equal to $6.02 \times 10^{23}$ water molecules. The pH is calculated as the negative of the base 10 logarithm of this concentration. The log10 of $1 \times 10^{-7}$ is -7.0, and the negative of this number (indicated by the "p" of "pH") yields a pH of 7.0, which is also known as neutral pH. The pH inside of human cells and blood are examples of two areas of the body where near-neutral pH is maintained.

Non-neutral pH readings result from dissolving acids or bases in water. Using the negative logarithm to generate positive integers, high concentrations of hydrogen ions yield a low pH number, whereas low levels of hydrogen ions result in a high pH. An acid is a substance that increases the concentration of hydrogen ions ($H^+$) in a solution, usually by having one of its hydrogen atoms dissociate. A base provides either hydroxide ions ($OH^–$) or other negatively charged ions that combine with hydrogen ions, reducing their concentration in the solution and thereby raising the pH. In cases where the base releases hydroxide ions, these ions bind to free hydrogen ions, generating new water molecules.

The stronger the acid, the more readily it donates $H^+$. For example, hydrochloric acid (HCl) completely dissociates into hydrogen and chloride ions and is highly acidic, whereas the acids in tomato juice or vinegar do not completely dissociate and are considered weak acids. Conversely, strong bases are those substances that readily donate $OH^–$ or take up hydrogen ions. Sodium hydroxide (NaOH) and many household cleaners are highly alkaline and give up $OH^–$ rapidly when placed in water, thereby raising the pH. An example of a weak basic solution is seawater, which has a pH near 8.0, close enough to neutral pH that marine organisms adapted to this saline environment are able to thrive in it.

The pH scale is, as previously mentioned, an inverse logarithm and ranges from 0 to 14 (Figure **3.4.7**). Anything below 7.0 (ranging from 0.0 to 6.9) is acidic, and anything above 7.0 (from 7.1 to 14.0) is alkaline. Extremes in pH in either direction from 7.0 are usually considered inhospitable to life. The pH inside cells (6.8) and the pH in the blood (7.4) are both very close to neutral. However, the environment in the stomach is highly acidic, with a pH of 1 to 2. So how do the cells of the stomach survive in such an acidic environment? How do they homeostatically maintain the near neutral pH inside them? The answer is that they cannot do it and are constantly dying. New stomach cells are constantly produced to replace dead ones, which are digested by the stomach acids. It is estimated that the lining of the human stomach is completely replaced every seven to ten days.

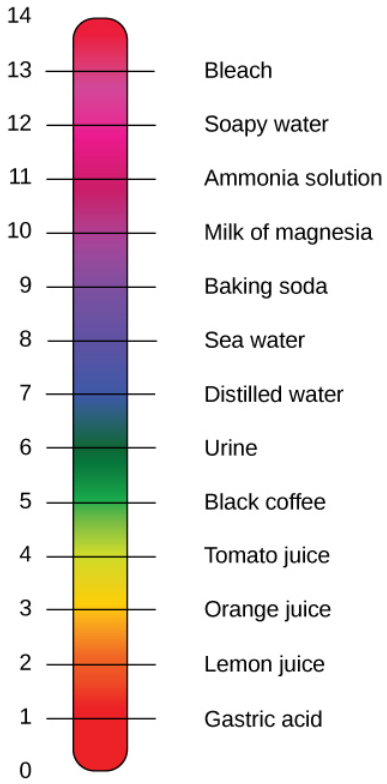



Figure **3.4.7**: The pH scale measures the concentration of hydrogen ions ($H^+$) in a solution. (credit: modification of work by Edward Stevens)

Link to Learning



Watch this video for a straightforward explanation of pH and its logarithmic scale.

So how can organisms whose bodies require a near-neutral pH ingest acidic and basic substances (a human drinking orange juice, for example) and survive? Buffers are the key. Buffers readily absorb excess $H^+$ or $OH^-$, keeping the pH of the body carefully maintained in the narrow range required for survival. Maintaining a constant blood pH is critical to a person's well-being. The buffer maintaining the pH of human blood involves carbonic acid ($H_2CO_3$), bicarbonate ion ($HCO_3^-$), and carbon dioxide ($CO_2$). When bicarbonate ions combine with free hydrogen ions and become carbonic acid, hydrogen ions are removed, moderating pH

changes. Similarly, as shown in Figure **3.4.8**, excess carbonic acid can be converted to carbon dioxide gas and exhaled through the lungs. This prevents too many free hydrogen ions from building up in the blood and dangerously reducing the blood's pH. Likewise, if too much $OH^-$ is introduced into the system, carbonic acid will combine with it to create bicarbonate, lowering the pH. Without this buffer system, the body's pH would fluctuate enough to put survival in jeopardy.

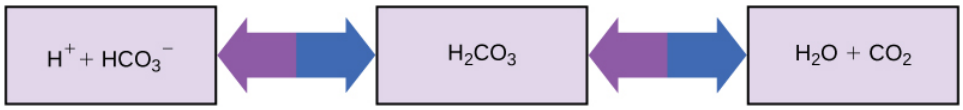



Figure **3.4.8**: This diagram shows the body's buffering of blood pH levels. The blue arrows show the process of raising pH as more $CO_2$ is made. The purple arrows indicate the reverse process: the lowering of pH as more bicarbonate is created.

Other examples of buffers are antacids used to combat excess stomach acid. Many of these over-the-counter medications work in the same way as blood buffers, usually with at least one ion capable of absorbing hydrogen and moderating pH, bringing relief to those that suffer "heartburn" after eating. The unique properties of water that contribute to this capacity to balance pH—as well as water's other characteristics—are essential to sustaining life on Earth.

Link to Learning

To learn more about water. Visit the U.S. Geological Survey Water Science for Schools All About Water! website.

## Summary

Water has many properties that are critical to maintaining life. It is a polar molecule, allowing for the formation of hydrogen bonds. Hydrogen bonds allow ions and other polar molecules to dissolve in water. Therefore, water is an excellent solvent. The hydrogen bonds between water molecules cause the water to have a high heat capacity, meaning it takes a lot of added heat to raise its temperature. As the temperature rises, the hydrogen bonds between water continually break and form anew. This allows for the overall temperature to remain stable, although energy is added to the system. Water also exhibits a high heat of vaporization, which is key to how organisms cool themselves by the evaporation of sweat. Water's cohesive forces allow for the property of surface tension, whereas its adhesive properties are seen as water rises inside capillary tubes. The pH value is a measure of hydrogen ion concentration in a solution and is one of many chemical characteristics that is highly regulated in living organisms through homeostasis. Acids and bases can change pH values, but buffers tend to moderate the changes they cause. These properties of water are intimately connected to the biochemical and physical processes performed by living organisms, and life would be very different if these properties were altered, if it could exist at all.

## Footnotes

1. 1 W. Humphrey W., A. Dalke, and K. Schulten, "VMD—Visual Molecular Dynamics," *Journal of Molecular Graphics* 14 (1996): 33-38.
2. 2 W. Humphrey W., A. Dalke, and K. Schulten, "VMD—Visual Molecular Dynamics," *Journal of Molecular Graphics* 14 (1996): 33-38.

## Glossary

**acid**

molecule that donates hydrogen ions and increases the concentration of hydrogen ions in a solution

**adhesion**

attraction between water molecules and other molecules

**base**
molecule that donates hydroxide ions or otherwise binds excess hydrogen ions and decreases the concentration of hydrogen ions in a solution

**buffer**
substance that prevents a change in pH by absorbing or releasing hydrogen or hydroxide ions

**calorie**
amount of heat required to change the temperature of one gram of water by one degree Celsius

**capillary action**
occurs because water molecules are attracted to charges on the inner surfaces of narrow tubular structures such as glass tubes, drawing the water molecules to the sides of the tubes

**cohesion**
intermolecular forces between water molecules caused by the polar nature of water; responsible for surface tension

**dissociation**
release of an ion from a molecule such that the original molecule now consists of an ion and the charged remains of the original, such as when water dissociates into $H^+$ and $OH^-$

**evaporation**
separation of individual molecules from the surface of a body of water, leaves of a plant, or the skin of an organism

**heat of vaporization of water**
high amount of energy required for liquid water to turn into water vapor

**hydrophilic**
describes ions or polar molecules that interact well with other polar molecules such as water

**hydrophobic**
describes uncharged non-polar molecules that do not interact well with polar molecules such as water

**litmus paper**
(also, pH paper) filter paper that has been treated with a natural water-soluble dye that changes its color as the pH of the environment changes so it can be used as a pH indicator

**pH paper**
see litmus paper

**pH scale**
scale ranging from zero to 14 that is inversely proportional to the concentration of hydrogen ions in a solution

**solvent**
substance capable of dissolving another substance

**specific heat capacity**
the amount of heat one gram of a substance must absorb or lose to change its temperature by one degree Celsius

**sphere of hydration**
when a polar water molecule surrounds charged or polar molecules thus keeping them dissolved and in solution

**surface tension**

tension at the surface of a body of liquid that prevents the molecules from separating; created by the attractive cohesive forces between the molecules of the liquid

---

This page titled 3.4: Water is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **2.2: Water** by OpenStax is licensed CC BY 4.0.

# 3.5: Themes and Concepts of Biology

**Learning Objectives**

- Identify and describe the properties of life
- Recognize and interpret a phylogenetic tree
- List examples of different sub disciplines in biology

Biology is the science that studies life, but what exactly is life? This may sound like a silly question with an obvious response, but it is not always easy to define life. For example, a branch of biology called virology studies viruses, which exhibit some of the characteristics of living entities but lack others. It turns out that although viruses can attack living organisms, cause diseases, and even reproduce, they do not meet the criteria that biologists use to define life. Consequently, virologists are not biologists, strictly speaking. Similarly, some biologists study the early molecular evolution that gave rise to life; since the events that preceded life are not biological events, these scientists are also excluded from biology in the strict sense of the term.

From its earliest beginnings, biology has wrestled with three questions: What are the shared properties that make something "alive"? And once we know something is alive, how do we find meaningful levels of organization in its structure? And, finally, when faced with the remarkable diversity of life, how do we organize the different kinds of organisms so that we can better understand them? As new organisms are discovered every day, biologists continue to seek answers to these and other questions.

## Properties of Life

All living organisms share several key characteristics or functions: order, sensitivity or response to the environment, reproduction, adaptation, growth and development, regulation, homeostasis, energy processing, and evolution. When viewed together, these nine characteristics serve to define life.

### Order

Organisms are highly organized, coordinated structures that consist of one or more cells. Even very simple, single-celled organisms are remarkably complex: inside each cell, atoms make up molecules; these in turn make up cell organelles and other cellular inclusions. In multicellular organisms (Figure **3.5.1**), similar cells form tissues. Tissues, in turn, collaborate to create organs (body structures with a distinct function). Organs work together to form organ systems.



Figure **3.5.1**: A toad represents a highly organized structure consisting of cells, tissues, organs, and organ systems. (credit: "Ivengo"/Wikimedia Commons)

### Sensitivity or Response to Stimuli

Organisms respond to diverse stimuli. For example, plants can bend toward a source of light, climb on fences and walls, or respond to touch (Figure **3.5.2**). Even tiny bacteria can move toward or away from chemicals (a process called *chemotaxis*) or light (*phototaxis*). Movement toward a stimulus is considered a positive response, while movement away from a stimulus is considered a negative response.



Figure **3.5.2**: The leaves of this sensitive plant (*Mimosa pudica*) will instantly droop and fold when touched. After a few minutes, the plant returns to normal. (credit: Alex Lomas)

Link to Learning



Video: Watch this video to see how plants respond to a stimulus—from opening to light, to wrapping a tendril around a branch, to capturing prey.

## Reproduction

Single-celled organisms reproduce by first duplicating their DNA, and then dividing it equally as the cell prepares to divide to form two new cells. Multicellular organisms often produce specialized reproductive germline cells that will form new individuals. When reproduction occurs, genes containing DNA are passed along to an organism's offspring. These genes ensure that the offspring will belong to the same species and will have similar characteristics, such as size and shape.

## Growth and Development

Organisms grow and develop following specific instructions coded for by their genes. These genes provide instructions that will direct cellular growth and development, ensuring that a species' young (Figure **3.5.3**) will grow up to exhibit many of the same characteristics as its parents.



Figure **3.5.3**: Although no two look alike, these kittens have inherited genes from both parents and share many of the same characteristics. (credit: Rocky Mountain Feline Rescue)

## Regulation

Even the smallest organisms are complex and require multiple regulatory mechanisms to coordinate internal functions, respond to stimuli, and cope with environmental stresses. Two examples of internal functions regulated in an organism are nutrient transport and blood flow. Organs (groups of tissues working together) perform specific functions, such as carrying oxygen throughout the body, removing wastes, delivering nutrients to every cell, and cooling the body.

## Homeostasis

In order to function properly, cells need to have appropriate conditions such as proper temperature, pH, and appropriate concentration of diverse chemicals. These conditions may, however, change from one moment to the next. Organisms are able to maintain internal conditions within a narrow range almost constantly, despite environmental changes, through homeostasis (literally, “steady state”)—the ability of an organism to maintain constant internal conditions. For example, an organism needs to regulate body temperature through a process known as thermoregulation. Organisms that live in cold climates, such as the polar bear (Figure **3.5.4**), have body structures that help them withstand low temperatures and conserve body heat. Structures that aid in this type of insulation include fur, feathers, blubber, and fat. In hot climates, organisms have methods (such as perspiration in humans or panting in dogs) that help them to shed excess body heat.



Figure **3.5.4**: Polar bears (*Ursus maritimus*) and other mammals living in ice-covered regions maintain their body temperature by generating heat and reducing heat loss through thick fur and a dense layer of fat under their skin. (credit: “longhorndave”/Flickr)

## Energy Processing

All organisms use a source of energy for their metabolic activities. Some organisms capture energy from the sun and convert it into chemical energy in food; others use chemical energy in molecules they take in as food (Figure **3.5.5**).



Figure **3.5.5**: The California condor (*Gymnogyps californianus*) uses chemical energy derived from food to power flight. California condors are an endangered species; this bird has a wing tag that helps biologists identify the individual. (credit: Pacific Southwest Region U.S. Fish and Wildlife Service)

## The Diversity of Life

The fact that biology, as a science, has such a broad scope has to do with the tremendous diversity of life on earth. The source of this diversity is evolution, the process of gradual change during which new species arise from older species. Evolutionary biologists study the evolution of living things in everything from the microscopic world to ecosystems.

The evolution of various life forms on Earth can be summarized in a phylogenetic tree (Figure **3.5.6**). A phylogenetic tree is a diagram showing the evolutionary relationships among biological species based on similarities and differences in genetic or physical traits or both. A phylogenetic tree is composed of nodes and branches. The internal nodes represent ancestors and are points in evolution when, based on scientific evidence, an ancestor is thought to have diverged to form two new species. The length of each branch is proportional to the time elapsed since the split.

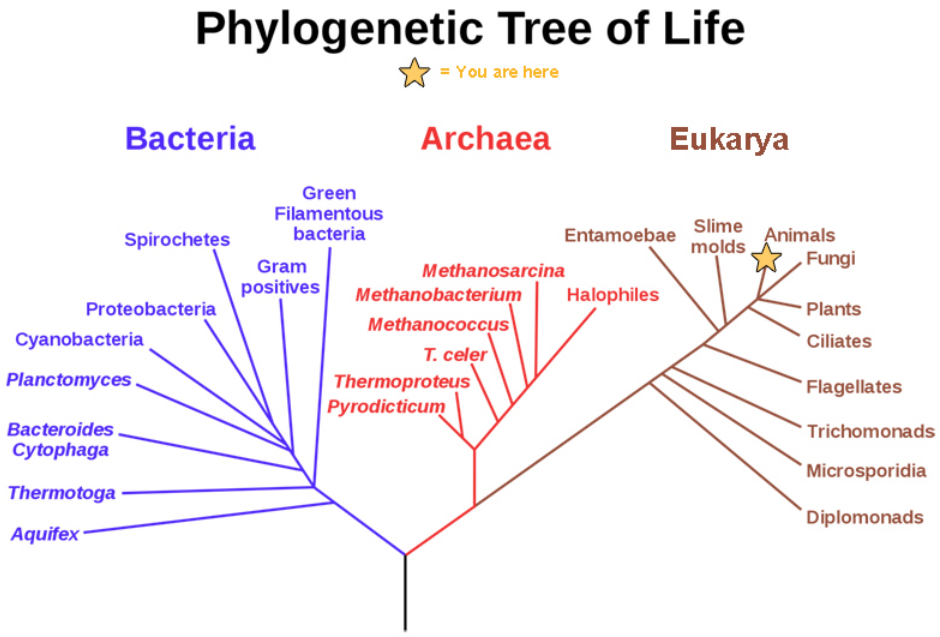



Figure 3.5.6: This phylogenetic tree was constructed by microbiologist Carl Woese using data obtained from sequencing ribosomal RNA genes. The tree shows the separation of living organisms into three domains: Bacteria, Archaea, and Eukarya. Bacteria and Archaea are prokaryotes, single-celled organisms lacking intracellular organelles. (credit: Eric Gaba; NASA Astrobiology Institute)

## Evolution Connection: Carl Woese and the Phylogenetic Tree

In the past, biologists grouped living organisms into five kingdoms: animals, plants, fungi, protists, and bacteria. The organizational scheme was based mainly on physical features, as opposed to physiology, biochemistry, or molecular biology, all of which are used by modern systematics. The pioneering work of American microbiologist Carl Woese in the early 1970s has shown, however, that life on Earth has evolved along three lineages, now called domains—Bacteria, Archaea, and Eukarya. The first two are prokaryotic cells with microbes that lack membrane-enclosed nuclei and organelles. The third domain contains the eukaryotes and includes unicellular microorganisms together with the four original kingdoms (excluding bacteria). Woese defined Archaea as a new domain, and this resulted in a new taxonomic tree (Figure 3.5.6). Many organisms belonging to the Archaea domain live under extreme conditions and are called extremophiles. To construct his tree, Woese used genetic relationships rather than similarities based on morphology (shape).

Woese's tree was constructed from comparative sequencing of the genes that are universally distributed, present in every organism, and conserved (meaning that these genes have remained essentially unchanged throughout evolution). Woese's approach was revolutionary because comparisons of physical features are insufficient to differentiate between the prokaryotes that appear fairly similar in spite of their tremendous biochemical diversity and genetic variability (Figure 3.5.7). The comparison of homologous DNA and RNA sequences provided Woese with a sensitive device that revealed the extensive variability of prokaryotes, and which justified the separation of the prokaryotes into two domains: bacteria and archaea.

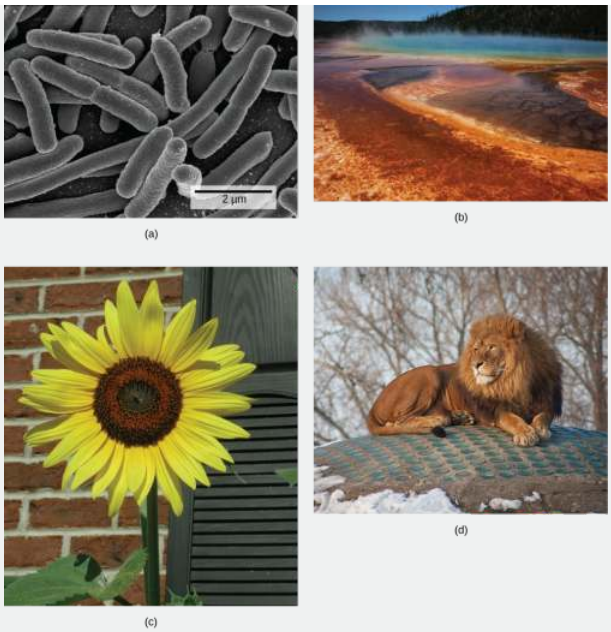



Figure **3.5.7**: These images represent different domains. The (a) bacteria in this micrograph belong to Domain Bacteria, while the (b) extremophiles (not visible) living in this hot vent belong to Domain Archaea. Both the (c) sunflower and (d) lion are part of Domain Eukarya. (credit a: modification of work by Drew March; credit b: modification of work by Steve Jurvetson; credit c: modification of work by Michael Arrighi; credit d: modification of work by Leszek Leszcynski)

## Branches of Biological Study

The scope of biology is broad and therefore contains many branches and subdisciplines. Biologists may pursue one of those subdisciplines and work in a more focused field. For instance, molecular biology and biochemistry study biological processes at the molecular and chemical level, including interactions among molecules such as DNA, RNA, and proteins, as well as the way they are regulated. Microbiology, the study of microorganisms, is the study of the structure and function of single-celled organisms. It is quite a broad branch itself, and depending on the subject of study, there are also microbial physiologists, ecologists, and geneticists, among others.

### Career Connection: Forensic Scientist

Forensic science is the application of science to answer questions related to the law. Biologists as well as chemists and biochemists can be forensic scientists. Forensic scientists provide scientific evidence for use in courts, and their job involves examining trace materials associated with crimes. Interest in forensic science has increased in the last few years, possibly because of popular television shows that feature forensic scientists on the job. Also, the development of molecular techniques and the establishment of DNA databases have expanded the types of work that forensic scientists can do. Their job activities are primarily related to crimes against people such as murder, rape, and assault. Their work involves analyzing samples such as hair, blood, and other body fluids and also processing DNA (Figure **3.5.8**) found in many different environments and materials. Forensic scientists also analyze other biological evidence left at crime scenes, such as insect larvae or pollen grains. Students who want to pursue careers in forensic science will most likely be required to take chemistry and biology courses as well as some intensive math courses.



Figure **3.5.8**: This forensic scientist works in a DNA extraction room at the U.S. Army Criminal Investigation Laboratory at Fort Gillem, GA. (credit: United States Army CID Command Public Affairs)

Another field of biological study, neurobiology, studies the biology of the nervous system, and although it is considered a branch of biology, it is also recognized as an interdisciplinary field of study known as neuroscience. Because of its interdisciplinary nature, this subdiscipline studies different functions of the nervous system using molecular, cellular, developmental, medical, and computational approaches.



Figure **3.5.9**: Researchers work on excavating dinosaur fossils at a site in Castellón, Spain. (credit: Mario Modesto)

Paleontology, another branch of biology, uses fossils to study life's history (Figure **3.5.9**). Zoology and botany are the study of animals and plants, respectively. Biologists can also specialize as biotechnologists, ecologists, or physiologists, to name just a few areas. This is just a small sample of the many fields that biologists can pursue.

Biology is the culmination of the achievements of the natural sciences from their inception to today. Excitingly, it is the cradle of emerging sciences, such as the biology of brain activity, genetic engineering of custom organisms, and the biology of evolution that uses the laboratory tools of molecular biology to retrace the earliest stages of life on earth. A scan of news headlines—whether reporting on immunizations, a newly discovered species, sports doping, or a genetically-modified food—demonstrates the way biology is active in and important to our everyday world.

## Levels of Ecological Study

Ecology is the study of the interactions of living organisms with their environment. One core goal of ecology is to understand the distribution and abundance of living things in the physical environment. Attainment of this goal requires the integration of scientific disciplines inside and outside of biology, such as biochemistry, physiology, evolution, biodiversity, molecular biology, geology, and climatology. Some ecological research also applies aspects of chemistry and physics, and it frequently uses mathematical models.

Link to Learning

Climate change can alter where organisms live, which can sometimes directly affect human health. Watch the PBS video "Feeling the Effects of Climate Change" in which researchers discover a pathogenic organism living far outside of its normal range.

Ecologists interested in the factors that influence the survival of an endangered species might use mathematical models to predict how current conservation efforts affect endangered organisms. To produce a sound set of management options, a conservation biologist needs to collect accurate data, including current population size, factors affecting reproduction (like physiology and behavior), habitat requirements (such as plants and soils), and potential human influences on the endangered population and its habitat (which might be derived through studies in sociology and urban ecology). Within the discipline of ecology, researchers work at four specific levels, sometimes discretely and sometimes with overlap: organism, population, community, and ecosystem (Figure 3.5.10).

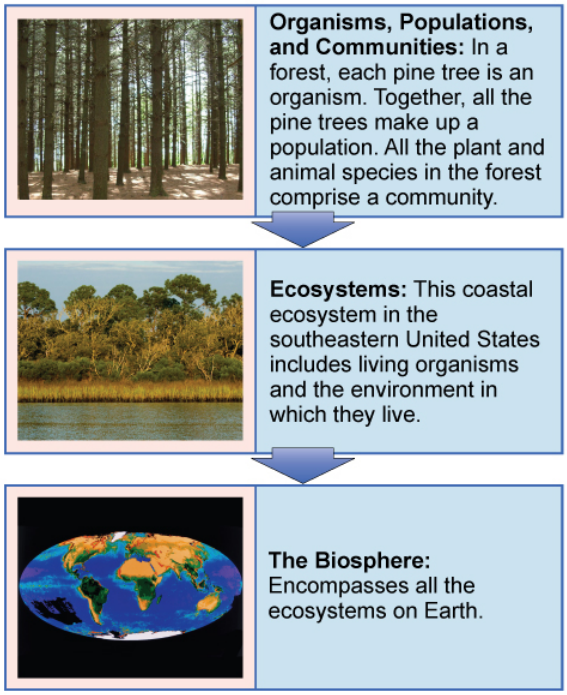



Figure 3.5.10: Ecologists study within several biological levels of organization. (credit "organisms": modification of work by "Crystl"/Flickr; credit "ecosystems": modification of work by Tom Carlisle, US Fish and Wildlife Service Headquarters; credit "biosphere": NASA)

## Organismal Ecology

Researchers studying ecology at the organismal level are interested in the adaptations that enable individuals to live in specific habitats. These adaptations can be morphological, physiological, and behavioral. For instance, the Karner blue butterfly (*Lycaeides melissa samuelis*) (Figure 3.5.11) is considered a specialist because the females preferentially oviposit (that is, lay eggs) on wild lupine. This preferential adaptation means that the Karner blue butterfly is highly dependent on the presence of wild lupine plants for its continued survival.



Figure **3.5.11**: The Karner blue butterfly (*Lycaeides melissa samuelis*) is a rare butterfly that lives only in open areas with few trees or shrubs, such as pine barrens and oak savannas. It can only lay its eggs on lupine plants. (credit: modification of work by J & K Hollingsworth, USFWS)

After hatching, the larval caterpillars emerge and spend four to six weeks feeding solely on wild lupine (Figure **3.5.12**). The caterpillars pupate (undergo metamorphosis) and emerge as butterflies after about four weeks. The adult butterflies feed on the nectar of flowers of wild lupine and other plant species. A researcher interested in studying Karner blue butterflies at the organismal level might, in addition to asking questions about egg laying, ask questions about the butterflies' preferred temperature (a physiological question) or the behavior of the caterpillars when they are at different larval stages (a behavioral question).



Figure **3.5.12**: The wild lupine (*Lupinus perennis*) is the host plant for the Karner blue butterfly.

## Population Ecology

A population is a group of interbreeding organisms that are members of the same species living in the same area at the same time. (Organisms that are all members of the same species are called conspecifics.) A population is identified, in part, by where it lives, and its area of population may have natural or artificial boundaries: natural boundaries might be rivers, mountains, or deserts, while examples of artificial boundaries include mowed grass, manmade structures, or roads. The study of population ecology focuses on the number of individuals in an area and how and why population size changes over time. Population ecologists are particularly interested in counting the Karner blue butterfly, for example, because it is classified as federally endangered. However, the distribution and density of this species is highly influenced by the distribution and abundance of wild lupine. Researchers might ask questions about the factors leading to the decline of wild lupine and how these affect Karner blue butterflies. For example, ecologists know that wild lupine thrives in open areas where trees and shrubs are largely absent. In natural settings, intermittent wildfires regularly remove trees and shrubs, helping to maintain the open areas that wild lupine requires. Mathematical models can be used to understand how wildfire suppression by humans has led to the decline of this important plant for the Karner blue butterfly.

## Community Ecology

A biological community consists of the different species within an area, typically a three-dimensional space, and the interactions within and among these species. Community ecologists are interested in the processes driving these interactions and their consequences. Questions about conspecific interactions often focus on competition among members of the same species for a limited resource. Ecologists also study interactions among various species; members of different species are called heterospecifics. Examples of heterospecific interactions include predation, parasitism, herbivory, competition, and pollination. These interactions can have regulating effects on population sizes and can impact ecological and evolutionary processes affecting diversity.

For example, Karner blue butterfly larvae form mutualistic relationships with ants. Mutualism is a form of a long-term relationship that has coevolved between two species and from which each species benefits. For mutualism to exist between individual organisms, each species must receive some benefit from the other as a consequence of the relationship. Researchers have shown that there is an increase in the probability of survival when Karner blue butterfly larvae (caterpillars) are tended by ants. This might be because the larvae spend less time in each life stage when tended by ants, which provides an advantage for the larvae. Meanwhile, the Karner blue butterfly larvae secrete a carbohydrate-rich substance that is an important energy source for the ants. Both the Karner blue larvae and the ants benefit from their interaction.

## Ecosystem Ecology

Ecosystem ecology is an extension of organismal, population, and community ecology. The ecosystem is composed of all the biotic components (living things) in an area along with the abiotic components (non-living things) of that area. Some of the abiotic components include air, water, and soil. Ecosystem biologists ask questions about how nutrients and energy are stored and how they move among organisms and the surrounding atmosphere, soil, and water.

The Karner blue butterflies and the wild lupine live in an oak-pine barren habitat. This habitat is characterized by natural disturbance and nutrient-poor soils that are low in nitrogen. The availability of nutrients is an important factor in the distribution of the plants that live in this habitat. Researchers interested in ecosystem ecology could ask questions about the importance of limited resources and the movement of resources, such as nutrients, though the biotic and abiotic portions of the ecosystem.

### Career Connection: Ecologist

A career in ecology contributes to many facets of human society. Understanding ecological issues can help society meet the basic human needs of food, shelter, and health care. Ecologists can conduct their research in the laboratory and outside in natural environments (Figure **3.5.13**). These natural environments can be as close to home as the stream running through your campus or as far away as the hydrothermal vents at the bottom of the Pacific Ocean. Ecologists manage natural resources such as white-tailed deer populations (*Odocoileus virginianus*) for hunting or aspen (*Populus* spp.) timber stands for paper production. Ecologists also work as educators who teach children and adults at various institutions including universities, high schools, museums, and nature centers. Ecologists may also work in advisory positions assisting local, state, and federal policymakers to develop laws that are ecologically sound, or they may develop those policies and legislation themselves. To become an ecologist requires an undergraduate degree, usually in a natural science. The undergraduate degree is often followed

by specialized training or an advanced degree, depending on the area of ecology selected. Ecologists should also have a broad background in the physical sciences, as well as a sound foundation in mathematics and statistics.



Figure **3.5.13**: This landscape ecologist is releasing a black-footed ferret into its native habitat as part of a study. (credit: USFWS Mountain Prairie Region, NPS)

Link to Learning

Visit this site to see Stephen Wing, a marine ecologist from the University of Otago, discuss the role of an ecologist and the types of issues ecologists explore.

## Contributors and Attributions

- Modified from the orignial by Kyle Whittinghill (University of Pittsburgh)
- 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).
- 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).
- **1.2: Themes and Concepts of Biology** by OpenStax, is licensed CC BY
- **44.1: The Scope of Ecology** by OpenStax, is licensed CC BY

This page titled 3.5: Themes and Concepts of Biology is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **1.2: Themes and Concepts of Biology** is licensed CC BY 4.0.
- **44.1: The Scope of Ecology** by OpenStax is licensed CC BY 4.0.

# CHAPTER OVERVIEW

## 4: Evolution and the Origin of Species

The theory of evolution is the unifying theory of biology, meaning it is the framework within which biologists ask questions about the living world. Its power is that it provides direction for predictions about living things that are borne out in experiment after experiment. The Ukrainian-born American geneticist Theodosius Dobzhansky famously wrote that “nothing makes sense in biology except in the light of evolution." He meant that the tenet that all life has evolved and diversified from a common ancestor is the foundation from which we approach all questions in biology.


4.1: Understanding Evolution

4.2: Formation of New Species


### Contributors and Attributions

- 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).


This page titled 4: Evolution and the Origin of Species is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

# 4.1: Understanding Evolution

**Learning Objectives**

- Describe how the present-day theory of evolution was developed
- Define adaptation
- Explain convergent and divergent evolution
- Describe homologous and vestigial structures
- Discuss misconceptions about the theory of evolution

**Evolution** by natural selection describes a mechanism for how species change over time. That species change had been suggested and debated well before Darwin began to explore this idea. The view that species were static and unchanging was grounded in the writings of Plato, yet there were also ancient Greeks who expressed evolutionary ideas. In the eighteenth century, ideas about the evolution of animals were reintroduced by the naturalist Georges-Louis Leclerc Comte de Buffon who observed that various geographic regions have different plant and animal populations, even when the environments are similar. It was also accepted that there were extinct species.

During this time, James Hutton, a Scottish naturalist, proposed that geological change occurred gradually by the accumulation of small changes from processes operating like they are today over long periods of time. This contrasted with the predominant view that the geology of the planet was a consequence of catastrophic events occurring during a relatively brief past. Hutton's view was popularized in the nineteenth century by the geologist Charles Lyell who became a friend to Darwin. Lyell's ideas were influential on Darwin's thinking: Lyell's notion of the greater age of Earth gave more time for gradual change in species, and the process of change provided an analogy for gradual change in species. In the early nineteenth century, Jean-Baptiste Lamarck published a book that detailed a mechanism for evolutionary change. This mechanism is now referred to as an inheritance of acquired characteristics by which modifications in an individual are caused by its environment, or the use or disuse of a structure during its lifetime, could be inherited by its offspring and thus bring about change in a species. While this mechanism for evolutionary change was discredited, Lamarck's ideas were an important influence on evolutionary thought.

## Charles Darwin and Natural Selection

In the mid-nineteenth century, the actual mechanism for evolution was independently conceived of and described by two naturalists: Charles Darwin and Alfred Russel Wallace. Importantly, each naturalist spent time exploring the natural world on expeditions to the tropics. From 1831 to 1836, Darwin traveled around the world on *H.M.S. Beagle*, including stops in South America, Australia, and the southern tip of Africa. Wallace traveled to Brazil to collect insects in the Amazon rainforest from 1848 to 1852 and to the Malay Archipelago from 1854 to 1862. Darwin's journey, like Wallace's later journeys to the Malay Archipelago, included stops at several island chains, the last being the Galápagos Islands west of Ecuador. On these islands, Darwin observed species of organisms on different islands that were clearly similar, yet had distinct differences. For example, the ground finches inhabiting the Galápagos Islands comprised several species with a unique beak shape (Figure **4.1.1**). The species on the islands had a graded series of beak sizes and shapes with very small differences between the most similar. He observed that these finches closely resembled another finch species on the mainland of South America. Darwin imagined that the island species might be species modified from one of the original mainland species. Upon further study, he realized that the varied beaks of each finch helped the birds acquire a specific type of food. For example, seed-eating finches had stronger, thicker beaks for breaking seeds, and insect-eating finches had spear-like beaks for stabbing their prey.

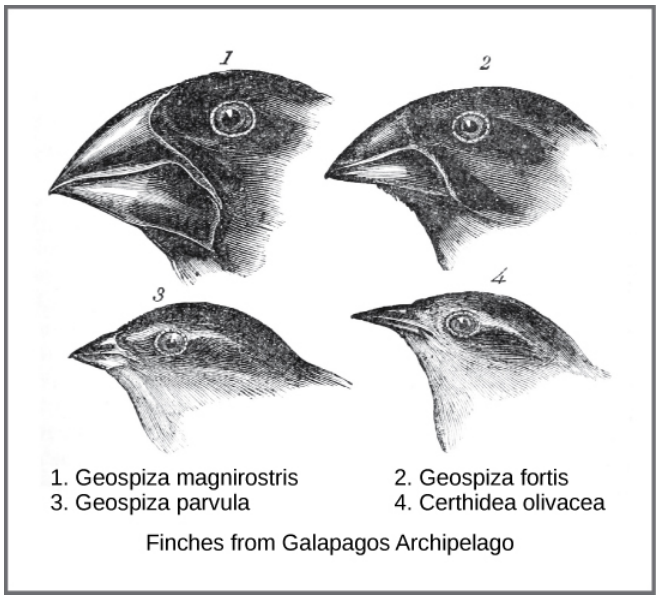


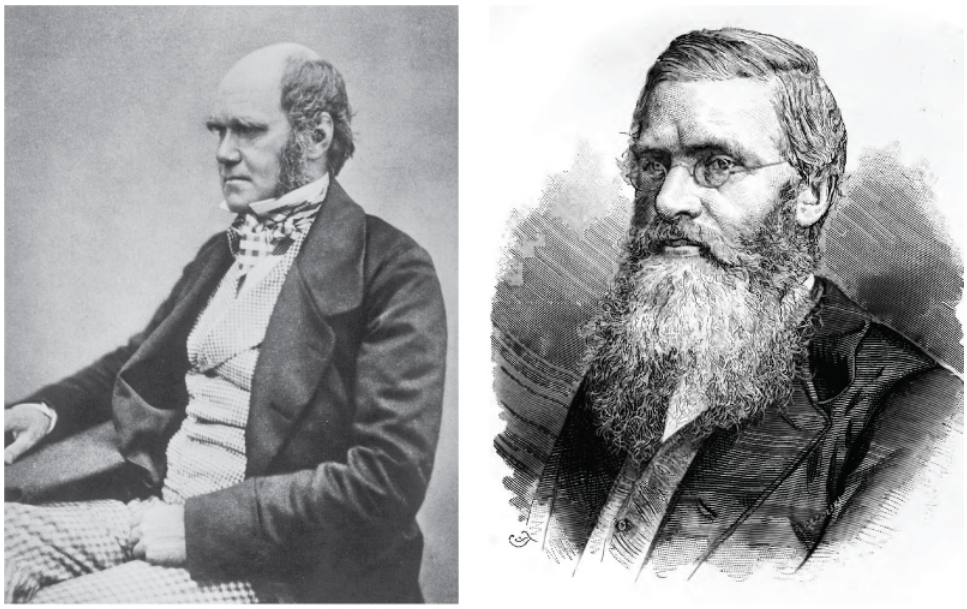

Figure **4.1.1**: Darwin observed that beak shape varies among finch species. He postulated that the beak of an ancestral species had adapted over time to equip the finches to acquire different food sources.

Wallace and Darwin both observed similar patterns in other organisms and they independently developed the same explanation for how and why such changes could take place. Darwin called this mechanism natural selection. Natural selection, also known as "survival of the fittest," is the more prolific reproduction of individuals with favorable traits that survive environmental change because of those traits; this leads to evolutionary change.

For example, a population of giant tortoises found in the Galapagos Archipelago was observed by Darwin to have longer necks than those that lived on other islands with dry lowlands. These tortoises were "selected" because they could reach more leaves and access more food than those with short necks. In times of drought when fewer leaves would be available, those that could reach more leaves had a better chance to eat and survive than those that couldn't reach the food source. Consequently, long-necked tortoises would be more likely to be reproductively successful and pass the long-necked trait to their offspring. Over time, only long-necked tortoises would be present in the population.

Natural selection, Darwin argued, was an inevitable outcome of three principles that operated in nature. First, most characteristics of organisms are inherited, or passed from parent to offspring. Although no one, including Darwin and Wallace, knew how this happened at the time, it was a common understanding. Second, more offspring are produced than are able to survive, so resources for survival and reproduction are limited. The capacity for reproduction in all organisms outstrips the availability of resources to support their numbers. Thus, there is competition for those resources in each generation. Both Darwin and Wallace's understanding of this principle came from reading an essay by the economist Thomas Malthus who discussed this principle in relation to human populations. Third, offspring vary among each other in regard to their characteristics and those variations are inherited. Darwin and Wallace reasoned that offspring with inherited characteristics which allow them to best compete for limited resources will survive and have more offspring than those individuals with variations that are less able to compete. Because characteristics are inherited, these traits will be better represented in the next generation. This will lead to change in populations over generations in a process that Darwin called descent with modification. Ultimately, natural selection leads to greater adaptation of the population to its local environment; it is the only mechanism known for adaptive evolution.

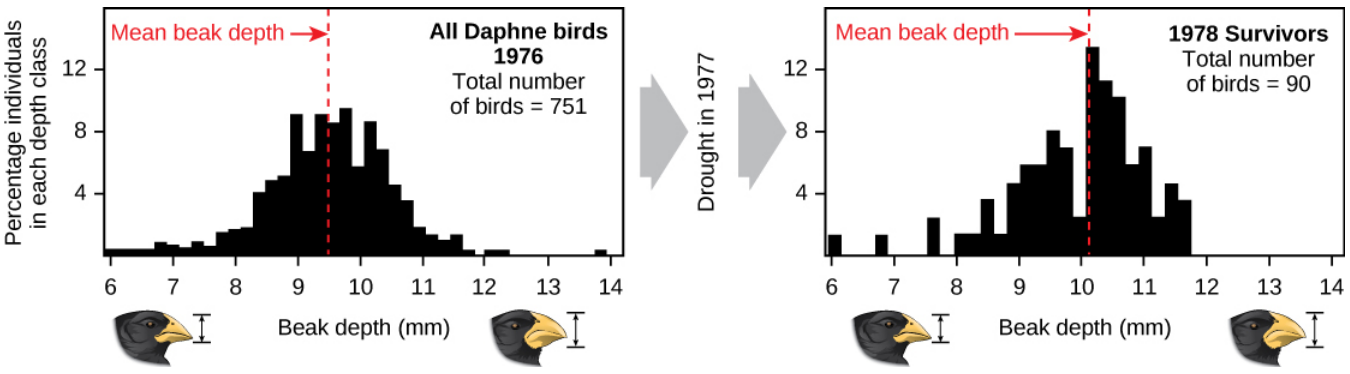
Papers by Darwin and Wallace (Figure **4.1.2**) presenting the idea of natural selection were read together in 1858 before the Linnean Society in London. The following year Darwin's book, *On the Origin of Species*, was published. His book outlined in considerable detail his arguments for evolution by natural selection.



(a) (b)

Figure **4.1.2**: Both (a) Charles Darwin and (b) Alfred Wallace wrote scientific papers on natural selection that were presented together before the Linnean Society in 1858.

Demonstrations of evolution by natural selection are time consuming and difficult to obtain. One of the best examples has been demonstrated in the very birds that helped to inspire Darwin's theory: the Galápagos finches. Peter and Rosemary Grant and their colleagues have studied Galápagos finch populations every year since 1976 and have provided important demonstrations of natural selection. The Grants found changes from one generation to the next in the distribution of beak shapes with the medium ground finch on the Galápagos island of Daphne Major. The medium ground finch feeds on seeds. The birds have inherited variation in the bill shape with some birds having wide deep bills and others having thinner bills. Large-billed birds feed more efficiently on large, hard seeds, whereas smaller billed birds feed more efficiently on small, soft seeds. During 1977, a drought period altered vegetation on the island. After this period, the number of seeds declined dramatically: the decline in small, soft seeds was greater than the decline in large, hard seeds. The large-billed birds were able to survive better than the small-billed birds the following year. The year following the drought when the Grants measured beak sizes in the much-reduced population, they found that the average bill size was larger ((Figure **4.1.3**)). This was clear evidence for natural selection (differences in survival) of bill size caused by the availability of seeds. The Grants had studied the inheritance of bill sizes and knew that the surviving large-billed birds would tend to produce offspring with larger bills, so the selection would lead to evolution of bill size.




**((Figure 4.1.3)):** A drought on the Galápagos island of Daphne Major in 1977 reduced the number of small seeds available to finches, causing many of the small-beaked finches to die. This caused an increase in the finches' average beak size between 1976 and 1978.

During a period in which rainfall was higher than normal because of an El Niño, the large hard seeds that large-billed birds ate were reduced in number; however, there was an abundance of the small soft seeds which the small-billed birds ate. Therefore,

survival and reproduction were much better in the following years for the small-billed birds. In the years following this El Niño, the Grants measured beak sizes in the population and found that the average bill size was smaller. Since bill size is an inherited trait, parents with smaller bills had more offspring and the size of bills had evolved to be smaller. As conditions improved in 1987 and larger seeds became more available, the trend toward smaller average bill size ceased. Subsequent studies by the Grants have demonstrated selection on and evolution of bill size in this species in response to changing conditions on the island. The evolution has occurred both to larger bills, as in this case, and to smaller bills when large seeds became rare.

### Career Connection: Field Biologist

Many people hike, explore caves, scuba dive, or climb mountains for recreation. People often participate in these activities hoping to see wildlife. Experiencing the outdoors can be incredibly enjoyable and invigorating. What if your job was to be outside in the wilderness? Field biologists by definition work outdoors in the "field." The term field in this case refers to any location outdoors, even under water. A field biologist typically focuses research on a certain species, group of organisms, or a single habitat (Figure **4.1.4**).



Figure **4.1.4**: A field biologist tranquilizes a polar bear for study. (credit: Karen Rhode)

One objective of many field biologists includes discovering new species that have never been recorded. Not only do such findings expand our understanding of the natural world, but they also lead to important innovations in fields such as medicine and agriculture. Plant and microbial species, in particular, can reveal new medicinal and nutritive knowledge. Other organisms can play key roles in ecosystems or be considered rare and in need of protection. When discovered, these important species can be used as evidence for environmental regulations and laws.

## Processes and Patterns of Evolution

Natural selection can only take place if there is **variation**, or differences, among individuals in a population. Importantly, these differences must have some genetic basis; otherwise, the selection will not lead to change in the next generation. This is critical because variation among individuals can be caused by non-genetic reasons such as an individual being taller because of better nutrition rather than different genes.

**Genetic diversity** in a population comes from two main mechanisms: **mutation** and **sexual reproduction**. Mutation, a change in DNA, is the ultimate source of new **alleles**, or new genetic variation in any population. The genetic changes caused by mutation can have one of three outcomes on the phenotype. A mutation affects the phenotype of the organism in a way that gives it reduced fitness—lower likelihood of survival or fewer offspring. A mutation may produce a phenotype with a beneficial effect on fitness. And, many mutations will also have no effect on the fitness of the phenotype; these are called neutral mutations. Mutations may also have a whole range of effect sizes on the fitness of the organism that expresses them in their phenotype, from a small effect to a great effect. Sexual reproduction also leads to genetic diversity: when two parents reproduce, unique combinations of alleles assemble to produce the unique genotypes and thus phenotypes in each of the offspring.

A **heritable trait** that helps the survival and reproduction of an organism in its present environment is called an **adaptation**. Scientists describe groups of organisms becoming adapted to their environment when a change in the range of genetic variation occurs over time that increases or maintains the "fit" of the population to its environment. The webbed feet of platypuses are an

adaptation for swimming. The snow leopards' thick fur is an adaptation for living in the cold. The cheetahs' fast speed is an adaptation for catching prey.

Whether or not a trait is favorable depends on the environmental conditions at the time. The same traits are not always selected because environmental conditions can change. For example, consider a species of plant that grew in a moist climate and did not need to conserve water. Large leaves were selected because they allowed the plant to obtain more energy from the sun. Large leaves require more water to maintain than small leaves, and the moist environment provided favorable conditions to support large leaves. After thousands of years, the climate changed, and the area no longer had excess water. The direction of natural selection shifted so that plants with small leaves were selected because those populations were able to conserve water to survive the new environmental conditions.

The evolution of species has resulted in enormous variation in form and function. Sometimes, evolution gives rise to groups of organisms that become tremendously different from each other. When two species evolve in diverse directions from a common point, it is called divergent evolution. Such divergent evolution can be seen in the forms of the reproductive organs of flowering plants which share the same basic anatomies; however, they can look very different as a result of selection in different physical environments and adaptation to different kinds of pollinators (Figure **4.1.5**).



(a) (b)

Figure **4.1.5**: Flowering plants evolved from a common ancestor. Notice that the (a) dense blazing star (*Liatrus spicata*) and the (b) purple coneflower (*Echinacea purpurea*) vary in appearance, yet both share a similar basic morphology. (credit a: modification of work by Drew Avery; credit b: modification of work by Cory Zanker)

In other cases, similar phenotypes evolve independently in distantly related species. For example, flight has evolved in both bats and insects, and they both have structures we refer to as wings, which are adaptations to flight. However, the wings of bats and insects have evolved from very different original structures. This phenomenon is called convergent evolution, where similar traits evolve independently in species that do not share a recent common ancestry. The two species came to the same function, flying, but did so separately from each other.

These physical changes occur over enormous spans of time and help explain how evolution occurs. Natural selection acts on individual organisms, which in turn can shape an entire species. Although natural selection may work in a single generation on an individual, it can take thousands or even millions of years for the genotype of an entire species to evolve. It is over these large time spans that life on earth has changed and continues to change.

## Evidence of Evolution

The evidence for evolution is compelling and extensive. Looking at every level of organization in living systems, biologists see the signature of past and present evolution. Darwin dedicated a large portion of his book, *On the Origin of Species*, to identifying patterns in nature that were consistent with evolution, and since Darwin, our understanding has become clearer and broader.

### Fossils

Fossils provide solid evidence that organisms from the past are not the same as those found today, and fossils show a progression of evolution. Scientists determine the age of fossils and categorize them from all over the world to determine when the organisms

lived relative to each other. The resulting fossil record tells the story of the past and shows the evolution of form over millions of years (Figure **4.1.6**). For example, scientists have recovered highly detailed records showing the evolution of humans and horses (Figure **4.1.6 – 7**). The whale flipper shares a similar morphology to appendages of birds and mammals (Figure **4.1.8**) indicating that these species share a common ancestor.




(a) (b)

Figure **4.1.6**: In this (a) display, fossil hominids are arranged from oldest (bottom) to newest (top). As hominids evolved, the shape of the skull changed. An artist's rendition of (b) extinct species of the genus *Equus* reveals that these ancient species resembled the modern horse (*Equus ferus*) but varied in size.

The fossil record of horses in North America is especially rich and many contain transition fossils: those showing intermediate anatomy between earlier and later forms (Figure **4.1.7**). The fossil record extends back to a dog-like ancestor some 55 million years ago that gave rise to the first horse-like species 55 to 42 million years ago in the genus *Eohippus*. The series of fossils tracks the change in anatomy resulting from a gradual drying trend that changed the landscape from a forested one to a prairie. Successive fossils show the evolution of teeth shapes and foot and leg anatomy to a grazing habit, with adaptations for escaping predators, for example in species of *Mesohippus* found from 40 to 30 million years ago. Later species showed gains in size, such as those of *Hipparion*, which existed from about 23 to 2 million years ago. The fossil record shows several adaptive radiations in the horse lineage, which is now much reduced to only one genus, *Equus*, with several species.

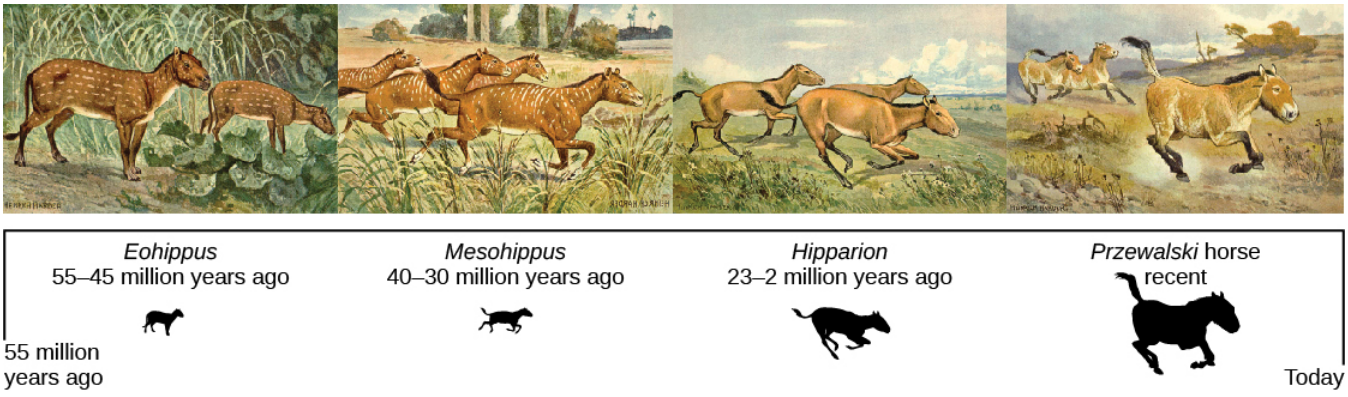



Figure **4.1.7**: This illustration shows an artist's renderings of these species derived from fossils of the evolutionary history of the horse and its ancestors. The species depicted are only four from a very diverse lineage that contains many branches, dead ends, and adaptive radiations. One of the trends, depicted here is the evolutionary tracking of a drying climate and increase in prairie versus forest habitat reflected in forms that are more adapted to grazing and predator escape through running. Przewalski's horse is one of a few living species of horse.

Another type of evidence for evolution is the presence of structures in organisms that share the same basic form. For example, the bones in the appendages of a human, dog, bird, and whale all share the same overall construction (Figure **4.1.8**). That similarity results from their origin in the appendages of a common ancestor. Over time, evolution led to changes in the shapes and sizes of these bones in different species, but they have maintained the same overall layout, evidence of descent from a common ancestor. Scientists call these synonymous parts homologous structures. Some structures exist in organisms that have no apparent function at all, and appear to be residual parts from a past ancestor. For example, some snakes have pelvic bones despite having no legs because they descended from reptiles that did have legs. These unused structures without function are called vestigial structures. Other examples of vestigial structures are wings on flightless birds (which may have other functions), leaves on some cacti, traces of pelvic bones in whales, and the sightless eyes of cave animals.

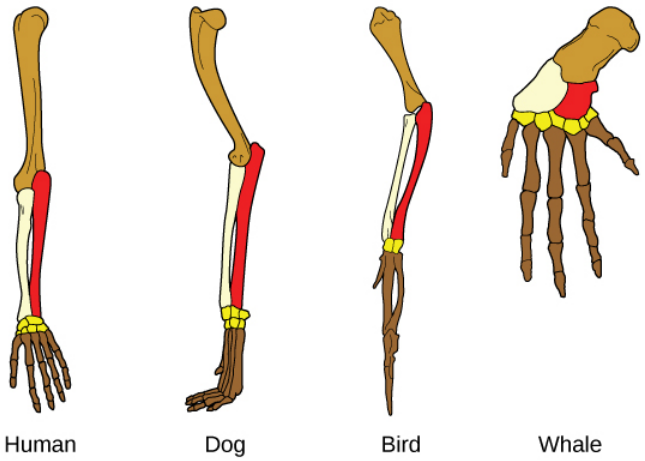



Figure **4.1.8**: The similar construction of these appendages indicates that these organisms share a common ancestor.

**Link to Learning**

Visit this interactive site to guess which bones structures are homologous and which are analogous, and see examples of evolutionary adaptations to illustrate these concepts.

Another evidence of evolution is the convergence of form in organisms that share similar environments. For example, species of unrelated animals, such as the arctic fox and ptarmigan, living in the arctic region have been selected for seasonal white phenotypes during winter to blend with the snow and ice (Figure **4.1.9**). These similarities occur not because of common ancestry, but because of similar selection pressures—the benefits of not being seen by predators.



(a) (b)

Figure **4.1.9**: The white winter coat of the (a) arctic fox and the (b) ptarmigan's plumage are adaptations to their environments. (credit a: modification of work by Keith Morehouse)

Embryology, the study of the development of the anatomy of an organism to its adult form, also provides evidence of relatedness between now widely divergent groups of organisms. Mutational tweaking in the embryo can have such magnified consequences in the adult that embryo formation tends to be conserved. As a result, structures that are absent in some groups often appear in their embryonic forms and disappear by the time the adult or juvenile form is reached. For example, all vertebrate embryos, including humans, exhibit gill slits and tails at some point in their early development. These disappear in the adults of terrestrial groups but are maintained in adult forms of aquatic groups such as fish and some amphibians. Great ape embryos, including humans, have a tail structure during their development that is lost by the time of birth.

## Biogeography

The geographic distribution of organisms on the planet follows patterns that are best explained by evolution in conjunction with the movement of tectonic plates over geological time. Broad groups that evolved before the breakup of the supercontinent Pangaea (about 200 million years ago) are distributed worldwide. Groups that evolved since the breakup appear uniquely in regions of the planet, such as the unique flora and fauna of northern continents that formed from the supercontinent Laurasia and of the southern continents that formed from the supercontinent Gondwana. The presence of members of the plant family Proteaceae in Australia, southern Africa, and South America is best explained by their presence prior to the southern supercontinent Gondwana breaking up (Figure **4.1.10**).

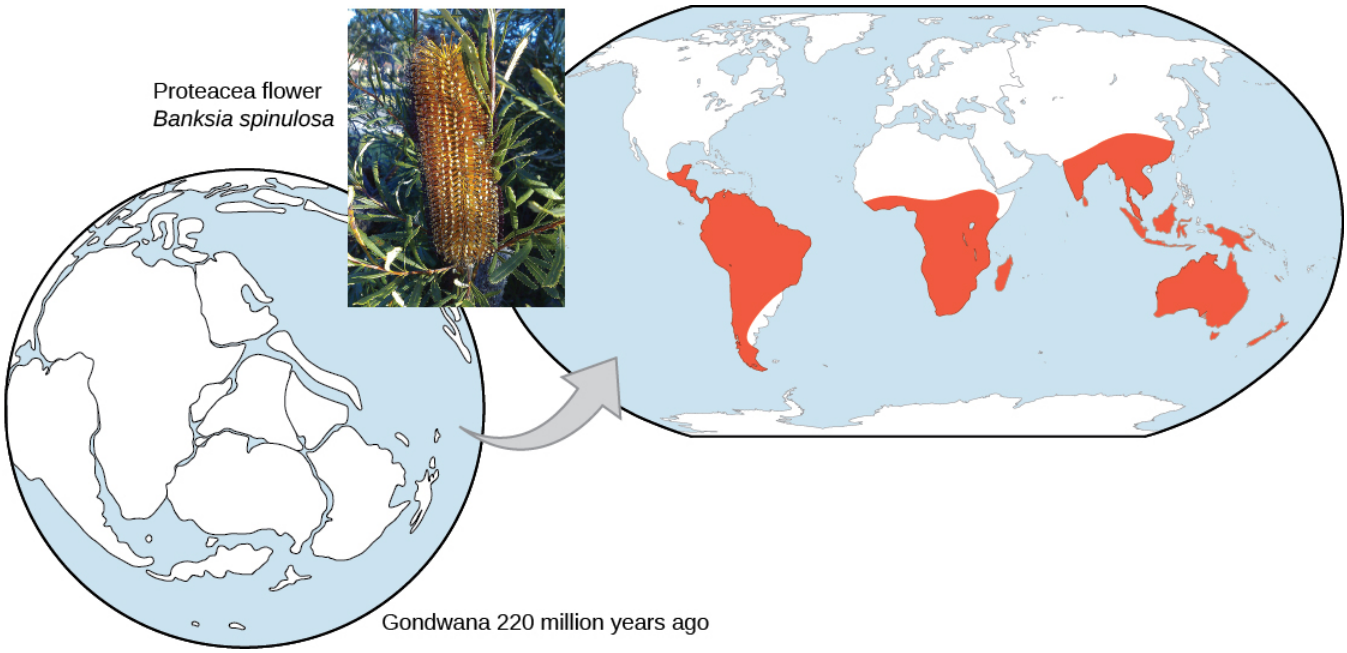



**Figure 4.1.10:** The Proteacea family of plants evolved before the supercontinent Gondwana broke up. Today, members of this plant family are found throughout the southern hemisphere (shown in red). (credit "Proteacea flower": modification of work by "dorofofoto"/Flickr)

The great diversification of marsupials in Australia and the absence of other mammals reflect Australia's long isolation. Australia has an abundance of endemic species—species found nowhere else—which is typical of islands whose isolation by expanses of water prevents species from migrating. Over time, these species diverge evolutionarily into new species that look very different from their ancestors that may exist on the mainland. The marsupials of Australia, the finches on the Galápagos, and many species on the Hawaiian Islands are all unique to their one point of origin, yet they display distant relationships to ancestral species on mainlands.

### Molecular Biology

Like anatomical structures, the structures of the molecules of life reflect descent with modification. Evidence of a common ancestor for all of life is reflected in the universality of DNA as the genetic material and in the near universality of the genetic code and the machinery of DNA replication and expression. Fundamental divisions in life between the three domains are reflected in major structural differences in otherwise conservative structures such as the components of ribosomes and the structures of membranes. In general, the relatedness of groups of organisms is reflected in the similarity of their DNA sequences—exactly the pattern that would be expected from descent and diversification from a common ancestor.

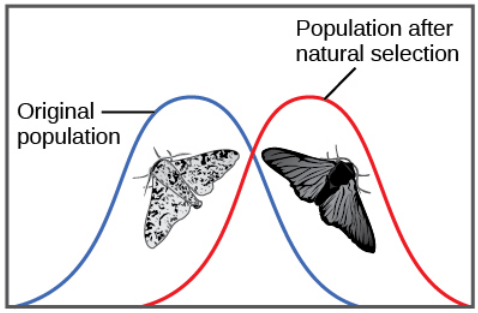
DNA sequences have also shed light on some of the mechanisms of evolution. For example, it is clear that the evolution of new functions for proteins commonly occurs after gene duplication events that allow the free modification of one copy by mutation, selection, or drift (changes in a population's gene pool resulting from chance), while the second copy continues to produce a functional protein.

## Population Genetics

Recall that a gene for a particular character may have several variants, or alleles, that code for different traits associated with that character. For example, in the ABO blood type system in humans, three alleles determine the particular blood-type protein on the surface of red blood cells. Each individual in a population of diploid organisms can only carry two alleles for a particular gene, but more than two may be present in the individuals that make up the population. Mendel followed alleles as they were inherited from parent to offspring. In the early twentieth century, biologists began to study what happens to all the alleles in a population in a field of study known as population genetics.

Until now, we have defined evolution as a change in the characteristics of a population of organisms, but behind that phenotypic change is genetic change. In population genetic terms, evolution is defined as a change in the frequency of an allele in a population. Using the ABO system as an example, the frequency of one of the alleles, $I^A$, is the number of copies of that allele divided by all the copies of the ABO gene in the population. For example, a study in Jordan found a frequency of $I^A$ to be 26.1 percent.[2] The $I^B$, $I^0$ alleles made up 13.4 percent and 60.5 percent of the alleles respectively, and all of the frequencies add up to 100 percent. A change in this frequency over time would constitute evolution in the population.

There are several ways the allele frequencies of a population can change. One of those ways is natural selection. If a given allele confers a phenotype that allows an individual to have more offspring that survive and reproduce, that allele, by virtue of being inherited by those offspring, will be in greater frequency in the next generation. Since allele frequencies always add up to 100 percent, an increase in the frequency of one allele always means a corresponding decrease in one or more of the other alleles. Highly beneficial alleles may, over a very few generations, become "fixed" in this way, meaning that every individual of the population will carry the allele. Similarly, detrimental alleles may be swiftly eliminated from the gene pool, the sum of all the alleles in a population. Part of the study of population genetics is tracking how selective forces change the allele frequencies in a population over time, which can give scientists clues regarding the selective forces that may be operating on a given population. The studies of changes in wing coloration in the peppered moth from mottled white to dark in response to soot-covered tree trunks and then back to mottled white when factories stopped producing so much soot is a classic example of studying evolution in natural populations (Figure **4.1.11**).




Light-colored peppered moths are better camouflaged against a pristine environment; likewise, dark-colored peppered moths are better camouflaged against a sooty environment. Thus, as the Industrial Revolution progressed in nineteenth-century England, the color of the moth population shifted from light to dark.

Figure **4.1.11**: As the Industrial Revolution caused trees to darken from soot, darker colored peppered moths were better camouflaged than the lighter colored ones, which caused there to be more of the darker colored moths in the population.

In the early twentieth century, English mathematician Godfrey Hardy and German physician Wilhelm Weinberg independently provided an explanation for a somewhat counterintuitive concept. Hardy's original explanation was in response to a misunderstanding as to why a "dominant" allele, one that masks a recessive allele, should not increase in frequency in a population until it eliminated all the other alleles. The question resulted from a common confusion about what "dominant" means, but it forced Hardy, who was not even a biologist, to point out that if there are no factors that affect an allele frequency those frequencies will remain constant from one generation to the next. This principle is now known as the Hardy-Weinberg equilibrium. The theory states that a population's allele and genotype frequencies are inherently stable—unless some kind of evolutionary force is acting on the population, the population would carry the same alleles in the same proportions generation after generation. Individuals would, as a whole, look essentially the same and this would be unrelated to whether the alleles were dominant or recessive. The four most important evolutionary forces, which will disrupt the equilibrium, are natural selection, mutation, genetic drift, and migration into or out of a population. A fifth factor, nonrandom mating, will also disrupt the Hardy-Weinberg equilibrium but only by shifting genotype frequencies, not allele frequencies. In nonrandom mating, individuals are more likely to mate with like individuals (or unlike individuals) rather than at random. Since nonrandom mating does not change allele frequencies, it does not cause evolution directly. Natural selection has been described. Mutation creates one allele out of another one and changes an allele's frequency by a small, but continuous amount each generation. Each allele is generated by a low, constant mutation rate that will slowly increase the allele's frequency in a population if no other forces act on the allele. If natural selection acts against the allele, it will be removed from the population at a low rate leading to a frequency that results from a balance between selection and mutation. This is one reason that genetic diseases remain in the human population at very low frequencies. If the allele is favored by selection, it will increase in frequency. Genetic drift causes random changes in allele frequencies when populations are small. Genetic drift can often be important in evolution, as discussed in the next section. Finally, if two populations of a species have different allele frequencies, migration of individuals between them will cause frequency changes in both populations. As it happens, there is no

population in which one or more of these processes are not operating, so populations are always evolving, and the Hardy-Weinberg equilibrium will never be exactly observed. However, the Hardy-Weinberg principle gives scientists a baseline expectation for allele frequencies in a non-evolving population to which they can compare evolving populations and thereby infer what evolutionary forces might be at play. The population is evolving if the frequencies of alleles or genotypes deviate from the value expected from the Hardy-Weinberg principle.

Darwin identified a special case of natural selection that he called sexual selection. Sexual selection affects an individual's ability to mate and thus produce offspring, and it leads to the evolution of dramatic traits that often appear maladaptive in terms of survival but persist because they give their owners greater reproductive success. Sexual selection occurs in two ways: through male–male competition for mates and through female selection of mates. Male–male competition takes the form of conflicts between males, which are often ritualized, but may also pose significant threats to a male's survival. Sometimes the competition is for territory, with females more likely to mate with males with higher quality territories. Female choice occurs when females choose a male based on a particular trait, such as feather colors, the performance of a mating dance, or the building of an elaborate structure. In some cases male–male competition and female choice combine in the mating process. In each of these cases, the traits selected for, such as fighting ability or feather color and length, become enhanced in the males. In general, it is thought that sexual selection can proceed to a point at which natural selection against a character's further enhancement prevents its further evolution because it negatively impacts the male's ability to survive. For example, colorful feathers or an elaborate display make the male more obvious to predators.

## Genetic Drift

Another way a population's allele frequencies can change is genetic drift (Figure **4.1.12**), which is simply the effect of chance. Genetic drift is most important in small populations. Drift would be completely absent in a population with infinite individuals, but, of course, no population is this large. Genetic drift occurs because the alleles in an offspring generation are a random sample of the alleles in the parent generation. Alleles may or may not make it into the next generation due to chance events including mortality of an individual, events affecting finding a mate, and even the events affecting which gametes end up in fertilizations. If one individual in a population of ten individuals happens to die before it leaves any offspring to the next generation, all of its genes—a tenth of the population's gene pool—will be suddenly lost. In a population of 100, that 1 individual represents only 1 percent of the overall gene pool; therefore, it has much less impact on the population's genetic structure and is unlikely to remove all copies of even a relatively rare allele.

Imagine a population of ten individuals, half with allele *A* and half with allele *a* (the individuals are haploid). In a stable population, the next generation will also have ten individuals. Choose that generation randomly by flipping a coin ten times and let heads be *A* and tails be *a*. It is unlikely that the next generation will have exactly half of each allele. There might be six of one and four of the other, or some different set of frequencies. Thus, the allele frequencies have changed and evolution has occurred. A coin will no longer work to choose the next generation (because the odds are no longer one half for each allele). The frequency in each generation will drift up and down on what is known as a random walk until at one point either all *A* or all *a* are chosen and that allele is fixed from that point on. This could take a very long time for a large population. This simplification is not very biological, but it can be shown that real populations behave this way. The effect of drift on frequencies is greater the smaller a population is. Its effect is also greater on an allele with a frequency far from one half. Drift will influence every allele, even those that are being naturally selected.

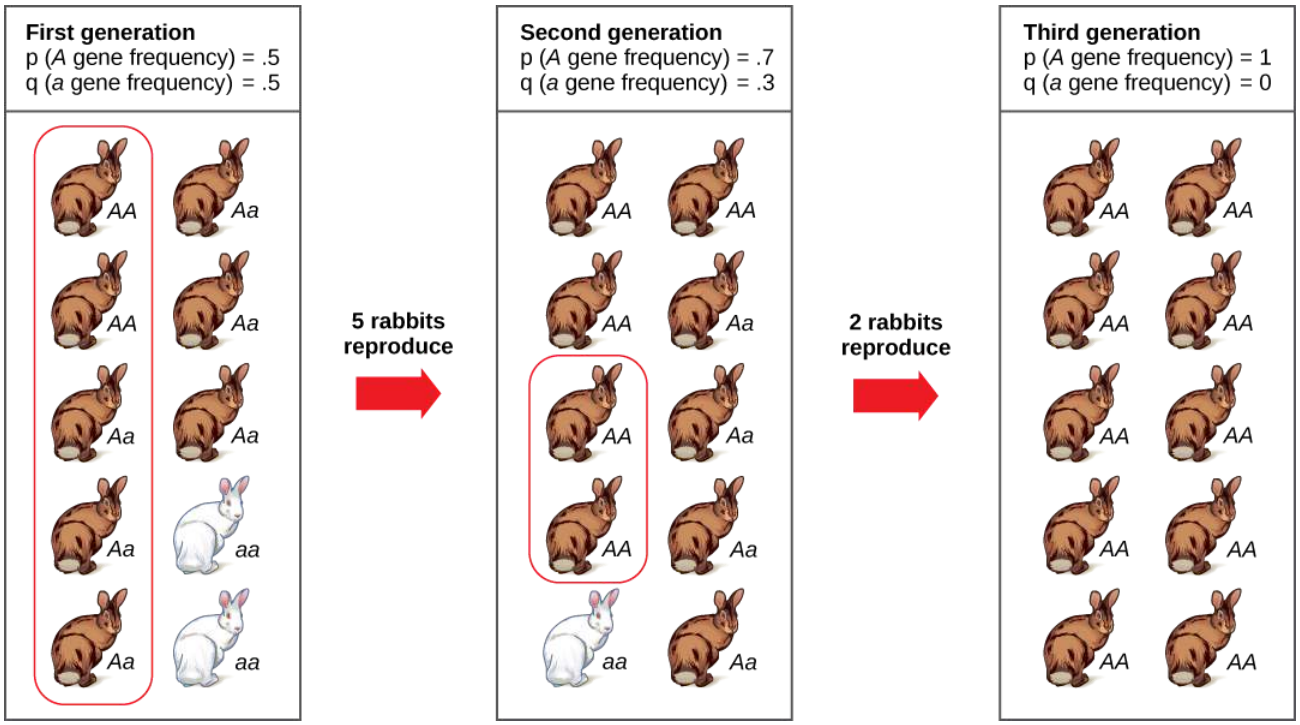



Figure **4.1.12**: Genetic drift in a population can lead to the elimination of an allele from a population by chance. In each generation, a random set of individuals reproduces to produce the next generation. The frequency of alleles in the next generation is equal to the frequency of alleles among the individuals reproducing.

Do you think genetic drift would happen more quickly on an island or on the mainland?

Genetic drift can also be magnified by natural or human-caused events, such as a disaster that randomly kills a large portion of the population, which is known as the bottleneck effect that results in a large portion of the genome suddenly being wiped out (Figure **4.1.13**). In one fell swoop, the genetic structure of the survivors becomes the genetic structure of the entire population, which may be very different from the pre-disaster population. The disaster must be one that kills for reasons unrelated to the organism's traits, such as a hurricane or lava flow. A mass killing caused by unusually cold temperatures at night, is likely to affect individuals differently depending on the alleles they possess that confer cold hardiness.

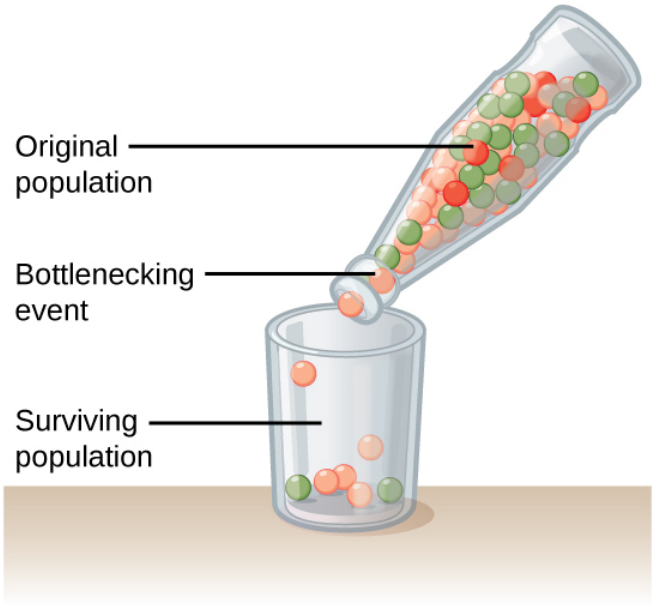



Figure **4.1.13**: A chance event or catastrophe can reduce the genetic variability within a population.

Another scenario in which populations might experience a strong influence of genetic drift is if some portion of the population leaves to start a new population in a new location, or if a population gets divided by a physical barrier of some kind. In this situation, those individuals are unlikely to be representative of the entire population which results in the founder effect. The founder effect occurs when the genetic structure matches that of the new population's founding fathers and mothers. The founder effect is believed to have been a key factor in the genetic history of the Afrikaner population of Dutch settlers in South Africa, as evidenced by mutations that are common in Afrikaners but rare in most other populations. This is likely due to a higher-than-normal proportion of the founding colonists, which were a small sample of the original population, carried these mutations. As a result, the population expresses unusually high incidences of Huntington's disease (HD) and Fanconi anemia (FA), a genetic disorder known to cause bone marrow and congenital abnormalities, and even cancer.[3]

Visit this site to learn more about genetic drift and to run simulations of allele changes caused by drift.

### Gene Flow

Another important evolutionary force is gene flow, or the flow of alleles in and out of a population resulting from the migration of individuals or gametes (Figure **4.1.14**). While some populations are fairly stable, others experience more flux. Many plants, for example, send their seeds far and wide, by wind or in the guts of animals; these seeds may introduce alleles common in the source population to a new population in which they are rare.

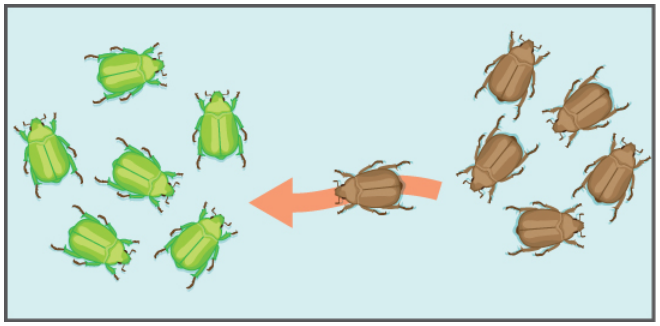


Figure **4.1.14**: Gene flow can occur when an individual travels from one geographic location to another and joins a different population of the species. In the example shown here, the brown allele is introduced into the green population.

## Misconceptions of Evolution

Although the theory of evolution generated some controversy when it was first proposed, it was almost universally accepted by biologists, particularly younger biologists, within 20 years after publication of *On the Origin of Species*. Nevertheless, the theory of evolution is a difficult concept and misconceptions about how it works abound

**Link to Learning**

This site addresses some of the main misconceptions associated with the theory of evolution.

**Evolution Is Just a Theory**

Critics of the theory of evolution dismiss its importance by purposefully confounding the everyday usage of the word "theory" with the way scientists use the word. In science, a "theory" is understood to be a body of thoroughly tested and verified explanations for a set of observations of the natural world. Scientists have a theory of the atom, a theory of gravity, and the theory of relativity, each of which describes understood facts about the world. In the same way, the theory of evolution describes facts about the living world. As such, a theory in science has survived significant efforts to discredit it by scientists. In contrast, a "theory" in common vernacular is a word meaning a guess or suggested explanation; this meaning is more akin to the scientific concept of "hypothesis." When critics of evolution say evolution is "just a theory," they are implying that there is little evidence supporting it and that it is still in the process of being rigorously tested. This is a mischaracterization.

**Individuals Evolve**

Evolution is the change in genetic composition of a population over time, specifically over generations, resulting from differential reproduction of individuals with certain alleles. Individuals do change over their lifetime, obviously, but this is called development

and involves changes programmed by the set of genes the individual acquired at birth in coordination with the individual's environment. When thinking about the evolution of a characteristic, it is probably best to think about the change of the average value of the characteristic in the population over time. For example, when natural selection leads to bill-size change in medium-ground finches in the Galápagos, this does not mean that individual bills on the finches are changing. If one measures the average bill size among all individuals in the population at one time and then measures the average bill size in the population several years later, this average value will be different as a result of evolution. Although some individuals may survive from the first time to the second, they will still have the same bill size; however, there will be many new individuals that contribute to the shift in average bill size.

**Evolution Explains the Origin of Life**

It is a common misunderstanding that evolution includes an explanation of life's origins. Conversely, some of the theory's critics believe that it cannot explain the origin of life. The theory does not try to explain the origin of life. The theory of evolution explains how populations change over time and how life diversifies the origin of species. It does not shed light on the beginnings of life including the origins of the first cells, which is how life is defined. The mechanisms of the origin of life on Earth are a particularly difficult problem because it occurred a very long time ago, and presumably it just occurred once. Importantly, biologists believe that the presence of life on Earth precludes the possibility that the events that led to life on Earth can be repeated because the intermediate stages would immediately become food for existing living things.

However, once a mechanism of inheritance was in place in the form of a molecule like DNA either within a cell or pre-cell, these entities would be subject to the principle of natural selection. More effective reproducers would increase in frequency at the expense of inefficient reproducers. So while evolution does not explain the origin of life, it may have something to say about some of the processes operating once pre-living entities acquired certain properties.

**Organisms Evolve on Purpose**

Statements such as "organisms evolve in response to a change in an environment" are quite common, but such statements can lead to two types of misunderstandings. First, the statement must not be understood to mean that individual organisms evolve. The statement is shorthand for "a population evolves in response to a changing environment." However, a second misunderstanding may arise by interpreting the statement to mean that the evolution is somehow intentional. A changed environment results in some individuals in the population, those with particular phenotypes, benefiting and therefore producing proportionately more offspring than other phenotypes. This results in change in the population if the characteristics are genetically determined.

It is also important to understand that the variation that natural selection works on is already in a population and does not arise in response to an environmental change. For example, applying antibiotics to a population of bacteria will, over time, select a population of bacteria that are resistant to antibiotics. The resistance, which is caused by a gene, did not arise by mutation because of the application of the antibiotic. The gene for resistance was already present in the gene pool of the bacteria, likely at a low frequency. The antibiotic, which kills the bacterial cells without the resistance gene, strongly selects individuals that are resistant, since these would be the only ones that survived and divided. Experiments have demonstrated that mutations for antibiotic resistance do not arise as a result of antibiotic.

In a larger sense, evolution is not goal directed. Species do not become "better" over time; they simply track their changing environment with adaptations that maximize their reproduction in a particular environment at a particular time. Evolution has no goal of making faster, bigger, more complex, or even smarter species, despite the commonness of this kind of language in popular discourse. What characteristics evolve in a species are a function of the variation present and the environment, both of which are constantly changing in a non-directional way. What trait is fit in one environment at one time may well be fatal at some point in the future. This holds equally well for a species of insect as it does the human species.

## References


1. 1 Charles Darwin, *Journal of Researches into the Natural History and Geology of the Countries Visited during the Voyage of H.M.S. Beagle Round the World, under the Command of Capt. Fitz Roy, R.N*, 2nd. ed. (London: John Murray, 1860), http://www.archive.org/details/journalofresea00darw.
2. 2 Sahar S. Hanania, Dhia S. Hassawi, and Nidal M. Irshaid, "Allele Frequency and Molecular Genotypes of ABO Blood Group System in a Jordanian Population," *Journal of Medical Sciences* 7 (2007): 51-58, doi:10.3923/jms.2007.51.58

3. 1 A. J. Tipping et al., “Molecular and Genealogical Evidence for a Founder Effect in Fanconi Anemia Families of the Afrikaner Population of South Africa,” *PNAS* 98, no. 10 (2001): 5734-5739, doi: 10.1073/pnas.091402398.

## Contributors and Attributions

- Modified from the Original by Kyle Whittinghill (University of Pittsburgh)
- 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).
- 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).
- **18.1: Understanding Evolution** by OpenStax, is licensed CC BY
- 11.3 Evidence of Evolution by OpenStax, is licensed CC BY
- 11.2 Mechanisms of Evolution by OpenStax, is licensed CC BY
- 11.1: Discovering How Populations Change by OpenStax, is licensed CC BY

This page titled 4.1: Understanding Evolution is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **18.1: Understanding Evolution** by OpenStax is licensed CC BY 4.0.

# 4.2: Formation of New Species

**Skills to Develop**

- Define species and describe how species are identified as different
- Describe genetic variables that lead to speciation
- Identify prezygotic and postzygotic reproductive barriers
- Explain allopatric and sympatric speciation
- Describe adaptive radiation

Although all life on earth shares various genetic similarities, only certain organisms combine genetic information by sexual reproduction and have offspring that can then successfully reproduce. Scientists call such organisms members of the same biological species.

## Species and the Ability to Reproduce

A species is a group of individual organisms that interbreed and produce fertile, viable offspring. According to this definition, one species is distinguished from another when, in nature, it is not possible for matings between individuals from each species to produce fertile offspring.

Members of the same species share both external and internal characteristics, which develop from their DNA. The closer relationship two organisms share, the more DNA they have in common, just like people and their families. People's DNA is likely to be more like their father or mother's DNA than their cousin or grandparent's DNA. Organisms of the same species have the highest level of DNA alignment and therefore share characteristics and behaviors that lead to successful reproduction.

Species' appearance can be misleading in suggesting an ability or inability to mate. For example, even though domestic dogs (*Canis lupus familiaris*) display phenotypic differences, such as size, build, and coat, most dogs can interbreed and produce viable puppies that can mature and sexually reproduce (Figure **4.2.1**).



(a) (b) (c)

Figure **4.2.1**: The (a) poodle and (b) cocker spaniel can reproduce to produce a breed known as (c) the cockapoo. (credit a: modification of work by Sally Eller, Tom Reese; credit b: modification of work by Jeremy McWilliams; credit c: modification of work by Kathleen Conklin)

In other cases, individuals may appear similar although they are not members of the same species. For example, even though bald eagles (*Haliaeetus leucocephalus*) and African fish eagles (*Haliaeetus vocifer*) are both birds and eagles, each belongs to a separate species group (Figure **4.2.2**). If humans were to artificially intervene and fertilize the egg of a bald eagle with the sperm of an African fish eagle and a chick did hatch, that offspring, called a hybrid (a cross between two species), would probably be infertile —unable to successfully reproduce after it reached maturity. Different species may have different genes that are active in development; therefore, it may not be possible to develop a viable offspring with two different sets of directions. Thus, even though hybridization may take place, the two species still remain separate.



(a) (b)

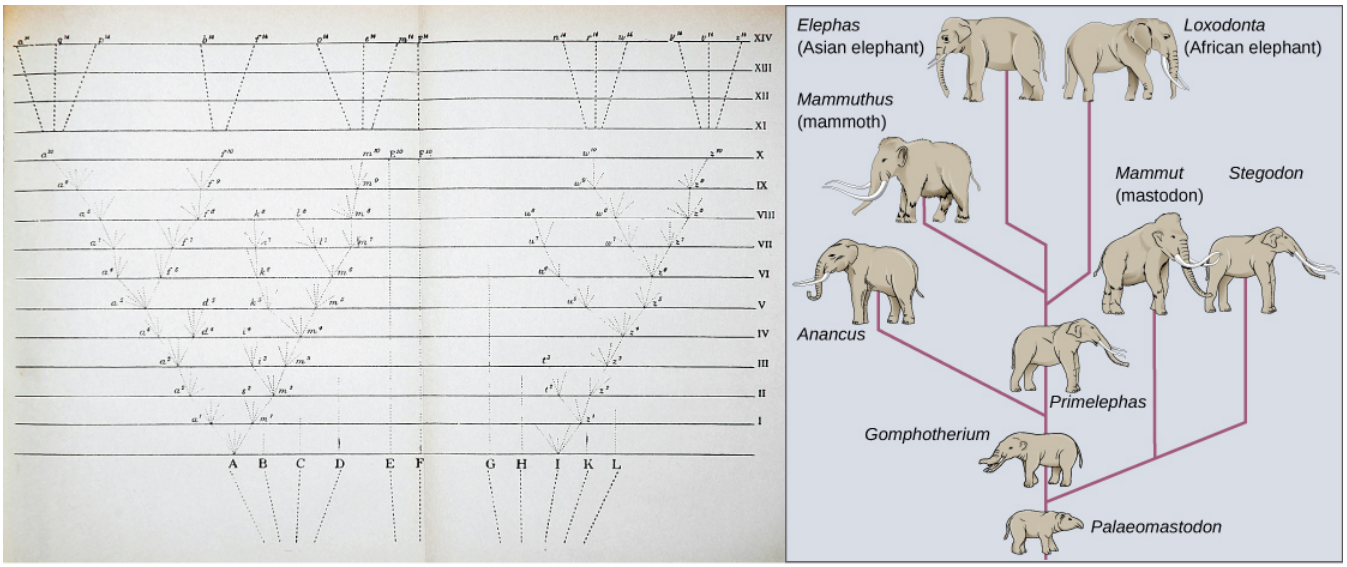
Figure **4.2.2**: The (a) African fish eagle is similar in appearance to the (b) bald eagle, but the two birds are members of different species. (credit a: modification of work by Nigel Wedge; credit b: modification of work by U.S. Fish and Wildlife Service)

Populations of species share a gene pool: a collection of all the variants of genes in the species. Again, the basis to any changes in a group or population of organisms must be genetic for this is the only way to share and pass on traits. When variations occur within a species, they can only be passed to the next generation along two main pathways: asexual reproduction or sexual reproduction. The change will be passed on asexually simply if the reproducing cell possesses the changed trait. For the changed trait to be passed on by sexual reproduction, a gamete, such as a sperm or egg cell, must possess the changed trait. In other words, sexually-reproducing organisms can experience several genetic changes in their body cells, but if these changes do not occur in a sperm or egg cell, the changed trait will never reach the next generation. Only heritable traits can evolve. Therefore, reproduction plays a paramount role for genetic change to take root in a population or species. In short, organisms must be able to reproduce with each other to pass new traits to offspring.

## Speciation

The biological definition of species, which works for sexually reproducing organisms, is a group of actually or potentially interbreeding individuals. There are exceptions to this rule. Many species are similar enough that hybrid offspring are possible and may often occur in nature, but for the majority of species this rule generally holds. In fact, the presence in nature of hybrids between similar species suggests that they may have descended from a single interbreeding species, and the speciation process may not yet be completed.

Given the extraordinary diversity of life on the planet there must be mechanisms for speciation: the formation of two species from one original species. Darwin envisioned this process as a branching event and diagrammed the process in the only illustration found in *On the Origin of Species* (Figure **4.2.3**a). Compare this illustration to the diagram of elephant evolution (Figure **4.2.3**b), which shows that as one species changes over time, it branches to form more than one new species, repeatedly, as long as the population survives or until the organism becomes extinct.




(a) (b)

Figure **4.2.3**: The only illustration in Darwin's *On the Origin of Species* is (a) a diagram showing speciation events leading to biological diversity. The diagram shows similarities to phylogenetic charts that are drawn today to illustrate the relationships of species. (b) Modern elephants evolved from the Palaeomastodon, a species that lived in Egypt 35–50 million years ago.

For speciation to occur, two new populations must be formed from one original population and they must evolve in such a way that it becomes impossible for individuals from the two new populations to interbreed. Biologists have proposed mechanisms by which this could occur that fall into two broad categories. Allopatric speciation (allo- = "other"; -patric = "homeland") involves geographic separation of populations from a parent species and subsequent evolution. Sympatric speciation (sym- = "same"; -patric = "homeland") involves speciation occurring within a parent species remaining in one location.

Biologists think of speciation events as the splitting of one ancestral species into two descendant species. There is no reason why there might not be more than two species formed at one time except that it is less likely and multiple events can be conceptualized as single splits occurring close in time.

## Allopatric Speciation

A geographically continuous population has a gene pool that is relatively homogeneous. Gene flow, the movement of alleles across the range of the species, is relatively free because individuals can move and then mate with individuals in their new location. Thus, the frequency of an allele at one end of a distribution will be similar to the frequency of the allele at the other end. When populations become geographically discontinuous, that free-flow of alleles is prevented. When that separation lasts for a period of time, the two populations are able to evolve along different trajectories. Thus, their allele frequencies at numerous genetic loci gradually become more and more different as new alleles independently arise by mutation in each population. Typically, environmental conditions, such as climate, resources, predators, and competitors for the two populations will differ causing natural selection to favor divergent adaptations in each group.

Isolation of populations leading to allopatric speciation can occur in a variety of ways: a river forming a new branch, erosion forming a new valley, a group of organisms traveling to a new location without the ability to return, or seeds floating over the ocean to an island. The nature of the geographic separation necessary to isolate populations depends entirely on the biology of the organism and its potential for dispersal. If two flying insect populations took up residence in separate nearby valleys, chances are, individuals from each population would fly back and forth continuing gene flow. However, if two rodent populations became divided by the formation of a new lake, continued gene flow would be unlikely; therefore, speciation would be more likely.

Biologists group allopatric processes into two categories: dispersal and vicariance. Dispersal is when a few members of a species move to a new geographical area, and vicariance is when a natural situation arises to physically divide organisms.

Scientists have documented numerous cases of allopatric speciation taking place. For example, along the west coast of the United States, two separate sub-species of spotted owls exist. The northern spotted owl has genetic and phenotypic differences from its close relative: the Mexican spotted owl, which lives in the south (Figure **4.2.4**).

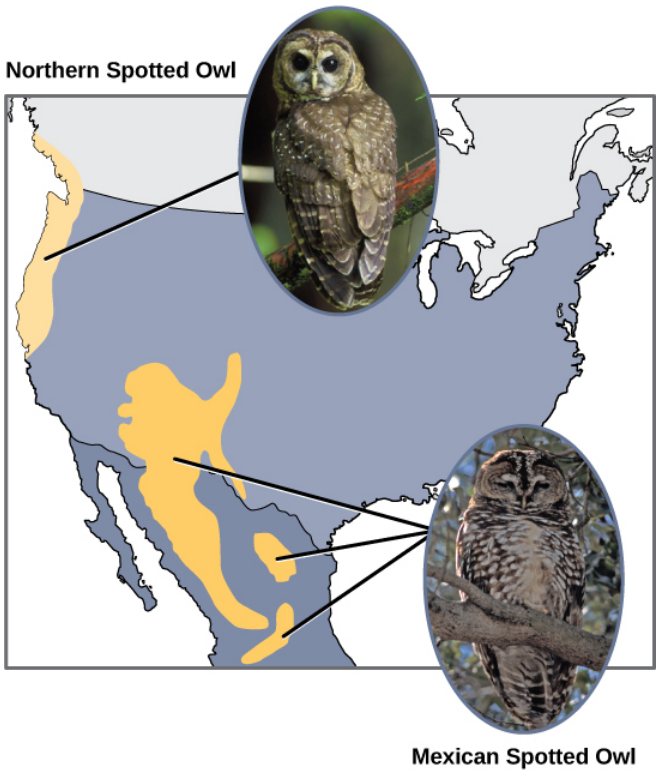



Figure **4.2.4**: The northern spotted owl and the Mexican spotted owl inhabit geographically separate locations with different climates and ecosystems. The owl is an example of allopatric speciation. (credit "northern spotted owl": modification of work by John and Karen Hollingsworth; credit "Mexican spotted owl": modification of work by Bill Radke)

Additionally, scientists have found that the further the distance between two groups that once were the same species, the more likely it is that speciation will occur. This seems logical because as the distance increases, the various environmental factors would likely have less in common than locations in close proximity. Consider the two owls: in the north, the climate is cooler than in the south; the types of organisms in each ecosystem differ, as do their behaviors and habits; also, the hunting habits and prey choices of the southern owls vary from the northern owls. These variances can lead to evolved differences in the owls, and speciation likely will occur.

## Adaptive Radiation

In some cases, a population of one species disperses throughout an area, and each finds a distinct niche or isolated habitat. Over time, the varied demands of their new lifestyles lead to multiple speciation events originating from a single species. This is called adaptive radiation because many adaptations evolve from a single point of origin; thus, causing the species to radiate into several new ones. Island archipelagos like the Hawaiian Islands provide an ideal context for adaptive radiation events because water surrounds each island which leads to geographical isolation for many organisms. The Hawaiian honeycreeper illustrates one example of adaptive radiation. From a single species, called the founder species, numerous species have evolved, including the six shown in Figure **4.2.5**.

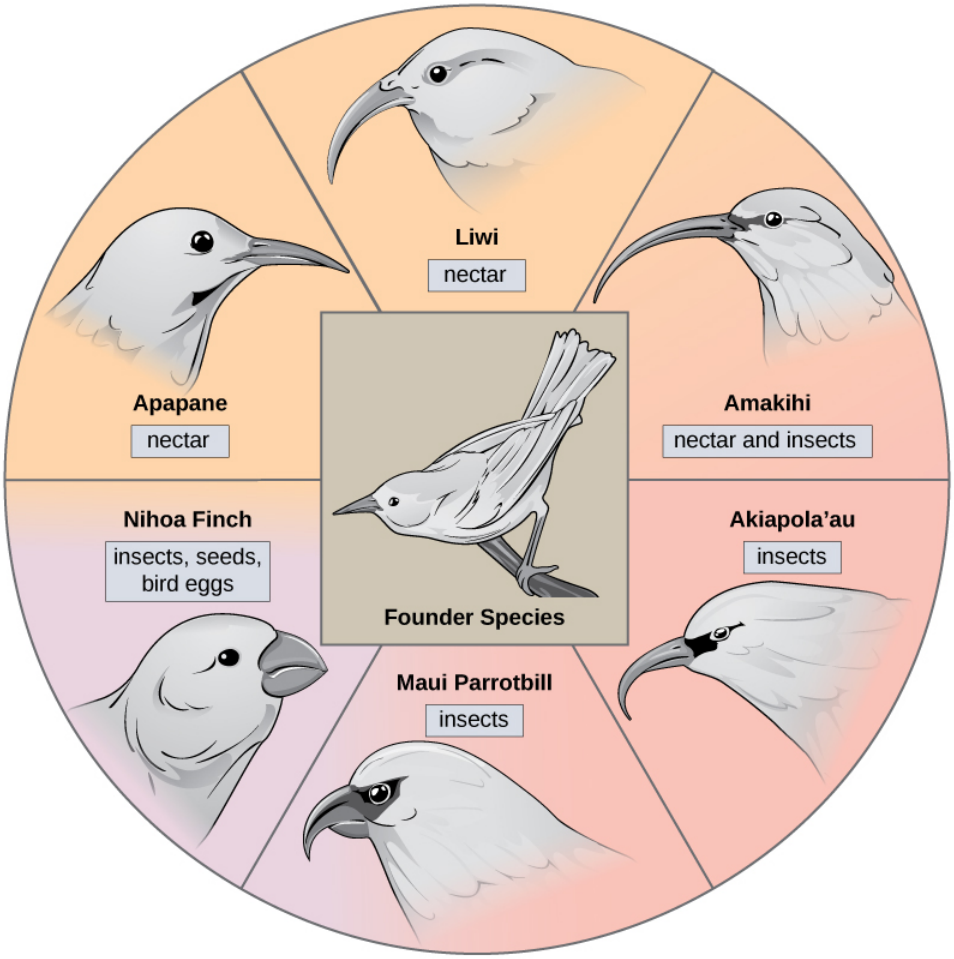



Figure 4.2.5: The honeycreeper birds illustrate adaptive radiation. From one original species of bird, multiple others evolved, each with its own distinctive characteristics.

Notice the differences in the species' beaks in Figure 4.2.5. Evolution in response to natural selection based on specific food sources in each new habitat led to evolution of a different beak suited to the specific food source. The seed-eating bird has a thicker, stronger beak which is suited to break hard nuts. The nectar-eating birds have long beaks to dip into flowers to reach the nectar. The insect-eating birds have beaks like swords, appropriate for stabbing and impaling insects. Darwin's finches are another example of adaptive radiation in an archipelago.

Link to Learning



Click through this interactive site to see how island birds evolved in evolutionary increments from 5 million years ago to today.

## Sympatric Speciation

Can divergence occur if no physical barriers are in place to separate individuals who continue to live and reproduce in the same habitat? The answer is yes. The process of speciation within the same space is called *sympatric speciation*; the prefix "sym" means same, so "sympatric" means "same homeland" in contrast to "allopatric" meaning "other homeland." A number of mechanisms for sympatric speciation have been proposed and studied.

One form of sympatric speciation can begin with a serious chromosomal error during cell division. In a normal cell division event chromosomes replicate, pair up, and then separate so that each new cell has the same number of chromosomes. However, sometimes the pairs separate and the end cell product has too many or too few individual chromosomes in a condition called **aneuploidy** (Figure 4.2.6).

Art Connection

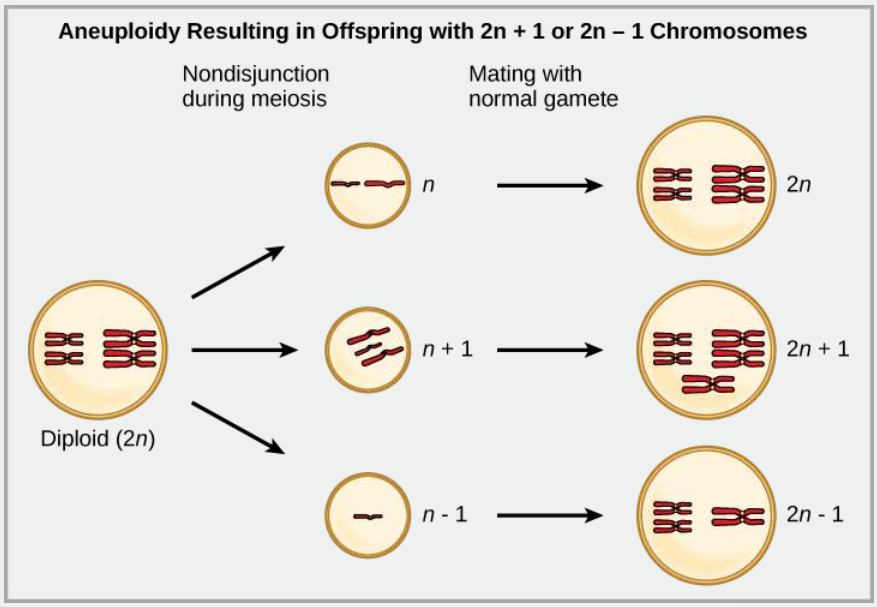



Figure 4.2.6: Aneuploidy results when the gametes have too many or too few chromosomes due to nondisjunction during meiosis. In the example shown here, the resulting offspring will have 2*n*+1 or 2*n*-1 chromosomes.

Which is most likely to survive, offspring with 2*n*+1 chromosomes or offspring with 2*n*-1 chromosomes?

Polyploidy is a condition in which a cell or organism has an extra set, or sets, of chromosomes. Scientists have identified two main types of polyploidy that can lead to reproductive isolation of an individual in the polyploidy state. Reproductive isolation is the inability to interbreed. In some cases, a polyploid individual will have two or more complete sets of chromosomes from its own species in a condition called autopolyploidy (Figure 4.2.7). The prefix "auto-" means "self," so the term means multiple chromosomes from one's own species. Polyploidy results from an error in meiosis in which all of the chromosomes move into one cell instead of separating.

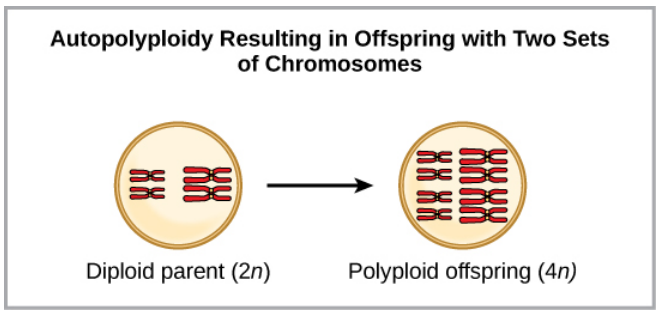



Figure 4.2.7: Autopolyploidy results when mitosis is not followed by cytokinesis.

For example, if a plant species with $2n = 6$ produces autopolyploid gametes that are also diploid ($2n = 6$, when they should be $n = 3$), the gametes now have twice as many chromosomes as they should have. These new gametes will be incompatible with the normal gametes produced by this plant species. However, they could either self-pollinate or reproduce with other autopolyploid plants with gametes having the same diploid number. In this way, sympatric speciation can occur quickly by forming offspring with $4n$ called a tetraploid. These individuals would immediately be able to reproduce only with those of this new kind and not those of the ancestral species.

The other form of polyploidy occurs when individuals of two different species reproduce to form a viable offspring called an allopolyploid. The prefix "allo-" means "other" (recall from allopatric): therefore, an allopolyploid occurs when gametes from two different species combine. Figure 4.2.8 illustrates one possible way an allopolyploid can form. Notice how it takes two generations, or two reproductive acts, before the viable fertile hybrid results.

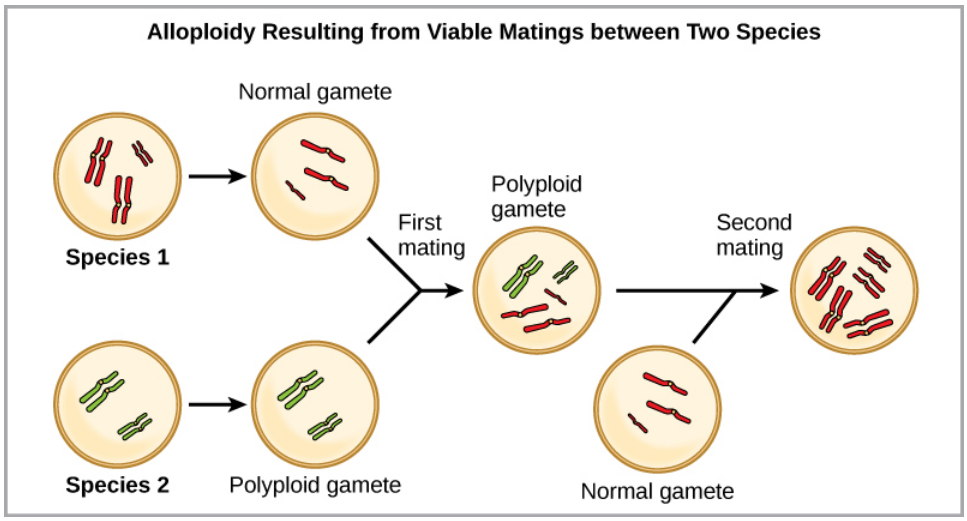



Figure 4.2.8: Alloploidy results when two species mate to produce viable offspring. In the example shown, a normal gamete from one species fuses with a polyploid gamete from another. Two matings are necessary to produce viable offspring.

The cultivated forms of wheat, cotton, and tobacco plants are all allopolyploids. Although polyploidy occurs occasionally in animals, it takes place most commonly in plants. (Animals with any of the types of chromosomal aberrations described here are unlikely to survive and produce normal offspring.) Scientists have discovered more than half of all plant species studied relate back to a species evolved through polyploidy. With such a high rate of polyploidy in plants, some scientists hypothesize that this mechanism takes place more as an adaptation than as an error.

## Reproductive Isolation

Given enough time, the genetic and phenotypic divergence between populations will affect characters that influence reproduction: if individuals of the two populations were to be brought together, mating would be less likely, but if mating occurred, offspring would be non-viable or infertile. Many types of diverging characters may affect the reproductive isolation, the ability to interbreed, of the two populations.

Reproductive isolation can take place in a variety of ways. Scientists organize them into two groups: prezygotic barriers and postzygotic barriers. Recall that a zygote is a fertilized egg: the first cell of the development of an organism that reproduces sexually. Therefore, a prezygotic barrier is a mechanism that blocks reproduction from taking place; this includes barriers that prevent fertilization when organisms attempt reproduction. A postzygotic barrier occurs after zygote formation; this includes organisms that don't survive the embryonic stage and those that are born sterile.

Some types of prezygotic barriers prevent reproduction entirely. Many organisms only reproduce at certain times of the year, often just annually. Differences in breeding schedules, called temporal isolation, can act as a form of reproductive isolation. For example, two species of frogs inhabit the same area, but one reproduces from January to March, whereas the other reproduces from March to May (Figure 4.2.9).



(a) (b)

Figure **4.2.9**: These two related frog species exhibit temporal reproductive isolation. (a) *Rana aurora* breeds earlier in the year than (b) *Rana boylii*. (credit a: modification of work by Mark R. Jennings, USFWS; credit b: modification of work by Alessandro Catenazzi)

In some cases, populations of a species move or are moved to a new habitat and take up residence in a place that no longer overlaps with the other populations of the same species. This situation is called habitat isolation. Reproduction with the parent species ceases, and a new group exists that is now reproductively and genetically independent. For example, a cricket population that was divided after a flood could no longer interact with each other. Over time, the forces of natural selection, mutation, and genetic drift will likely result in the divergence of the two groups (Figure **4.2.10**).

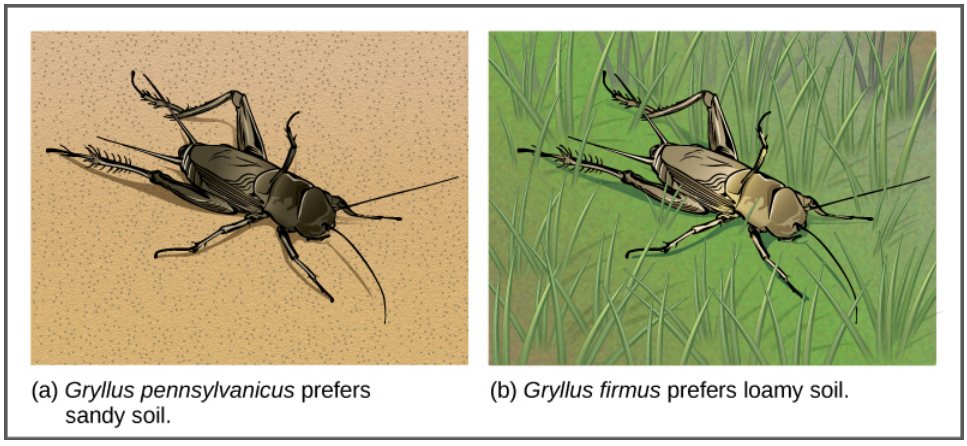



Figure **4.2.10**: Speciation can occur when two populations occupy different habitats. The habitats need not be far apart. The cricket (a) *Gryllus pennsylvanicus* prefers sandy soil, and the cricket (b) *Gryllus firmus* prefers loamy soil. The two species can live in close proximity, but because of their different soil preferences, they became genetically isolated.

Behavioral isolation occurs when the presence or absence of a specific behavior prevents reproduction from taking place. For example, male fireflies use specific light patterns to attract females. Various species of fireflies display their lights differently. If a male of one species tried to attract the female of another, she would not recognize the light pattern and would not mate with the male.

Other prezygotic barriers work when differences in their gamete cells (eggs and sperm) prevent fertilization from taking place; this is called a gametic barrier. Similarly, in some cases closely related organisms try to mate, but their reproductive structures simply do not fit together. For example, damselfly males of different species have differently shaped reproductive organs. If one species tries to mate with the female of another, their body parts simply do not fit together (Figure **4.2.11**).

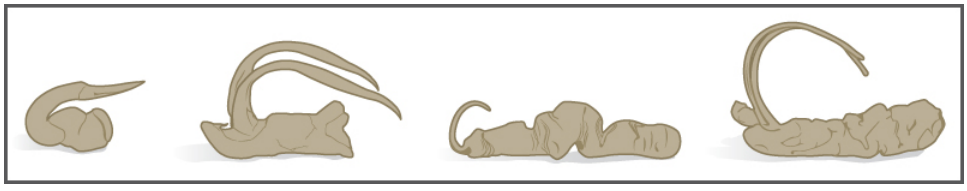


Figure **4.2.11**: The shape of the male reproductive organ varies among male damselfly species, and is only compatible with the female of that species. Reproductive organ incompatibility keeps the species reproductively isolated.

In plants, certain structures aimed to attract one type of pollinator simultaneously prevent a different pollinator from accessing the pollen. The tunnel through which an animal must access nectar can vary widely in length and diameter, which prevents the plant from being cross-pollinated with a different species (Figure 4.2.12).



(a) Honeybee drinking nectar from a foxglove flower

(b) Ruby-throated hummingbird drinking nectar from a trumpet creeper flower

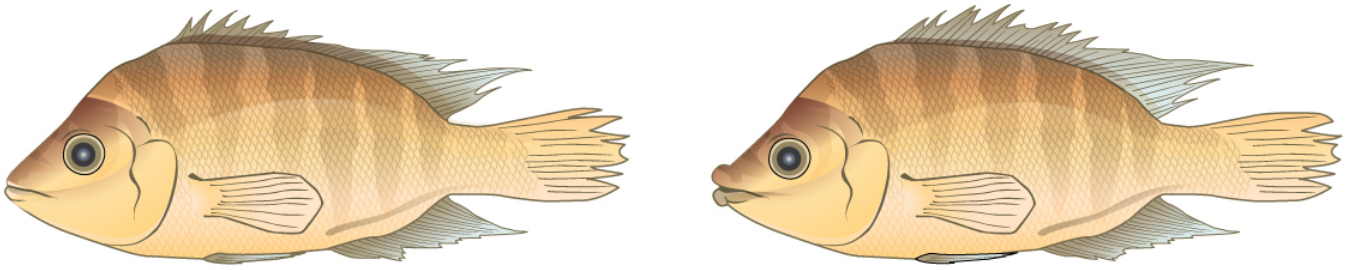
Figure 4.2.12: Some flowers have evolved to attract certain pollinators. The (a) wide foxglove flower is adapted for pollination by bees, while the (b) long, tube-shaped trumpet creeper flower is adapted for pollination by humming birds.

When fertilization takes place and a zygote forms, postzygotic barriers can prevent reproduction. Hybrid individuals in many cases cannot form normally in the womb and simply do not survive past the embryonic stages. This is called hybrid inviability because the hybrid organisms simply are not viable. In another postzygotic situation, reproduction leads to the birth and growth of a hybrid that is sterile and unable to reproduce offspring of their own; this is called hybrid sterility.

## Habitat Influence on Speciation

Sympatric speciation may also take place in ways other than polyploidy. For example, consider a species of fish that lives in a lake. As the population grows, competition for food also grows. Under pressure to find food, suppose that a group of these fish had the genetic flexibility to discover and feed off another resource that was unused by the other fish. What if this new food source was found at a different depth of the lake? Over time, those feeding on the second food source would interact more with each other than the other fish; therefore, they would breed together as well. Offspring of these fish would likely behave as their parents: feeding and living in the same area and keeping separate from the original population. If this group of fish continued to remain separate from the first population, eventually sympatric speciation might occur as more genetic differences accumulated between them.

This scenario does play out in nature, as do others that lead to reproductive isolation. One such place is Lake Victoria in Africa, famous for its sympatric speciation of cichlid fish. Researchers have found hundreds of sympatric speciation events in these fish, which have not only happened in great number, but also over a short period of time. Figure 4.2.13 shows this type of speciation among a cichlid fish population in Nicaragua. In this locale, two types of cichlids live in the same geographic location but have come to have different morphologies that allow them to eat various food sources.



Thin-lipped cichlid

Thick-lipped cichlid

Figure 4.2.13: Cichlid fish from Lake Apoyeque, Nicaragua, show evidence of sympatric speciation. Lake Apoyeque, a crater lake, is 1800 years old, but genetic evidence indicates that the lake was populated only 100 years ago by a single population of cichlid fish. Nevertheless, two populations with distinct morphologies and diets now exist in the lake, and scientists believe these populations may be in an early stage of speciation.

## Summary

Speciation occurs along two main pathways: geographic separation (allopatric speciation) and through mechanisms that occur within a shared habitat (sympatric speciation). Both pathways isolate a population reproductively in some form. Mechanisms of reproductive isolation act as barriers between closely related species, enabling them to diverge and exist as genetically independent species. Prezygotic barriers block reproduction prior to formation of a zygote, whereas postzygotic barriers block reproduction after fertilization occurs. For a new species to develop, something must cause a breach in the reproductive barriers. Sympatric speciation can occur through errors in meiosis that form gametes with extra chromosomes (polyploidy). Autopolyploidy occurs within a single species, whereas allopolyploidy occurs between closely related species.

## Art Connections

Figure **4.2.6**: Which is most likely to survive, offspring with $2n+1$ chromosomes or offspring with $2n-1$ chromosomes?

**Answer**

Loss of genetic material is almost always lethal, so offspring with $2n+1$ chromosomes are more likely to survive.

## Glossary

**adaptive radiation**
speciation when one species radiates out to form several other species

**allopatric speciation**
speciation that occurs via geographic separation

**allopolyploid**
polyploidy formed between two related, but separate species

**aneuploidy**
condition of a cell having an extra chromosome or missing a chromosome for its species

**autopolyploid**
polyploidy formed within a single species

**behavioral isolation**
type of reproductive isolation that occurs when a specific behavior or lack of one prevents reproduction from taking place

**dispersal**
allopatric speciation that occurs when a few members of a species move to a new geographical area

**gametic barrier**
prezygotic barrier occurring when closely related individuals of different species mate, but differences in their gamete cells (eggs and sperm) prevent fertilization from taking place

**habitat isolation**
reproductive isolation resulting when populations of a species move or are moved to a new habitat, taking up residence in a place that no longer overlaps with the other populations of the same species

**hybrid**
offspring of two closely related individuals, not of the same species

**postzygotic barrier**
reproductive isolation mechanism that occurs after zygote formation

**prezygotic barrier**

reproductive isolation mechanism that occurs before zygote formation

**reproductive isolation**

situation that occurs when a species is reproductively independent from other species; this may be brought about by behavior, location, or reproductive barriers

**speciation**

formation of a new species

**species**

group of populations that interbreed and produce fertile offspring

**sympatric speciation**

speciation that occurs in the same geographic space

**temporal isolation**

differences in breeding schedules that can act as a form of prezygotic barrier leading to reproductive isolation

**vicariance**

allopatric speciation that occurs when something in the environment separates organisms of the same species into separate groups

---

This page titled 4.2: Formation of New Species is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **18.2: Formation of New Species** by OpenStax is licensed CC BY 4.0.

# CHAPTER OVERVIEW

## 5: Cycling of Matter in the Earth System

Like a generating plant, plants and animals also must take in energy from the environment and convert it into a form that their cells can use. Energy enters an organism's body in one form and is converted into another form that can fuel the organism's life functions. In the process of photosynthesis, plants and other photosynthetic producers take in energy in the form of light (solar energy) and convert it into chemical energy, glucose, which stores this energy in its chemical bonds. Then, a series of metabolic pathways, collectively called cellular respiration, extracts the energy from the bonds in glucose and converts it into a form that all living things can use—both producers, such as plants, and consumers, such as animals.


5.1: Introduction - Systems, Cycles, Reservoirs, and Fluxes
5.2: Water (Hydrologic) Cycle
5.3: Carbon Cycle
5.4: Photosynthesis
5.5: Nitrogen Cycle
5.6: Phosphorus Cycle
5.7: Eutrophication and Dead Zones
5.8: Sulfur Cycle


### Contributors and Attributions


- 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).



This page titled 5: Cycling of Matter in the Earth System is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

# 5.1: Introduction - 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***. Earth system science includes many aspects of Environmental Science including hydrologic cycles and biogeochemical cycles which we will discuss in this chapter.

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. We can think about cycles at global scales or at the scale of a single organism. It is important to remember that while matter and energy are processed in cycles, they aren't necessarily moving in a simple circle and don't really have a beginning or an end.

Energy flows directionally through ecosystems, entering as entering as sunlight (or inorganic molecules for chemoautotrophs) and leaving as heat during the transfers between trophic levels. Rather than flowing through an ecosystem, the matter that makes up living 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.

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 (Figure 5.1.1). The cycling of these elements is interconnected. 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**. Today, **anthropogenic** (human) activities are altering all major ecosystems and the biogeochemical cycles they drive.

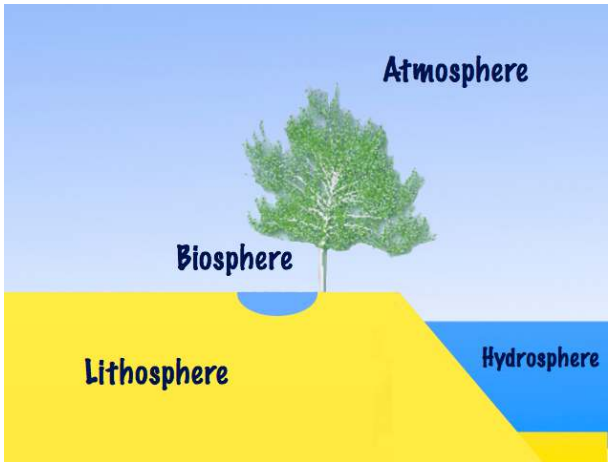



Figure 5.1.1: **Importance of the hydrosphere**: Earth has a hydrosphere, where water movement and storage occurs. It is important for leaching certain components of organic matter into rivers, lakes, and oceans, and is a reservoir for carbon.

The most important biogeochemical cycle, 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 ($CO_2$) is the second most important of the 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.) 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. 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.

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*. This part of a cycle that holds an element or water for a short period of time is sometimes also called an **exchange pool,** or **pool**. For example, the atmosphere is an exchange pool for water. It usually holds water (in the form of water vapor) for just a few days. Some other examples of reservoirs or pools for water you will encounter in the course include 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.

## Contributors and Attributions

Modified by Kyle Whittinghill (University of Pittsburgh) from the following sources

- Section 10.1 from *The Environment of the Earth's Surface* by John Southard Professor Emeritus (Earth, Atmospheric and Planetary Sciences) at Massachusetts Institute of Technology Sourced from MIT OpenCourseware
- 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
- **6.6: Water Cycle** by CK-12: Biology Concepts, is licensed CC BY-NC
- General Microbiology **Provided by**: Boundless.com. **License**: ***CC BY-SA: Attribution-ShareAlike***

---

This page titled 5.1: Introduction - Systems, Cycles, Reservoirs, and Fluxes is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) .

- **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/.

# 5.2: Water (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 **5.2.1 – 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.

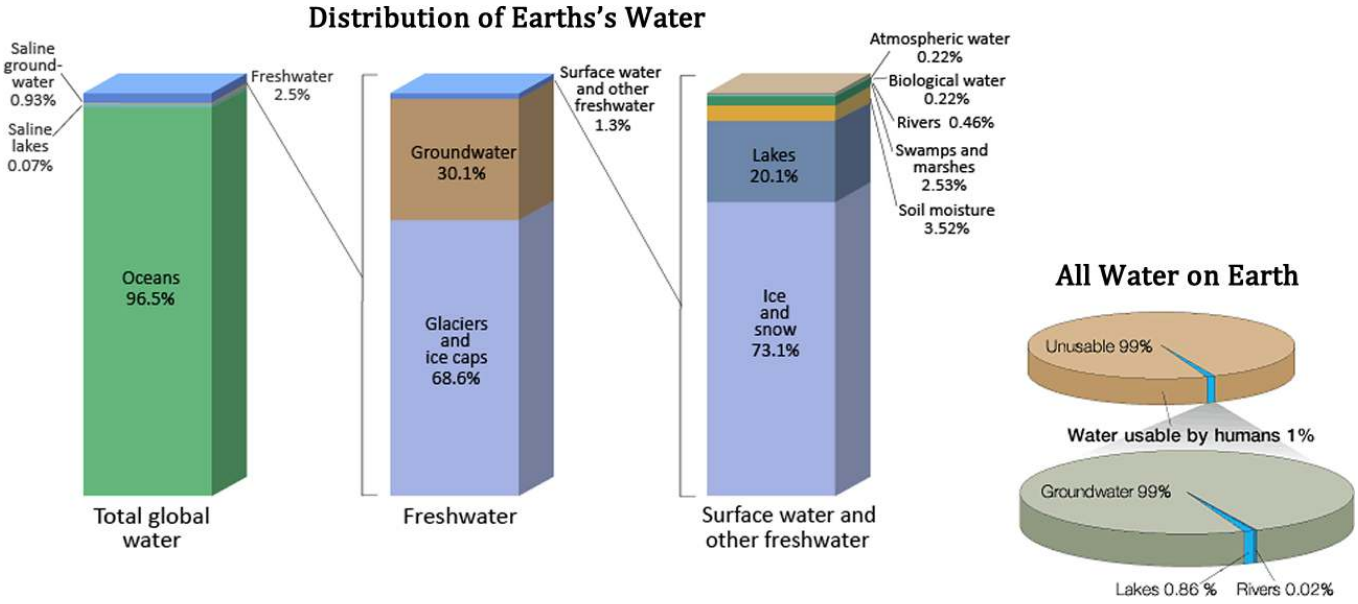



**Figure 5.2.1: Earth's Water Reservoirs** Bar chart Distribution of Earth's water including total global water, fresh water, and surface water and other fresh water and Pie chart Water usable by humans and sources of usable water. Only 2.5 percent of water on Earth is fresh water, and less than 1 percent 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 **5.2.2** 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.

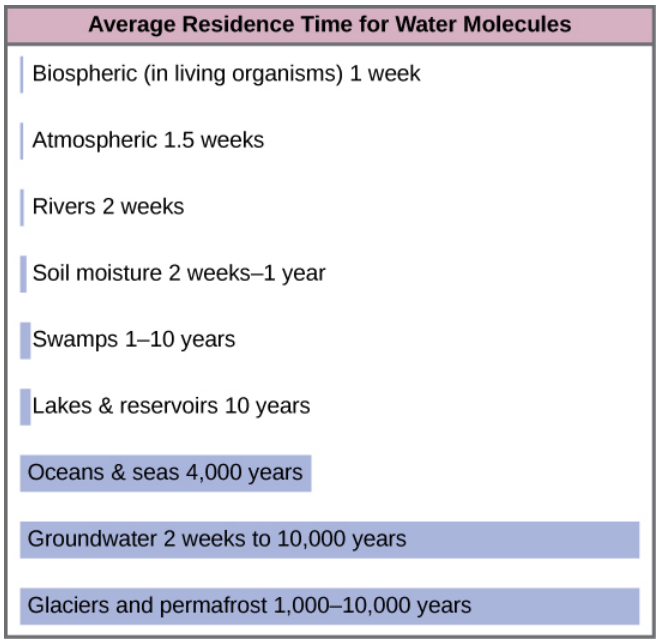



**Figure 5.2.2:** Average residence time that water remains in each reservoir. Water remains in organisms for about one week, in the atmosphere for 1.5 weeks, in rivers for two weeks, as soil moisture from two weeks to a year, in swamps for 1-10 years, in lakes for 10 years, in oceans and seas for 4,000 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 **5.2.4**. The processes include the following:

- **evaporation** and **sublimation**
- **transpiration**
- **condensation** and **precipitation**
- subsurface water flow
- surface **runoff** and snowmelt
- streamflow

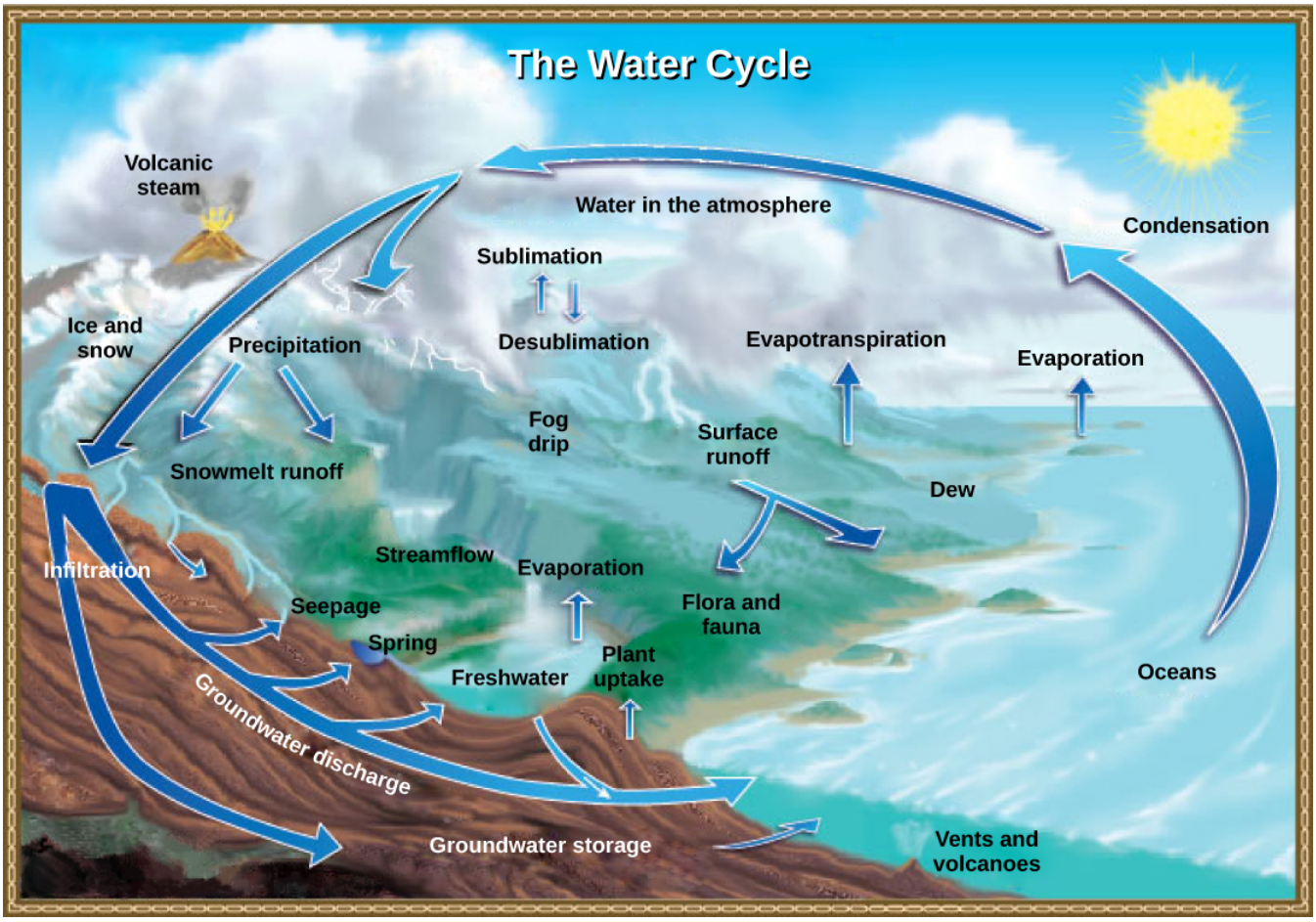



Figure **5.2.4**: Water from the land and oceans enters the atmosphere by evaporation or sublimation, where it condenses into clouds and falls as rain or snow. Precipitated water may enter freshwater bodies or infiltrate the soil. The cycle is complete when surface or groundwater reenters the ocean. (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 5.2.5). 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.

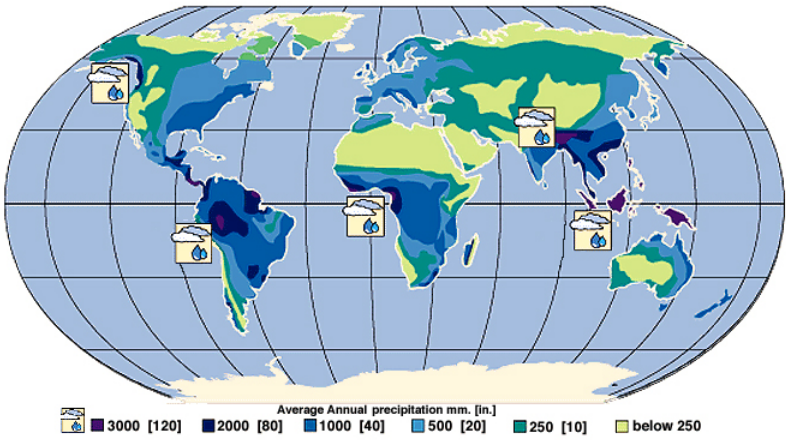
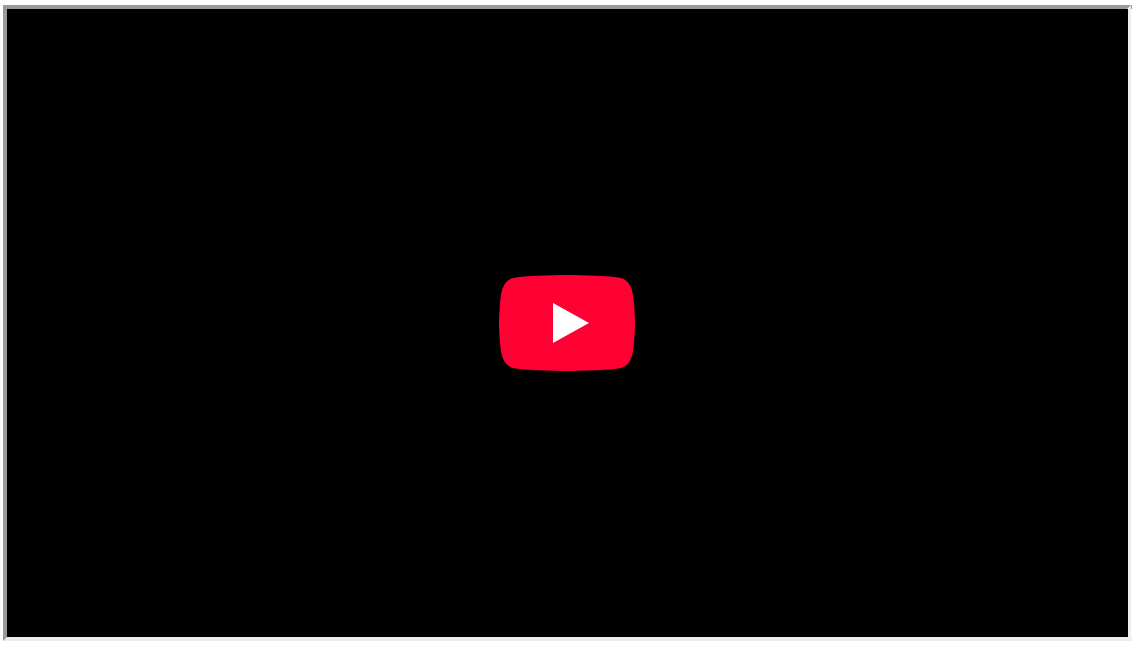
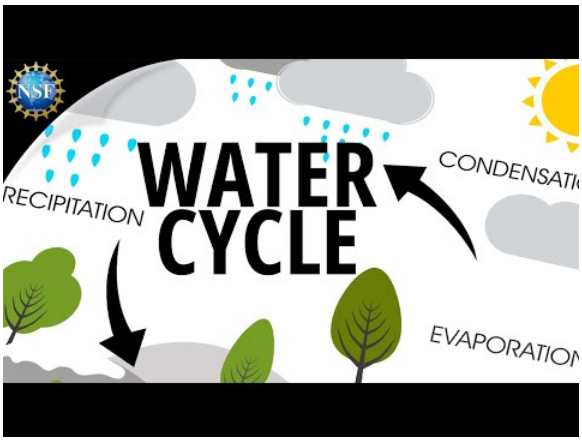



**Figure 5.2.5: World Rainfall Map** The false-color map above shows the amount of rain that falls around the world. Areas of high rainfall include Central and South America, western Africa, and Southeast Asia. Since these areas receive so much rainfall, they are where most of the world's rainforests 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 **5.2.6**). 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 5.2.6:** 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 5.2.7). 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 5.2.7: *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***

---

5.2: Water (Hydrologic) Cycle is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

- **20.2: Biogeochemical Cycles** by OpenStax is licensed CC BY 4.0.

# 5.3: Carbon Cycle

## Key Points

- Carbon is present in all organic molecules; carbon compounds contain large amounts of energy, which humans use as fuel.
- The biological carbon cycle is the rapid exchange of carbon among living things; autotrophs use carbon dioxide produced by heterotrophs to produce glucose and oxygen, which are then utilized by heterotrophs.
- The biogeochemical cycle occurs at a much slower rate than the biological cycle since carbon is stored in carbon reservoirs for long periods of time.
- Carbon dioxide from the atmosphere dissolves in water, combining with water molecules to form carbonic acid, which then ionizes to carbonate and bicarbonate ions.
- Most of the carbon in the ocean is in the form of bicarbonate ions, which can combine with seawater calcium to form calcium carbonate ($CaCO3$), a major component of marine organism shells.
- Carbon can enter the soil as a result of the decomposition of living organisms, the weathering of rocks, the eruption of volcanoes, and other geothermal systems.

## The Carbon Cycle

Carbon is the fourth most abundant element in living organisms. Carbon is present in all organic molecules, and its role in the structure of macromolecules is of primary importance to living organisms. Carbon compounds contain energy, and many of these compounds from plants and algae have remained stored as fossilized carbon, which humans use as fuel. Since the 1800s, the use of fossil fuels has accelerated. As global demand for Earth’s limited fossil fuel supplies has risen since the beginning of the Industrial Revolution, the amount of carbon dioxide in our atmosphere has increased as the fuels are burned. This increase in carbon dioxide has been associated with climate change and is a major environmental concern worldwide.

The carbon cycle is most easily studied as two interconnected subcycles: one dealing with rapid carbon exchange among living organisms and the other dealing with the long-term cycling of carbon through geologic processes. The entire carbon cycle is shown in Figure **5.3.1**. The overall effect is that carbon is constantly recycled in the dynamic processes taking place in the atmosphere, at the surface and in the crust of the earth. The vast majority of carbon resides as inorganic minerals in crustal rocks. Other **reservoirs** of carbon, places where carbon accumulates, include the oceans and atmosphere. Some of the carbon atoms in your body today may long ago have resided in a dinosaur's body, or perhaps were once buried deep in the Earth's crust as carbonate rock minerals.

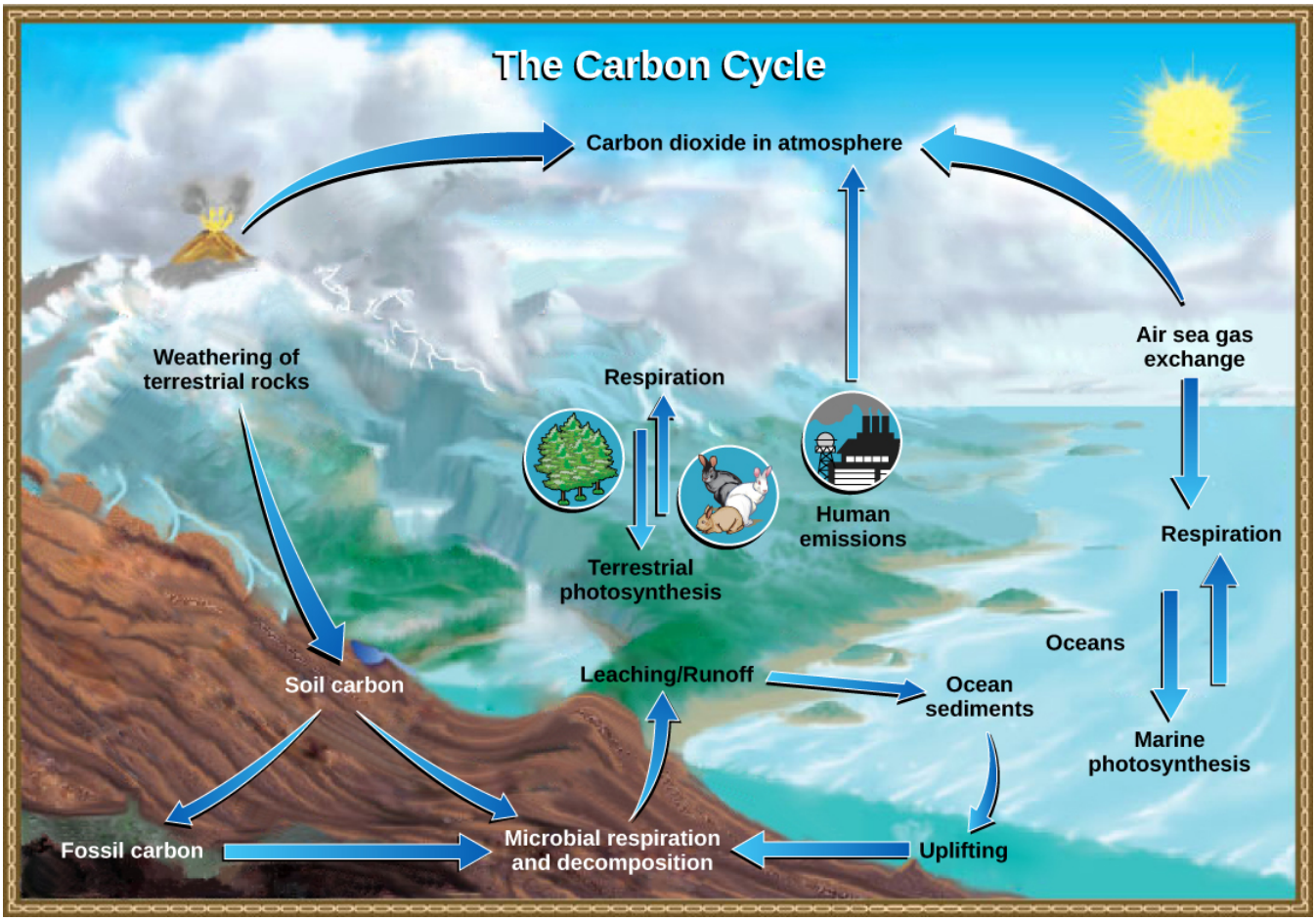



Figure **5.3.1**: Carbon dioxide gas exists in the atmosphere and is dissolved in water. Photosynthesis converts carbon dioxide gas to organic carbon, and respiration cycles the organic carbon back into carbon dioxide gas. Long-term storage of organic carbon occurs when matter from living organisms is buried deep underground and becomes fossilized. Volcanic activity and, more recently, human emissions bring this stored carbon back into the carbon cycle. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

## Carbon Cycles Quickly between Organisms and the Atmosphere

Cells run on the chemical energy found mainly in carbohydrate molecules, and the majority of these molecules are produced by one process: **photosynthesis.** Through photosynthesis, certain organisms convert solar energy (sunlight) into chemical energy, which is then used to build other organic molecules like complex carbohydrates (such as starch), proteins and lipids. The energy stored in the bonds to hold these molecules together is released when an organism breaks down food. Cells then use this energy to perform work, such as movement. The energy that is harnessed from photosynthesis enters the ecosystems of our planet continuously and is transferred from one organism to another. Therefore, directly or indirectly, the process of photosynthesis provides most of the energy required by living things on Earth. Photosynthesis also results in the release of oxygen into the atmosphere. In short, to eat and breathe humans depend almost entirely on the organisms that carry out photosynthesis.




(a) (b) (c)

Figure 5.3.2: (a) Plants, (b) algae, and (c) certain bacteria, called cyanobacteria, are can carry out photosynthesis. Algae can grow over enormous areas in water, at times completely covering the surface. (credit a: Steve Hillebrand, U.S. Fish and Wildlife Service; credit b: “eutrophication&hypoxia”/Flickr; credit c: NASA; scale-bar data from Matt Russell)

Some organisms can carry out photosynthesis, whereas others cannot. An **autotroph** is an organism that can produce its own food. The Greek roots of the word autotroph mean “self” (auto) “feeder” (troph). Plants are the best-known autotrophs, but others exist, including certain types of bacteria and algae (Figure 5.3.2). Oceanic algae contribute enormous quantities of food and oxygen to global food chains. Carbon dioxide is the basic building block that most autotrophs use to build multi-carbon, high-energy compounds, such as glucose. Most terrestrial autotrophs obtain their carbon dioxide directly from the atmosphere, while marine autotrophs acquire it in the dissolved form (bicarbonate, $HCO_3^-$).

**Heterotrophs** are organisms incapable of photosynthesis that must therefore obtain energy and carbon from food by consuming other organisms. The Greek roots of the word *heterotroph* mean “other” (*hetero*) “feeder” (*troph*), meaning that their food comes from other organisms. Even if the organism being consumed is another animal, it traces its stored energy back to autotrophs and the process of photosynthesis. Heterotrophs acquire the high-energy carbon compounds from the autotrophs by consuming them and breaking them down by respiration to obtain cellular energy, such as ATP. The most efficient type of respiration, aerobic respiration, requires oxygen obtained from the atmosphere or dissolved in water. Thus, there is a constant exchange of oxygen and carbon dioxide between the autotrophs (which need the carbon) and the heterotrophs (which need the oxygen). Humans are heterotrophs, as are all animals and fungi. A deer obtains energy by eating plants. A wolf eating a deer obtains energy that originally came from the plants eaten by that deer (Figure 5.3.3). Using this reasoning, all food eaten by humans can be traced back to autotrophs that carry out photosynthesis.



Figure 5.3.3: The energy stored in carbohydrate molecules from photosynthesis passes through the food chain. The predator that eats these deer is getting energy that originated in the photosynthetic vegetation that the deer consumed. (credit: Steve VanRiper, U.S. Fish and Wildlife Service)

Plants, animals, and other organisms break down organic molecules during the process of **aerobic cellular respiration**, which consumes oxygen and releases energy, water and carbon dioxide. Carbon dioxide is returned to the atmosphere during gaseous exchange. Another process by which organic material is recycled is the decomposition of dead organisms. During this process,

bacteria and fungi break down the complex organic compounds. Decomposers may do respiration, releasing carbon dioxide, or other processes that release methane ($CH_4$).

Photosynthesis and respiration are actually reciprocal to one another with regard to the cycling of carbon: photosynthesis removes carbon dioxide from the atmosphere and respiration returns it (Figure **5.3.4**). A significant disruption of one process can therefore affect the amount of carbon dioxide in the atmosphere.

| Photosynthesis Equation | | | | | | |
|---|---|---|---|---|---|---|
| Carbon dioxide | + | Water | SUNLIGHT → | Sugar | + | Oxygen |
| $6CO_2$ | | $6H_2O$ | | $C_6H_{12}O_6$ | | $6O_2$ |

Figure **5.3.4**: This equation means that six molecules of carbon dioxide ($CO_2$) combine with six molecules of water ($H_2O$) in the presence of sunlight. This produces one molecule of glucose ($C_6H_{12}O_6$) and six molecules of oxygen ($O_2$).

Cellular respiration is only one process that releases carbon dioxide. Physical processes, such as the eruption of volcanoes and release from **hydrothermal vents** (openings in the ocean floor) add carbon dioxide to the atmosphere. Additionally, the **combustion** of wood and fossil fuels releases carbon dioxide. The level of carbon dioxide in the atmosphere is greatly influenced by the reservoir of carbon in the oceans. The exchange of carbon between the atmosphere and water reservoirs influences how much carbon is found in each.

## Carbon Cycles Slowly between Land and the Ocean

The movement of carbon through land, water, and air is complex, and, in many cases, it occurs much more slowly geologically than the movement between living organisms. Carbon is stored for long periods in what are known as carbon reservoirs, which include the atmosphere, bodies of liquid water (mostly oceans), ocean sediment, soil, rocks (including fossil fuels), and Earth's interior.

As stated, the atmosphere is a major reservoir of carbon in the form of carbon dioxide that is essential to the process of photosynthesis. The level of carbon dioxide in the atmosphere is greatly influenced by the reservoir of carbon in the oceans. The exchange of carbon between the atmosphere and water reservoirs influences how much carbon is found in each, and each one affects the other reciprocally. Atmospheric carbon dioxide also dissolves in the ocean, reacting with water molecules to form carbonic acid and then dissociating into bicarbonate ($HCO_3^-$) and carbonate ions ($CO_3^{2-}$). More than 90 percent of the carbon in the ocean is found as bicarbonate ions. Some of these ions combine with calcium ions in the seawater to form calcium carbonate ($CaCO_3$), a major component of the shells of marine organisms. These organisms eventually form sediments on the ocean floor. Over geologic time, the calcium carbonate forms limestone, which comprises the largest carbon reservoir on Earth.

| |
|---|
| Step 1: $CO_2$ (atmospheric) ⇌ $CO_2$ (dissolved) |
| Step 2: $CO_2$ (dissolved) + $H_2O$ ⇌ $H_2CO_3$ (carbonic acid) |
| Step 3: $H_2CO_3$ ⇌ + $H^+$ + $HCO_3^-$ (biocarbonate ion) |
| Step 4: $HCO_3^-$ ⇌ $H^+$ + $CO_3^{2-}$ (carbonate ion) |

Figure **5.3.1**: **Formation of bicarbonate**: Carbon dioxide reacts with water to form bicarbonate and carbonate ions.

On land, carbon is stored in soil as organic carbon as a result of the decomposition of living organisms or from **weathering** of terrestrial rock and minerals. This carbon can be leached into the water reservoirs by surface runoff. Partially decomposed plants and algae are sometimes buried and compressed between layers of sediments. After millions of years fossil fuels such as coal, oil, and natural gas are formed. Fossil fuels are considered a non-renewable resource because their use far exceeds their rate of formation. A non-renewable resource is either regenerated very slowly or not at all.

Another way for carbon to enter the atmosphere is from land (including land beneath the surface of the ocean) by the eruption of volcanoes and other geothermal systems. Carbon sediments from the ocean floor are taken deep within Earth by the process of **subduction**: the movement of one tectonic plate beneath another. The ocean sediments are subducted by the actions of **plate**

**tectonics**, melted and then returned to the surface during volcanic activity. Carbon is released as carbon dioxide when a volcano erupts or from volcanic hydrothermal vents.Plate tectonics can also cause **uplifting**, returning ocean sediments to land.

## Importance of the Carbon Cycle

The carbon cycle is crucially important to the biosphere. If not for the recycling processes, carbon might long ago have become completely sequestered in crustal rocks and sediments, and life would no longer exist (Figure 5.3.5). Photosynthesis not only makes energy and carbon available to higher trophic levels, but it also releases gaseous oxygen ($O_2$). Gaseous oxygen is necessary for cellular respiration to occur. Photosynthetic bacteria were likely the first organisms to perform photosynthesis, dating back 2-3 billion years ago. Thanks to their activity, and a diversity of present-day photosynthesizing organisms, Earth's atmosphere is currently about 21% $O_2$. Also, this $O_2$ is vital for the creation of the ozone layer, which protects life from harmful ultraviolet radiation emitted by the sun. Ozone ($O_3$) is created from the breakdown and reassembly of $O_2$.



Figure 5.3.5: Decomposers will break down the organic compounds in this fallen tree at Cliffs of the Neuse State Park in Wayne County, North Carolina, releasing carbon dioxide into the atmosphere. Decomposition ensures that carbon dioxide will be available in the atmosphere for photosynthetic organisms, which then provide carbon for consumers. Image by Gerry Dincher (CC-BY-SA).

The global carbon cycle contributes substantially to the provisioning ecosystem services upon which humans depend. We harvest approximately 25% of the total plant biomass that is produced each year on the land surface to supply food, fuel wood and fiber from croplands, pastures and forests. In addition, the global carbon cycle plays a key role in regulating ecosystem services because it significantly influences climate via its effects on atmospheric $CO_2$ concentrations.

## Human Alteration of the Carbon Cycle

Atmospheric $CO_2$ concentration increased from 280 parts per million (ppm) to 413 ppm between the start of industrial revolution in the late eighteenth century and 2020. This reflected a new flux in the global carbon cycle—anthropogenic $CO_2$ emissions—where humans release $CO_2$ into the atmosphere by burning fossil fuels and changing land use. Fossil fuel burning takes carbon from coal, gas, and oil reserves, where it would be otherwise stored on very long time scales, and introduces it into the active carbon cycle. Land use change releases carbon from soil and plant biomass pools into the atmosphere, particularly through the process of deforestation for wood extraction or conversion of land to agriculture. Carbon dioxide is also added to the atmosphere by the breeding and raising of livestock. In 2018, the additional flux of carbon into the atmosphere from anthropogenic sources was estimated to be 36.6 gigatons of carbon (GtC = 1 billion tons of carbon)—a significant disturbance to the natural carbon cycle that had been in balance for several thousand years previously. High levels of carbon dioxide in the atmosphere cause warming that results in climate change.

## 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)
- Carbon 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
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.
- General Microbiology **Provided by**: Boundless.com. **License**: ***CC BY-SA: Attribution-ShareAlike***

---

5.3: Carbon Cycle is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

- **20.2: Biogeochemical Cycles** by OpenStax is licensed CC BY 4.0.

# 5.4: Photosynthesis

**Learning Objectives**

- Explain the relevance of photosynthesis to other living things
- Describe the main structures involved in photosynthesis
- Identify the substrates and products of photosynthesis
- Summarize the process of photosynthesis

## Main Structures and Summary of Photosynthesis

Photosynthesis is a multi-step process that requires sunlight, carbon dioxide (which is low in energy), and water as substrates (Figure 5.4.1). After the process is complete, it releases oxygen and produces glyceraldehyde-3-phosphate (GA3P), simple carbohydrate molecules (which are high in energy) that can subsequently be converted into glucose, sucrose, or any of dozens of other sugar molecules. These sugar molecules contain energy and the energized carbon that all living things need to survive.

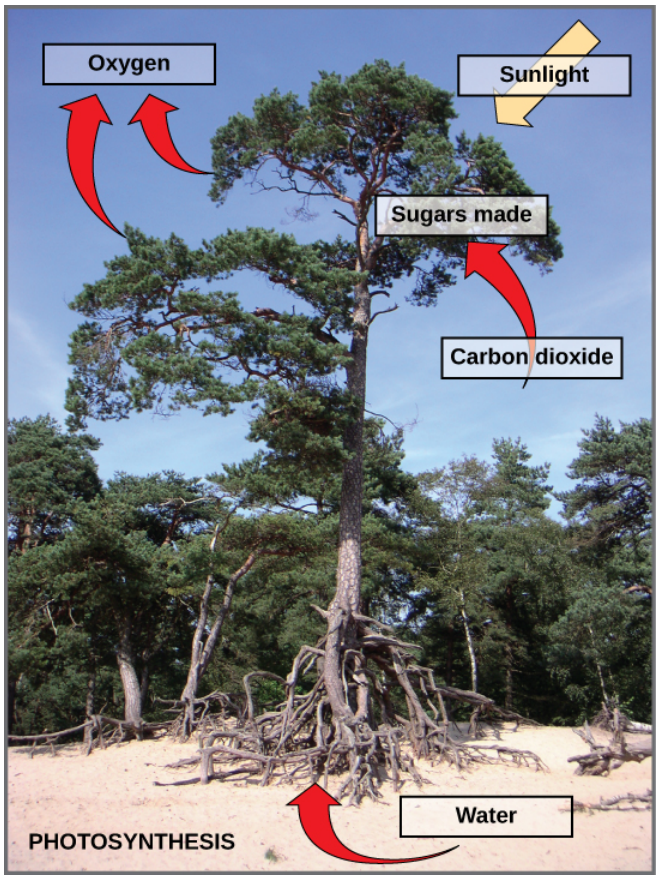



Figure 5.4.1: Photosynthesis uses solar energy, carbon dioxide, and water to produce energy-storing carbohydrates. Oxygen is generated as a waste product of photosynthesis.

The following is the chemical equation for photosynthesis (Figure 5.4.2):

| Photosynthesis Equation | | | | | | |
|---|---|---|---|---|---|---|
| Carbon dioxide | + | Water | SUNLIGHT → | Sugar | + | Oxygen |
| $6CO_2$ | | $6H_2O$ | | $C_6H_{12}O_6$ | | $6O_2$ |

Figure 5.4.2: The basic equation for photosynthesis is deceptively simple. In reality, the process takes place in many steps involving intermediate reactants and products. Glucose, the primary energy source in cells, is made from two three-carbon GA3Ps.

Although the equation looks simple, the many steps that take place during photosynthesis are actually quite complex. Before learning the details of how photoautotrophs turn sunlight into food, it is important to become familiar with the structures involved.

In plants, photosynthesis generally takes place in leaves, which consist of several layers of cells. The process of photosynthesis occurs in a middle layer called the mesophyll. The gas exchange of carbon dioxide and oxygen occurs through small, regulated openings called stomata (singular: stoma), which also play roles in the regulation of gas exchange and water balance. The stomata are typically located on the underside of the leaf, which helps to minimize water loss. Each stoma is flanked by guard cells that regulate the opening and closing of the stomata by swelling or shrinking in response to osmotic changes.

In all autotrophic eukaryotes, photosynthesis takes place inside an organelle called a chloroplast. For plants, chloroplast-containing cells exist in the mesophyll. Chloroplasts have a double membrane envelope (composed of an outer membrane and an inner membrane). Within the chloroplast are stacked, disc-shaped structures called thylakoids. Embedded in the thylakoid membrane is chlorophyll, a pigment (molecule that absorbs light) responsible for the initial interaction between light and plant material, and numerous proteins that make up the electron transport chain. The thylakoid membrane encloses an internal space called the thylakoid lumen. As shown in Figure 5.4.3, a stack of thylakoids is called a granum, and the liquid-filled space surrounding the granum is called stroma or "bed" (not to be confused with stoma or "mouth," an opening on the leaf epidermis).

Art Connection

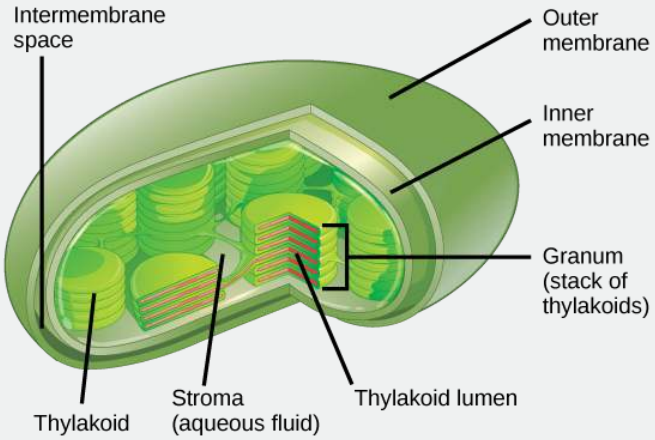



Figure 5.4.3: Photosynthesis takes place in chloroplasts, which have an outer membrane and an inner membrane. Stacks of thylakoids called grana form a third membrane layer.

## The Two Parts of Photosynthesis

Photosynthesis takes place in two sequential stages: the light-dependent reactions and the light independent-reactions. In the light-dependent reactions, energy from sunlight is absorbed by chlorophyll and that energy is converted into stored chemical energy. In the light-independent reactions, the chemical energy harvested during the light-dependent reactions drive the assembly of sugar molecules from carbon dioxide. Therefore, although the light-independent reactions do not use light as a reactant, they require the products of the light-dependent reactions to function. In addition, several enzymes of the light-independent reactions are activated by light. The light-dependent reactions utilize certain molecules to temporarily store the energy: These are referred to as energy carriers. The energy carriers that move energy from light-dependent reactions to light-independent reactions can be thought of as "full" because they are rich in energy. After the energy is released, the "empty" energy carriers return to the light-dependent

reaction to obtain more energy. Figure 5.4.4 illustrates the components inside the chloroplast where the light-dependent and light-independent reactions take place.

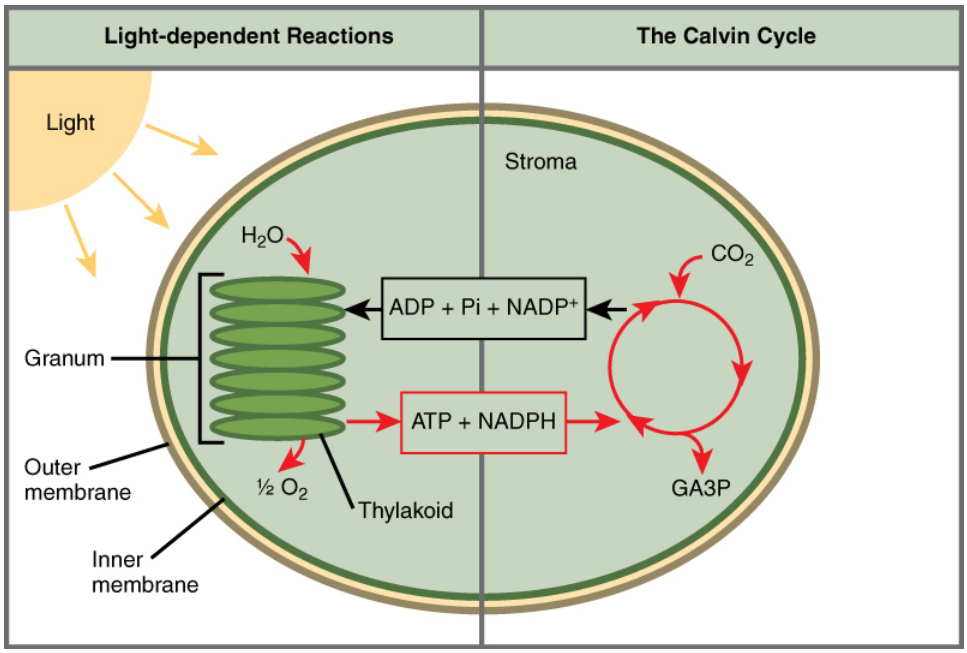



Figure 5.4.4: Photosynthesis takes place in two stages: light dependent reactions and the Calvin cycle. Light-dependent reactions, which take place in the thylakoid membrane, use light energy to make ATP and NADPH. The Calvin cycle, which takes place in the stroma, uses energy derived from these compounds to make GA3P from $CO_2$.

## The Light-Dependent Reactions of Photosynthesis

Light energy initiates the process of photosynthesis when pigments absorb the light. Different kinds of organic pigments, whether in the human retina or the chloroplast thylakoid, have each evolved to absorb only certain wavelengths (colors) of light. Pigments reflect or transmit the wavelengths they cannot absorb, making them appear in the corresponding color. Plants pigment molecules absorb only light in the wavelength range of 700 nm to 400 nm; plant physiologists refer to this range for plants as photosynthetically active radiation.

Chlorophylls and carotenoids are the two major classes of photosynthetic pigments found in plants and algae; each class has multiple types of pigment molecules. There are five major chlorophylls: *a*, *b*, *c* and *d* and a related molecule found in prokaryotes called bacteriochlorophyll. Chlorophyll *a* and chlorophyll *b* are found in higher plant chloroplasts and will be the focus of the following discussion. With dozens of different forms, carotenoids are a much larger group of pigments. The carotenoids 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)—are used as advertisements to attract seed dispersers. In photosynthesis, carotenoids function as photosynthetic pigments that are very efficient molecules for the disposal of excess energy. When a leaf is exposed to full sun, the light-dependent reactions are required to process an enormous amount of energy; if that energy is not handled properly, it can do significant damage. Therefore, many carotenoids reside in the thylakoid membrane, absorb excess energy, and safely dissipate that energy as heat.

Each type of pigment can be identified by the specific pattern of wavelengths it absorbs from visible light, which is the absorption spectrum. The graph in Figure 5.4.5 shows the absorption spectra for chlorophyll *a*, chlorophyll *b*, and a type of carotenoid pigment called β-carotene (which absorbs blue and green light). Notice how each pigment has a distinct set of peaks and troughs, revealing a highly specific pattern of absorption. 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.

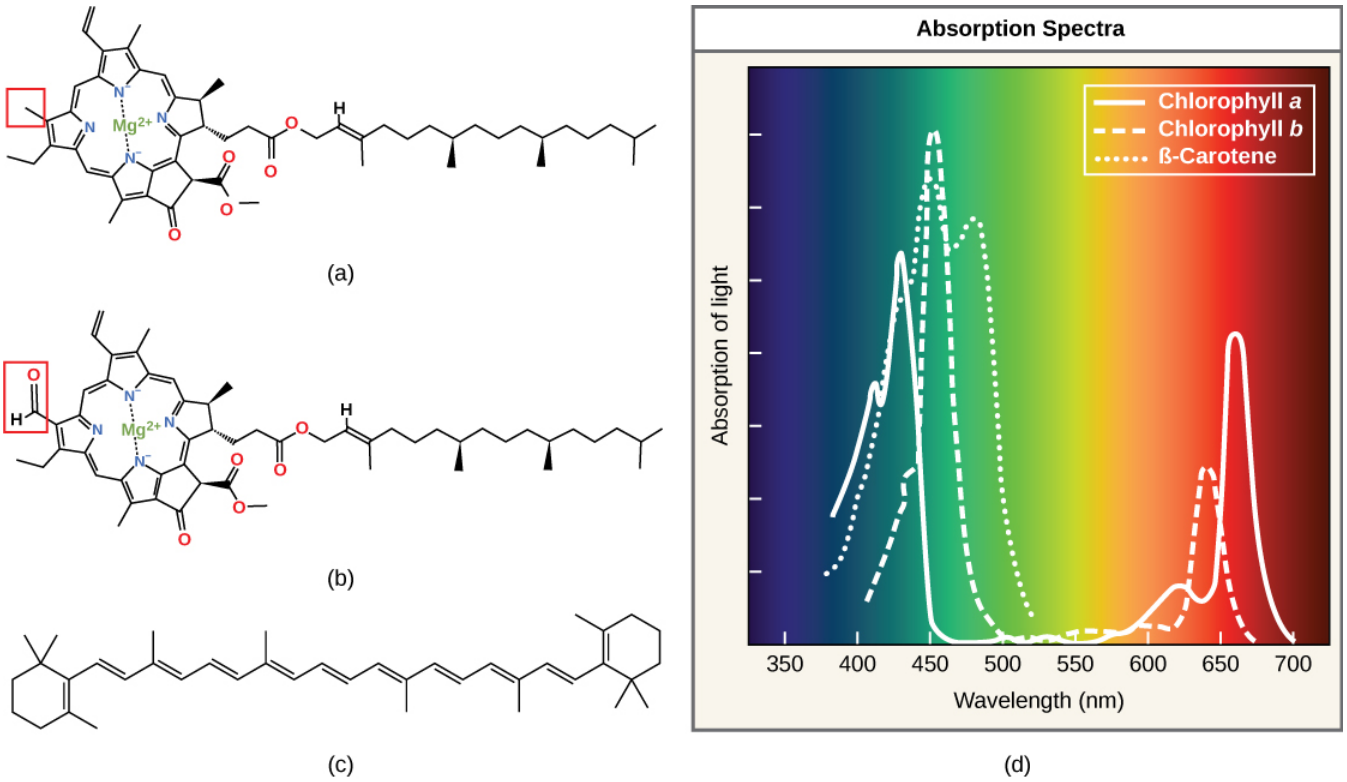



Figure 5.4.5: (a) Chlorophyll *a*, (b) chlorophyll *b*, and (c) *β*-carotene are hydrophobic organic pigments found in the thylakoid membrane. Chlorophyll *a* and *b*, which are identical except for the part indicated in the red box, are responsible for the green color of leaves. *β*-carotene is responsible for the orange color in carrots. Each pigment has (d) a unique absorbance spectrum.

The overall function of light-dependent reactions is to convert solar energy into chemical energy in the form of NADPH and ATP. Protein complexes and pigment molecules work together to produce NADPH and ATP. The actual step that converts light energy into chemical energy takes place in a multiprotein complex called a photosystem, two types of which are found embedded in the thylakoid membrane, photosystem II (PSII) and photosystem I (PSI) (Figure 5.4.7). The two complexes differ on the basis of what they oxidize (that is, the source of the low-energy electron supply) and what they reduce (the place to which they deliver their energized electrons). Both photosystems have the same basic structure; a number of antenna proteins to which the chlorophyll molecules are bound surround the reaction center where the photochemistry takes place. In short, the light energy has now been captured by biological molecules but is not stored in any useful form yet. The 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.

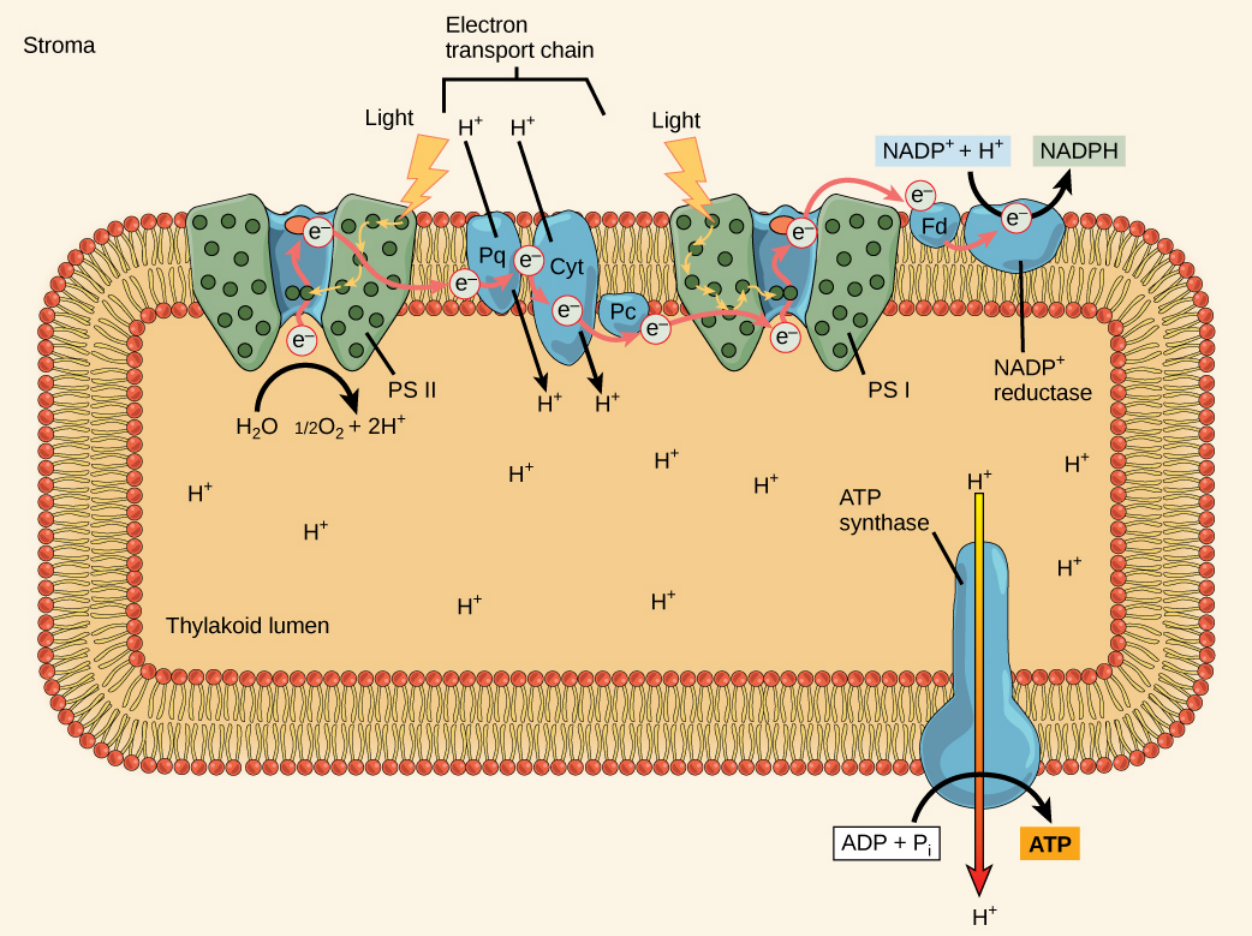



**Figure 5.4.7:** A photosystem consists of a light-harvesting complex and a reaction center. Pigments in the light-harvesting complex pass light energy to two special chlorophyll a molecules in the reaction center. The light excites an electron from the chlorophyll a pair, which passes to the primary electron acceptor. The excited electron must then be replaced. In the photosystem II (PSII) reaction center, energy from sunlight is used to extract electrons from water. The electrons travel through the chloroplast electron transport chain to photosystem I (PSI), which reduces $NADP^+$ to NADPH. The electron transport chain moves protons across the thylakoid membrane into the lumen. At the same time, splitting of water adds protons to the lumen, and reduction of NADPH removes protons from the stroma. The net result is a low pH in the thylakoid lumen, and a high pH in the stroma. ATP synthase uses this electrochemical gradient to make ATP.

The reaction center in PSII contains a pair of chlorophyll *a* molecules with a special property. Those two chlorophylls can undergo oxidation upon excitation; they can actually give up an electron. It is at this step in photosynthesis, that light energy is converted into an excited electron. All of the subsequent steps involve getting that electron onto the energy carrier NADPH for delivery to the Calvin cycle where the electron is deposited onto carbon for long-term storage in the form of a carbohydrate. The reaction center of PSII extracts low-energy electrons from water, splitting two molecules of water to release four electrons, four hydrogen atoms, and one molecule of diatomic $O_2$ gas. The reaction center of PSII then delivers its high-energy electrons, one at the time, to the primary electron acceptor, and through the electron transport chain.

The cytochrome complex, an enzyme composed of two protein complexes, transfers the electrons from the carrier molecule plastoquinone (Pq) to the protein plastocyanin (Pc), thus enabling both the transfer of protons across the thylakoid membrane and the transfer of electrons from PSII to PSI (Figure 5.4.7). As electrons move through the proteins that reside between PSII and PSI, they lose energy. That energy is used to move hydrogen atoms from the stromal side of the membrane to the thylakoid lumen. Those hydrogen atoms, plus the ones produced by splitting water, accumulate in the thylakoid lumen and create a concentration gradient. The passive diffusion of hydrogen ions from high concentration (in the thylakoid lumen) to low concentration (in the stroma) through a specialized protein channel called the ATP synthase. The energy released by the hydrogen ion stream as it passes

through the thylakoid membrane allows ATP synthase to attach a third phosphate group to ADP, which forms a molecule of ATP (Figure 5.4.7).

Because the electrons have lost energy prior to their arrival at PSI, they must be re-energized by PSI, hence, another photon is absorbed by the PSI antenna. That energy is relayed to the PSI reaction center which is oxidized and sends a high-energy electron to $NADP^+$ to form NADPH. Thus, PSII captures the energy to create proton gradients to make ATP, and PSI captures the energy to reduce $NADP^+$ into NADPH.

## The Light-Independent Reactions of Photosynthesis

After the energy from the sun is converted into chemical energy temporarily stored in the bonds of ATP and NADPH molecules, the cell has the fuel needed to build carbohydrate molecules for long-term energy storage. The products of the light-dependent reactions, ATP and NADPH, have lifespans in the range of millionths of seconds, whereas the products of the light-independent reactions (carbohydrates and other forms of reduced carbon) can survive for hundreds of millions of years. In plants, carbon dioxide ($CO_2$) enters the leaves through stomata, where it diffuses over short distances through intercellular spaces until it reaches the mesophyll cells. Once in the mesophyll cells, $CO_2$ diffuses into the stroma of the chloroplast—the site of light-independent reactions of photosynthesis (Figure 5.4.8). The light-independent reactions (also known as the **Calvin cycle**) can be organized into three basic stages: fixation, reduction, and regeneration.

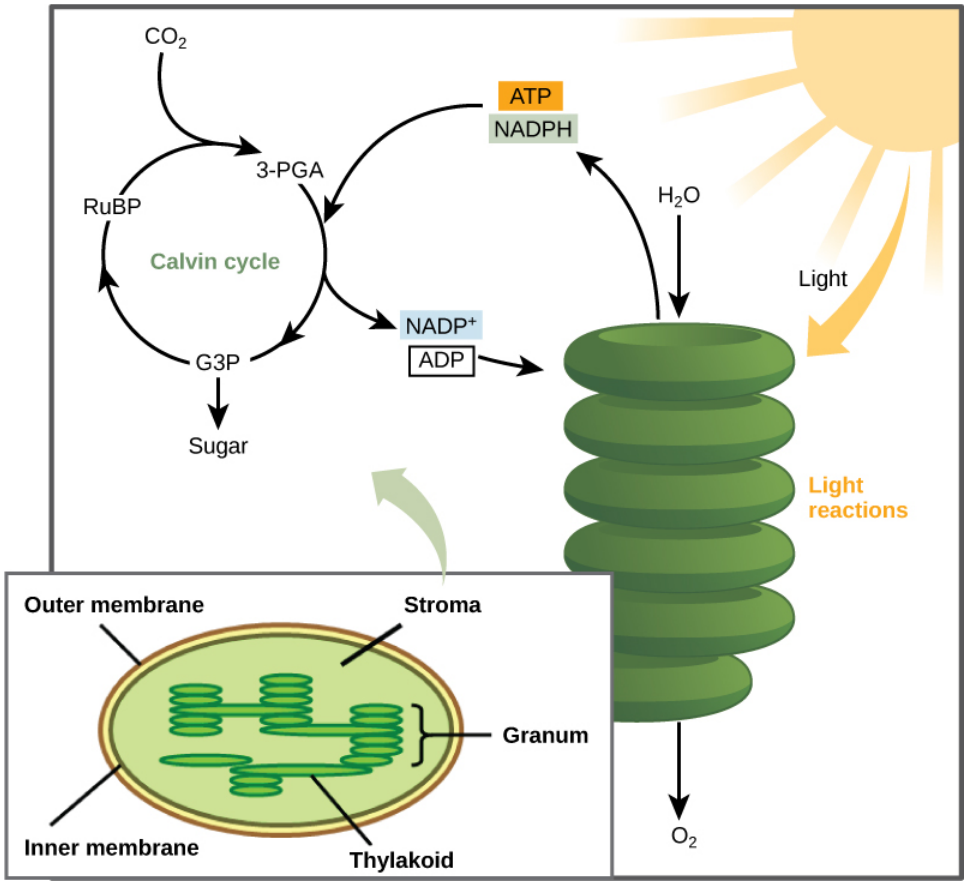



Figure 5.4.8: Light-dependent reactions harness energy from the sun to produce ATP and NADPH. These energy-carrying molecules travel into the stroma where the Calvin cycle reactions take place.

**Stage 1: Fixation**:

In the stroma, 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 three molecules of **ribulose-1,5-bisphosphate** (**RuBP**), as shown in Figure 5.4.1.25.4.1.2. RuBP has five atoms of carbon, flanked by two phosphates.

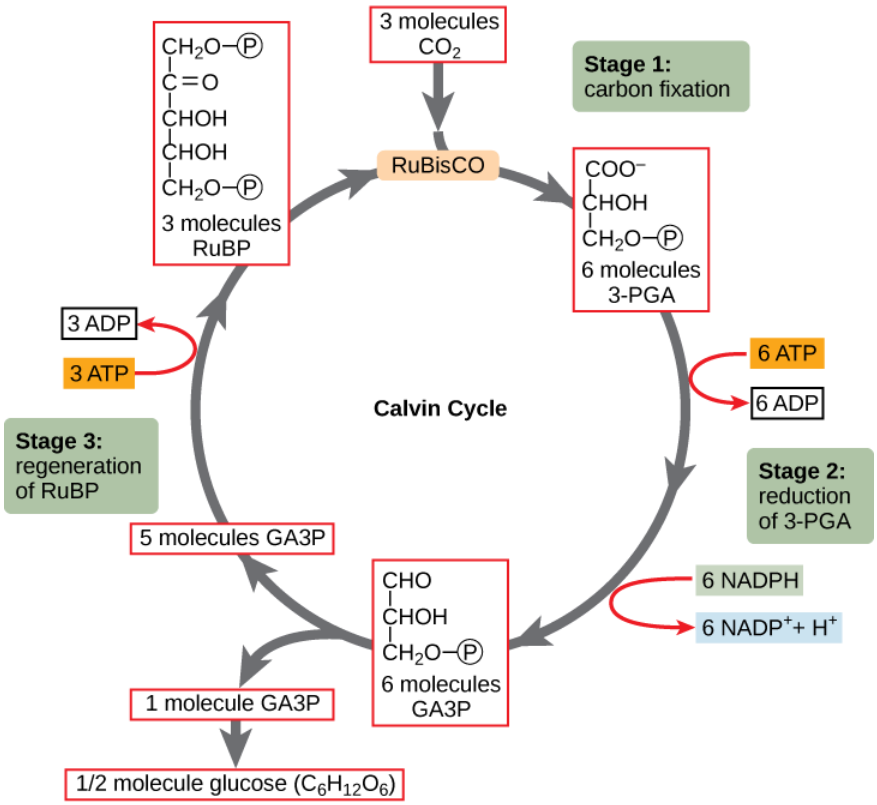



Figure **5.4.9**:: The light-independent reactions (Calvin cycle) has three stages. In stage 1, the enzyme RuBisCO adds carbon dioxide to RuBP, which immediately splits, producing two three-carbon 3-PGA molecules. In stage 2, two NADPH and two ATP are used to reduce 3-PGA to GA3P. In stage 3, RuBP, the molecule that starts the cycle, is regenerated so that the cycle can continue. One ATP is used in the process. 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.

RuBisCO catalyzes a reaction between $CO_2$ and RuBP. For each $CO_2$ molecule that reacts with one RuBP, two molecules of another compound, **3-phosphoglycerate** (3-PGA), form. 3-PGA has three carbon atoms and one phosphate. Each turn of the cycle involves only one RuBP and one carbon dioxide and forms two molecules of 3-PGA. The number of carbon atoms remains the same, as the atoms move to form new bonds during the reactions (3 atoms from 3 $CO_2$ + 15 atoms from 3 RuBP = 18 atoms in 3 atoms of 3-PGA). This process is called **carbon fixation**, because $CO_2$ is "fixed" from an inorganic form into organic molecules.

**Stage 2: Reduction:**

ATP and NADPH are used to convert the six molecules of 3-PGA into six molecules of a chemical called **glyceraldehyde 3-phosphate** (G3P). That is a reduction reaction because it involves the gain of electrons by 3-PGA. Recall that **reduction** is the gain of an electron by an atom or molecule. Six molecules of both ATP and NADPH are used. For ATP, energy is released with the loss of the terminal phosphate atom, converting it into ADP; for NADPH, both energy and a hydrogen atom are lost, converting it into $NADP^+$. Both of these molecules return to the nearby light-dependent reactions to be reused and re-energized.

**Stage 3: Regeneration:**

Interestingly, at this point, only one of the G3P molecules leaves the light-independent reactions and is sent to the cytoplasm to contribute to the formation of other compounds needed by the plant. Because the G3P exported from the chloroplast has three carbon atoms, it takes three "turns" of the cycle to fix enough net carbon to export one G3P. But each turn makes two G3P, thus three turns make six G3P. One of these six is exported while the remaining five G3P molecules remain in the cycle and are used to regenerate RuBP, which enables the system to prepare for more $CO_2$ to be fixed. Three more molecules of ATP are used in these regeneration reactions.

Because the carbohydrate molecule has six carbon atoms, it takes six turns of the Calvin cycle to make one carbohydrate molecule (one for each carbon dioxide molecule fixed). The remaining G3P molecules regenerate RuBP, which enables the system to prepare

for the carbon-fixation step. ATP is also used in the regeneration of RuBP. These six turns require energy input from at total of 12 ATP molecules and 12 NADPH molecules in the reduction step and a total of 6 ATP molecules in the regeneration step.

## Photosynthesis in Prokaryotes

The two parts of photosynthesis—the light-dependent reactions and the Calvin cycle—have been described, as they take place in chloroplasts. However, prokaryotes, such as cyanobacteria, lack membrane-bound organelles. Prokaryotic photosynthetic autotrophic organisms have infoldings of the plasma membrane for chlorophyll attachment and photosynthesis (Figure 5.4.10). It is here that organisms like cyanobacteria can carry out photosynthesis.

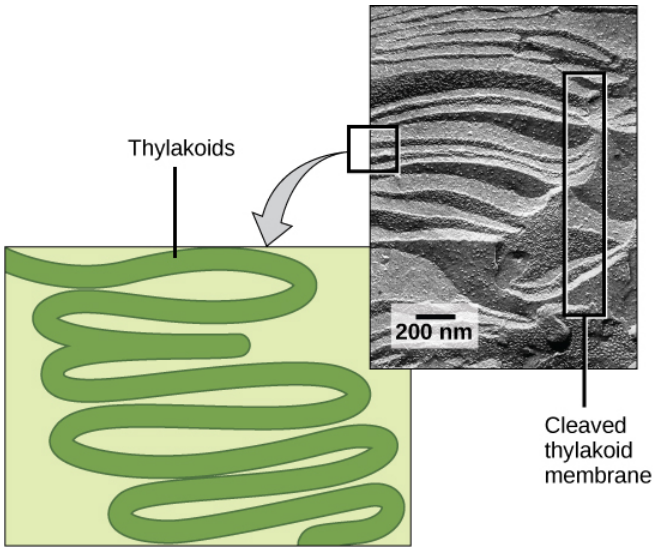



Figure 5.4.10: A photosynthetic prokaryote has infolded regions of the plasma membrane that function like thylakoids. Although these are not contained in an organelle, such as a chloroplast, all of the necessary components are present to carry out photosynthesis. (credit: scale-bar data from Matt Russell)

## Photorespiration and $C_4$ and CAM Plants

Different plant species have adaptations that allow them to do different variations of the light-independent reactions. These are called **photosynthetic pathways**. Plants are classified as $C_3$, $C_4$, or CAM depending on their use of these pathways, but note that some plants can switch photosynthetic pathways depending on environmental conditions. The process for light-independent reactions described in the previous section was the **$C_3$ pathway**: the compound formed during fixation (3-PGA) has three carbon atoms.

Photosynthesis in dry-climate plants (Figure 5.4.11) has evolved with adaptations that conserve water. In the harsh dry heat, every drop of water and precious energy must be used to survive. Two adaptations have evolved in such plants. In one form, a more efficient use of $CO_2$ allows plants to photosynthesize even when $CO_2$ is in short supply, as when the stomata are closed on hot days. The other adaptation performs preliminary reactions of the Calvin cycle at night, because opening the stomata at this time conserves water due to cooler temperatures. In addition, this adaptation has allowed plants to carry out low levels of photosynthesis without opening stomata at all, an extreme mechanism to face extremely dry periods.



Figure 5.4.11: Living in the harsh conditions of the desert has led plants like this cactus to evolve variations in reactions outside the Calvin cycle. These variations increase efficiency and help conserve water and energy. (credit: Piotr Wojtkowski)

As its name suggests, **ribulose-1,5-bisphosphate carboxylase/oxygenase** (**RuBisCO**) catalyzes two different reactions. The first is adding $CO_2$ to ribulose-1,5- bisphosphate (RuBP) — the **carboxylase** activity. The second is adding $O_2$ to RuBP — the **oxygenase** activity. The oxygenase activity of RuBisCO forms the three-carbon molecule 3-phosphoglycerate (3-PGA), just as in the light-independent reactions, and the two-carbon molecule glycolate. The glycolate enters peroxisomes, where it uses $O_2$ to form intermediates that enter mitochondria where they are broken down to $CO_2$. So this process uses $O_2$ and liberates $CO_2$ as aerobic cellular respiration does, which is why it is called **photorespiration**. It undoes the work of photosynthesis, which is to build sugars. Which action of RuBisCO predominates depends on the relative concentrations of $O_2$ and $CO_2$ with high $CO_2$, low $O_2$ favoring the carboxylase action and high $O_2$, low $CO_2$ favoring the oxygenase action. The light reactions of photosynthesis liberate oxygen, and more oxygen dissolves in the cytosol of the cell at higher temperatures. Therefore, high light intensities and high temperatures (above ~ 30°C) favor the second reaction and result in photorespiration.

One solution to photorespiration is for plants to open their stomata to release $O_2$ and obtain $CO_2$. However, if conditions are hot or dry, this will result in too much water loss (**transpiration**). For this reason, $C_3$ plants, which only do the $C_3$ pathway and do not use the $C_4$ pathway to prevent photorespiration (see below), do best in cool, moist areas. Rice and potatoes are examples of $C_3$ plants.

### $C_4$ Plants

Many angiosperms have developed adaptations which minimize the losses to photorespiration. They all use a supplementary method of $CO_2$ uptake which initially forms a four-carbon molecule compared to the two three-carbon molecules that are initially formed in the $C_3$ pathway. Hence, these plants are called **$C_4$ plants**. Note that $C_4$ plants will eventually conduct the light-independent reactions ($C_3$ pathway), but they form a four-carbon molecule first. $C_4$ plants have structural changes in their leaf anatomy so that synthesizing the four-carbon sugar (the $C_4$ pathway) and resuming the light-independent reactions ($C_3$ pathways) are separated in different parts of the leaf with RuBisCO sequestered where the $CO_2$ level is high and the $O_2$ level low. After

entering through stomata, $CO_2$ diffuses into a mesophyll cell (Figure **5.4.12**). Being close to the leaf surface, these cells are exposed to high levels of $O_2$, but they have no RuBisCO so cannot start photorespiration (nor the light-independent reactions).

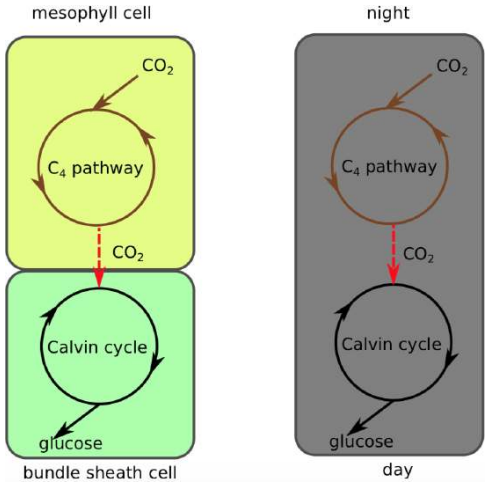



Figure **5.4.12**: C4 plants conduct the $C_4$ pathway in the mesophyll cells and the Calvin cycle ($C_3$ pathway) in the bundle sheath cells, meaning they spatially separate the two (left). CAM plants (right) conduct the $C_4$ pathway at night and the Calvin cycle ($C_3$ pathway) during the day, resulting an a temporal separation of the two.

Instead, the $CO_2$ is inserted into a three-carbon compound called **phosphoenolpyruvic acid** (**PEP**) forming the four-carbon compound **oxaloacetic acid**. Oxaloacetic acid is converted into malic acid or aspartic acid (both have 4 carbons), which is transported through plasmodesmata into a **bundle sheath cell**. Bundle sheath cells are deep in the leaf, so atmospheric oxygen cannot diffuse easily to them (Figure **5.4.13**). Additionally, they often have thylakoids with reduced photosystem II complexes (the one that produces $O_2$). Both of these features keep oxygen levels low in bundle sheath cells, which is where the four-carbon compound is broken down into carbon dioxide, which enters the light-independent reactions ($C_3$ pathway) to form sugars and pyruvic acid, which is transported back to a mesophyll cell where it is converted back into **PEP**.

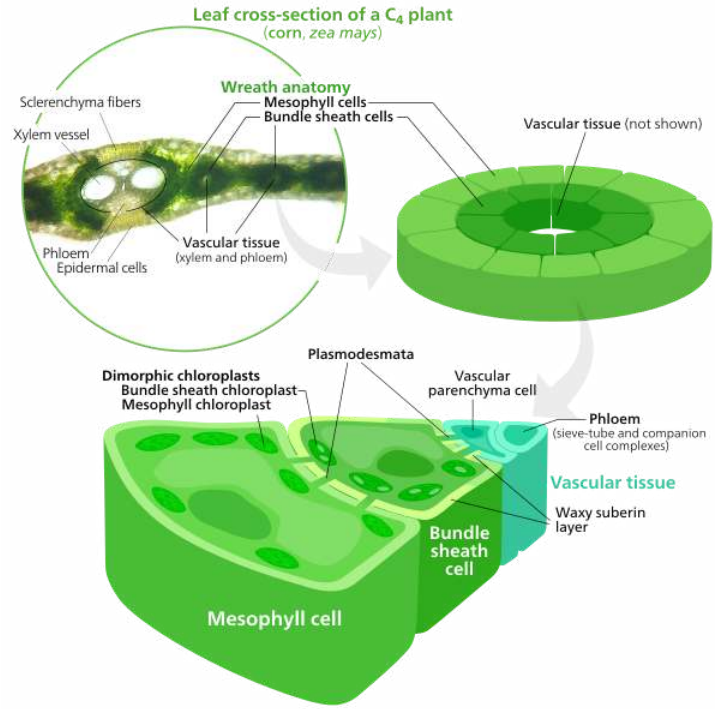



Figure 5.4.13: Cross section and diagram of a $C_4$ plant, showing mesophyll cells surrounding bundle sheath cells in concentric circles. The bundle sheath cells are larger and have more chloroplasts than in other plants. This arrangement is called Kranz anatomy (wreath anatomy). Vascular bundles contain xylem (vessel elements and tracheids) and phloem (sieve-tube and companion cell complexes). A vascular parenchyma is also shown in the vascular bundle. Waxy suberin borders the bundle sheath cells. The chloroplasts in the mesophyll cells are distinct from those in the bundle sheath cells, and the chloroplasts are thus called dimorphic. Sclerenchyma fibers are just above and below the vascular bundle, and epidermal cells surround the entire leaf. Image by Kelvinsong (CC-BY-SA).

These $C_4$ plants are well adapted to (and likely to be found in) habitats with high daytime temperatures and intense sunlight. Because they use the $C_4$ pathway to prevent photorespiration, they do not have to open their stomata to the same extent as $C_3$ plants and can thus conserve water. Some examples crabgrass, corn (maize), sugarcane, and sorghum. Although comprising only ~3% of the angiosperms by species, $C_4$ plants are responsible for ~25% of all the photosynthesis on land.

## CAM plants

CAM stands for **c**rassulacean **a**cid **m**etabolism because it was first studied in members of the plant family Crassulaceae. **CAM plants** also do the $C_4$ pathway. However, instead of segregating the $C_4$ and $C_3$ pathways in different parts of the leaf, CAM plants separate them in **time** instead (Table 5.4.1). As a result, CAM plants do not need to open their stomata in the daytime to reduce photorespiration because they have already formed a four-carbon molecule at night that can be broken down to release carbon dioxide during the day.

Table 5.4.1: Activities of CAM plants at night and in the morning.

| Night | Morning |
|---|---|
| • CAM plants take in $CO_2$ through their open stomata (they tend to have reduced numbers of them).<br>• The $CO_2$ joins with PEP to form the four-carbon oxaloacetic acid.<br>• This is converted to four-carbon **malic acid** that accumulates during the night in the central vacuole of the cells. | • The stomata close (thus conserving moisture as well as reducing the inward diffusion of oxygen).<br>• The accumulated malic acid leaves the vacuole and is broken down to release $CO_2$.<br>• The $CO_2$ is taken up into the light-independent reactions ($C_3$ pathway). |

CAM plants thus thrive in conditions of high daytime temperatures, intense sunlight, and low soil moisture. Some examples of CAM plants include cacti (Figure 5.4.14), pineapples, all epiphytic bromeliads, sedums, and the "ice plant" that invade the California coast line.



Figure **5.4.14**: Cultivated cacti in the Singapore Botanic Gardens. Image by Calvin Teo (CC-BY-SA).

## Further "Reading"

For more on the Calvin Cycle see this Calvin Cycle Video from TEDEd or this video Phyotosynthesis Animation from HHMI Biointeractive.

## Contributors and Attributions

Modified by Kyle Whittinghill (University of Pittsburgh) and Melissa Ha from

- 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).
- 8.3 Using Light Energy to Make Organic Molecules from *Biology 2e* by OpenStax (licensed CC-BY). Access for free at openstax.org.
- 3.3 Enzymatic Stage from *Introduction to Botany* by Alexey Shipunov (public domain)
- **13.6: Light-independent Reactions** by Melissa Ha, Maria Morrow, & Kammy Algiers, is licensed CC BY-SA
- **8.1: Overview of Photosynthesis** by OpenStax, is licensed CC BY
- **8.2: The Light-Dependent Reactions of Photosynthesis** by OpenStax, is licensed CC BY
- **5.3: The Calvin Cycle** by OpenStax, is licensed CC BY
- **13.7: Photorespiration and Photosynthetic Pathways** by Melissa Ha, Maria Morrow, & Kammy Algiers, is licensed CC BY-SA
- 16.2E Photorespiration and C4 Plants from *Biology* by John W. Kimball (licensed CC-BY)
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.

---

This page titled 5.4: Photosynthesis is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **8.1: Overview of Photosynthesis** by OpenStax is licensed CC BY 4.0.
- **8.2: The Light-Dependent Reactions of Photosynthesis** by OpenStax is licensed CC BY 4.0.
- **5.3: The Calvin Cycle** by OpenStax is licensed CC BY 4.0.

# 5.5: 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 **5.5.1**). Nitrogen enters the living world through **nitrogen fixation (**Figure **5.5.1 – 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 **5.5.1 – 2**).

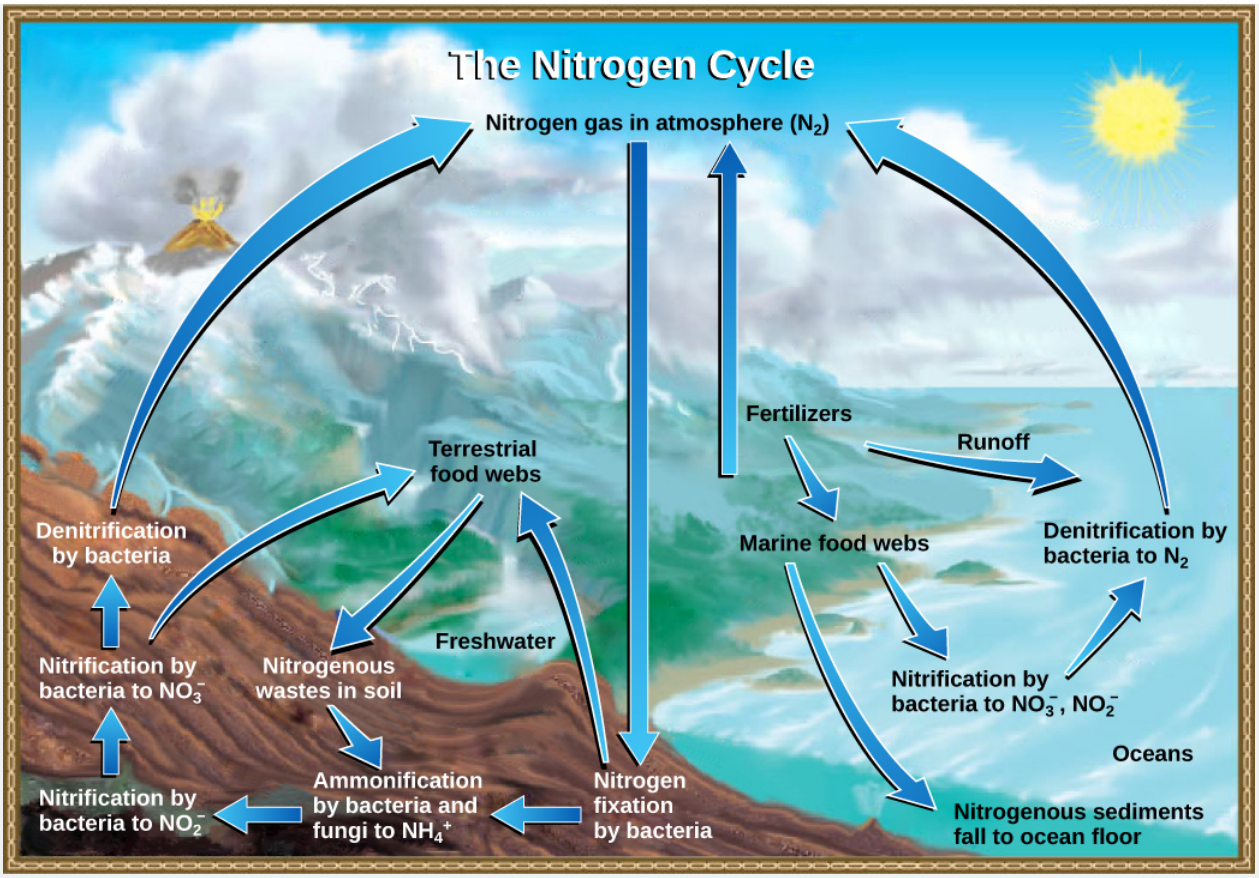



Figure **5.5.1**: Nitrogen enters the living world from the atmosphere through nitrogen-fixing bacteria. This nitrogen and nitrogenous waste from animals is then processed back into gaseous nitrogen by soil bacteria, which also supply terrestrial food webs with the organic nitrogen they need. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

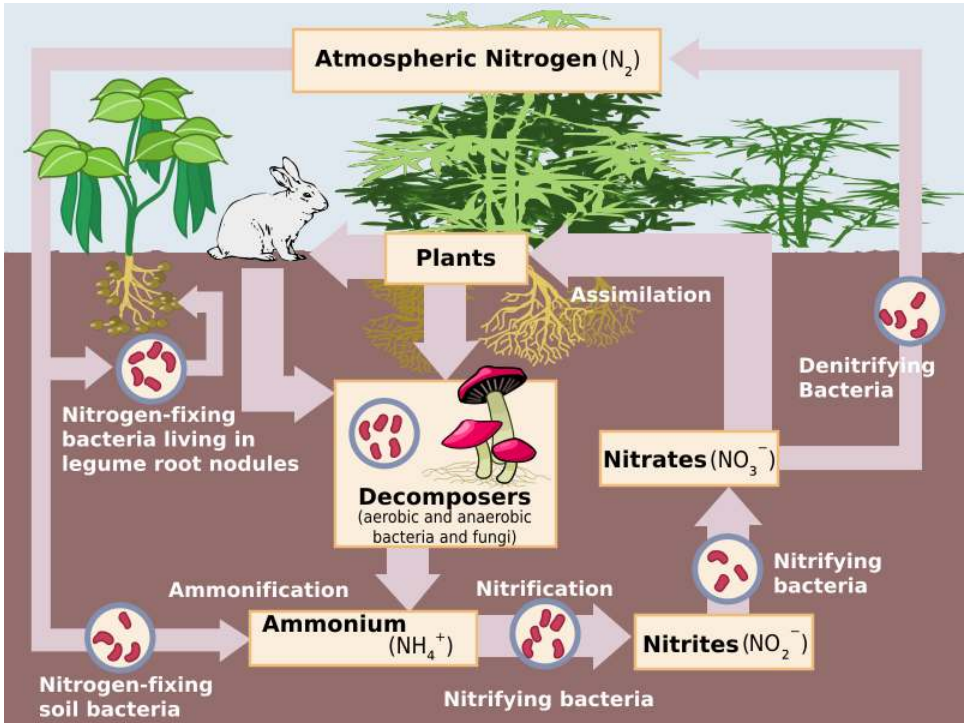



Figure 5.5.2: In the nitrogen cycle, nitrogen-fixing bacteria in the soil or legume root nodules convert nitrogen gas ($N_2$) from the atmosphere to ammonium ($NH_4^+$). Nitrification occurs when bacteria convert ammonium to nitrites ($NO_2^-$) and then to nitrates ($NO_3^-$). Nitrates re-enter the atmosphere 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 ($NH_3$). Ammonia can be used directly as fertilizer, but most of it is further processed to urea and ammonium nitrate ($NH_4NO_3$).

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 5.5.3). *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 5.5.3: 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 5.5.2). 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 5.5.2, 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 matter in 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 5.5.4). 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 5.5.4:** 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***

---

5.5: Nitrogen Cycle is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

- **20.2: Biogeochemical Cycles** by OpenStax is licensed CC BY 4.0.

# 5.6: 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 5.6.1). 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.

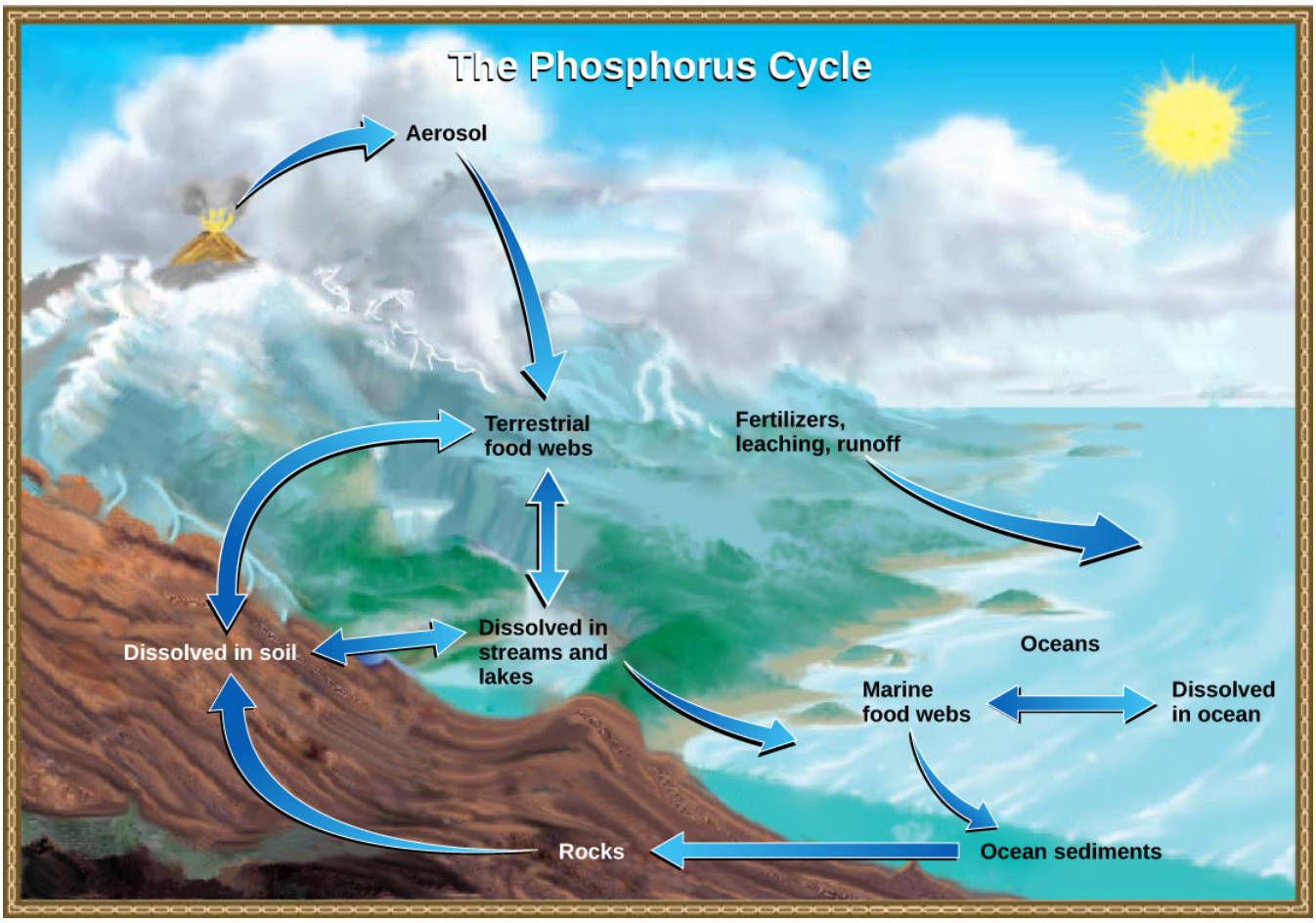



Figure **5.6.5**: In nature, phosphorus exists as the phosphate ion ($PO_4^{3-}$). Weathering of rocks and volcanic activity releases phosphate into the soil, water, and air, where it becomes available to terrestrial food webs. Phosphate enters the oceans in surface runoff, groundwater flow, and river 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***

---

5.6: Phosphorus Cycle is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

- **20.2: Biogeochemical Cycles** by OpenStax is licensed CC BY 4.0.

# 5.7: Eutrophication and Dead Zones

## 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 **5.7.1**). 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 **5.7.1**). 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.

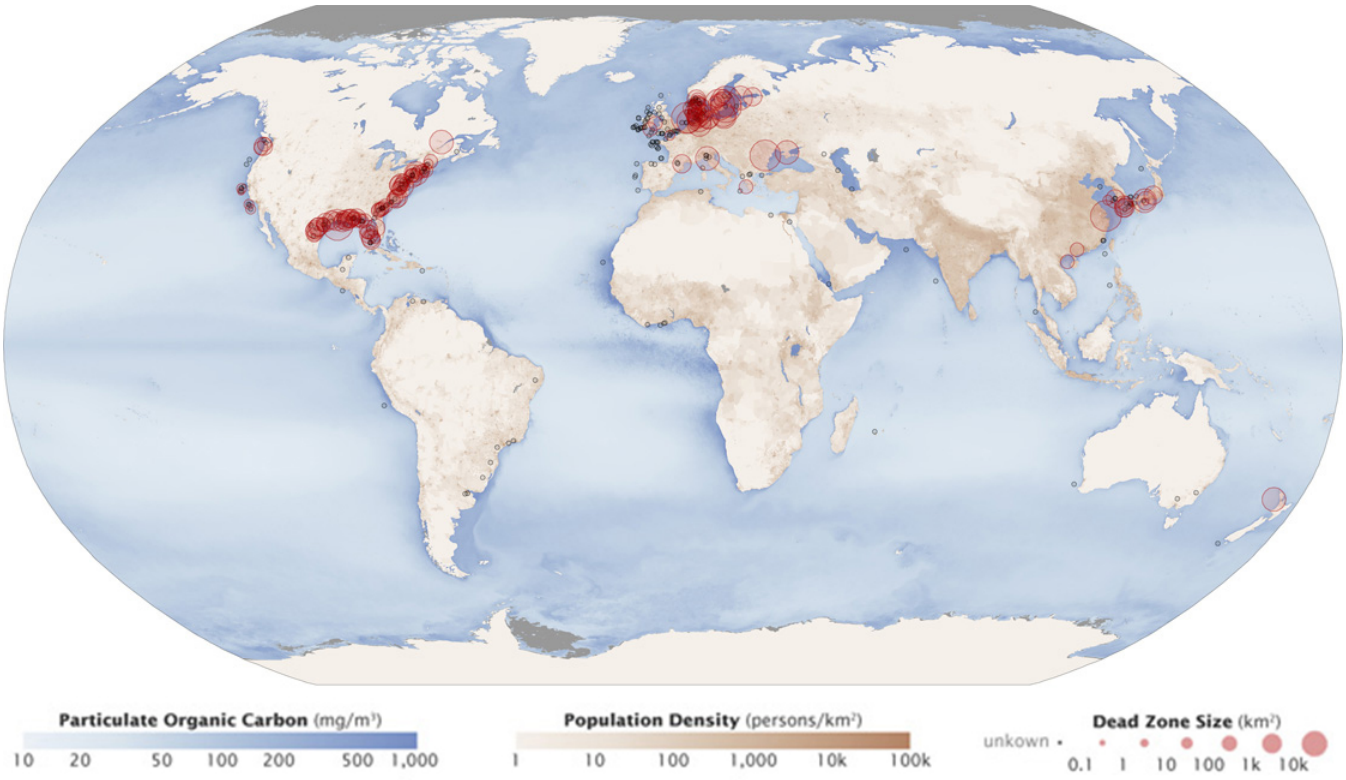



Figure 5.7.1: Dead zones occur when phosphorus and nitrogen from fertilizers cause excessive growth of microorganisms, which depletes oxygen and kills fauna. This map shows dead zones around the world in 2008. Worldwide, large dead zones are found in coastal areas of high population density. (credit: 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 5.7.1), 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 5.7.2). 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 5.7.2: 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

---

This page titled 5.7: Eutrophication and Dead Zones is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 5.8: 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 **5.8.1**, 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.

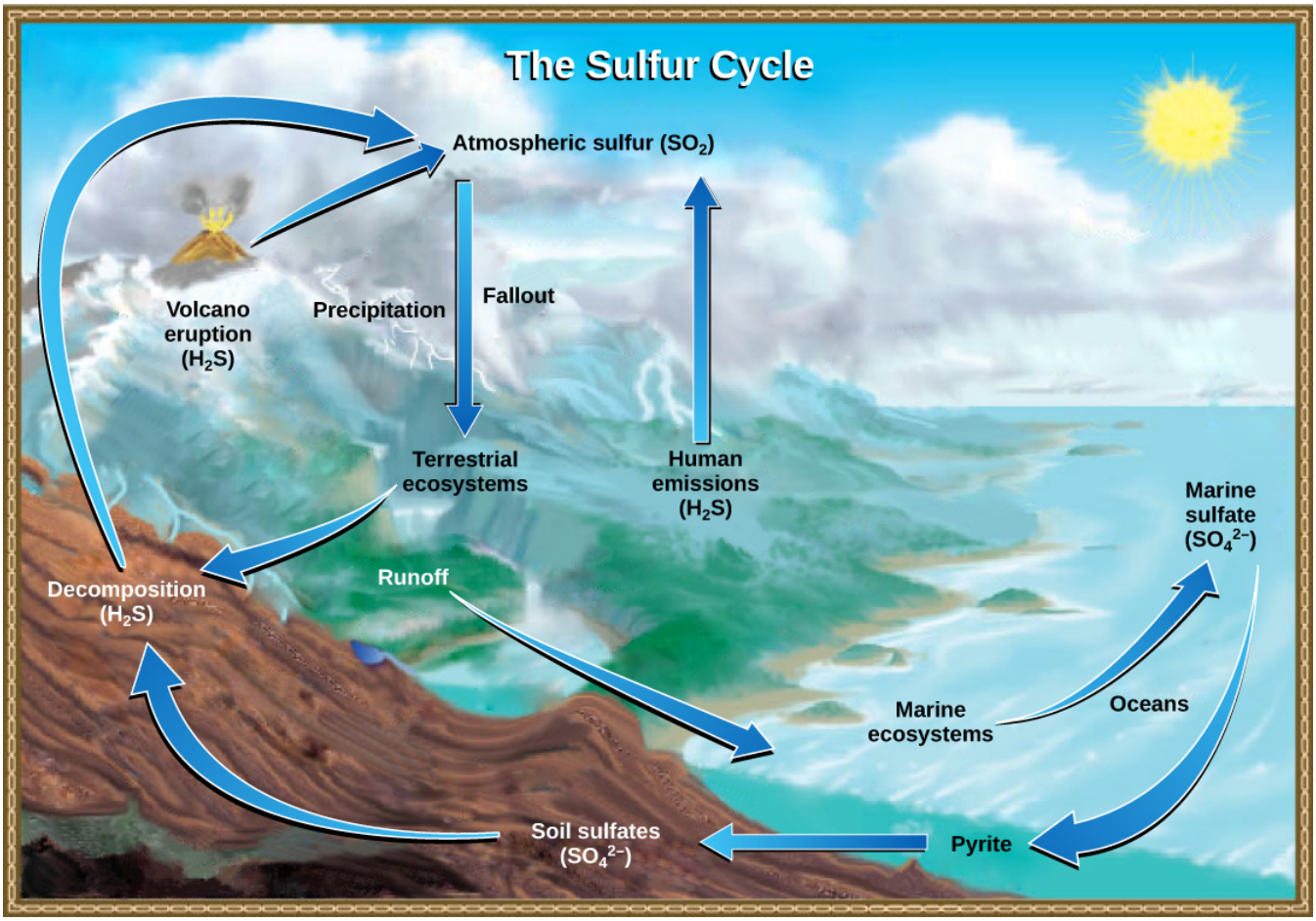



Figure **5.8.1**: The sulfur cycle. Sulfur dioxide ($SO_2$) from the atmosphere is dissolved in precipitation as weak sulfuric acid or falls directly to Earth as fallout. This releases sulfates ($SO_4^{2-}$) into the soil and water. Soil sulfates can be carried as runoff into the water. Marine sulfate 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 **5.8.2**). 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 5.8.2: 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 5.8.3), and Pompeii worms ((Figure 5.8.4).

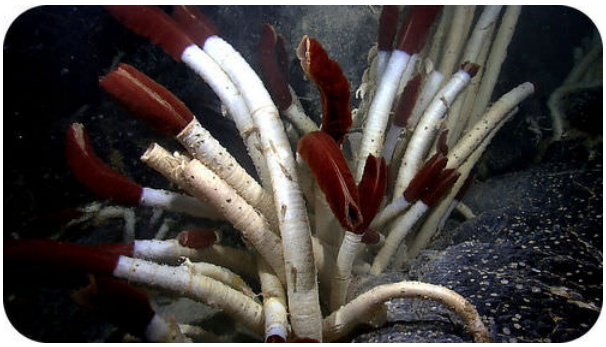


**Figure 5.8.3:** Tubeworms deep in the Galapagos Rift get their energy from chemosynthetic bacteria. Tubeworms have no mouth, eyes or stomach. Their survival depends on a symbiotic relationship with the billions of bacteria that live inside them. These bacteria convert the chemicals that shoot out of the hydrothermal vents into food for the worm.

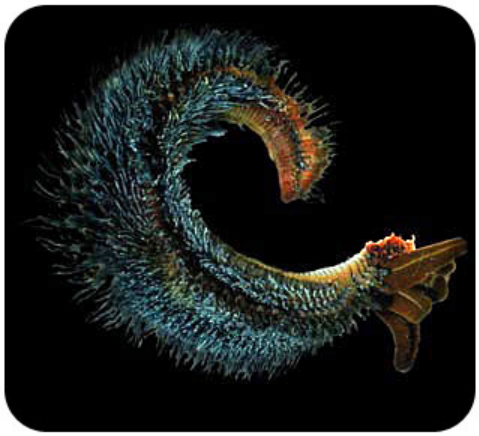


**Figure 5.8.3: Is it possible to live in temperatures over 175°F?** It is if you're a Pompeii worm. The Pompeii worm, the most heat-tolerant animal on Earth, lives in the deep ocean at super-heated hydrothermal vents. Covering this deep-sea worm's back is a fleece of bacteria. These microbes contain 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***

---

5.8: Sulfur Cycle is shared under a CC BY-NC-SA 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 CK-12 Foundation is licensed CK-12. Original source: http://www.ck12.org/book/CK-12-Biology-Concepts.

# CHAPTER OVERVIEW

## 6: Geology


6.1: Introduction to Geology
6.2: Fundamentals of Plate Tectonics
6.3: The Rock Cycle
6.4: Geologic Hazards
6.4.1: Earthquakes
6.4.2: Volcanoes
6.4.3: Mass Wasting
6.4.4: Floods
6.5: Geologic Resources
6.5.1: Mineral Resources
6.5.2: Mining
6.5.3: Fossil Fuels- Formation and Mining
6.6: Soils



6: Geology is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 6.1: Introduction to Geology

## What is Geology?

In its broadest sense, geology is the study of Earth—its interior and its exterior surface, the minerals, rocks and other materials that are around us, the processes that have resulted in the formation of those materials, the water that flows over the surface and through the ground, the changes that have taken place over the vastness of geological time, and the changes that we can anticipate will take place in the near future. Geology is a science, meaning that we use deductive reasoning and scientific methods to understand geological problems. It is, arguably, the most integrated of all of the sciences because it involves the understanding and application of all of the other sciences: physics, chemistry, biology, mathematics, astronomy, and others. But unlike most of the other sciences, geology has an extra dimension, that of time—deep time—billions of years of it. Geologists study the evidence that they see around them, but in most cases, they are observing the results of processes that happened thousands, millions, and even billions of years in the past. Those were processes that took place at incredibly slow rates—millimeters per year to centimeters per year—but because of the amount of time available, they produced massive results.

Geology is displayed on a grand scale in mountainous regions, perhaps nowhere better than the Rocky Mountains in Canada (Figure **6.1.1**). The peak on the right is Rearguard Mountain, which is a few kilometers northeast of Mount Robson, the tallest peak in the Canadian Rockies (3,954 meters). The large glacier in the middle of the photo is the Robson Glacier. The river flowing from Robson Glacier drains into Berg Lake in the bottom right. There are many geological features portrayed here. The sedimentary rock that these mountains are made of formed in ocean water over 500 million years ago. A few hundred million years later, these beds were pushed east for tens to hundreds of kilometers by tectonic plate convergence and also pushed up to thousands of meters above sea level. Over the past two million years this area—like most of the rest of Canada—has been repeatedly glaciated, and the erosional effects of those glaciations are obvious. The Robson Glacier is now only a small remnant of its size during the Little Ice Age of the 15th to 18th centuries, and even a lot smaller that it was just over a century ago in 1908. The distinctive line on the slope on the left side of both photos shows the elevation of the edge of the glacier a few hundred years ago. Like almost all other glaciers in the world, it receded after the 18th century because of natural climate change, is now receding even more rapidly because of human-caused climate change.



Figure **6.1.1**: Rearguard Mountain and Robson Glacier in Mount Robson Provincial Park, BC. Left: Robson Glacier in 2012. Right: Robson Glacier circa 1908.
(left): © Steven Earle. CC BY. (right): A.P. Coleman. Public domain. Source: Arthur P. Coleman Collection at Victoria University Library.

## How Earth Formed

The **Big Bang** (12-15 billion years ago) produced the atoms of hydrogen and helium. Elements as heavy as lithium followed the Big Bang within minutes. Eventually, under the pervasive influence of gravity, the hydrogen and helium were aggregated into immense masses that became increasingly compressed under enormously high pressure and temperature. When the pressure and temperature were sufficiently intense, nuclear fusion reactions began to occur within the masses, at which point they had become young stars. Stars such as red giants fused hydrogen and helium nuclei to form elements from carbon (the foundation of life) to

calcium (now our bones and teeth). Supernova explosions formed and ejected heavier elements such as iron (for red blood cells). Hydrogen and helium are still, however, the most abundant elements, comprising more than 99.9% of the mass of the universe.

The Sun is an ordinary star, one of billions of billions that exist in the universe. The Sun, its eight orbiting planets, plus miscellaneous comets, meteors, asteroids, and other materials (such as space dust) are collectively known as the solar system. This particular region of the universe is organized and held together by a balance of the attractive force of gravitation and counteracting influences associated with rotation and orbiting (these same forces, along with continuing expansion from the initial big bang, also organize the universe). The age of the solar system (and of Earth) is at least 4.6 billion years. The solar system began as a rotating cloud of stardust. About 4.6 billion years ago, a nearby star exploded and sent a shock wave through the dust cloud, increasing its rate of spin. As a result, most of the mass became concentrated in the middle, forming the sun. Smaller concentrations of mass rotating around the center formed the planets, including Earth.

You can watch a video showing how Earth formed at this link:http://www.youtube.com/watch?v=-x8-KMR0nx8.

Earth is the third-closest planet to the Sun. Earth is a dense planet, as are other so-called terrestrial planets located relatively close to the Sun: Mercury, Venus, and Mars. The mass of these planets consists almost entirely of heavier elements such as iron, nickel, magnesium, aluminum, and silicon. These inner planets were formed by a selective condensing of heavier elements out of the primordial planetary nebula (the disk of gases and other matter that slowly rotated around the Sun during the early stages of formation of its solar system). This happened because the inner planets were subjected to relatively intense heating by solar radiation, which caused lighter gases such as hydrogen and helium to disperse further away, to the extent that they ended up mostly in the outer, cooler planets. Meanwhile the terrestrial planets retained heavier elements. Consequently, the more distant planets in the solar system, such as Jupiter and Saturn, are relatively large, gaseous, and diffuse in character. Most of their volume is composed of an extensive atmosphere of hydrogen and helium, although these planets may contain heavier elements in their core.

At first, Earth was molten and lacked an atmosphere and oceans. The main source of heat at that time was probably the decay of naturally-occurring radioactive elements. Gradually, the planet cooled and formed a solid crust. As the planet continued to cool, volcanoes released gases, which eventually formed an **atmosphere**. The early atmosphere contained ammonia, methane,water vapor, and carbon dioxide but only a trace of oxygen. As the atmosphere became denser, clouds formed and rain fell. Water from rain, and perhaps from comets and asteroidsthat struck Earth as well, eventually formed the oceans. The ancient atmosphere and oceans represented by the picture in Figure **6.1.2** would be toxic to today's life, but they set the stage for life to begin.



**Figure 6.1.2:** Ancient Earth. This is how ancient Earth may have looked after its atmosphere and oceans formed.

## Geological Time

In 1788, after many years of geological study, James Hutton, one of the great pioneers of geology, wrote the following about the age of Earth: *The result, therefore, of our present enquiry is, that we find no vestige of a beginning — no prospect of an end.*[1] Of course he wasn't exactly correct, there was a beginning and there will be an end to Earth, but what he was trying to express is that geological time is so vast that we humans, who typically live for less than a century, have no means of appreciating how much geological time there is. Hutton didn't even try to assign an age to Earth, but we now know that it is approximately 4,570 million years old. Using the scientific notation for geological time, that is 4,570 **Ma** (for *mega annum* or "millions of years") or 4.57 **Ga**

(for *gigaannum* or billions of years). More recent dates can be expressed in **ka** (*kiloannum*); for example, the last cycle of glaciation ended at approximately 11.7 ka or 11,700 years ago.

In order to describe the time relationships between rock formations and fossils, scientists developed a relative **geologic time scale** in which the earth's history is divided and subdivided into time divisions. The three eons (**Phanerozoic**, **Proterozoic**, and **Archean**) represent the largest time divisions (measured in billions of years). They in turn are subdivided into **Eras**, **Periods** and **Epochs**. Major discontinuities in the geologic record and in the corresponding biological (fossil) record are chosen as boundary lines between the different time segments. For example, the Cretaceous-Tertiary boundary (65 million years ago) marks a sudden mass extinction of species, including the dinosaurs. Through the use of modern quantitative techniques, some rocks and organic matter can be accurately dated using the decay of naturally-occurring radioactive isotopes. Therefore, absolute ages can be assigned to some parts of the geologic time scale.

A version of the geological time scale is included as Figure **6.1.3**. Unlike time scales you'll see in other places, or even later in this book, this time scale is linear throughout its length, meaning that 50 Ma during the **Cenozoic** is the same thickness as 50 Ma during the **Hadean**—in each case about the height of the "M" in Ma. The Pleistocene glacial epoch began at about 2.6 Ma, which is equivalent to half the thickness of the thin grey line at the top of the yellow bar marked "Cenozoic." Most other time scales have earlier parts of Earth's history compressed so that more detail can be shown for the more recent parts. That makes it difficult to appreciate the extent of geological time.

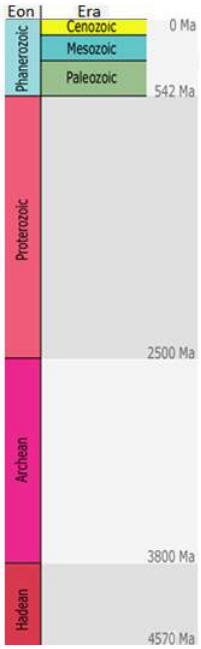



Figure **6.1.3** The geological time scale. The Hadean eon (3800 Ma to 4570 Ma), Archean eon (2500 Ma to 3800 Ma), and Proterozoic eon (542 Ma to 2500 Ma) make up 88% of geological time. The Phanerozoic eon makes up the last 12% of geological time. The Phanerozoic eon (0 Ma to 542 Ma) contains the Paleozoic, Mesozoic, and Cenozoic eras. © Steven Earle. CC BY.

To create some context, the **Phanerozoic** Eon (the last 542 million years) is named for the time during which visible (*phaneros*) life (*zoi*) is present in the geological record. In fact, large organisms—those that leave fossils visible to the naked eye—have existed for a little longer than that, first appearing around 600 Ma, or a span of just over 13% of geological time. Animals have been on land for 360 million years, or 8% of geological time. Mammals have dominated since the demise of the dinosaurs around 65 Ma, or 1.5% of geological time, and the genus *Homo* has existed since approximately 2.8 Ma, or 0.06% (1/1,600th) of geological time.

Geologists (and geology students) need to understand geological time. That doesn't mean memorizing the geological time scale; instead, it means getting your mind around the concept that although most geological processes are extremely slow, very large and important things can happen if such processes continue for enough time.

For example, the Atlantic Ocean between Nova Scotia and northwestern Africa has been getting wider at a rate of about 2.5 centimeters (cm) per year. Imagine yourself taking a journey at that rate—it would be impossibly and ridiculously slow. And yet,

since it started to form at around 200 Ma (just 4% of geological time), the Atlantic Ocean has grown to a width of over 5,000 kilometers (km)!

A useful mechanism for understanding geological time is to scale it all down into one year. The origin of the solar system and Earth at 4.57 Ga would be represented by January 1, and the present year would be represented by the last tiny fraction of a second on New Year's Eve. At this scale, each day of the year represents 12.5 million years; each hour represents about 500,000 years; each minute represents 8,694 years; and each second represents 145 years. Some significant events in Earth's history, as expressed on this time scale, are summarized on Table **6.1.1**.

Table **6.1.1** A summary of some important geological dates expressed as if all of geological time was condensed into one year.

[Skip Table]

| Event | Approximate Date | Calendar Equivalent |
|---|---|---|
| Formation of oceans and continents | 4.5 to 4.4 Ga | January |
| Evolution of the first primitive life forms | 3.8 Ga | early March |
| Formation of British Columbia's oldest rocks | 2.0 Ga | July |
| Evolution of the first multi-celled animals | 0.6 Ga or 600 Ma | November 15 |
| Animals first crawled onto land | 360 Ma | December 1 |
| Vancouver Island reached North America and the Rocky Mountains were formed | 90 Ma | December 25 |
| Extinction of the non-avian dinosaurs | 65 Ma | December 26 |
| Beginning of the Pleistocene ice age | 2 Ma or 2000 ka | 8 p.m., December 31 |
| Retreat of the most recent glacial ice from southern Canada | 14 ka | 11:58 p.m., December 31 |
| Arrival of the first people in British Columbia | 10 ka | 11:59 p.m., December 31 |
| Arrival of the first Europeans on the west coast of what is now Canada | 250 years ago | 2 seconds before midnight, December 31 |

## References


1. Hutton, J, 1788. Theory of the Earth; or an investigation of the laws observable in the composition, dissolution, and restoration of land upon the Globe. Transactions of the Royal Society of Edinburgh. ↵


## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- **5.4: How Earth Formed** by CK-12: Biology Concepts, is licensed CC BY-NC
- What is Geology and Geeological Time from Physical Geology by Steven Earle (licensed under a Creative Commons Attribution 4.0 International License)
- The Solid Earth from *AP Environmental Science* by University of California College Prep
- The Physical World from *Environmental Science: A Canadian Perspective* by Bill Freedman (Creative Commons Attribution NonCommercial)


This page titled 6.1: Introduction to Geology is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by CK-12 Foundation.

- **5.4: How Earth Formed** by CK-12 Foundation is licensed CK-12. Original source: http://www.ck12.org/book/CK-12-Biology-Concepts.

# 6.2: Fundamentals of Plate Tectonics

**Plate tectonics** is the model or theory that has been used for the past 60 years to understand and explain how the Earth works—more specifically the origins of continents and oceans, of folded rocks and mountain ranges, of earthquakes and volcanoes, and of continental drift.

Key to understanding plate tectonics is an understanding of Earth's internal structure, which is illustrated in Figure **6.2.1**. Earth's **core** consists mostly of iron. The outer core is hot enough for the iron to be liquid. The inner core—although even hotter—is under so much pressure that it is solid. The **mantle** is made up of iron and magnesium **silicate** minerals. The bulk of the mantle surrounding the outer core is solid rock, but is plastic enough to be able to flow slowly. The outermost part of the mantle is rigid. The **crust**—composed mostly of granite on the continents and mostly of basalt beneath the oceans—is also rigid. The crust and outermost rigid mantle together make up the **lithosphere**. The lithosphere is divided into about 20 **tectonic plates** that move in different directions on Earth's surface.

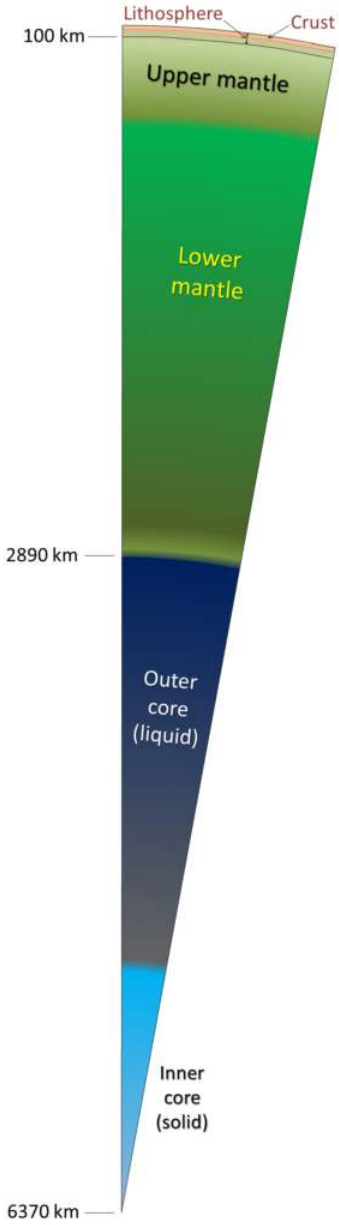



Figure **6.2.1** The components of the interior of the Earth. © Steven Earle. CC BY.

The **core** extends down to the center of the earth, a depth of about 6400 km from the surface (Figure **6.2.1**). The core makes up 16 percent of the volume of the earth and about 31 percent of the mass. It can be divided into two regions: a solid inner core and a liquid outer core. The inner core is probably mostly metallic iron alloyed with a small amount of nickel, as its density is somewhat greater than that of pure metallic iron. The outer core is similar in composition, but probably also contains small amounts of lighter elements, such as sulfur and oxygen, because its density is slightly less than that of pure metallic iron. The presence of the lighter elements depresses the freezing point and is probably responsible for the outer core's liquid state.

The **mantle** is the largest layer in the earth, making up about 82 percent of the volume and 68 percent of the mass of the earth (Figure **6.2.1**). The mantle is dominated by magnesium and iron-rich (mafic) minerals. Heat from the core of the earth is transported to the crustal region by large-scale convection in the mantle. Near the top of the mantle is a region of partially melted rock called the **asthenosphere**. Numerous small-scale convection currents occur here as hot **magma** (i.e., molten rock) rises and cooler magma sinks due to differences in density.The mantle extends down to a depth of 2900 km where the core begins.

The **crust** is the thinnest layer in the earth, extending downward from the surface to an average depth of 35 km, making up only 1 percent of the mass and 2 percent of the volume (Figure **6.2.1**). Relative to the rest of the earth, the crust is rich in elements such as silicon, aluminum, calcium, sodium and potassium. Crustal materials are very diverse, consisting of more than 2000 minerals. The less dense crust floats upon the mantle in two forms: the **continental crust** and the **oceanic crust**. The oceanic crust, which contains more **mafic minerals** is thinner and denser than the continental crust which contains minerals richer in silicon and aluminum. The thick continental crust has deep buoyant roots that help to support the higher elevations above. The crust contains the mineral resources and the fossil fuels used by humans.

An important property of Earth (and other planets) is that the temperature increases with depth, from close to 0°C at the surface to about 7000°C at the centre of the core. In the crust, the rate of temperature increase is about 30°C every kilometre. This is known as the **geothermal gradient**.

Heat is continuously flowing outward from Earth’s interior, and the transfer of heat from the core to the mantle causes convection in the mantle (Figure **6.2.2**). This convection is the primary driving force for the movement of tectonic plates. At places where convection currents in the mantle are moving upward, new lithosphere forms (at ocean ridges), and the plates move apart (diverge). Where two plates are converging (and the convective flow is downward), one plate will be **subducted** (pushed down) into the mantle beneath the other. Many of Earth’s major earthquakes and volcanoes are associated with convergent boundaries.

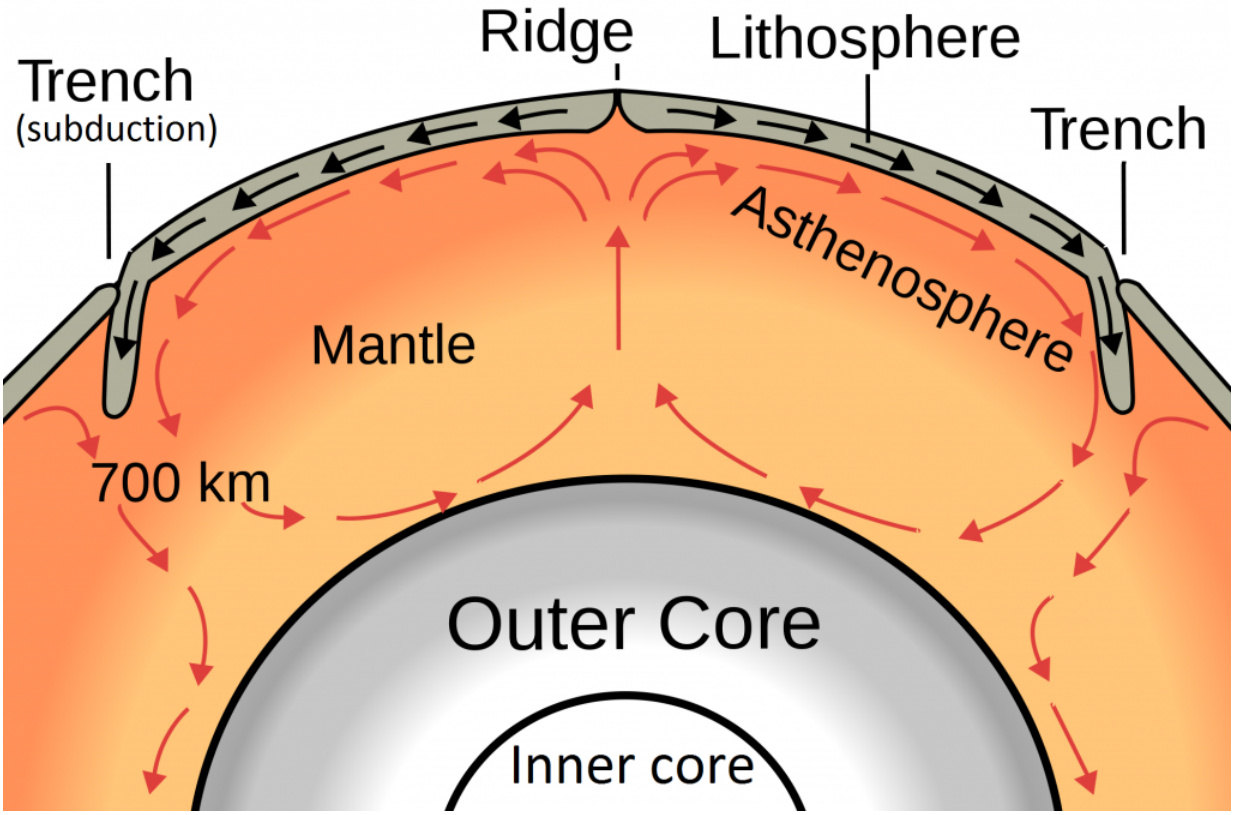



Figure 6.2.2 Depiction of the convection in the mantle and it's relationship to plate motion. Oceanic Spreading by Surachit. Public domain.

Earth's major tectonic plates and the directions and rates at which they are diverging at sea-floor ridges, are shown in Figure 6.2.3. A convergent (colliding) plate boundary occurs when two plates collide. If the convergent boundary involves two continental plates, the crust is compressed into high mountain ranges such as the Himalayas. If an oceanic plate and a continental plate collide, the oceanic crust (because it is more dense) is subducted under the continental crust. The region where subduction takes place is called a **subduction zone** and usually results in a deep ocean trench such as the **"Mariana Trench"** in the western Pacific ocean. The subducted crust melts and the resultant magma can rise to the surface and form a volcano. A **divergent plate** boundary occurs when two plates move away from each other. Magma upwelling from the mantle region is forced through the resulting cracks, forming new crust. The mid-ocean ridge in the Atlantic ocean is a region where new crustal material continually forms as plates diverge. Volcanoes can also occur at divergent boundaries. The island of Iceland is an example of such an occurrence. A third type of plate boundary is the **transform boundary**. This occurs when two plates slide past one another. This interaction can build up strain in the adjacent crustal regions, resulting in earthquakes when the strain is released. The San Andreas Fault in California is an example of a transform plate boundary.

Using either a map of the tectonic plates from the Internet or Figure 6.2.3 determine which tectonic plate you are on right now, approximately how fast it is moving, and in what direction. How far has that plate moved relative to Earth's core since you were born?

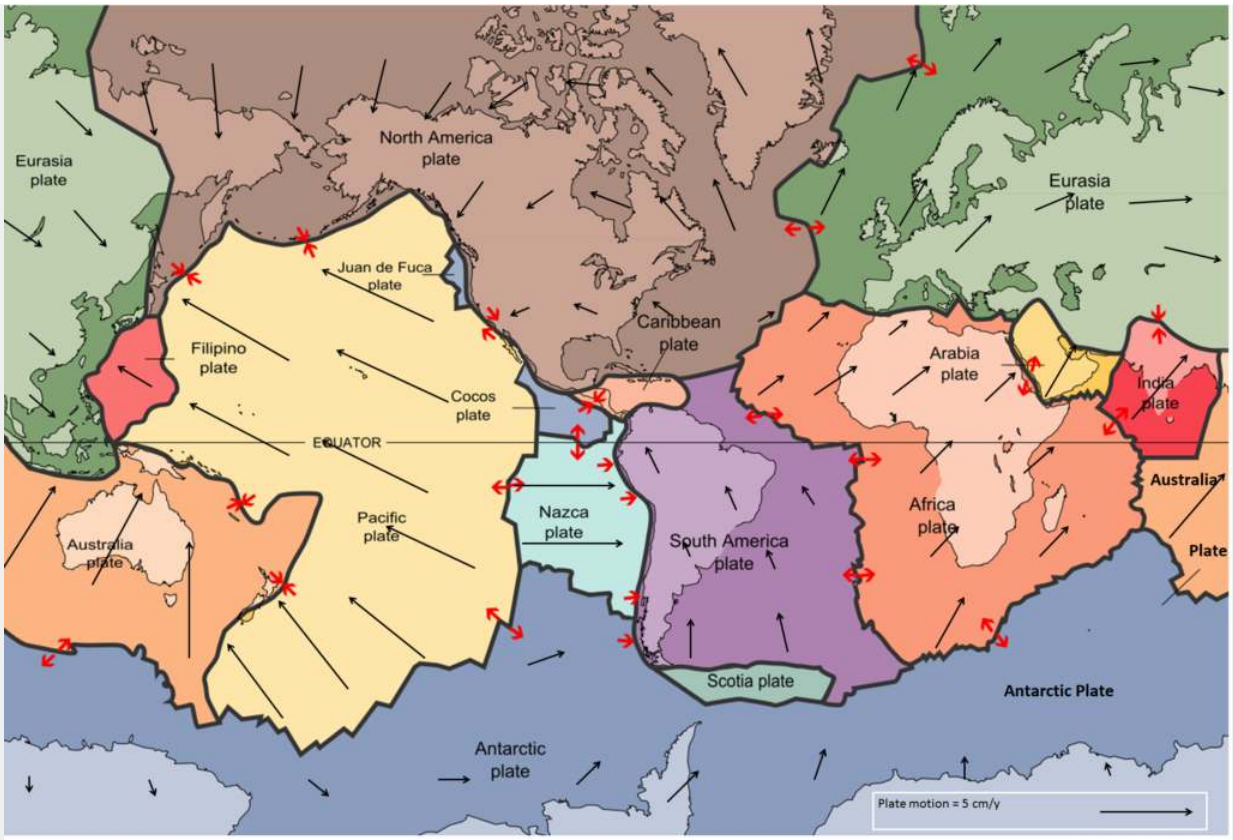



Figure **6.2.3** A map showing 15 of the Earth's tectonic plates and the approximate rates and directions of plate motions. Tectonic Plates by USGS. Public domain. Adapted by Steven Earle.

## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- Fundamentals of Plate Tectonics from *Physical Geology* by Steven Earle (licensed under a Creative Commons Attribution 4.0 International License)
- The Solid Earth from *AP Environmental Science* by University of California College Prep

This page titled 6.2: Fundamentals of Plate Tectonics is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Steven Earle (BCCampus) .

# 6.3: The Rock Cycle

The rock components of the crust are slowly but constantly being changed from one form to another and the processes involved are summarized in the **rock cycle** (Figure **6.3.3**). The rock cycle is driven by two forces: (1) Earth's internal heat engine, which moves material around in the core and the mantle and leads to slow but significant changes within the crust, and (2) the hydrological cycle, which is the movement of water, ice, and air at the surface, and is powered by the sun.

However, before we begin our discussion of the rock cycle, let's clarify a very important point: *rocks and minerals are not the same thing!* A **mineral** is a naturally occurring, inorganic, solid element or compound, with a definite composition or compositional range and a regular internal crystal structure. A mineral is a naturally occurring combination of specific elements that are arranged in a particular repeating three-dimensional structure or **lattice**.[1] The mineral **halite** is shown as an example in Figure **6.3.1**.

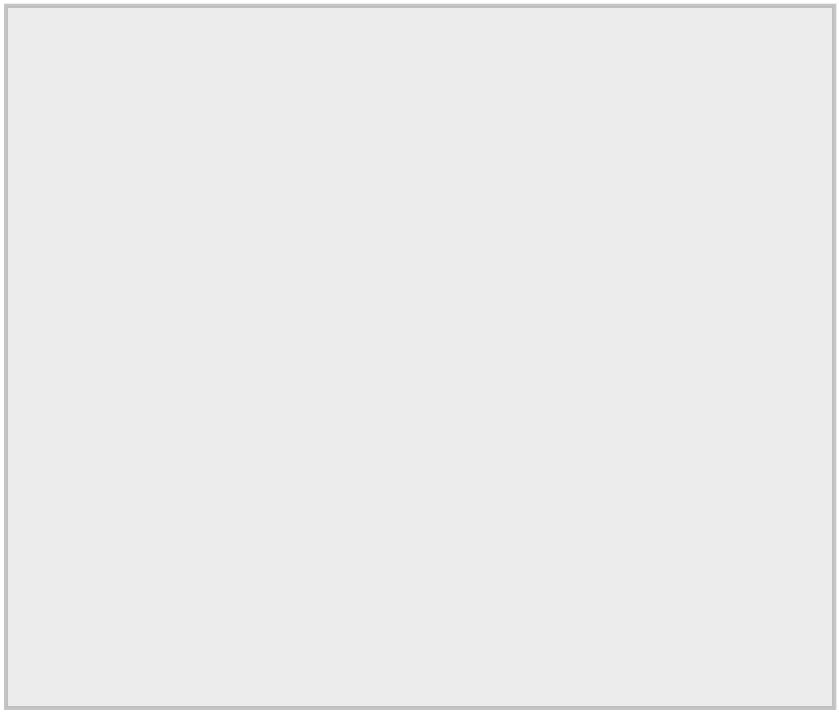


Figure **6.3.1**: The lattice structure and composition of the mineral halite (common table salt). © Steven Earle. CC BY

In this case, atoms of sodium (Na: purple) alternate with atoms of chlorine (Cl: green) in all three dimensions, and the angles between the bonds are all 90°. Even in a tiny crystal, like the ones in your salt shaker, the lattices extend in all three directions for thousands of repetitions. Halite always has this composition and this structure.

A **rock** is a solid, cohesive aggregate of one or more minerals or mineral materials. The key difference between a rock and a mineral is the 'regular internal crystal structure' of a mineral. For example, Quartz is a mineral ($SiO_2$), while Granite is a rock. If you were to look at a quartz crystal at the molecular level, the entire crystal would be made up of identical building blocks ($SiO_2$ molecules). Granite, on the other hand is not homogenous. A close-up view of **granite**, a common rock, is shown in Figure **6.3.2**. Although a hand-sized piece of granite may have thousands of individual mineral crystals in it, there are typically only a few different minerals, as shown here. Therefore, one chunk of a piece of granite, when viewed at the molecular level would appear very different from another chunk viewed at the molecular level because the first chunk might be a quartz crystal, while the second chunk might be another type of crystal.

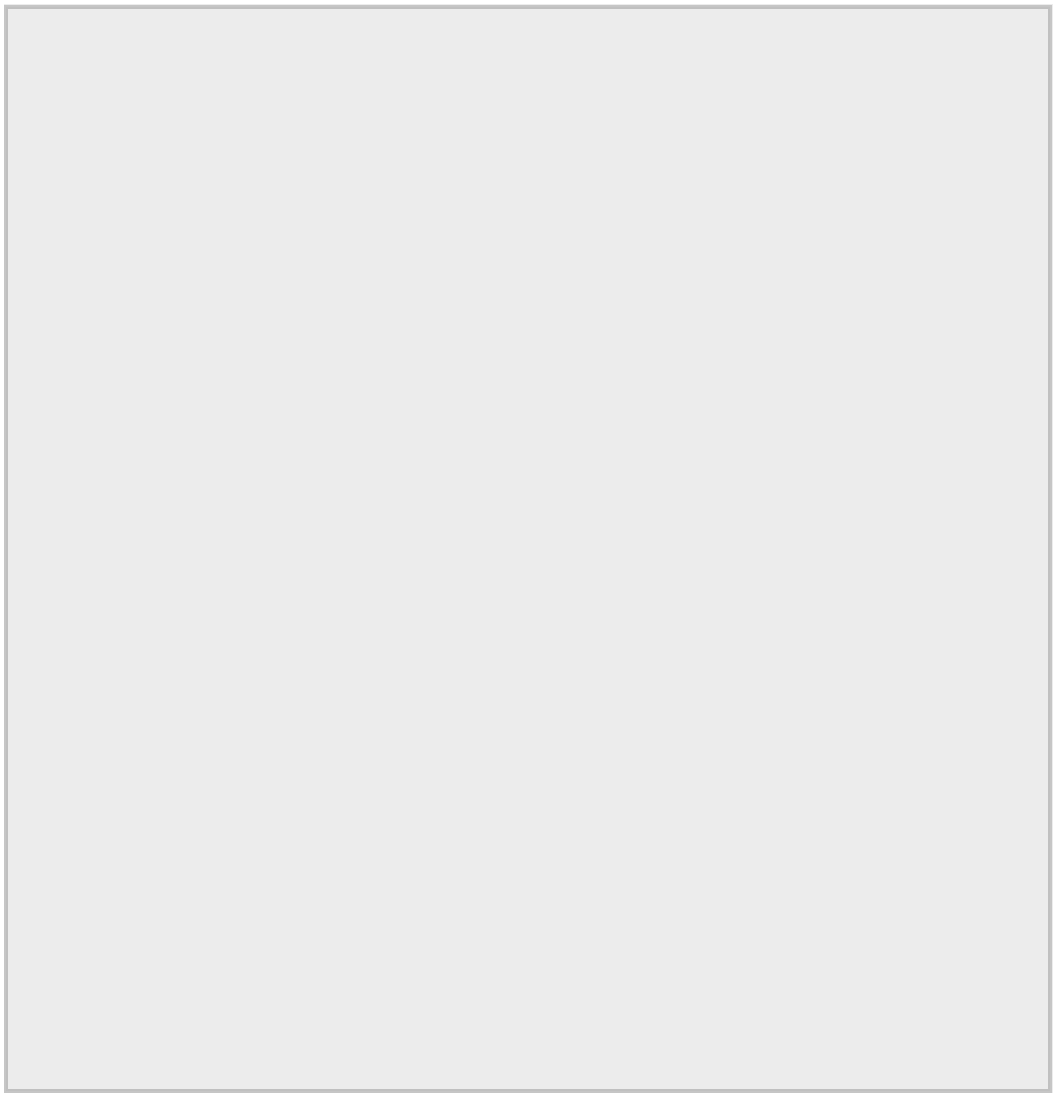


Figure **6.3.2**: A close-up view of the rock granite and some of the minerals that it typically contains (H = hornblende (amphibole), Q = quartz and F = feldspar). The crystals range from about 0.1 to 3 millimeters (mm) in diameter. Most are irregular in outline, but some are rectangular.

Rocks are categorized by how they are formed. Rocks that are formed directly from liquid rock, or magma that wells up from deep in the earth are called **igneous rocks.** Rocks that form at, or just below the surface of the earth from weathered and decomposed bits of other rocks are called **sedimentary rocks**. Igneous or sedimentary rocks which are subjected to intense pressure or temperature (but not enough to melt them) form **metamorphic rocks.**

The rock cycle is still active on Earth because our core is hot enough to keep the mantle moving, our atmosphere is relatively thick, and we have liquid water (Figure **6.3.3**). On some other planets or their satellites, such as the Moon, the rock cycle is virtually dead because the core is no longer hot enough to drive mantle convection and there is no atmosphere or liquid water.

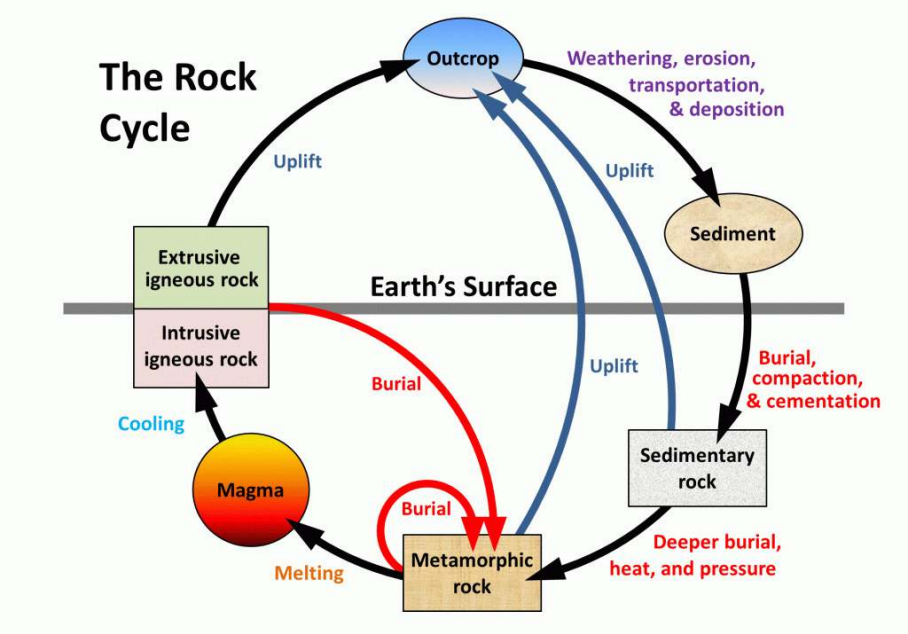



Figure **6.3.3**: A schematic view of the rock cycle. The rock cycle takes place both above and below the Earth's surface. The rock deepest beneath the earth's surface, and under extreme heat and pressure, is metamorphic rock. This metamorphic rock can melt and become magma. When magma cools below the earth's surface, it becomes "intrusive igneous rock." If magma cools above the earth's surface, it is "extrusive igneous rock" and becomes part of the outcrop. The outcrop is subject to weathering and erosion, and can be moved and redeposited around the earth by forces such as water and wind. As the outcrop is eroded, it becomes sediment which can be buried, compacted, and cemented beneath the Earth's surface to become sedimentary rock. As sedimentary rock gets buried deeper and comes under increased heat and pressure, it returns to its original state as metamorphic rock. Rocks in the rock cycle do not always make a complete loop. It is possible for sedimentary rock to be uplifted back above the Earth's surface and for intrusive and extrusive igneous rock to be reburied and become metamorphic rock. © Steven Earle. CC BY.

In describing the rock cycle, we can start anywhere we like, although it's convenient to start with magma. As we'll see in more detail below, magma is rock that is hot to the point of being entirely molten, with a temperature of between about 800° and 1300°C, depending on the composition and the pressure.



Figure **6.3.4** Magma forming pahoehoe basalt at Kilauea Volcano, Hawaii. © Steven Earle. CC BY.

Magma can either cool slowly within the crust (over centuries to millions of years)—forming **intrusive igneous rock**, or erupt onto the surface and cool quickly (within seconds to years)—forming **extrusive igneous rock** (volcanic rock) (Figure **6.3.4**). Intrusive igneous rock typically crystallizes at depths of hundreds of meters to tens of kilometers below the surface. The size of crystals in igneous rocks is inversely related to the speed at which the rock cooled. Therefore, extrusive igneous rocks have very small crystals or no crystals at all because they cooled so rapidly. Basalt is a common example of an extrusive igneous rock. Because intrusive ingeneous rocks cool very slowly large (1 cm to 1 m) crystals can form. Granite is an example of an intrusive igneous rock. Most of the Sierra Nevada and Cascade mountain ranges are made up of granite. To change its position in the rock cycle, intrusive igneous rock has to be uplifted and then exposed by the erosion of the overlying rocks.

Through the various plate-tectonics-related processes of mountain building, all types of rocks are uplifted and exposed at the surface. Once exposed, they are weathered, both physically (by mechanical breaking of the rock) and chemically (by weathering of the minerals), and the weathering products—mostly small rock and mineral fragments—are eroded, transported, and then deposited as **sediments**. Sediment is a loose term for anything ranging from fist-sized chunks of rock in a stream bed to gravel, sand and mud. Dead organic material is also sediment. Basically, all the unconsolidated material you see at the surface of the earth is sediment. Transportation and deposition of sediments occur through the action of glaciers, streams, waves, wind, and other agents, and sediments are deposited in rivers, lakes, deserts, and the ocean.

Unless they are re-eroded and moved along, sediments will eventually be buried by more sediments. At depths of hundreds of meters or more, they become compressed and cemented into **sedimentary rock** (See Figure **6.3.5** for example). There are two types of sedimentary rocks. The first is the *clastic sedimentary rock*. Over time (geologic time) new sediment covers up old sediment, and there is enough pressure on the sediment to *lithify* it. *Lithification* is the process of compaction, cementation and hardening of sediments into sedimentary rocks. Clastic sedimentary rocks are composed of fragments of other rocks. They are often stacked in distinct layers with the oldest material on the bottom. An example of a common clastic sedimentary rock is sandstone, such as the sandstone layers of the Grand Canyon. The second type of sedimentary rock, the *chemical sedimentary rock*, forms when water seeps through sediments and dissolves certain minerals, then deposits the minerals in a new location. An example of this is deposits of gypsum or halite (table salt) which often form in the desert after rain. Another example of a chemical sedimentary rock is limestone. Limestone is precipitated calcium carbonate ($CaCO_3$). Limestone is the primary component of much of the central United States. One problem with chemical sedimentary rocks is that it is very easy for water to re-dissolve them. Sink holes form where water has dissolved limestone in the ground, forming a cave which then collapses. Sink holes are a common hazard of areas that have thick limestone deposits.



Figure **6.3.5** Cretaceous-aged marine sandstone overlying marine mudstone, Gabriola Island, B.C. © Steven Earle. CC BY.

Again through various means, largely resulting from plate-tectonic forces, different kinds of rocks are either uplifted, to be re-eroded, or buried deeper within the crust where they are heated up, squeezed, and changed into **metamorphic rock** (Figure **6.3.6**). Metamorphic rocks have been subjected to such severe pressures and/or temperatures that the chemical elements that make up the minerals in the rock rearrange themselves to form new minerals. Metamorphic rocks reach the surface the same way that intrusive igneous rocks reach the surface, through uplift, weathering and erosion. Metamorphism (the process that creates metamorphic rocks) also changes the appearance of landscapes. You can often see twisted and folded layers of rocks in metamorphic regions. Gneiss, for example, is a metamorphic rock derived from granite, while marble is derived from limestone, and slate from shale.



Figure **6.3.6** Metamorphosed and folded Triassic-aged limestone, Quadra Island, B.C. © Steven Earle. CC BY.

# Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- The Rock Cycle and Minerals and Rocks from Physical Geology by Steven Earle (licensed under a Creative Commons Attribution 4.0 International License)
- The Rock Cycle by K. Allison Lenkeit-Meezan (Foothill College)
- The Physical World from *Environmental Science: A Canadian Perspective* by Bill Freedman (Creative Commons Attribution NonCommercial)

---

This page titled 6.3: The Rock Cycle is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Steven Earle (BCCampus) .

# 6.4: Geologic Hazards

6.4: Geologic Hazards is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 6.4.1: Earthquakes

An earthquake is the shaking caused by the **rupture** (breaking) and subsequent displacement of rocks (one body of rock moving with respect to another) beneath Earth's surface. A body of rock that is under stress becomes deformed. When the rock can no longer withstand the deformation, it breaks and the two sides slide past each other. The region where the rupture occurs is called the **focus**. This is often deep below the surface of the crust. The point on the surface directly above the focus is called the **epicenter**. Destructive waves propagate outward from the region of the quake, traveling throughout the earth. The magnitude of an earthquake is a measure of the total amount of energy released. The first step in determining the magnitude is to measure the propagated waves using a device called a **seismograph**. Based on this information, the earthquake is given a number classification on a modified **Richter scale**. The scale is logarithmic, so a difference of one unit means a difference of ten-fold in wave intensity, which corresponds to an energy difference of 32-fold. The intensity of an earthquake is an indicator of the effect of an earthquake at a particular locale. The effect depends not only on the magnitude of the earthquake, but also the types of subsurface materials and the structure and design of surface structures.

Earthquakes generally occur along breaks in the rock mass known as **faults**, and most occur in regions near plate boundaries, although not all earthquakes happen on pre-existing faults. Some 80 percent of all earthquakes occur near convergent plate boundaries, triggered by the interaction of the plates. Earthquakes are also often associated with volcanic activity due to the movement of sub-surface magma.

An intense earthquake can cause great damage to buildings, and the collapsing structures, fires, and other destruction can take a great toll on people. In 1556 an earthquake struck Shanxi Province in China and caused about 830,000 deaths, making it the most deadly earthquake in recorded history. The most famous catastrophic earthquake in North America was the San Francisco event in 1906, caused by slippage along the San Andreas Fault, which killed 503 people and resulted in tremendous physical damage. However, other earthquakes during the twentieth century resulted in much greater losses of human lives, including one in 1976 that killed 242,000 people in Tangshan, China; another in 1927 that killed 200,000 in Nan-Shan, China; and one in Tokyo–Yokohama, Japan, that killed 200,000 in 1926. Notable recent earthquakes include one in Kobe, Japan (1995) that killed 5,500 people, another in Kashmir (2005) that killed 79,000 people, one in Sichuan, China (2008) that killed 70,000 people, and one in the Tohuku region of Japan (2011) that generated a tsunami (seismic sea wave) that killed 16,000 people. The events in San Francisco (1906) and Tokyo (1926) affected large cities. The powerful tremors caused great damage, partly because of weak architectural designs that were unable to withstand the strong forces. In both cases, however, about 90% of the actual destruction resulted from fires. Earthquakes can also cause soil to lose some of its mechanical stability, resulting in destructive landslides and subsidence (sinking) of land and buildings.

When an earthquake occurs under the ocean, it can trigger a destructive tidal wave known as a **tsunami**. A tsunami may be almost indiscernible in deep water of the open ocean, but it can become gigantic when the wave reaches shallow water and piles up to heights that can swamp coastal villages and towns. In 1946, a large earthquake centered on Umiak Island in the Aleutian Islands (Alaska) caused a tsunami to strike Hawaii, 4,500 km away, with an 18 m crest. The greatest tsunami of recent times was triggered by an undersea, so-called "megathrust" earthquake on December 26, 2004. Its epicenter was located about 40 km off the coast of Aceh in northern Sumatra, an Indonesian island, and it registered a massive 9.2 on the Richter scale (making it the largest earthquake in 40 years) (Figure **6.4.1.1**). The tremor generated an immense tsunami (actually, a close series of individual waves) that gathered to a height up to 30 m when it impacted shallow-sloping coasts of countries fringing the Indian Ocean. The colossal tsunami waves were moving at speeds of about 60 kmhr when they impacted the shore. They caused widespread devastation and more than 225,000 deaths by drowning and injuries caused by floating debris and collapsing buildings. The hardest-hit places were Sumatra (which suffered at least 168,000 dead and missing), Sri Lanka (35,000), the eastern coast and islands of India (18,000), and Thailand (8,000). At least 7,000 of the deaths were tourists from developed countries who were visiting coastal resorts during their holiday break. In addition to the mortality, tens of millions of people were displaced from their homes and livelihoods by the flooding. In many of the worst-hit places, the damage was made much more severe because of increased coastal vulnerability caused by the removal of previously abundant mangrove forest, mostly to develop tourist resorts and brackish ponds for shrimp aquaculture. Where mangroves remained intact, the coastal forest provided a sea-wall that helped to absorb much of the force of the tsunami, providing a measure of protection to areas further inland.



Figure **6.4.1.1**: A devastated village on the coast of Aceh, Sumatra, after the killer tsunami of December 26, 2004. Source: image by P.A. McDaniel, United States Navy, ID 050102-N-9593M-040; https://commons.wikimedia.org/wiki/File%3AUS_Navy_050102-N-9593M-040_A_village_near_the_coast_of_Sumatra_lays_in_ruin_after_the_Tsunami_that_struck_South_East_Asia.jpg

## Earthquake Processes

The engineering principle of **elastic deformation**, which can be used to understand earthquakes, is illustrated in Figure **6.4.1.2**. The stress applied to a rock—typically because of ongoing plate movement—results in strain or deformation of the rock (Figure **6.4.1.2** b). Because most rock is strong (unlike loose sand, for example), it can withstand a significant amount of deformation without breaking. But every rock has a deformation limit and will rupture (break) once that limit is reached. At that point, in the case of rocks within the crust, the rock breaks and there is displacement along the **rupture surface** (Figure **6.4.1.2**c). The magnitude of the earthquake depends on the extent of the area that breaks (the area of the rupture surface) and the average amount of displacement (sliding).

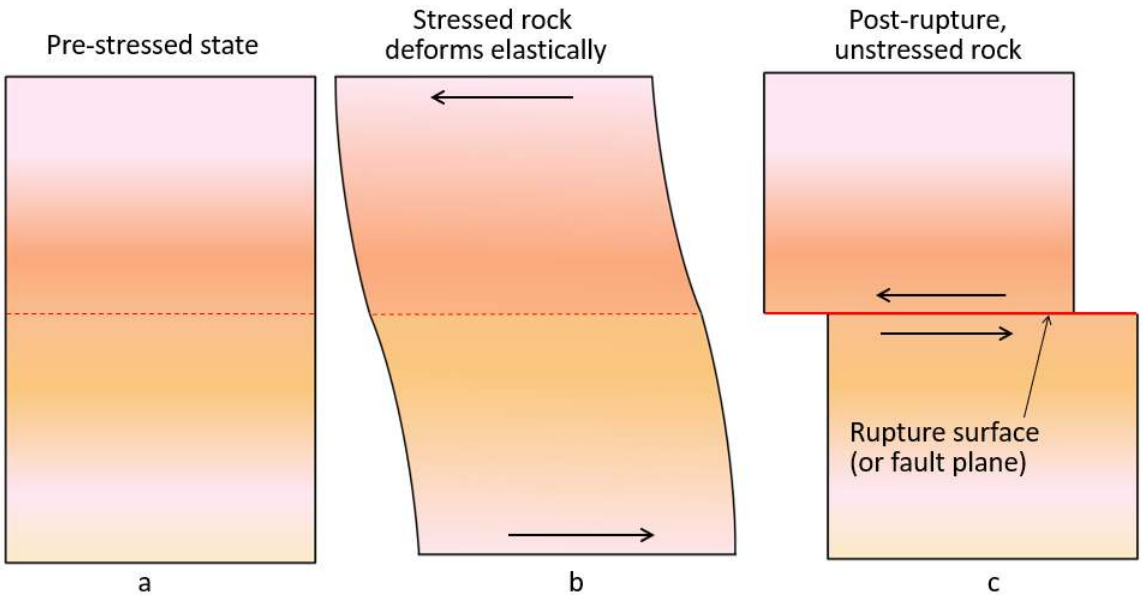



Figure **6.4.1.2** Depiction of the concept of elastic deformation and rupture. The plate boundary is shown as a dashed red line. In b the two plates are moving as shown by the arrows, but they are locked against each along the plate boundary so they are both deforming and the rocks are stressed. In c there has been a rupture along the boundary and the stress is released. © Steven Earle. CC BY

The concept of a rupture surface, which is critical to understanding earthquakes, is illustrated in Figure **6.4.1.3**. An earthquake does not happen at a point, it happens over an area within a plane, although not necessarily a *flat* plane. Within the area of the rupture surface, the amount of displacement is variable (Figure **6.4.1.3**), and, by definition, it decreases to zero at the edges of the rupture surface because the rock beyond that point isn't displaced at all. The extent of a rupture surface and the amount of displacement will depend on a number of factors, including the type and strength of the rock, and the degree to which it was stressed beforehand.

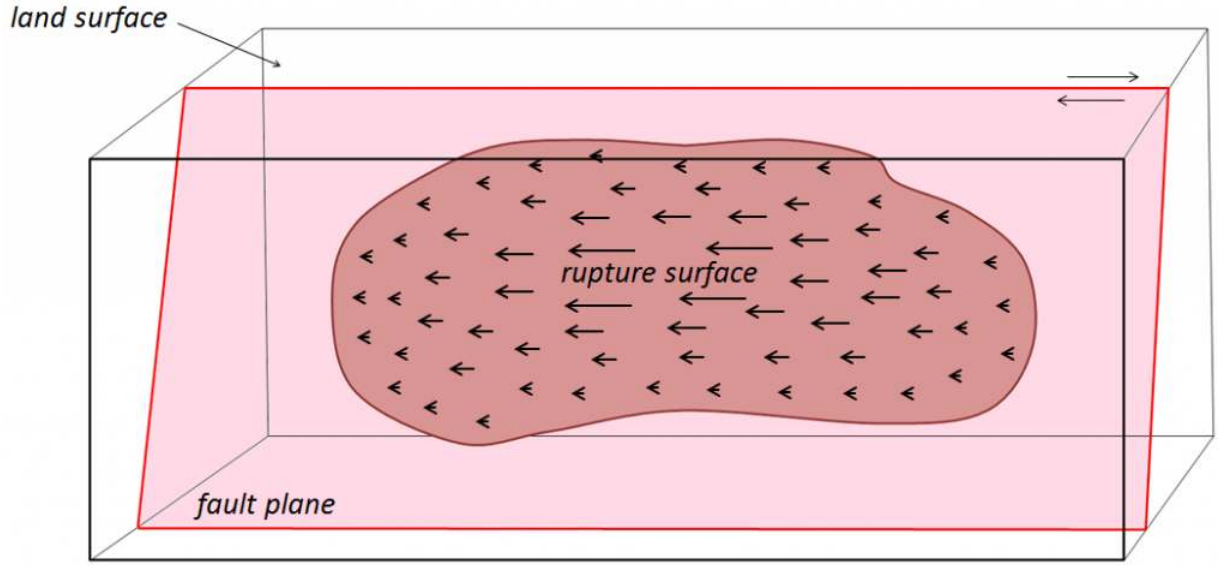



Figure **6.4.1.3** A rupture surface (dark pink), on a steeply dipping fault plane (light pink). The diagram represents a part of the crust that may be up to tens or hundreds of kilometers long. The rupture surface is the part of the fault plane along which displacement occurred. In this example, the near side of the fault is moving to the left, and the lengths of the arrows within the rupture surface represent relative amounts of displacement. © Steven Earle. CC BY

Earthquake rupture doesn't happen all at once; it starts at a single point and spreads rapidly from there. Depending on the extent of the rupture surface, the propagation of failures out from the point of initiation is typically completed within seconds to several tens of seconds (Figure **6.4.1.4**). The initiation point isn't necessarily in the centre of the rupture surface; it may be close to one end, near the top, or near the bottom.

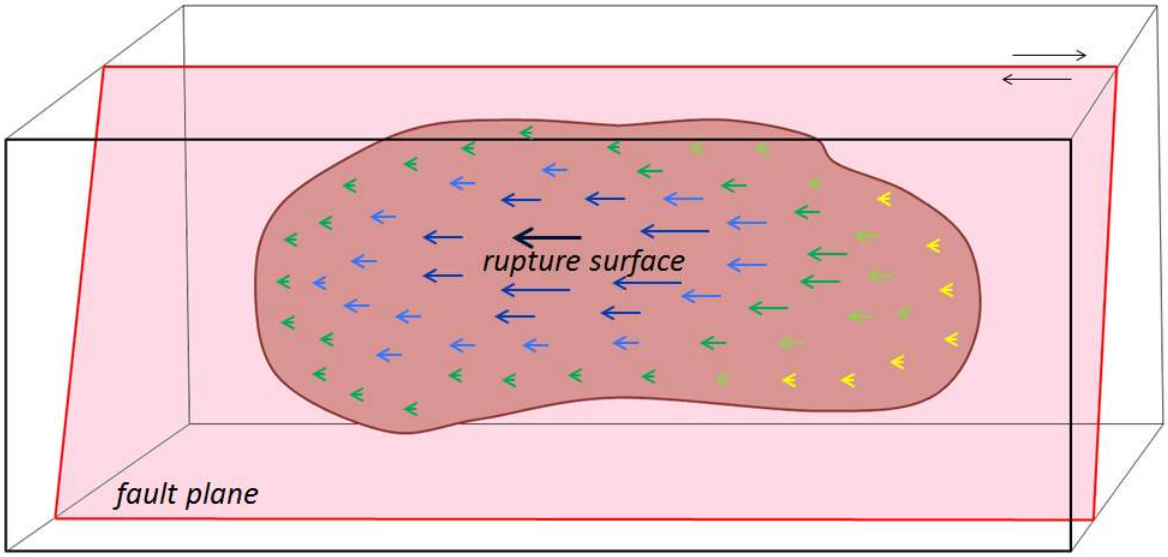



Figure **6.4.1.4** Propagation of failure on a rupture surface. In this case, the failure starts at the dark blue heavy arrow in the centre and propagates outward, reaching the left side first (green arrows) and the right side last (yellow arrows). © Steven Earle. CC BY

Figure **6.4.1.5** shows the distribution of immediate **aftershocks** associated with the 1989 Loma Prieta earthquake. Panel (b) is a section along the San Andreas Fault; this view is equivalent to what is shown in Figures **6.4.1.3** and **6.4.1.4**. The area of red dots is the rupture surface; each red dot is a specific aftershock that was recorded on a seismometer. The hexagon labelled "main earthquake" represents the first or main shock. When that happened, the rock at that location broke and was displaced. That released the stress on that particular part of the fault, but it resulted in an increase of the stress on other nearby parts of the fault, and contributed to a cascade of smaller ruptures (aftershocks), in this case, over an area about 60 kilometers long and 15 kilometers wide.

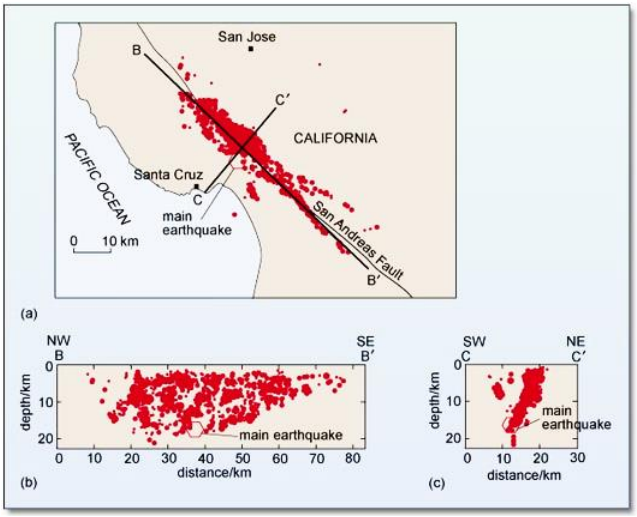



Figure **6.4.1.5** Distribution of the aftershocks of the 1989 M 6.9 Loma Prieta earthquake (a: plan view, b: section along the fault, c: section across the fault.) © Open University. CC BY-SA.

So, what exactly is an aftershock then? An aftershock is an earthquake just like any other, but it is one that can be shown to have been triggered by **stress transfer** from a preceding earthquake. Within a few tens of seconds of the main Loma Prieta earthquake, there were hundreds of smaller aftershocks; their distribution defines the area of the rupture surface.

Aftershocks can be of any magnitude. Most are smaller than the earthquake that triggered them, but they can be bigger. The aftershocks shown in Figure **6.4.1.4** all happened within seconds or minutes of the main shock, but aftershocks can be delayed for hours, days, weeks, or even years. As already noted, aftershocks are related to stress transfer. For example, the main shock of the Loma Prieta earthquake triggered aftershocks in the immediate area, which triggered more in the surrounding area, eventually extending for 30 kilometers along the fault in each direction and for 15 kilometers toward the surface. But the earthquake as a

whole also changed the stress on adjacent parts of the San Andreas Fault. This effect, which has been modelled for numerous earthquakes and active faults around the world, is depicted in Figure **6.4.1.6**. Stress was reduced in the area of the rupture (blue), but was increased at either end of the rupture surface (red and yellow).

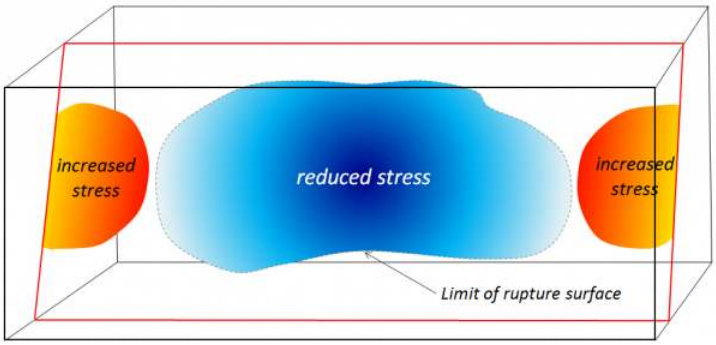



Figure **6.4.1.6** Depiction of stress changes related to an earthquake. Stress decreases in the area of the rupture surface, but increases on adjacent parts of the fault. © Steven Earle. CC BY. Based on data from 2010 Laguna Salada earthquake by Stein and Toda.

Stress transfer isn't necessarily restricted to the fault along which an earthquake happened. It will affect the rocks in general around the site of the earthquake and may lead to increased stress on other faults in the region. The effects of stress transfer don't necessarily show up right away. Segments of faults are typically in some state of stress, and the transfer of stress from another area is only rarely enough to push a fault segment beyond its limits to the point of rupture. The stress that is added by stress transfer accumulates along with the ongoing buildup of stress from plate motion and eventually leads to another earthquake.

## Episodic Tremor and Slip

Episodic tremor and slip (ETS) is periodic slow sliding along part of a subduction boundary. It does not produce recognizable earthquakes, but does produce seismic tremor (rapid seismic vibrations on a seismometer). It was first discovered on the Vancouver Island part of the Cascadia subduction zone by Geological Survey of Canada geologists Herb Dragert and Garry Rogers.[1]

The boundary between the subducting Juan de Fuca Plate and the North America Plate can be divided into three segments (Figure **6.4.1.7**). The cold upper part of the Juan de Fuca Plate boundary is locked. The plates are stuck and don't move, except with very large earthquakes that happen *approximately* every 500 years (the last one was approximately M9 on January 26, 1700). The warm lower part of the boundary is sliding continuously because the warm rock is weaker. The central part of the boundary isn't cold enough to be stuck, but isn't warm enough to slide continuously. Instead it slips episodically, approximately every 14 months for about 2 weeks, moving a few centimeters each time.

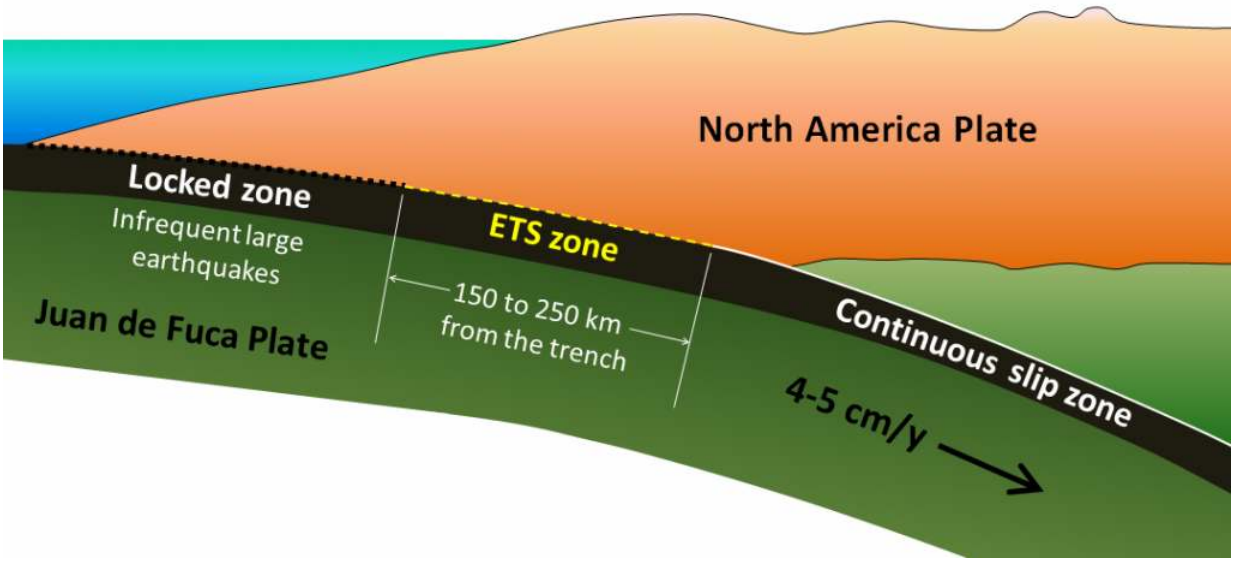



Figure **6.4.1.7** The boundary between the subducting Juan de Fuca Plate and the North America Plate is locked in the upper part, slides continuously in the lower part and slides episodically in the middle part. © Steven Earle. CC BY

You might be inclined to think that it's a good thing that there is periodic slip on this part of the plate because it releases some of the tension and reduces the risk of a large earthquake. In fact, the opposite is likely the case. The movement along the ETS part of the plate boundary acts like a medium-sized earthquake and leads to stress transfer to the adjacent locked part of the plate. Approximately every 14 months, during the two-week ETS period, there is a transfer of stress to the shallow locked part of the Cascadia subduction zone, and therefore an increased chance of a large earthquake.

Since 2003, ETS processes have also been observed on subduction zones in Mexico, New Zealand and Japan.

## References


1. Rogers, G. and Dragert, H., 2003, Episodic tremor and slip on the Cascadia subduction zone: the chatter of silent slip, Science, V. 300, p. 1942-1943. ↵


## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- What is an Earthquake from Physical Geology by Steven Earle (licensed under a Creative Commons Attribution 4.0 International License)
- The Physical World from *Environmental Science: A Canadian Perspective* by Bill Freedman (Creative Commons Attribution NonCommercial)

---


This page titled 6.4.1: Earthquakes is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Steven Earle (BCCampus) .

# 6.4.2: Volcanoes

A volcano is any location where magma comes to the surface, or has done so within the past several million years (Figure **6.4.2.1**). This can include eruptions on the ocean floor (or even under the water of lake), where they are called **subaqueous eruptions**, or on land, where they are called **subaerial eruptions**. Volcanoes commonly occur near plate boundaries where the motion of the plates has created cracks in the lithosphere through which the magma can flow. About eighty percent of volcanoes occur at convergent plate boundaries where subducted material melts and rises through cracks in the crust. Not all volcanic eruptions produce the volcanic mountains with which we are familiar; in fact most of Earth's volcanism takes place along the spreading ridges on the sea floor and does not produce volcanic mountains at all — not even sea-floor mountains.



***Figure* 6.4.2.1*:*** *Mt. Garibaldi (background left, looking from the north) with Garibaldi Lake in the foreground. Mt. Garibaldi, near Squamish B.C., is one of Canada's tallest (2,678 m) and most recently active volcanoes. [SE photo] The volcanic peak in the centre is Mt. Price and the dark flat–topped peak is The Table. All three of these volcanoes were active during the last glaciation. Mt. Garibaldi last erupted approximately 10,000 years ago.* © Steven Earle. CC BY

The study of volcanoes is critical to our understanding of the geological evolution of Earth, and to our understanding of significant changes in climate. But, most important of all, understanding volcanic eruptions allows us to save lives and property. Over the past few decades, volcanologists have made great strides in their ability to forecast volcanic eruptions and predict the consequences — this has already saved thousands of lives.

The largest eruptions can literally explode a volcanic mountain, ejecting immense quantities of material into the environment and causing enormous damage and loss of life. For example, an eruption of Mount Vesuvius in the year 79 CE (Common Era) buried the Roman city of Pompeii, killing almost all of its inhabitants. A 1902 explosion of Mont Pelée on the Caribbean island of Martinique killed 30,000 people. Another famous Indonesian eruption was that of Krakatau in the Sunda Strait in 1883, which ejected 18-21 $km^3$ of material as high as 50-80 km into the atmosphere. The 30-m tsunami associated with this eruption killed about 36,000 people in coastal villages.

The greatest eruption of modern times was that of Tambora, a volcano in Indonesia that exploded in 1815 and blew more than 300 $km^3$ of material into the atmosphere (including the top 1,300 m of the mountain). Some of the finer particulates of this massive eruption were blown into the upper atmosphere (the stratosphere), causing an increase in Earth's reflectivity that resulted in global cooling. The year 1816 became known as the "year without a summer" in Europe and North America because of its unusually cool and wet weather, including frost and snowfall during the summer months.

Large volcanic eruptions can also disturb great expanses of forest and other ecosystems. For instance, the 1980 explosion of Mount St. Helen's in the state of Washington blew down about 21,000 hectares of coniferous forest and otherwise damaged another 40,000 ha. Mudslides also devastated large areas, and a vast region was covered by particulate debris (known as tephra) that settled from the atmosphere.

Some volcanoes produce chronic lava flows and venting of gases. These volcanoes tend to form distinctive, cone-shaped mountains from their accumulated lava, which solidifies into finely crystalline, glassy rocks. An active example of this spectacular process is Mount Kilauea in Hawaii, which sometimes erupts continuously for years. The slowly flowing lava from these volcanoes can destroy buildings and vegetation but is not otherwise dangerous because people and animals can avoid the molten streams.

## Types of Volcanoes

Volcanoes can be classified according to the type and form of their ejecta. The basic types are: composite volcanoes, shield volcanoes, cinder cones, and lava domes. **Composite volcanoes** are steep-sided, symmetrical cones built of multiple layers of viscous lava and ash. Most composite volcanoes have a crater at the summit which contains the central vent. Lavas flow from

breaks in the crater wall or from cracks on the flanks of the cone. Mt Fuji in Japan and Mt Ranier in Washington are examples of composite volcanoes.

**Shield volcanoes** are built almost entirely of highly fluid (low viscosity) lava flows. They form slowly from numerous flows that spread out over a wide area from a central vent. The resultant structure is a broad, gently sloping cone with a profile like a warrior's shield. Mt Kilauea in Hawaii is an example of a shield volcano.

**Cinder cones** are the simplest type of volcano. They form when lava blown violently into the area breaks into small fragments that solidify and fall as cinders. A steep-sided cone shape is formed around the vent, with a crater at the summit. Sunset Crater in Arizona is a cinder cone that formed less than a thousand years ago, disrupting the lives of the native inhabitants of the region.

**Lava domes** are formed when highly viscous lava is extruded from a vent and forms a rounded, steep-sided dome. The lava piles up around and on the vent instead of flowing away, mostly growing by expansion from within. Lava domes commonly occur within the craters or on the flanks of composite volcanoes.

## Volcanic Hazards

There are two classes of volcanic hazards, direct and indirect. Direct hazards are forces that directly kill or injure people, or destroy property or wildlife habitat. Indirect hazards are volcanism-induced environmental changes that lead to distress, famine, or habitat degradation. It is estimated that indirect effects of volcanism have accounted for approximately 8 million deaths during historical times, while direct effects have accounted for fewer than 200,000, or 2.5% of the total. Some of the more important types of volcanic hazards are summarized in Table 6.4.2.1.

Table 6.4.2.1 A summary of the important volcanic hazards

[Skip Table]

| Type | Description | Risk |
|---|---|---|
| Tephra emissions | Small particles of volcanic rock emitted into the atmosphere | • Respiration problems for some individuals<br>• Significant climate cooling leading to crop failure and famine<br>• Damage to aircraft |
| Gas emissions | The emission of gases before, during, and after an eruption | • Climate cooling leading to crop failure and famine<br>• In some cases, widespread poisoning |
| Pyroclastic density current | A very hot (several 100°C) mixture of gases and volcanic fragments (tephra) that flows rapidly (up to 100s of kilometers per hour (km/h)) down the side of a volcano | Extreme hazard — destroys anything in the way |
| Pyroclastic fall | Vertical fall of tephra in the area surrounding an eruption | • Thick tephra coverage of areas close to the eruption (1 km to 10s of kilometers)<br>• Collapsed roofs |
| Lahar | A flow of mud and debris down a channel leading away from a volcano, triggered either by an eruption or a severe rain event | Severe risk of destruction for anything within the channel—lahar mud flows can move at 10s of km/h |
| Sector collapse/ debris avalanche | The failure of part of a volcano, either due to an eruption or for some other reason, leading to the failure of a large portion of the volcano | Severe risk of destruction for anything in the path of the debris avalanche |
| Lava flow | The flow of lava away from a volcanic vent | People and infrastructure at risk, but lava flows tend to be slow (less than km/h) and are relatively easy to avoid |

## Volcanic Gas and Tephra Emissions

Large volumes of **tephra** (rock fragments, mostly **pumice**) and gases are emitted during major **plinian eruptions** (large explosive eruptions with hot gas and tephra columns extending into the stratosphere) at composite volcanoes, and a large volume of gas is released during some very high-volume effusive eruptions. One of the major effects is cooling of the climate by 1° to 2°C for several months to a few years because the dust particles and tiny droplets and particles of sulfur compounds block the sun. The last significant event of this type was in 1991 and 1992 following the large eruption of Mount Pinatubo in the Philippines. A temperature decrease of 1° to 2°C may not seem like very much, but that is the global average amount of cooling, and cooling was much more severe in some regions and at some times.

Over an eight-month period in 1783 and 1784, a massive effusive eruption took place at the Laki volcano in Iceland. Although there was relatively little volcanic ash involved, a massive amount of sulfur dioxide was released into the atmosphere, along with a significant volume of hydrofluoric acid (HF). The sulphate **aerosols** that formed in the atmosphere led to dramatic cooling in the northern hemisphere. There were serious crop failures in Europe and North America, and a total of 6 million people are estimated to have died from famine and respiratory complications. In Iceland, poisoning from the HF resulted in the death of 80% of sheep, 50% of cattle, and the ensuing famine, along with HF poisoning, resulted in more than 10,000 human deaths—about 25% of the population.

Volcanic ash can also have serious implications for aircraft because it can destroy jet engines. For example, over 5 million airline passengers had their travel disrupted by the 2010 Eyjafjallajökull volcanic eruption in Iceland.

## Pyroclastic Density Currents

In a typical explosive eruption at a composite volcano, the tephra and gases are ejected with explosive force and are hot enough to be forced high up into the atmosphere. As the eruption proceeds, and the amount of gas in the rising magma starts to decrease, parts will become heavier than air, and they can then flow downward along the flanks of the volcano (Figure **6.4.2.2**). As they descend, they cool more and flow faster, reaching speeds up to several hundred kilometers per hour. A **pyroclastic density current** (PDC) consists of tephra ranging in size from boulders to microscopic shards of glass (made up of the edges and junctions of the bubbles of shattered pumice), plus gases (dominated by water vapor, but also including other gases). The temperature of this material can be as high as 1000°C. Among the most famous PDCs are the one that destroyed Pompeii in the year 79 CE, killing an estimated 18,000 people, and the one that destroyed the town of St. Pierre, Martinique, in 1902, killing an estimated 30,000.

The buoyant upper parts of pyroclastic density currents can flow over water, in some cases for several kilometers. The 1902 St. Pierre PDC flowed out into the city's harbour and destroyed several wooden ships anchored there.



Figure **6.4.2.2** The plinian eruption of Mount Mayon, Philippines. in 1984. Although most of the eruption column is ascending into the atmosphere, there are pyroclastic density currents flowing down the sides of the volcano in several places.
Source: "Pyroclastic flows at Mayon Volcano" by USGS. Public domain.

## Pyroclastic Fall

Most of the tephra from an explosive eruption ascends high into the atmosphere, and some of it is distributed around Earth by high-altitude winds. The larger components (larger than 0.1 mm) tend to fall relatively close to the volcano, and the amount produced by large eruptions can cause serious damage and casualties. The large 1991 eruption of Mount Pinatubo in the Philippines resulted in the accumulation of tens of centimeters of ash in fields and on rooftops in the surrounding populated region. Heavy typhoon rains that hit the island at the same time added to the weight of the tephra, leading to the collapse of thousands of roofs and to at least 300 of the 700 deaths attributed to the eruption.

## Lahar

A **lahar** is any mudflow or **debris flow** that is related to a volcano. Most are caused by melting snow and ice during an eruption, as was the case with the lahar that destroyed the Colombian town of Armero in 1985 (described earlier). Lahars can also happen when there is no volcanic eruption, and one of the reasons is that, as we've seen, composite volcanoes tend to be weak and easily eroded.

In October 1998, category 5 hurricane Mitch slammed into the coast of central America. Damage was extensive and 19,000 people died, not so much because of high winds but because of intense rainfall—some regions received almost 2 m of rain over a few days! Mudflows and debris flows occurred in many areas, especially in Honduras and Nicaragua. An example is at the Casita Volcano in Nicaragua, where the heavy rains weakened rock and volcanic debris on the upper slopes, resulting in a debris flow that rapidly built in volume as it raced down the steep slope, and then ripped through the towns of El Porvenir and Rolando Rodriguez

killing more than 2,000 people (Figure **6.4.2.3**). El Porvenir and Rolando Rodriguez were new towns that had been built without planning approval in an area that was known to be at risk of lahars.



Figure **6.4.2.3** Part of the path of the lahar from Casita Volcano, October 30, 1998.
Source: "Casita Volcano" by USGS. Public domain.

### Sector Collapse and Debris Avalanche

In the context of volcanoes, **sector collapse** or flank collapse is the catastrophic failure of a significant part of an existing volcano, creating a large debris avalanche. This hazard was first recognized with the failure of the north side of Mount St. Helens immediately prior to the large eruption on May 18, 1980. In the weeks before the eruption a large bulge had formed on the side of the volcano, the result of magma transfer from depth into a satellite magma body within the mountain itself. Early on the morning of May 18, a moderate earthquake struck nearby; this is thought to have destabilized the bulge, leading to Earth's largest ever observed slope failure. The failure of this part of the volcano exposed the underlying satellite magma chamber, causing it to explode sideways, which then exposed the conduit leading to the magma chamber below. The resulting plinian eruption—with a 24 kilometre high eruption column—lasted for nine hours.

In August 2010, a massive part of the flank of B.C.'s Mount Meager gave way and about 48 million cubic meters ($m^3$) of rock rushed down the valley, one of the largest slope failures in Canada in historical times. More than 25 slope failures have taken place at Mount Meager in the past 8,000 years, some of them more than 10 times larger than the 2010 failure.

### Lava Flows

Lava flows at volcanoes like Kilauea do not advance very quickly, and in most cases, people can get out of the way. Of course, it is more difficult to move infrastructure, and so buildings and roads are typically the main casualties of lava flows.

## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- Volcanism and Volcanic Hazards from Physical Geology by Steven Earle (licensed under a Creative Commons Attribution 4.0 International License)
- The Solid Earth from *AP Environmental Science* by University of California College Prep
- The Physical World from *Environmental Science: A Canadian Perspective* by Bill Freedman (Creative Commons Attribution NonCommercial)

This page titled 6.4.2: Volcanoes is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Steven Earle (BCCampus) .

# 6.4.3: Mass Wasting

Early in the morning on January 9, 1965, 47 million cubic metres of rock broke away from the steep upper slopes of Johnson Peak (16 km southeast of Hope) and roared 2,000 m down the mountain, gouging out the contents of a small lake at the bottom, and continuing a few hundred metres up the other side (Figure 15.1). Four people, who had been stopped on the highway by a snow avalanche, were killed. Many more might have become victims, except that a Greyhound bus driver, en route to Vancouver, turned his bus around on seeing the avalanche. The rock failed along weakened foliation planes of the metamorphic rock on Johnson Peak, in an area that had been eroded into a steep slope by glacial ice. There is no evidence that it was triggered by any specific event, and there was no warning that it was about to happen. Even if there had been warning, nothing could have been done to prevent it. There are hundreds of similar situations throughout British Columbia.



Figure 6.4.3.1: *The site of the 1965 Hope Slide as seen in 2014. The initial failure is thought to have taken place along the foliation planes and sill within the area shown in the inset.* © Steven Earle. CC BY

What can we learn from the Hope Slide? In general, we cannot prevent most mass wasting, and significant effort is required if an event is to be predicted with any level of certainty. Understanding the geology is critical to understanding mass wasting. Although failures are inevitable in a region with steep slopes, larger ones happen less frequently than smaller ones, and the consequences vary depending on the downslope conditions, such as the presence of people, buildings, roads, or fish-bearing streams.

An important reason for learning about mass wasting is to understand the nature of the materials that fail, and how and why they fail so that we can minimize risks from similar events in the future. For this reason, we need to be able to classify mass-wasting events, and we need to know the terms that geologists, engineers, and others use to communicate about them.

**Mass wasting**, which is synonymous with “slope failure,” is the failure and downslope movement of rock or unconsolidated materials in response to gravity. The term “landslide” is almost synonymous with mass wasting, but not quite because some people reserve “landslide” for relatively rapid slope failures, while others do not.

## Classification of Mass Wasting Events

It’s important to classify slope failures so that we can understand what causes them and learn how to mitigate their effects. The three criteria used to describe slope failures are:

- The type of material that failed (typically either bedrock or unconsolidated sediment)
- The mechanism of the failure (how the material moved)
- The rate at which it moved

The type of motion is the most important characteristic of a slope failure, and there are three different types of motion:

- If the material drops through the air, vertically or nearly vertically, it’s known as a **fall**.

- If the material moves as a mass along a sloping surface (without internal motion within the mass), it's a **slide.**
- If the material has internal motion, like a fluid, it's a **flow**.

Unfortunately it's not normally that simple. Many slope failures involve two of these types of motion, some involve all three, and in many cases, it's not easy to tell how the material moved. The types of slope failure that we'll cover here are summarized in Table **6.4.3.1**.

Table **6.4.3.1**: Classification of slope failures based on type of material and type of motion

[Skip Table]

| Failure Type | Type of Material | Type of Motion | Rate of Motion |
|---|---|---|---|
| Rock fall | Rock fragments | Vertical or near-vertical fall (plus bouncing in many cases) | Very fast (Greater than 10s of meters per second) |
| Rock slide | A large rock body | Motion as a unit along a planar surface (translational sliding) | Typically very slow (millimeters per year to centimeters per year), but some can be faster |
| Rock avalanche | A large rock body that slides and then breaks into small fragments | Flow (at high speeds the mass of rock fragments is suspended on a cushion of air) | Very fast (Greater than tens of meters per second) |
| Creep or solifluction | Soil or other overburden; in some cases, mixed with ice | Flow (although sliding motion may also occur) | Very slow (millimeters per year to centimeters per year) |
| Slump | Thick deposits (a metre to 10s of meters) of unconsolidated sediment | Motion as a unit along a curved surface (rotational sliding) | Slow (centimeters per year to meters per year) |
| Mudflow | Loose sediment with a significant component of silt and clay | Flow (a mixture of sediment and water moves down a channel) | Moderate to fast (centimeters per second to meters per second) |
| Debris flow | Sand, gravel, and larger fragments | Flow (similar to a mudflow, but typically faster) | Fast (meters per second) |

## Rock Fall

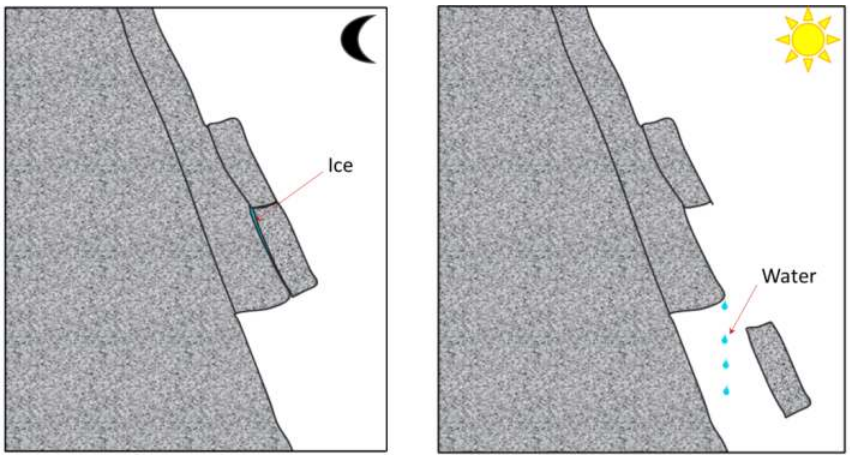



Figure **6.4.3.2** The contribution of freeze-thaw to rock fall. © Steven Earle. CC BY

Rock fragments can break off relatively easily from steep bedrock slopes, most commonly due to frost-wedging in areas where there are many freeze-thaw cycles per year. If you've ever hiked along a steep mountain trail on a cool morning, you might have heard the occasional fall of rock fragments onto a **talus slope**. This happens because the water between cracks freezes and expands

overnight, and then when that same water thaws in the morning sun, the fragments that had been pushed beyond their limit by the ice fall to the slope below (Figure **6.4.3.2**).

A typical talus slope, near Keremeos in southern B.C., is shown in Figure **6.4.3.3**. In December 2014, a large block of rock split away from a cliff in this same area. It broke into smaller pieces that tumbled down the slope and crashed into the road, smashing the concrete barriers and gouging out large parts of the pavement. Luckily no one was hurt.



Figure **6.4.3.3** Left: A talus slope near Keremeos, B.C., formed by rock fall from the cliffs above. Right: The results of a rock fall onto a highway west of Keremeos in December 2014. © Steven Earle. CC BY

## Rock Slide

A rock slide is the sliding motion of rock along a sloping surface. In most cases, the movement is parallel to a fracture, bedding, or metamorphic foliation plane, and it can range from very slow to moderately fast. The word **sackung** describes the very slow motion of a block of rock (millimeters per year to centimeters per year) on a slope. A good example is the Downie Slide north of Revelstoke, B.C., which is shown in Figure **6.4.3.4**. In this case, a massive body of rock is very slowly sliding down a steep slope along a plane of weakness that is approximately parallel to the slope. The Downie Slide, which was first recognized in the 1950s, prior to the construction of the Revelstoke Dam in the late 1970s, was moving very slowly at the time (a few centimeters per year). Geological engineers were concerned that the presence of water in the reservoir (visible in Figure **6.4.3.4**) could further weaken the plane of failure, leading to an acceleration of the motion. The result would have been a catastrophic failure into the reservoir that would have sent a wall of water over the dam and into the community of Revelstoke. During the construction of the dam they tunneled into the rock at the base of the slide and drilled hundreds of drainage holes upward into the plane of failure. This allowed water to drain out so that the pressure was reduced, which reduced the rate of movement of the sliding block. BC Hydro monitors this site continuously; the slide block is currently moving more slowly than it was prior to the construction of the dam.




Figure **6.4.3.4** The Downie Slide, a sackung, on the shore of the Revelstoke Reservoir (above the Revelstoke Dam). The head scarp is visible at the top and a side-scarp along the left side. © Google Earth. Used with permission for non-commerical purposes.



Figure **6.4.3.5** Site of the 2008 rock slide at Porteau Cove. Notice the prominent fracture set parallel to the surface of the slope. The slope has been stabilized with rock bolts (visible near to the top of the photo) and holes have been drilled into the rock to improve drainage (one is visible in the lower right). Risk to passing vehicles from rock fall has been reduced by hanging mesh curtains (background). © Steven Earle. CC BY Photo 2012

In the summer of 2008, a large block of rock slid rapidly from a steep slope above Highway 99 near Porteau Cove (between Horseshoe Bay and Squamish). The block slammed into the highway and adjacent railway and broke into many pieces. The highway was closed for several days, and the slope was subsequently stabilized with rock bolts and drainage holes. As shown in Figure **6.4.3.5**, the rock is fractured parallel to the slope, and this almost certainly contributed to the failure. However, it is not actually known what triggered this event as the weather was dry and warm during the preceding weeks, and there was no significant earthquake in the region.



Figure **6.4.3.6** The August 2010 Mount Meager rock avalanche, showing where the slide originated (red arrow, 4 km upstream) and its path down a steep narrow valley. The yellow arrows show how far up the valley the avalanche extended. “2010 Mt. Meager rock avalanche” © Isaac Earle. CC BY.

If a rock slides and then starts moving quickly (meters per second), the rock is likely to break into many small pieces, and at that point it turns into a **rock avalanche**, in which the large and small fragments of rock move in a fluid manner supported by a cushion of air within and beneath the moving mass. The 1965 Hope Slide (Figure **6.4.3.1**) was a rock avalanche, as was the famous 1903 Frank Slide in southwestern Alberta. The 2010 slide at Mount Meager (west of Lillooet) was also a rock avalanche, and rivals the Hope Slide as the largest slope failure in Canada during historical times (Figure **6.4.3.6**).

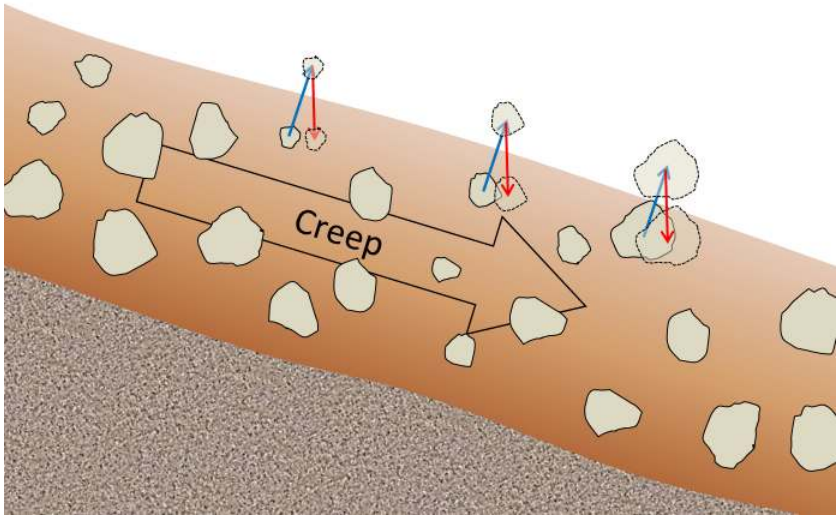



Figure 6.4.3.7 A depiction of the contribution of freeze-thaw to creep. The blue arrows represent uplift caused by freezing in the wet soil underneath, while the red arrows represent depression by gravity during thawing. The uplift is perpendicular to the slope, while the drop is vertical. © Steven Earle. CC BY.

The very slow—millimeters per year to centimeters per year—movement of soil or other unconsolidated material on a slope is known as creep. **Creep**, which normally only affects the upper several centimeters of loose material, is typically a type of very slow flow, but in some cases, sliding may take place. Creep can be facilitated by freezing and thawing because, as shown in Figure 6.4.3.7, particles are lifted perpendicular to the surface by the growth of ice crystals within the soil, and then let down vertically by gravity when the ice melts. The same effect can be produced by frequent wetting and drying of the soil. In cold environments, **solifluction** is a more intense form of freeze-thaw-triggered creep.

Creep is most noticeable on moderate-to-steep slopes where trees, fence posts, or grave markers are consistently leaning in a downhill direction. In the case of trees, they try to correct their lean by growing upright, and this leads to a curved lower trunk known as a “pistol butt.” An example is shown on Figure 6.4.3.8.



Figure 6.4.3.8 Pistol-butt shaped trees on a slope that is experiencing creep © Steven Earle. CC BY.

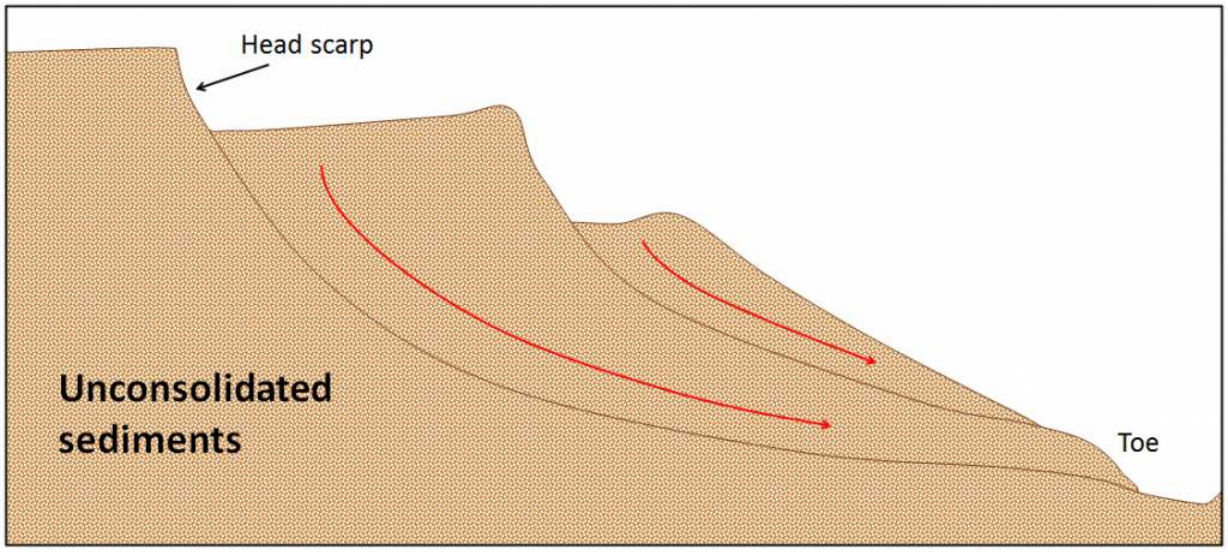



Figure **6.4.3.9** A depiction of the motion of unconsolidated sediments in an area of slumping. © Steven Earle. CC BY.

Slump is a type of slide (movement as a mass) that takes place within thick unconsolidated deposits (typically thicker than 10 meters). Slumps involve movement along one or more curved failure surfaces, with downward motion near the top and outward motion toward the bottom (Figure **6.4.3.9**). They are typically caused by an excess of water within these materials on a steep slope.



Figure **6.4.3.10** A slump along the banks of a small coulee near Lethbridge, Alberta. The main head-scarp is clearly visible at the top, and a second smaller one is visible about one-quarter of the way down. The toe of the slump is being eroded by the seasonal stream that created the coulee. © Steven Earle. CC BY. Photo 2005

An example of a slump in the Lethbridge area of Alberta is shown in Figure **6.4.3.10**. This feature has likely been active for many decades, and moves a little more whenever there are heavy spring rains and significant snowmelt runoff. The toe of the slump is failing because it has been eroded by the small stream at the bottom.

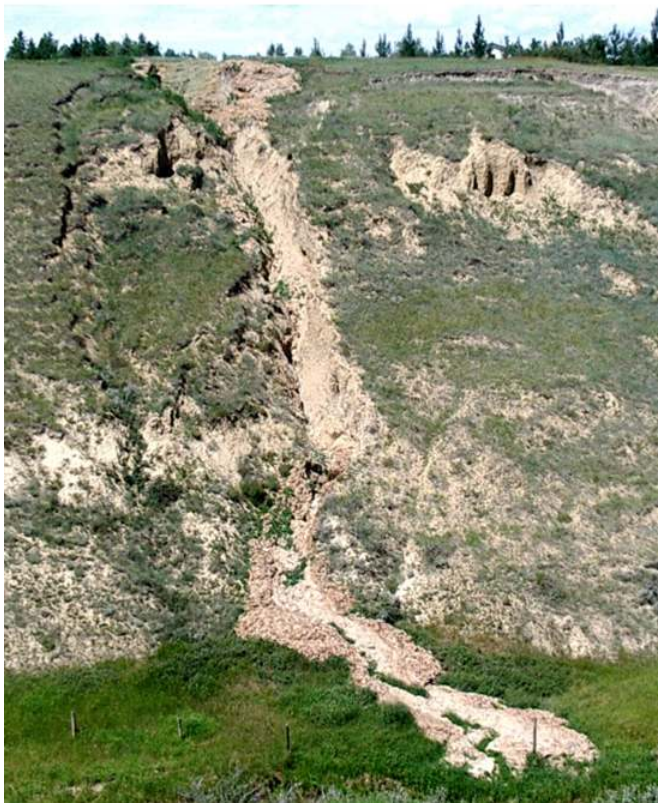


Figure **6.4.3.11** A slump (left) and an associated mudflow (centre) at the same location as Figure **6.4.3.10**, near Lethbridge, Alberta.
© Steven Earle. CC BY.

When a mass of sediment becomes completely saturated with water, the mass loses strength, to the extent that the grains are pushed apart, and it will flow, even on a gentle slope. This can happen during rapid spring snowmelt or heavy rains, and is also relatively common during volcanic eruptions because of the rapid melting of snow and ice. (A mudflow or debris flow on a volcano or during a volcanic eruption is a *lahar*.) If the material involved is primarily sand-sized or smaller, it is known as a mudflow, such as the one shown in Figure **6.4.3.11**.

If the material involved is gravel sized or larger, it is known as a debris flow. Because it takes more gravitational energy to move larger particles, a debris flow typically forms in an area with steeper slopes and more water than does a mudflow. In many cases, a debris flow takes place within a steep stream channel, and is triggered by the collapse of bank material into the stream. This creates a temporary dam, and then a major flow of water and debris when the dam breaks. This is the situation that led to the fatal debris flow at Johnsons Landing, B.C., in 2012. A typical west-coast debris flow is shown in Figure **6.4.3.12**. This event took place in November 2006 in response to very heavy rainfall. There was enough energy to move large boulders and to knock over large trees.



Figure **6.4.3.12** The lower part of debris flow within a steep stream channel near Buttle Lake, B.C., in November 2006. © Steven Earle. CC BY.

## Factors that Control Slope Stability

Mass wasting happens because tectonic processes have created uplift. Erosion, driven by gravity, is the inevitable response to that uplift, and various types of erosion, including mass wasting, have created slopes in the uplifted regions. Slope stability is ultimately determined by two factors: the angle of the slope and the strength of the materials on it.

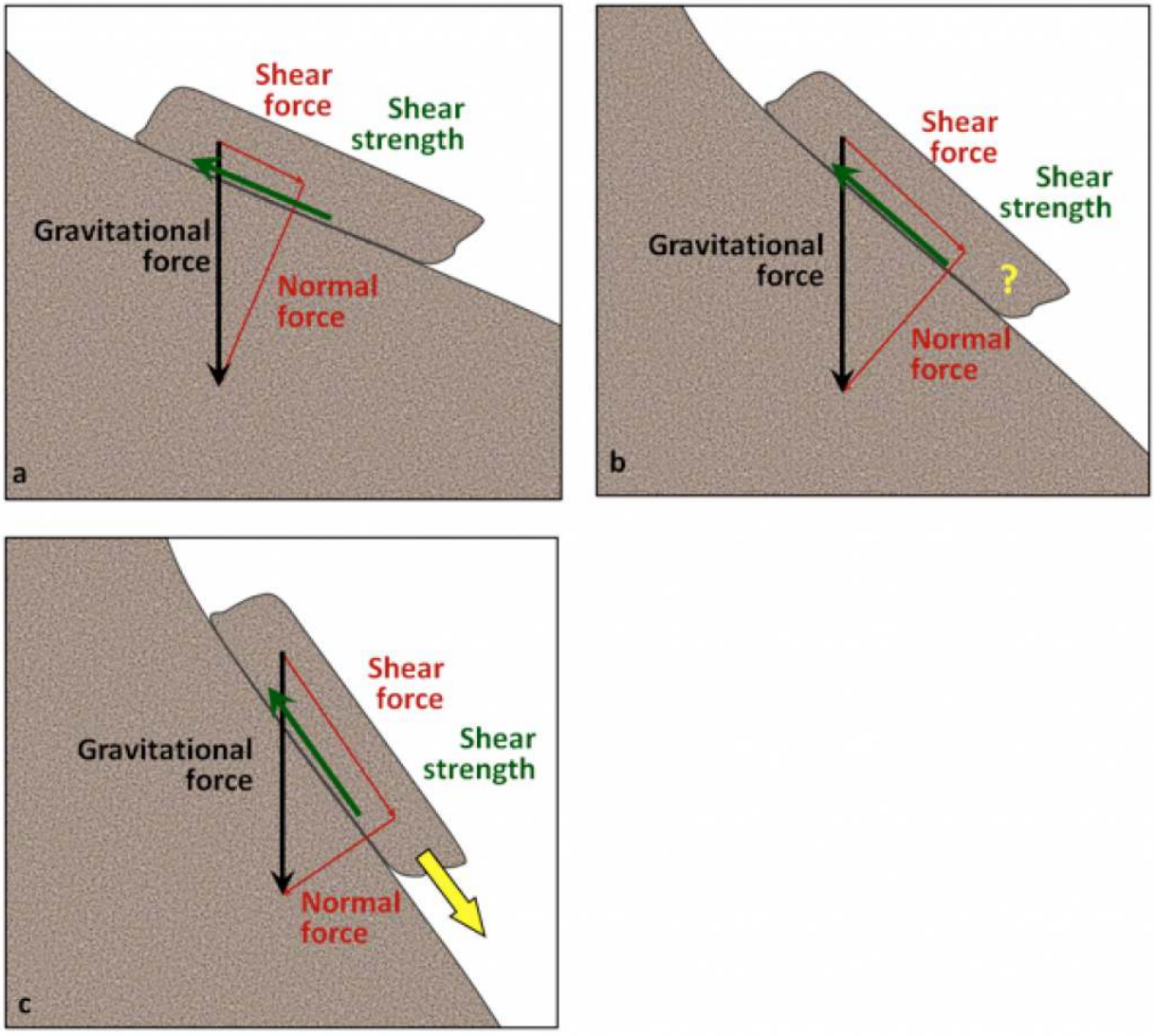



**Figure 6.4.3.13** Differences in the shear and normal components of the gravitational force on slopes with differing steepness. The gravitational force is the same in all three cases. In (a) the shear force is substantially less than the shear strength, so the block should be stable. In (b) the shear force and shear strength are about equal, so the block may or may not move. In (c) the shear force is substantially greater than the shear strength, so the block is very likely to move. © Steven Earle. CC BY

Figure 6.4.3.13 shows a block of rock situated on a rock slope. The block is being pulled toward Earth's centre (vertically down) by gravity. We can split the vertical gravitational force into two components relative to the slope: one pushing the block down the slope (the **shear force**), and the other pushing into the slope (the **normal force**). The shear force, which wants to push the block down the slope, has to overcome the strength of the connection between the block and the slope, which may be quite weak if the block has split away from the main body of rock, or may be very strong if the block is still a part of the rock. This is the **shear strength**, and in Figure 6.4.3.13a, it is greater than the shear force, so the block should not move. In Figure 6.4.3.13b the slope is steeper and the shear force is approximately equal to the shear strength. The block may or may not move under these circumstances. In Figure 6.4.3.13c, the slope is steeper still, so the shear force is considerably greater than the shear strength, and the block will very likely move.

As already noted, slopes are created by uplift followed by erosion. In areas with relatively recent uplift (such as most of British Columbia and the western part of Alberta), slopes tend to be quite steep. This is especially true where glaciation has taken place because glaciers in mountainous terrain create steep-sided valleys. In areas without recent uplift (such as central Canada), slopes are less steep because hundreds of millions of years of erosion (including mass wasting) has made them that way. However, as we'll see, some mass wasting can happen even on relatively gentle slopes.

The strength of the materials on slopes can vary widely. Solid rocks tend to be strong, but there is a very wide range of rock strength. If we consider just the strength of the rocks, and ignore issues like fracturing and layering, then most crystalline rocks—like granite, basalt, or gneiss—are very strong, while some metamorphic rocks—like schist—are moderately strong. Sedimentary rocks have variable strength. Dolostone and some limestone are strong, most sandstone and conglomerate are moderately strong, and some sandstone and all mudstones are quite weak.

Fractures, metamorphic foliation, or bedding can significantly reduce the strength of a body of rock, and in the context of mass wasting, this is most critical if the planes of weakness are parallel to the slope and least critical if they are perpendicular to the slope. This is illustrated in Figure **6.4.3.14**. At locations A and B the bedding is nearly perpendicular to the slope and the situation is relatively stable. At location D the bedding is nearly parallel to the slope and the situation is quite unstable. At location C the bedding is nearly horizontal and the stability is intermediate between the other two extremes.

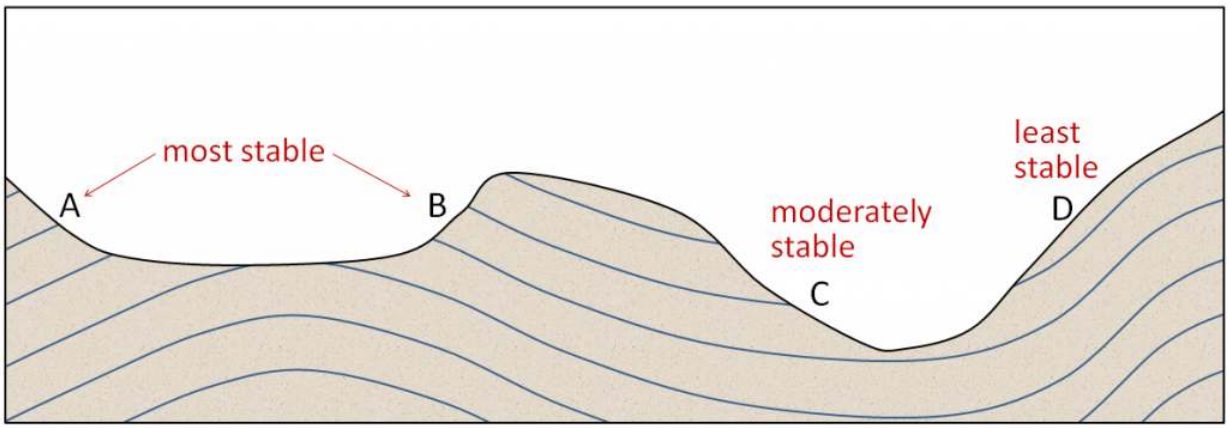



Figure **6.4.3.14** Relative stability of slopes as a function of the orientation of weaknesses (in this case bedding planes) relative to the slope orientations. © Steven Earle. CC BY

Internal variations in the composition and structure of rocks can significantly affect their strength. Schist, for example, may have layers that are rich in sheet silicates (mica or chlorite) and these will tend to be weaker than other layers. Some minerals tend to be more susceptible to weathering than others, and the weathered products are commonly quite weak (e.g., the clay formed from feldspar). The side of Johnson Peak that failed in 1965 (Hope Slide) is made up of chlorite schist (metamorphosed sea-floor basalt) that has feldspar-bearing sills within it (they are evident within the inset area of Figure **6.4.3.1**). The foliation and the sills are parallel to the steep slope. The schist is relatively weak to begin with, and the feldspar in the sills, which has been altered to clay, makes it even weaker.

Unconsolidated sediments are generally weaker than sedimentary rocks because they are not cemented and, in most cases, have not been significantly compressed by overlying materials. This binding property of sediment is sometimes referred to as *cohesion*. Sand and silt tend to be particularly weak, clay is generally a little stronger, and sand mixed with clay can be stronger still. The deposits that make up the cliffs at Point Grey in Vancouver include sand, silt, and clay overlain by sand. As shown in Figure **6.4.3.15** (left) the finer deposits are relatively strong (they maintain a steep slope), while the overlying sand is relatively weak, and has a shallower slope that has recently failed. Glacial till—typically a mixture of clay, silt, sand, gravel, and larger clasts—forms and is compressed beneath tens to thousands of meters of glacial ice so it can be as strong as some sedimentary rock (Figure **6.4.3.15**, right).



Figure **6.4.3.15** Left: Glacial outwash deposits at Point Grey, in Vancouver. The dark lower layer is made up of sand, silt, and clay. The light upper layer is well-sorted sand. Right: Glacial till on Quadra Island, B.C. The till is strong enough to have formed a near-vertical slope. © Steven Earle. CC BY

Apart from the type of material on a slope, the amount of water that the material contains is the most important factor controlling its strength. This is especially true for unconsolidated materials, like those shown in Figure **6.4.3.15**, but it also applies to bodies of rock. Granular sediments, like the sand at Point Grey, have lots of spaces between the grains. Those spaces may be completely dry (filled only with air); or moist (often meaning that some spaces are water filled, some grains have a film of water around them, and small amounts of water are present where grains are touching each other); or completely saturated (Figure **6.4.3.16**). Unconsolidated sediments tend to be strongest when they are moist because the small amounts of water at the grain boundaries hold the grains together with surface tension. Dry sediments are held together only by the friction between grains, and if they are well sorted or well rounded, or both, that cohesion is weak. Saturated sediments tend to be the weakest of all because the large amount of water actually pushes the grains apart, reducing the mount friction between grains. This is especially true if the water is under pressure.

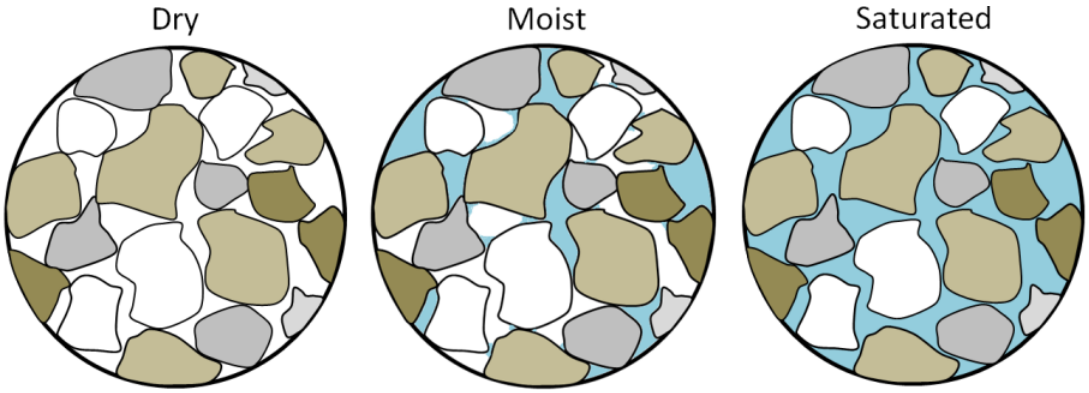



Figure **6.4.3.16** Depiction of dry, moist, and saturated sand. © Steven Earle. CC BY

Water will also reduce the strength of solid rock, especially if it has fractures, bedding planes, or clay-bearing zones. This effect is even more significant when the water is under pressure, which is why you'll often see holes drilled into rocks on road cuts to relieve this pressure. One of the hypotheses advanced to explain the 1965 Hope Slide is that the very cold conditions that winter caused small springs in the lower part of the slope to freeze over, preventing water from flowing out. It is possible that water pressure gradually built up within the slope, weakening the rock mass to the extent that the shear strength was no longer greater than the shear force.

Water also has a particular effect on clay-bearing materials. All clay minerals will absorb a little bit of water, and this reduces their strength. The **smectite** clays (such as the **bentonite** used in cat litter) can absorb a lot of water, and that water pushes the sheets apart at a molecular level and makes the mineral swell. Smectite that has expanded in this way has almost no strength; it is extremely slippery.

And finally, water can significantly increase the mass of the material on a slope, which increases the gravitational force pushing it down. A body of sediment that has 25% porosity and is saturated with water weighs approximately 13% more than it does when it is completely dry, so the gravitational shear force is also 13% higher. In the situation shown in Figure 15.2b, a 13% increase in the shear force could easily be enough to tip the balance between shear force and shear strength.

## Mass-Wasting Triggers

In the previous section, we talked about the shear force and the shear strength of materials on slopes, and about factors that can reduce the shear strength. Shear force is primarily related to slope angle, and this does not change quickly. But shear strength can change quickly for a variety of reasons, and events that lead to a rapid reduction in shear strength are considered to be **triggers** for mass wasting.

An increase in water content is the most common mass-wasting trigger. This can result from rapid melting of snow or ice, heavy rain, or some type of event that changes the pattern of water flow on the surface. Rapid melting can be caused by a dramatic increase in temperature (e.g., in spring or early summer) or by a volcanic eruption. Heavy rains are typically related to major storms. Changes in water flow patterns can be caused by earthquakes, previous slope failures that dam up streams, or human structures that interfere with runoff (e.g., buildings, roads, or parking lots). An example of this is the deadly January 2005 debris flow in North Vancouver in the Riverside Drive area. The debris flow happened during a rainy period, but was likely triggered by excess runoff related to the roads at the top of this slope and by landscape features, including a pool. The 2005 failure took place in an area that had failed previously, and a report written in 1980 recommended that the municipal authorities and residents take steps to address surface and slope drainage issues. Little was done to improve the situation.

In some cases, a *decrease* in water content can lead to failure. This is most common with clean sand deposits (e.g., the upper layer in Figure **6.4.3.15** (left)), which lose strength when there is no water to hold the grains together.

Freezing and thawing can also trigger some forms of mass wasting. More specifically, the thawing can release a block of rock that was attached to a slope by a film of ice.

One other process that can weaken a body of rock or sediment is shaking. The most obvious source of shaking is an earthquake, but shaking from highway traffic, construction, or mining will also do the job. Several deadly mass-wasting events (including snow avalanches) were triggered by the M7.8 earthquake in Nepal in April 2015.

Saturation with water and then seismic shaking led to the occurrence of thousands of slope failures in the Sapporo area of Hokkaido, Japan in September 2018, as shown on Figure **6.4.3.18**. The area was drenched with rain from tropical storm Jebi on September 4th. On September 6th it was shaken by a M6.6 earthquake which triggered debris flows in the water-saturated volcanic materials on steep slopes. There were 41 deaths related to the slope failures.




**Figure 6.4.3.18** Slope failures in the Sapporo area of Japan following a typhoon (Sept. 4th, 2018) and earthquake (Sept. 6th, 2018) (Before and after Landsat 8 images: top: July 2017, bottom: September 2018). Source: "Landslides in Hokkaido" by Lauren Dauphin, NASA Earth Observatory. Public domain.

## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- Mass Wasting, Factors that Control Slope Stability, and Classification of Mass Wasting from Physical Geology by Steven Earle (licensed under a Creative Commons Attribution 4.0 International License)

---

This page titled 6.4.3: Mass Wasting is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Steven Earle (BCCampus) .

# 6.4.4: Floods

## The Definition of a Flood

Here's a simple definition of a river flood: *the occurrence of a flow of such magnitude that it overtops the natural or artificial banks in a reach of river channel*. If a floodplain exists, here's another way of defining a flood: *any flow that spreads out over the floodplain*.

The *relative* volume of flood water is not large: on average, water discharged in excess of channel capacity constitutes about 5% of the total annual discharge of the given drainage basin. But the *absolute* volume of flood water is staggering. Hypothetical example: a two-inch rainstorm over a one-square-mile catchment the soils of which can absorb 0.5 inches would supply almost 3.5 million cubic feet of water is a couple of hours! (That's a cube 150 feet on a side.)

A flood is a *wave*: it has a wavelike shape (although very subdued, with low amplitude relative to wavelength), and it propagates downriver at some speed, typically less than the mean velocity of flow of water in the river.

## Floodplains

Here are two similar but not quite identical definitions of a floodplain, taken from the literature:

- the flat area adjacent to the river channel, constructed by the present river in the present climate and frequently subjected to overflow (Leopold, 1994)
- the flat area adjoining a river channel constructed by the river in the present climate and overflowed at times of high discharge (Dunne and Leopold, 1978)

The existence of a floodplain is testimony to the lateral migration of the river channel. With time, the shifting channel sweeps the entire width of the floodplain, eroding the "front" bank and depositing the "rear" bank as it goes (Figure **6.4.4.1 – 2**). If the climate, or the base level, changes in such a way that the river degrades (i.e., lowers its bed), the old floodplain is incised, forming a set of terraces, and a new, narrower floodplain develops. If the climate, or the base level, changes in such a way that the river aggrades, the floodplain typically (except in extreme cases) aggrades along with it.



**Figure 6.4.4.1:** Top: *The meandering channel of the Bonnell Creek, Nanoose, B.C. The stream is flowing toward the viewer. The sand and gravel point bar must have formed when the creek was higher and the flow faster than it was when the photo was taken.* © Steven Earle. CC BY.



**Figure 6.4.4.2:** *The meandering channel of the Nowitna River, Alaska. Numerous oxbow lakes are present and another meander cutoff will soon take place. [Oliver Kumis, http://commons.wikimedia.org/wiki/Fi...tna_river.jpg]*

What determines the height of the floodplain above the channel bed? That has to do with the relationship between *effectiveness* and *frequency*. Low discharges are common but not effective in sediment transport and channel shaping, whereas very high discharges are very effective at transporting sediment but very uncommon. The discharges with greatest effectiveness in maintaining channel and floodplain geometry are at some intermediate value: high but not extremely high, and common but not extremely common. For most rivers, bank-full discharges (those that spill over onto the floodplain) occur once or twice a year, on average. (It turns out that the discharge that is responsible for the maximum sediment discharge, in a given time period like a year, called the ***effective discharge,*** is very close to being equal to the ***bank-full discharge***.)

**What are floodplains like?**

- They range in width from several meters, on the smallest streams, to many tens of kilometers, on the largest rivers.
- They are not flat as a board: they have minor relief, in the form of sloughs and depressions left behind by the irregular shifting of the river channel.
- They are developed in their classic sense in connection with meandering rivers. Braided rivers work irregularly across the entire valley, such that in a sense the entire valley is “river bed” (Figure **6.4.4.3**).
- They are typically heavily vegetated in their natural state; in fact, the bank-stabilization effect of vegetation is one of the most important, if not *the* most important, factors in the existence of the floodplain in the first place.
- They vary in their “wetness”: usually they are dry and well above the local water table for most of the year, and are thus conducive to human use, but in areas with consistently high rainfall and/or strongly aggradational river regime ,there can be extensive swampy areas on the floodplain.



**Figure 6.4.4.3:** *Figure 13.19 The braided channel of the Kicking Horse River at Field, B.C.* © Steven Earle. CC BY.

## Stream Erosion and Deposition

Flowing water is a very important mechanism for both erosion and deposition. Water flow in a stream is primarily related to the stream's gradient (slope or "steepness"), but it is also controlled by the geometry of the stream channel. As shown in Figure 6.4.4.4, 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. In fact, the velocity just below the surface is typically a little 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 (Figure 6.4.4.1 *and* \(6.4.4.4).

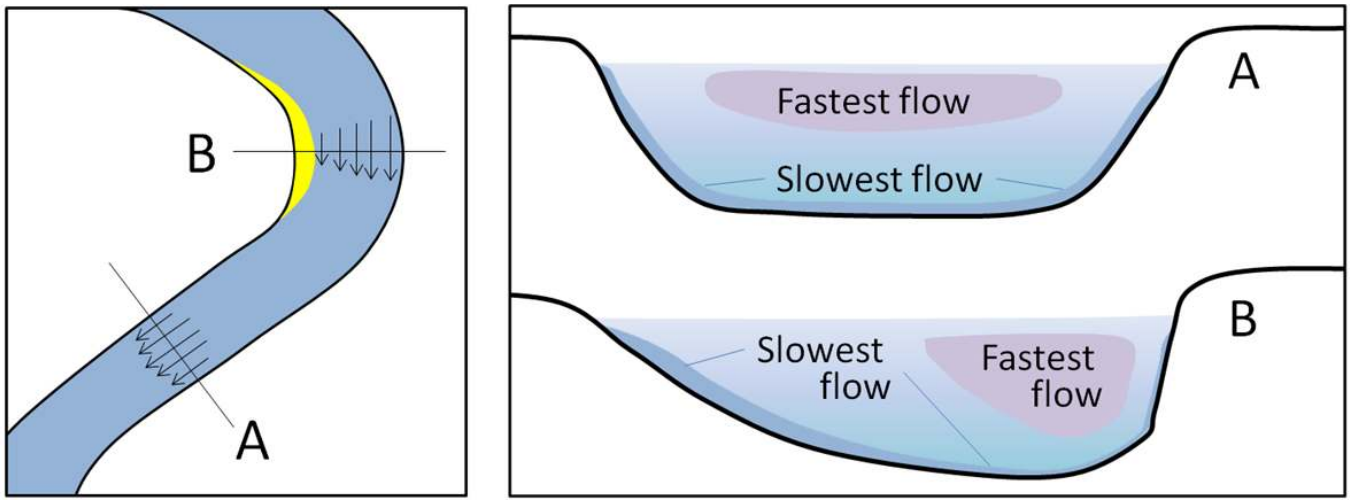



Figure 6.4.4.4: *The relative velocity of stream flow depending on whether the stream channel is straight or curved (left), and with respect to the water depth (right).* © Steven Earle. CC BY.

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 (e.g., $m^3$/second). During a flood, the water level always rises, so there is 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.

Figure 6.4.4.5 shows the nature of sediment transportation in a stream. 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. The faster the water is flowing, the larger the particles that can be kept in suspension and transported within the flowing water. It is important to be aware that a stream can both erode and deposit sediments of different sizes at the same time.

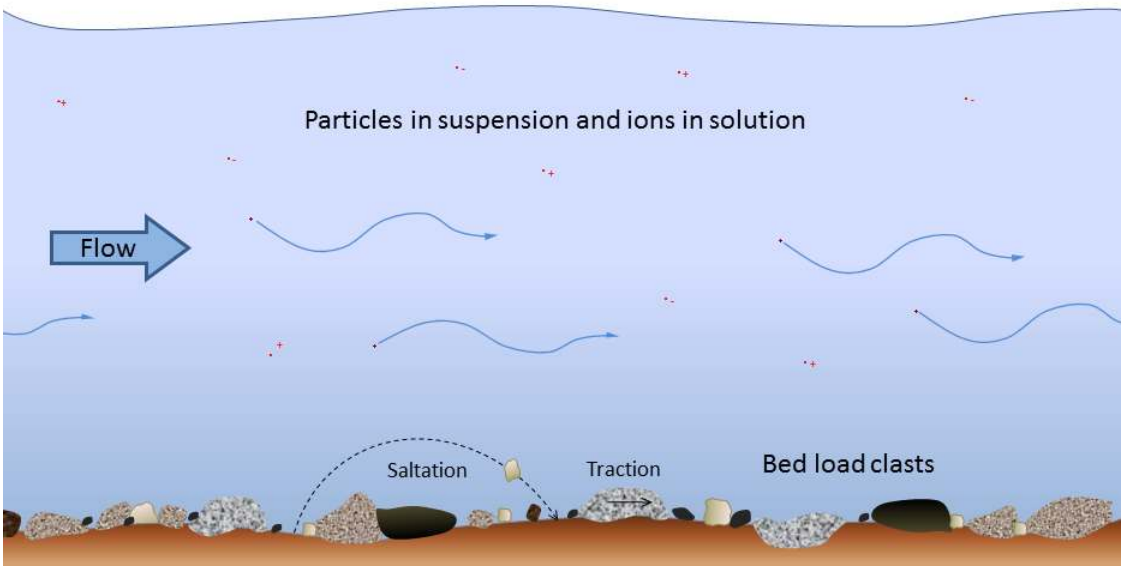



Figure \(\PageIndex{5}\): *Modes of transportation of sediments and dissolved ions (represented by red dots with + and – signs) in a stream.* © Steven Earle. CC BY.

As you know from intuition and from experience, 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, which represents (on average) about 15% of the mass of material transported, and includes ions such as calcium ($Ca^{+2}$) and chloride (Cl-) in solution. The solubility of these ions is not affected by flow velocity.

A stream typically reaches its greatest velocity when it is close to flooding over its banks. This is known as the bank-full stage, as shown in Figure \(\PageIndex{6}\). As soon as the flooding stream overtops its banks and occupies the wide 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**.

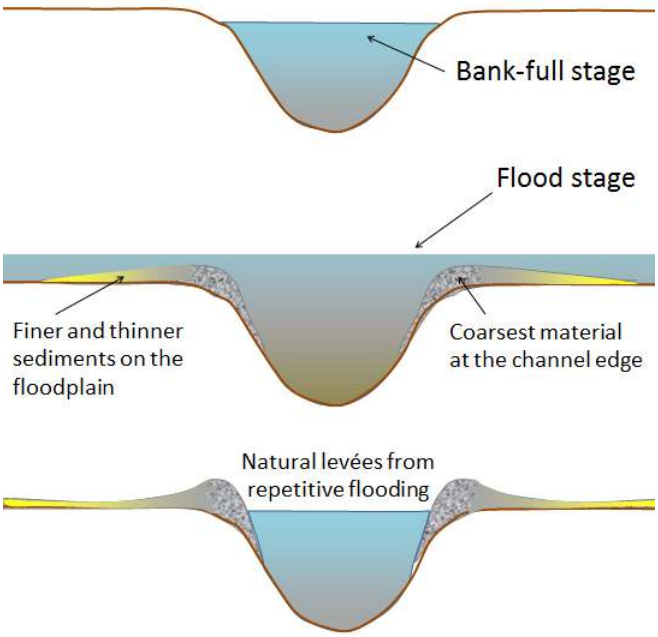


Figure \(\PageIndex{6}): *The development of natural levées during flooding of a stream. The sediments of the levée become increasingly fine away from the stream channel, and even finer sediments — clay, silt, and fine sand — are deposited across most of the flood plain.* © Steven Earle. CC BY.

## Important Aspects of Floods

There are several important aspects of floods:

- ***discharge****:* discharge is important because of its close relation to flow velocity: how erosive is the flood flow in terms of banks, levees, and river structures?
- ***stage****:* the height of the water determines flood damage on the floodplain (This is what the public thinks about most in connection with floods.) Also, it relates to the water-passing capacity of bridges, culverts, dam spillways, etc.
- ***volume****:* this relates to how much water is available to be caught and impounded in reservoirs for water supply and flood control.
- ***area inundated****:* this is determined by stage together with river-valley topography.
- ***flow velocity****:* Although the flow velocity is much less over the floodplain than in the main channel, it can be great enough to cause damage.

**What are the important effects or factors that govern the magnitude of a flood?**

• **Rainfall** in the drainage basin upstream of the given point on the given river. This boils down to both rainfall *intensity* (depth per unit time) and rainfall *duration*, and also the *area* of the drainage basin covered by the rainstorm. The worst scenario is: high intensity for a long time over a large area of the drainage basin.

• Preceding condition of **soil moisture** in the area of rainfall: dry, which lessens the runoff, or saturated, which maximizes the runoff? The difference can be as much as half an inch of equivalent rainfall. That's a big effect for small floods but only a small effect for large floods.

• **Channel storage**. Think in terms of what happens when a flood discharge from an upstream river enters a large lake. The lake has a large surface area, so it takes a long time for the flood flow to raise the lake elevation and therefore increase the discharge out of the lake. This storage effect is also present to a non-negligible extent along the river channel itself: As the flood proceeds downstream, it has to fill a greater and greater volume of space in the river channel itself. This effect *attenuates* the flood in the downstream direction: in effect, it *stretches out* the hydrograph and lowers its peak (in terms of either the stage hydrograph or the discharge hydrograph).

• **Channel geometry**. The narrower the channel, for a given flood discharge, the more the stage rises. When the stage reaches bank-full, the water spreads out over the floodplain, and the stage then increases far less rapidly with increasing discharge.

## Are Floods Getting Worse?

The answer to that question is certainly "yes", in terms of *dollar value of damage*, just because of greater human use of flood-prone areas near rivers. But the answer is yes also in terms of *higher stage*, because of the effect of more levee protection: the levees prevent the high discharges from spreading over the floodplain, so the rise in stage is confined to within the channel (Figure **6.4.4.7**).




Figure **6.4.4.7**: The stage of a flood flow for a given flood discharge. A) Before building of levees. B) After building of levees.

Is flood *discharge* increasing on average as well? There are two aspects to this question.

(1) In *urbanized* drainage basins, the percentage of area paved increases, thus decreasing the concentration time and tending to produce higher peaks in the hydrograph (stage and discharge) for a given rainfall. In rural drainage basins, bare-soil agriculture and clear cutting of forests have the same effect, although it seems not to be a really major effect. (Their effect on sediment yield is much greater, though.)

(2) Average rainfall in the drainage basin might be increasing because of *climate change* (although in much of the U.S., the prediction is that global warming will decrease precipitation rather than increase it). However, "storminess" (increased incidence of major rainstorms) is predicted to increase with climate change.

## Can Floods Be Controlled?

A better way of putting the question would be: how can flood damage be minimized, in the sense that a given measure affords more flood protection than it costs to institute?

(1) *Build higher and higher levees*. Yes, you can keep the floodwaters off the floodplain by building the levees higher. In a real sense, however, this is a self-defeating activity (Figure 5-50). Under natural conditions, the flood stage has built into it an inherently self-limiting effect: the floodplain area is so much greater than the channel area that, once the river goes beyond bank-full to inundate the floodplain, water-level rise is minor. (But remember that "minor" in a natural sense may be catastrophic for people living on the floodplain.) Building the levees higher restricts the river discharge to a much narrower width of the river valley, so the flood stage is much higher than would have been the case without the higher levees. Because of the effect of eliminating the floodplain-spreading effect, effective protection would necessitate impractically high levees. Moreover, if those high levees are overtopped, the flood damage to the surrounding floodplain area is likely to be much worse than if the levees had not been there.

(2) *Flood-control dams*. Clearly, the stage and discharge hydrographs can be reduced downstream of a given point on a river by building a water-storage dam there: catch the water so as to knock down the flood peak, and release the water slowly later in times of lower discharge. The problems, aside from the great cost of building dams, is that the effect lessens downstream, so a number of dams are needed. For decades, in the mid-1900s, there was a controversy over whether to build lots of small dams in upstream areas or a few big dams in downstream areas. It was finally concluded that *for effective flood control throughout the whole basin, both sets of dams would be needed.*

(3) *Land management.* Good agricultural practices (fallow planting; terrace agriculture; contour planting) should increase infiltration and thus lower runoff, and also make the concentration time longer. Careful consideration has shown, however, that the effect is measured in tenths of an inch of equivalent rainfall, so it would be effective against minor floods but not major floods. The great benefit is greatly decreased soil erosion—which is a different (but very important) matter.

(4) *Restrictions on human use of floodplains*. Clearly the best way to prevent flood damage is not to put anything susceptible to damage on floodplains! The expense of relocation is usually too great for this to be practical (although it's happening in some places; it's a policy issue rather than a scientific issue), but intermediate measures are effective: establish floodways and spreading areas that are used for such things as agriculture (of certain kinds) and recreation, but not permanent structures. This doesn't solve the problem for already urbanized areas, though.

## Can Floods Be Predicted?

Simple answer is just "no". Well, in a sense you *can* predict floods, but only after the rain has stopped falling! There are two aspects to this:

1. Given the rainstorm, predict the runoff in order to create a flood hydrograph
2. Given the flood hydrograph, watch the evolution of the hydrograph as the flood wave moves downstream. This technique is called ***flood routing***.

Various approximate techniques have been developed, and continue to be refined, for predicting the nature of the flood wave downstream of a rainstorm in a drainage basin. The assumptions and approximations are many: distribution of rainfall; extent of runoff vs. infiltration; time of concentration to create the hydrograph. It works okay for very small catchments but not so well for large catchments. It's a standard engineering technique for very small catchments (much less than one square mile). Once a flood is underway, however, there are good flood-routing techniques for predicting its future course down the river.

The other aspect of flood prediction has to do with the *statistics or probability of floods*. One assumes that, given the climate and the nature of the drainage basin upstream of a given point on a river, there is some underlying frequency distribution of floods, in terms of either stage or discharge. Any record of floods is then a sample population (in the parlance of statistics) from that underlying frequency distribution. So if you know the frequency distribution, you can give the probability of a given flood in a

given time period (that's not a prediction, but it's very useful in making a cost–benefit analysis or decision). However, the probability aspect of flood prediction assumes that the climate (specifically frequency of storms and average rainfall) and hydrology (partially controlled by development in the watershed) is not changing, which is not necessarily true with climate change or land use change. If you are interested in learning more about how fluvial hydrologists go about figuring out the probability of a given flood in a given time period in a given river system, go to the advanced topic at the end of this section.

## Effects of Urbanization

Urbanization can have major effects of the magnitude and frequency of floods. Here's an example ( Figure **6.4.4.8**): Seneca Creek, near Rockville, Maryland, drains an area of 100 square miles. Its watershed has been subject to much urbanization in recent decades. Take the record of discharge, 61 years long, from 1931 to 1991, and divide it into two equal parts, early and late. The average annual flood in the period 1931–1960 was 3000 cfs, in the period 1961–1991 it was 6000 cfs—twice as great! Likewise, bank-full discharge occurred 1.2 times per year in first period but 2.2 times per year in second period.

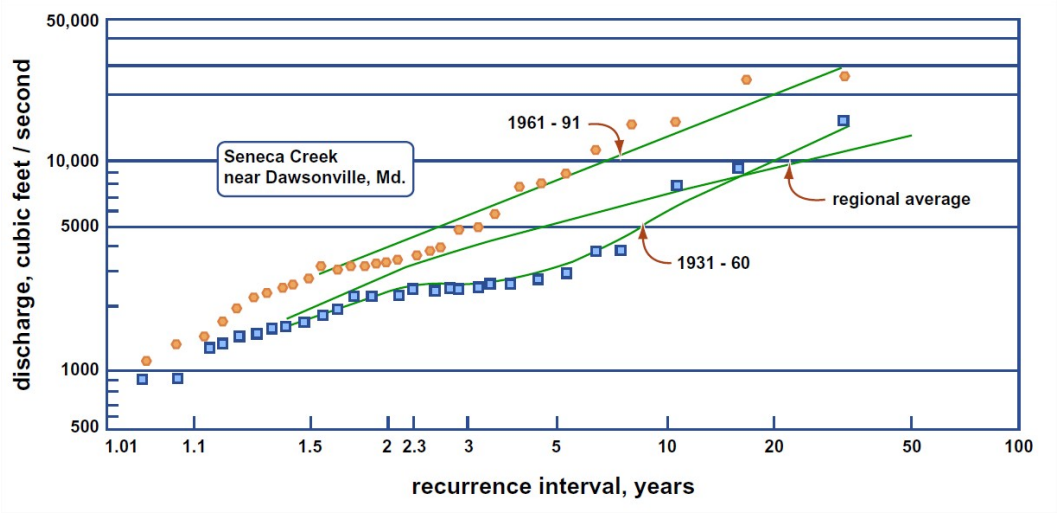



Figure **6.4.4.8**: Flood-frequency curve for Seneca Creek, Maryland. (From Leopold, 1994)

Rainfall was almost exactly the same in the two periods, so the great differences have to be accounted for entirely by the effects of urbanization effects. The stream is not artificially leveed, so presumably the effect has been caused by a decrease in concentration time; the more pavement, the faster the runoff, and the shorter the concentration time.

## Effects of Aggradation

There's a serious long-term problem in areas, especially in the lower reaches of large rivers, where crustal subsidence is non-negligible (i.e., the substrate beneath the active channel is moving downward relative to a datum like sea level) and/or where sea level itself is rising. (The former effect is more important than the latter.) The subsidence can be a combination of two effects:

- deep crustal subsidence
- subsidence caused by compaction of the sedimentary substrate of the river

The river channel aggrades to keep pace with subsidence, and in consequence it finds itself higher and higher above its floodplain —because, under conditions of net aggradation, deposition can be much faster in the channel, where the river is transporting most of its sediment, than over the floodplain. At certain times the river breaks through some particularly vulnerable point along its natural levees and finds its way out onto the floodplain, which is now at a significantly lower elevation, and establishes a whole new course for long distances down the floodplain. The process is called ***avulsion***; the river is said to ***avulse***. The river then proceeds to build up its channel bed and its levees, again to be high above the floodplain, and avulse once more. Avulsions often develop over a period of time rather than happening all at once: more and more of the river discharge is diverted through the gap in the levee, flood after flood. The end result, however, is the same.

To prevent avulsion, humans in avulsion-prone areas tend to build higher and higher levees, in the hope of preventing such an unfortunate event. The channel bed gets higher and higher relative to the surrounding floodplain ( Figure **6.4.4.9**).

This is happening today in the lowermost reaches of the Mississippi River. Much of the city of New Orleans lies well below the river. The Mississippi has an unfortunate preference for flowing down the Atchafalaya River to Morgan City. The U.S. Army Corps of Engineers maintains a gigantic headworks on the left bank of the river upstream of New Orleans to prevent the entire discharge of the Mississippi from going down the Atchafalaya; at present, about a third of the discharge leaves the Mississippi at that point. How long such a situation can be maintained is a matter of some controversy. If you would like to read an engaging account of the problem with the lower Mississippi, see the book by John McPhee (1989).

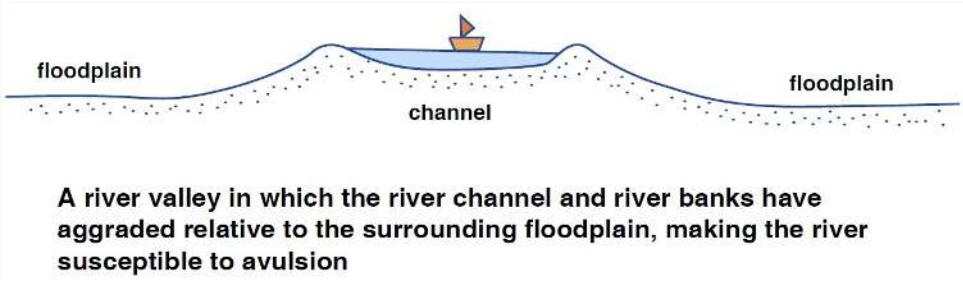



Figure **6.4.4.9**: A river valley in which the river channel and river banks have aggraded relative to the surrounding floodplain, making the river susceptible to avulsion.

China has an even bigger problem with the Huang He (in English, the Yellow River), because they have been fighting the river for far longer than we have been fighting the Mississippi. The bed of the Huang He at the city of Kaifeng is almost fifteen meters above the streets of the city, only several kilometers from the river! The Chinese are working hard to come up with a way of solving this problem in the long term; for the near term (a few more decades) the levees are considered to be adequate.

## Seasonality of River Discharge

The discharge levels of streams are highly variable depending on the time of year and on specific variations in the weather from one year to the next. In Canada, most streams show discharge variability similar to that of the Stikine River in northwestern B.C., as illustrated in Figure **6.4.4.10**. The Stikine River has its lowest discharge levels in the depths of winter when freezing conditions persist throughout most of its drainage basin. Discharge starts to rise slowly in May, and then rises dramatically through the late spring and early summer as a winter's worth of snow melts. For the year shown, the minimum discharge on the Stikine River was 56 cubic meters per second ($m^3/s$) in March, and the maximum was 37 times higher, 2,470 $m^3/s$, in May.

Figure **6.4.4.10**: Variations in discharge of the Stikine River during 2013.
© Steven Earle. CC BY. Based on from data at Water Survey of Canada, Environment Canada.

Streams in coastal areas of southern British Columbia show a very different pattern from those in most of the rest of the country because their drainage basins do not remain entirely frozen and because they receive a lot of rain (rather than snow) during the winter. The Qualicum River on Vancouver Island typically has its highest discharge levels in January or February and its lowest levels in late summer (Figure **6.4.4.11**). In 2013, the minimum discharge was 1.6 $m^3/s$, in August, and the maximum was 34 times higher, 53 $m^3/s$, in March.

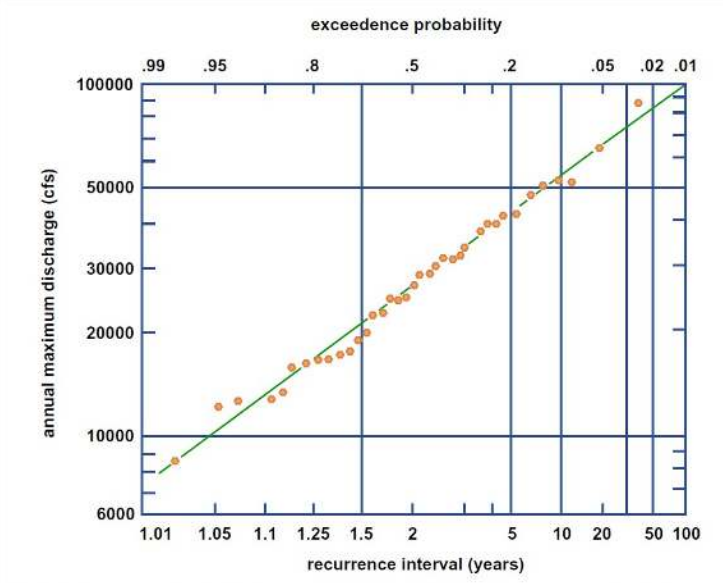
Figure **6.4.4.11**: Variations in discharge of the Qualicum River during 2013.
© Steven Earle. CC BY. Based on from data at Water Survey of Canada, Environment Canada.

When a stream's discharge increases, both the water level (**stage**) and the velocity increase as well. Rapidly flowing streams become muddy and large volumes of sediment are transported both in suspension and along the stream bed. In extreme situations, the water level reaches the top of the stream's banks (the **bank-full stage**, see Figure **6.4.4.4**), and if it rises any more, it floods the surrounding terrain. In the case of mature or old-age streams, this could include a vast area of relatively flat ground known as a flood plain, which is the area that is typically covered with water during a major flood. Because fine river sediments are deposited on flood plains, they are ideally suited for agriculture, and thus are typically occupied by farms and residences, and in many cases, by towns or cities. Such infrastructure is highly vulnerable to damage from flooding, and the people that live and work there are at risk.

## ADVANCED TOPIC: THE PROBABILITY APPROACH TO RIVER FLOODS

**1.** The important question that arises is: *what can be done about establishing the probability of very large floods*? This is extremely important, but it's fraught with difficulties. Here's how it's done, given a long record of discharge at some location along the river (Figure **6.4.4.12**):

- Determine the largest momentary discharge in each year.
- Rank these in order of decreasing discharge (the highest being number one, the lowest being equal to the number of years of record).
- Compute the ***recurrence interval*** (also called the ***return period***), in years: the average number of years within which a given event (i.e., a flood of given magnitude) will be equaled or exceeded), by the formula $T = (n+1)/m$, where $T$ is the recurrence interval, $n$ is the number of years in the record, and $m$ is the rank of the event (or, equivalently, the probability $p$ of a given event being equaled or exceeded in any given year; that's called the ***exceedence probability***).
- Plot the results in a graph with recurrence interval and/or exceedence probability on the horizontal axis and the annual maximum discharge on the vertical axis.
- Fit a smooth curve through the points in some way.




Flood-frequency curve plotted on logarithmic probability paper, Tana River at Garissa. Keny. 1934-1970. The scale at the top is the probability that the discharge is equaled or exceeded in any given year. The bottom scale, recurrence interval, is the average number of years in which the annual peak equals or exceeds the discharge given on the ordinate.

Figure **6.4.4.12**: An example of a flood-frequency curve. (From Leopold, 1994.)

**2.** The results are reliable for the mid-range of the graph, but they become very uncertain near the high-discharge end because of the small numbers of very large events. Fitting such a tail of a distribution is always tricky business, but it is important if one is to try to extrapolate to much longer times, which is one of the important goals of such an exercise. Various techniques have been in use for fitting such a curve. Several such distributions have been in use. The trouble with blind curve fitting is that one or two unrepresentative or outlying points at the high-discharge end can bias the curve significantly.

**3.** One valuable approach is to try to determine, from historical records, whether the greatest discharge in the period of record was exceeded in earlier times. If the greatest discharge in the period of record can be established to have been greater than any other for a number of years before the period of record, the period of record can be extended back in time and the ranking redone. This tends to shift the unrepresentatively large outlying point to the right and bring it more closely in line with the rest of the distribution.

### Contributors and Attributions

Modified by Kyle Whittinghill (University of Pittsburgh) from the following sources:

- **5.12: Floods** by John Southard, is licensed CC BY-NC-SA

- Stream Types, Stream Erosion and Deposition, and Flooding from Physical Geology by Steven Earle (licensed under a Creative Commons Attribution 4.0 International License)

---

This page titled 6.4.4: Floods is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by John Southard (MIT OpenCourseware) .

- **5.12: Floods** 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/.

# 6.5: Geologic Resources

6.5: Geologic Resources is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 6.5.1: Mineral Resources

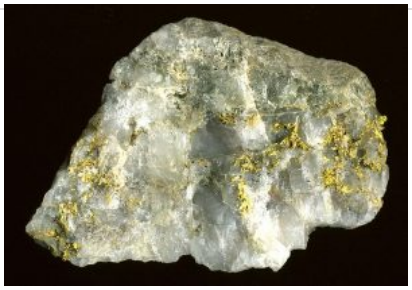


Figure **6.5.1.1**: Gold-bearing quartz vein from California.

Mineral resources, while principally nonrenewable, are generally placed in two main categories: **metallic** (containing metals) or **nonmetallic** (containing other useful materials). Metallic minerals are those from which valuable metals (e.g. iron, copper) can be extracted for commercial use. Metals that are considered geochemically abundant occur at crustal abundances of 0.1 percent or more (e.g. iron, aluminum, manganese, magnesium, titanium). Metals that are considered geochemically scarce occur at crustal abundances of less than 0.1 percent (e.g. nickel, copper, zinc, platinum metals). Some important metallic minerals are: hematite (a source of iron), bauxite (a source of aluminum), sphalerite (a source of zinc) and galena (a source of lead). Metallic minerals occasionally but rarely occur as a single element (e.g. native gold or copper). Most mining is focused on metallic minerals. A significant part of the advancement of human society has been developing the knowledge and technologies that yielded metal from the Earth and allowed the machines, buildings, and monetary systems that dominate our world today. The location and recovery of these metals have been a key facet of the study of geology since its inception.

While receiving much less attention, nonmetallic mineral resources (also known as industrial minerals) are just as vital to ancient and modern society as metallic minerals. Nonmetallic minerals are valuable, not for the metals they contain, but for their properties as chemical compounds. Because they are commonly used in industry, they are also often referred to as industrial minerals. They are classified according to their use. The most basic of these is building stone. Limestone, travertine, granite, slate, and marble are common building stones and have been quarried for centuries (Figure **6.5.1.2**). Some industrial minerals are also used for building materials (e.g. gypsum for plaster and kaolin for bricks). Some industrial minerals are used as sources of important chemicals (e.g. halite for sodium chloride and borax for borates). For everything made out of concrete or asphalt, we need sand and gravel. To make the cement that holds concrete together, we also need limestone. Others are used for making fertilizers (e.g. apatite for phosphate and sylvite for potassium). Still others are used as abrasives (e.g. diamond and corrundum). For the glass in our computer screens and for glass-sided buildings, we need silica sand plus sodium oxide ($Na_2O$), sodium carbonate ($Na_2CO_3$), and calcium oxide (CaO). For a wide range of applications (e.g., ceramics and many industrial processes), we also need various types of clay. Some nonmetallic mineral resources are not mineral specific; nearly any rock or mineral can be used. This is generally called aggregate and is used in concrete, roads, and foundations. Gravel is one of the more common aggregates. Quarried rock is also used in some applications where rounded gravel isn't suitable, such as the ballast (road bed) for railways, where crushed angular rock is needed.



Figure **6.5.1.2**: Carrara marble quarry in Italy, source to famous sculptures like Michelangelo's David.

## Mineral Deposits

Minerals are everywhere around us. For example, the ocean is estimated to contain more than 70 million tons of gold. Yet, it would be much too expensive to recover that gold because of its very low concentration in the water. Minerals must be concentrated into deposits to make their collection economically feasible. A mineral deposit containing one or more minerals that can be extracted profitably is called an **ore**. Many minerals are commonly found together (e.g. quartz and gold; molybdenum, tin and tungsten; copper, lead and zinc; platinum and palladium). Because various geologic processes can create local enrichments of minerals, mineral deposits can be classified according to the concentration process that formed them. The five basic types of mineral deposits are: **hydrothermal**, **magmatic**, **sedimentary**, **placer** and **residual**.

If the material can be mined at a profit, the body constitutes an **ore** deposit. Typically, the term ore is used for only metal-bearing minerals, though the concept of ore as a non-renewable resource can be applied to valuable concentrations of fossil fuels, building stones, and other non-metal deposits, even groundwater. Mineral **ores** are found in just a relatively few areas, because it takes a special set of circumstances to create them. Therefore, the signs of a mineral deposit are often small and difficult to recognize. Locating deposits requires experience and knowledge. Geologists can search for years before finding an economic mineral deposit. Deposit size, its mineral content, extracting efficiency, processing costs and market value of the processed minerals are all factors that determine if a mineral deposit can be profitably developed. For example, when the market price of copper increased significantly in the 1970s, some marginal or low-grade copper deposits suddenly became profitable ore bodies.

### Magmatic Deposits

**Magmatic** mineral deposits are formed when processes such as partial melting and fractional crystallization occur during the melting and cooling of rocks. **Layered intrusion** (typically ultramafic to mafic) can be host to deposits that contain copper, nickel, platinum-palladium-rhodium, and chromium. The Stillwater Complex in Montana is an example of an economic layered mafic intrusion [30]. Associated deposit types can contain chromium or titanium-vanadium. The largest magmatic deposits in the world are the chromite deposits in the Bushveld Igneous Complex in South Africa [31] (Figure **6.5.1.3**). Rocks of the Bushveld Igneous Complex have an areal extent larger than the state of Utah. The chromite occurs in layers, which resemble sedimentary layers, except this occurred within a crystallizing magma chamber.

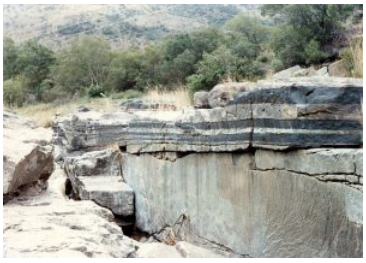


Figure **6.5.1.3**: Layered intrusion of dark chromium-bearing minerals, Bushveld Complex, South Africa

Water and other volatiles that are not incorporated into mineral crystals while a magma crystallizes become concentrated around the margins of these crystallizing magmas. Ions in these hot fluids are very mobile and can form exceptionally large crystals. Once crystallized, masses of these large crystals are called **pegmatites** that form from the concentration of magma fluids near the end of crystallization when nearly the entire magma body has crystallized (Figure **6.5.1.4**). In addition to minerals that are predominant in the main igneous mass, such as quartz, feldspar, and mica, pegmatite bodies may also contain very large crystals of unusual minerals that contain rare elements like beryllium, lithium, tantalum, niobium, and tin, as well as native elements like gold [32]. Such pegmatites are ores of these metals.

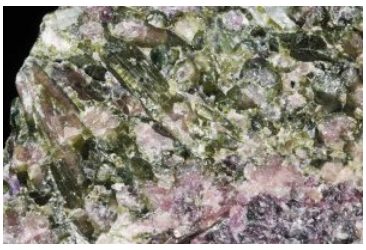


Figure **6.5.1.4**: This pegmatite from Brazil contains lithium-rich green elbaite (a tourmaline) and purple lepidolite (a mica).

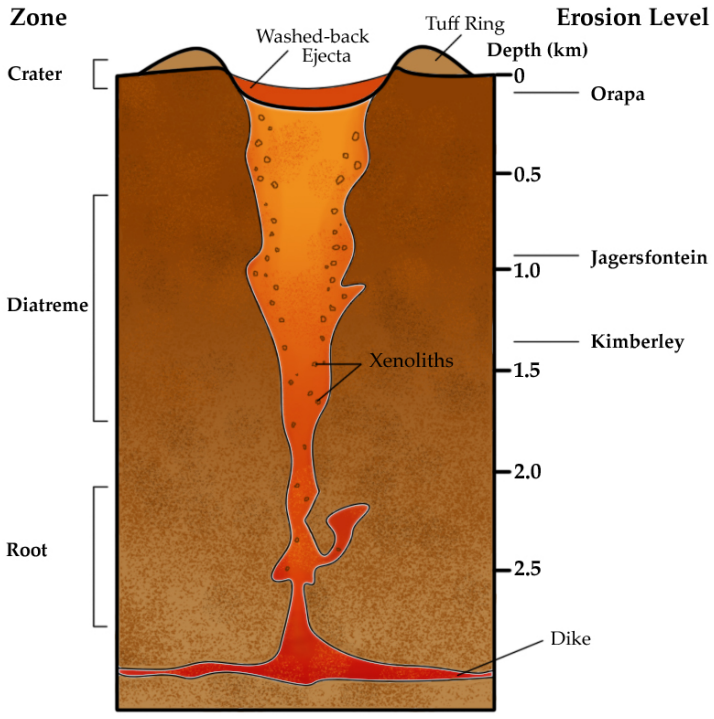



Figure 6.5.1.5: Schematic diagram of a kimberlite pipe.

An unusual magmatic process is a **kimberlite** pipe, which is a volcanic conduit that transports ultramafic magma from depths in the mantle to the surface (Figure 6.5.1.5). Diamonds, which are formed at great temperature and depth, are transported this way to locations where they can be mined. The process that emplaced these kimberlite (ultramafic) rocks is no longer common on Earth, and most of the known deposits are Archean [33].

## Hydrothermal Deposits

Hydrothermal mineral deposits are formed when minerals are deposited by hot, aqueous solutions flowing through fractures and pore spaces of crustal rock (Figure 6.5.1.6). Many famous ore bodies have resulted from **hydrothermal** deposition, including the tin mines in Cornwall, England and the copper mines in Arizona and Utah.

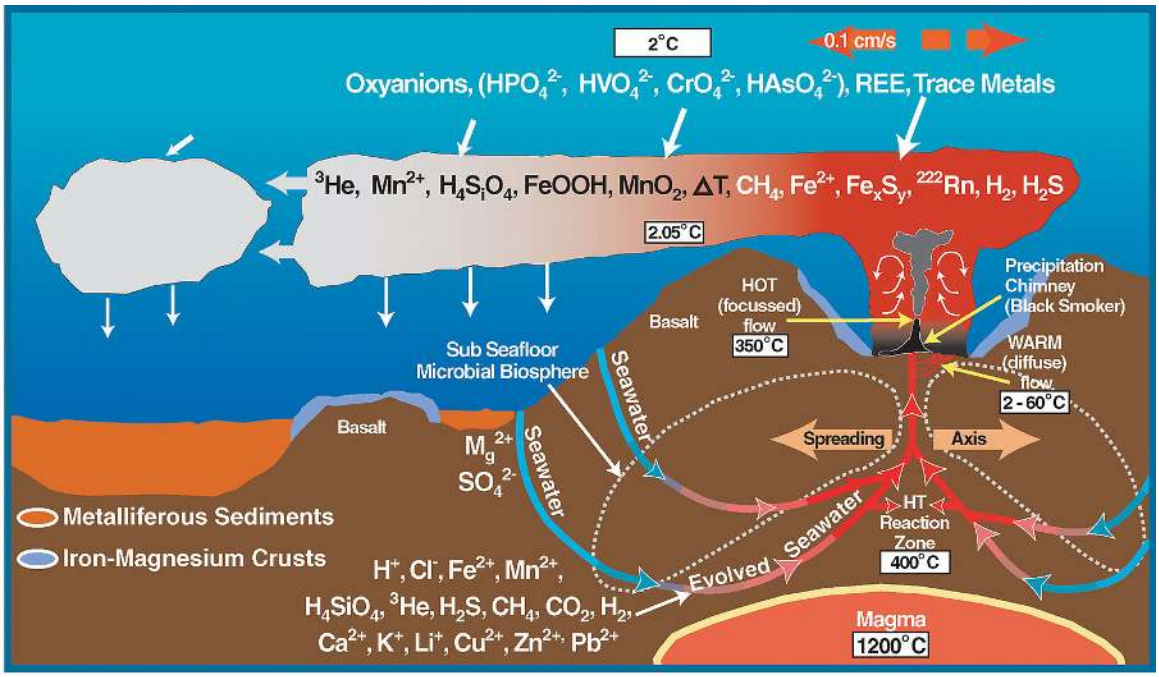



Figure 6.5.1.6: The complex chemistry around mid-ocean ridges.

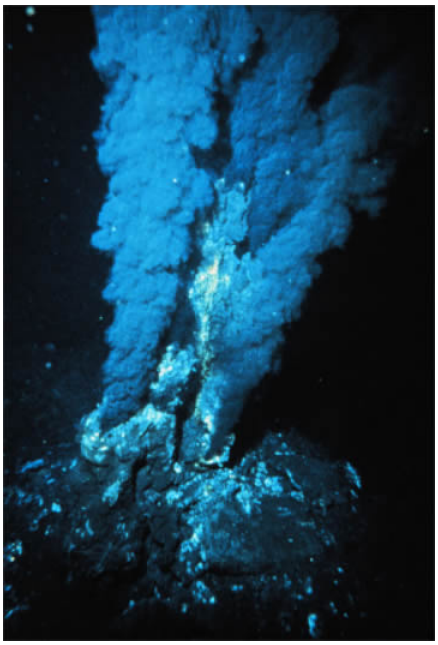


Figure 6.5.1.7: **Black Smoker** A billowing discharge of superheated mineral-rich water at an oceanic ridge, in the Atlantic Ocean. Black "smoke" is actually from metallic sulfide minerals that form modern ore deposits. *Source: P. Rona of U.S. National Oceanic and Atmospheric Administration via Wikimedia Commons*

The most active hydrothermal process today produces **volcanogenic massive sulfide** (VMS) deposits, which form from black smoker activity (Figure 6.5.1.7) near mid-ocean ridges all over the world, and commonly contain copper, zinc, lead, gold, and silver when found on the surface [34]. The largest of these deposits occur in Precambrian age rocks. The Jerome deposit in central Arizona is a good example.

Another type of deposit which draws on heated water from magma is a **porphyry** deposit (Figure 6.5.1.8). This is not to be confused with the igneous texture porphyritic, although the name is derived from the porphyritic texture that is nearly always present in the igneous rocks in a porphyry deposit. Several types of porphyry deposits exist: porphyry copper, porphyry

molybdenum, and porphyry tin. They are characterized by the presence of low-grade disseminated ore minerals closely associated with intermediate and felsic intrusive rocks over a very large area [35]. Porphyry deposits are typically the largest mines on Earth. One of the largest, richest, and possibly best-studied mines in the world is Utah's Bingham Canyon open-pit mine, which has had over 100 years of high production of several elements including copper, gold, molybdenum, and silver. Associated underground carbonate replacement deposits have produced lead, zinc, gold, silver, and copper [36]. Past open pit production at this mine was dominated by copper and gold from chalcopyrite and bornite. Gold occurs in minor quantities in the copper-bearing mineral, but the large scale of production makes Bingham Canyon one of the largest gold mines in the U.S. Future production may be more copper and molybdenum (molybdenite) from deeper underground mines.

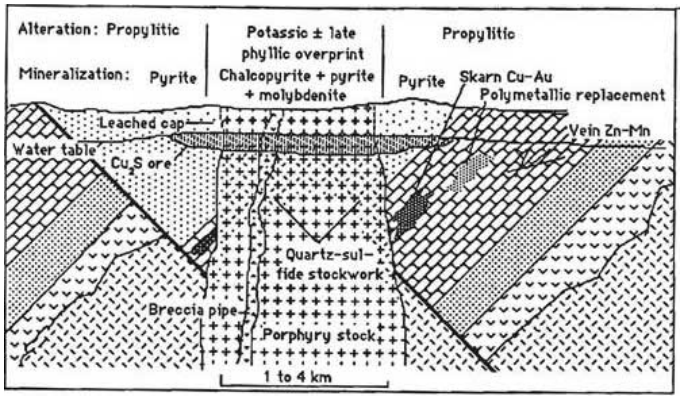



Figure 6.5.1.8: USGS schematic of a Porphyry copper deposit.

The majority of porphyry copper deposits owe their economic value to concentration by weathering processes occurring millions of years after the hosting intrusion called **supergene enrichment** (Figure 6.5.1.9). These occur once the hydrothermal event has ceased and the ore body has been uplifted, eroded, and exposed to oxidation [37]. When the upper pyrite-rich portion of the deposit is exposed to rain, pyrite in the oxidizing zone creates an extremely acid condition which dissolves copper out of copper minerals such as chalcopyrite, converting the chalcopyrite to iron oxides like hematite or goethite. The copper is carried downward in the solution until it arrives at the groundwater table and a reducing environment where the copper precipitates, converting primary copper minerals into secondary higher-copper content minerals. Chalcopyrite (35% Cu) is converted to bornite (63% Cu) and ultimately chalcocite (80% Cu). Without this enriched zone (2 to 5 times higher in copper content than the main deposit) most porphyry copper deposits would not be economic.



Figure 6.5.1.9: The Morenci porphyry is oxidized toward its top (as seen as red rocks in the wall of the mine), creating supergene enrichment.

If limestone or other calcareous sedimentary rocks are present adjacent to the magmatic body, then another type of ore deposit called a **skarn** deposit can form (Figure 6.5.1.10). These metamorphic rocks form as magma-derived, highly saline metalliferous fluids react with carbonate rocks, creating calcium-magnesium-silicate minerals like pyroxene, amphibole, and garnet, as well as high-grade zones of iron, copper, and zinc minerals and gold [38]. Intrusions that are genetically related to the intrusion that made the Bingham Canyon deposit have also produced copper-gold skarns that were mined by the early European settlers in Utah [39; 40]. Metamorphism of iron and/or sulfide deposits commonly results in an increase in grain size that makes separation of gangue from the desired sulfide or oxide minerals much easier.

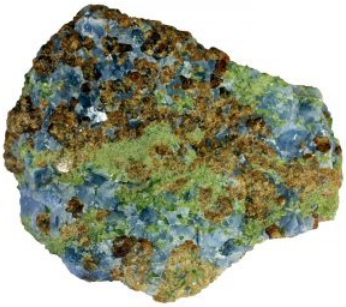


Figure 6.5.1.10: Garnet-augite skarn from Italy.

**Sediment-hosted disseminated gold** deposits consist of low concentrations of microscopic gold as inclusions and disseminated atoms in pyrite crystals (Figure 6.5.1.11). These are formed via low-level hydrothermal reactions (generally in the realm of diagenesis) that occur in certain rock types, namely muddy carbonates and limey mudstones. This hydrothermal alteration is generally far-removed from a magma source but can be found in extended rocks with a high geothermal gradient. The earliest locally mined deposit of this type was the Mercur deposit in the Oquirrh Mountains of Utah where almost one million ounces of gold were recovered between 1890 and 1917. In the 1960s a metallurgical process using cyanide was developed for these types of low-grade ores. These deposits are also called **Carlin-type** deposits because the disseminated deposit near Carlin, Nevada is where the new technology was first applied and because the first definitive scientific studies were conducted there [41]. Gold was introduced by hydrothermal fluids which reacted with silty calcareous rocks, removing carbonate, creating additional permeability, and adding silica and gold-bearing pyrite in the pore space between grains. The Betze-Post mine and the Gold Quarry mine on the "Carlin Trend" are two of the largest of the disseminated gold deposits in Nevada. Similar deposits, but not as large, have been found in China, Iran, and Macedonia [42].




Figure 6.5.1.11: In this rock, a pyrite cube has dissolved (as seen with the negative "corner" impression in the rock), leaving behind small specks of gold.

## Deposits from Sedimentary and Weathering Processes



Figure 6.5.1.12: Underground uranium mine near Moab, Utah.

Geochemical processes that occur at or near the surface without the aid of magma also concentrate metals, but to a lesser degree than hydrothermal processes. One of the main reactions is **redox** (short for reduction/oxidation) chemistry, which has to do with the amount of available oxygen in a system. Places where oxygen is plentiful, as in the atmosphere today, are considered oxidizing environments, while oxygen-poor environments are considered reducing. Uranium deposition is an example of redox mobilization. Uranium is soluble in oxidizing groundwater environments and precipitates as uraninite when reducing conditions are encountered. Many of the deposits across the Colorado Plateau (e.g. Moab, Utah) were formed by this method [43].

Redox reactions were also responsible for the creation of **banded iron formations** (BIFs), which are interbedded layers of iron oxide (hematite and magnetite), chert, and shale beds (Figure **6.5.1.13**). These deposits formed early in the Earth's history as the atmosphere was becoming oxygenated. Cyclic oxygenation of iron-rich waters initiated the precipitation of the iron beds. Because BIFs are generally Precambrian in age, they are only found in some of the older exposed rocks in the United States, in the upper peninsula of Michigan and northeastern Minnesota [44].



Figure **6.5.1.13**: *Banded iron formation from an unknown location in North America on display at a museum in Germany. The rock is about 2 m across. The dark grey layers are magnetite and the red layers are hematite. Chert is also present. [https://upload.wikimedia.org/wikiped...%28aka%29.jpg]*

Deep, saline, connate fluids (trapped in the pore spaces), within sedimentary basins may be highly metalliferous. When expelled outward and upward during basin compaction, these fluids may form lead and zinc deposits in limestone by replacement or by filling open spaces (caves, faults) and in sandstone by filling pore spaces. The most famous of these are called **Mississippi Valley-type** deposits [44] (Figure **6.5.1.14**). Also known as carbonate-hosted replacement deposits, they are large deposits of galena and sphalerite (lead and zinc ores) which form from fluids in the temperature range of 100 to 200°C. Although they are named for occurrences along the Mississippi River Valley in the United States, they are found worldwide.

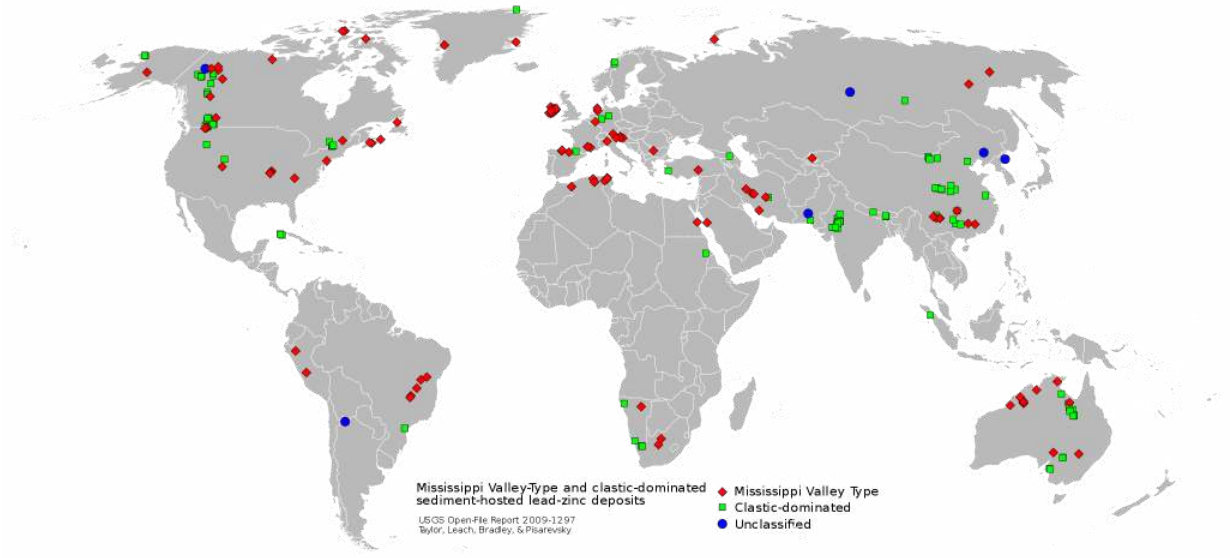



Figure **6.5.1.14**: Map of Mississippi-Valley type ore deposits.

**Sediment-hosted copper** deposits occurring in sandstones, shales, and marls are enormous in size and their contained resources are comparable to porphyry copper deposits. These were most-likely formed diagenetically by groundwater fluids in highly-permeable rocks [45]. Well-known examples are the Kupferschiefer in Europe, which has an areal coverage of >500,000 $Km_2$, and the Zambian Copper Belt in Africa.



Figure **6.5.1.15**: Apatite from Mexico.

Phosphorus is an essential element that occurs in the mineral apatite, which is found in trace amounts in common igneous rocks (Figure **6.5.1.15**). Phosphorite rock, which is formed in sedimentary environments in the ocean [50], contains abundant apatite and is mined to make fertilizer. Without phosphorus, life as we know it is not possible. Phosphorous is a major component of bone and a key component of DNA. Bone ash and guano are natural sources of phosphorus.

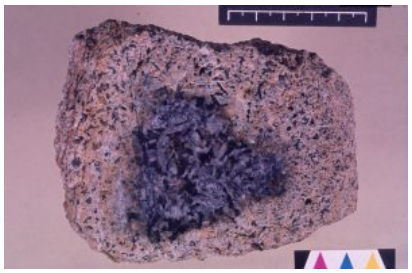


Figure **6.5.1.16**: A sample of bauxite. Note the unweathered igneous rock in the center.

**Residual** mineral deposits can form when weathering processes remove water soluble minerals from an area, leaving a concentration of less soluble minerals. The aluminum ore, **bauxite**, was originally formed in this manner under tropical weathering conditions [46] (Figure **6.5.1.16**). The best known bauxite deposit in the United States occurs in Arkansas. Aluminum concentrates in soils as feldspar and ferromagnesian minerals in igneous and metamorphic rocks undergo chemical weathering processes. Weathering of ultramafic rocks results in the formation of nickel-rich soils and weathering of magnetite and hematite in banded iron formation results in the formation of goethite, a friable mineral that is easily mined for its iron content.

At the earth's surface, the physical process of mass wasting or fluid movement concentrates high-density minerals by hydraulic sorting. When these minerals are concentrated in streams, rivers, and beaches, they are called **placer** deposits, whether in modern sands or ancient lithified rocks [47] (Figure **6.5.1.17**). Native gold, native platinum, zircon, ilmenite, rutile, magnetite, diamonds, and other gemstones can be found in placers. Humans have mimicked this natural process to recover gold manually by gold panning and by mechanized means such as dredging.



**Figure 6.5.1.17:** Lithified heavy mineral sand (dark layers) from a beach deposit in India.

The best types of **aggregate** (sand and gravel) resources are those that have been sorted by streams, and in Canada the most abundant and accessible fluvial deposits are associated with glaciation (Figure **6.5.1.18**). That doesn't include till of course, because it has too much silt and clay, but it does include glaciofluvial outwash, which is present in thick deposits in many parts of the country, similar to the one shown in Figure 20.15. In a typical gravel pit, these materials are graded on-site according to size and then used in a wide range of applications from constructing huge concrete dams to filling children's sandboxes. Sand is also used to make glass, but for most types of glass, it has to be at least 95% quartz (which the sandy layers shown in Figure 20.15 are definitely not), and for high-purity glass and the silicon wafers used for electronics, the source sand has to be over 98% quartz.



Figure **6.5.1.18**: *Sand and gravel in an aggregate pit near Nanaimo, BC. [SE]*

**Evaporite** deposits form in restricted basins, such as the Great Salt Lake or the Dead Sea, where evaporation of water exceeds the recharge of water into the basin [49] (Figure **6.5.1.19**). As the waters evaporate, soluble minerals are concentrated and become supersaturated, at which point they precipitate from the now highly-saline waters. If these conditions persist for long stretches of time, thick deposits of rock salt and rock gypsum and other minerals can accumulate.



Figure **6.5.1.19**: Salt-covered plain known as the Bonneville Salt Flats, Utah.

Evaporite minerals like halite are used in our food as common table salt. Salt was a vitally important economic resource prior to refrigeration as a food preservative. While still used in food, now it is mainly mined as a chemical agent, water softener, or a de-icer for roads. The largest salt mine in the world is at Goderich, Ontario, where salt is recovered from the 100 m thick Silurian Salina Formation. Gypsum ($CaSO_4.2H_20$) is a common nonmetallic mineral used as a building material, being the main component of drywall. It is also used as a fertilizer. Other evaporites include sylvite (potassium chloride) and bischofite (magnesium chloride), both of which are used in agriculture, medicine, food processing, and other applications. Potash, a group of highly soluble potassium-bearing evaporite minerals, is used as a fertilizer. In hyperarid locations, even rarer and more complex evaporites, like borax, trona, ulexite, and hanksite, are found and mined. They can be found in such localities as Searles Dry Lake and Death Valley, California, and in ancient evaporite deposits of the Green River Formation of Utah and Wyoming.

## References


30. Boudreau, A. E. The Stillwater Complex, Montana – Overview and the significance of volatiles. *Mineralogical Magazine* **80**, 585–637 (2016).

31. Willemse, J. The geology of the Bushveld Igneous Complex, the largest repository of magmatic ore deposits in the world. *Economic Geology Monograph* **4**, 1–22 (1969).

32. London, D. & Kontak, D. J. Granitic Pegmatites: Scientific Wonders and Economic Bonanzas. *Elements* **8**, 257–261 (2012).

33. Arndt, N. T. Chapter 1 Archean Komatiites. in *Developments in Precambrian Geology* (ed. K.C. Condie) **11**, 11–44 (Elsevier, 1994).

34. Barrie, C. T. Volcanic — associated massive sulfide deposits: processes and examples in modern and ancient settings. (1999). Available at: https://www.researchgate.net/profile/Michael_Perfit/publication/241276560_Geologic_petrologic_and_geochemical_relationships_between_magmatism_and_massive_sulfide_mineralization_along_the_eastern_Galapagos_Spreading_Center/links/02e7e51c8707bbfe9c000000.pdf. (Accessed: 2nd July 2016)

35. Richards, J. P. Tectono-Magmatic Precursors for Porphyry Cu-(Mo-Au) Deposit Formation. *Econ. Geol.* **98**, 1515–1533 (2003).

36. Hawley, C. C. *A Kennecott Story: Three Mines, Four Men, and One Hundred Years, 1887-1997*. (University of Utah Press, 2014).

37. Ague, J. J. & Brimhall, G. H. Geochemical modeling of steady state fluid flow and chemical reaction during supergene enrichment of porphyry copper deposits. *Econ. Geol.* **84**, 506–528 (1989).

38. Einaudi, M. T. & Burt, D. M. Introduction; terminology, classification, and composition of skarn deposits. *Econ. Geol.* **77**, 745–754 (1982).

39. Bromfield, C. S., Erickson, A. J., Haddadin, M. A. & Mehnert, H. H. Potassium-argon ages of intrusion, extrusion, and associated ore deposits, Park City mining district, Utah. *Econ. Geol.* **72**, 837–848 (1977).

40. James, L. P. *Geology, ore deposits, and history of the Big Cottonwood Mining District, Salt Lake County, Utah*. (Utah Geological and Mineral Survey, Utah Department of Natural Resources, 1979).

41. Hofstra, A. H. & Cline, J. S. Characteristics and models for Carlin-type gold deposits. *Reviews in Economic Geology* **13**, 163–220 (2000).

42. Rui-Zhong, H., Wen-Chao, S., Xian-Wu, B., Guang-Zhi, T. & Hofstra, A. H. Geology and geochemistry of Carlin-type gold deposits in China. *Miner. Deposita* **37**, 378–392 (2002).

43. Lehmann, I. P', Publ. *Bur. Centr. Seism. Internat. Serie A* **14**, 87–115 (1936).

44. Klein, C. Some Precambrian banded iron-formations (BIFs) from around the world: Their age, geologic setting, mineralogy, metamorphism, geochemistry, and origins. *Am. Mineral.* **90**, 1473–1499 (2005).

45. Hitzman, M., Kirkham, R., Broughton, D., Thorson, J. & Selley, D. The Sediment-Hosted Stratiform Copper Ore System. *Econ. Geol.* **100th**, (2005).

46. Bárdossy, G. & Aleva, G. J. J. *Lateritic bauxites*. **27**, (Elsevier Science Ltd, 1990).

47. Cabri, L. J., Harris, D. C. & Weiser, T. W. Mineralogy and distribution of platinum-group mineral (PGM) placer deposits of the world. *Explor. Min. Geol.* **2**, 73–167 (1996).

48. Nuss, P. & Eckelman, M. J. Life cycle assessment of metals: a scientific synthesis. *PLoS One* **9**, e101298 (2014).

49. Gordon, W. A. Distribution by latitude of Phanerozoic evaporite deposits. *J. Geol.* **83**, 671–684 (1975).

50. Delaney, M. L. Phosphorus accumulation in marine sediments and the oceanic phosphorus cycle. *Global Biogeochem. Cycles* **12**, 563–572 (1998).

## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- Metal Deposits and Industrial Minerals from Physical Geology by Steven Earle (licensed under a Creative Commons Attribution 4.0 International License)
- Minerals from *AP Environmental Science* by University of California College Prep
- Non-Renewable Resources from *Environmental Science: A Canadian Perspective* by Bill Freedman (Creative Commons Attribution NonCommercial)

- Mineral Resources and Mining from *An Introduction to Geology* by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, & Cam Mosher (Geology Faculty at Salt Lake Community College), download free from OpenGeology
- Mineral Resources: Formation, Mining, Environmental Impact from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin

---

This page titled 6.5.1: Mineral Resources is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, & Cam Mosher (OpenGeology) .

# 6.5.2: Mining

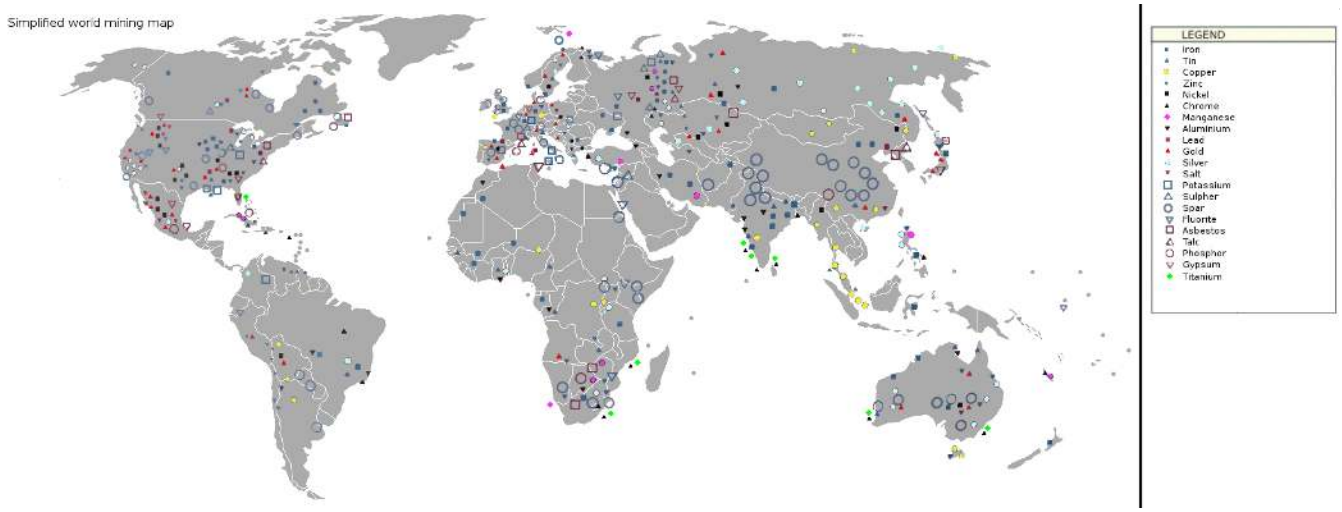



Figure **6.5.2.1**: Map of world mining areas.

**Mining** is defined as the extraction, from the Earth, of valuable material for societal use. Usually, this includes solid materials (e.g. gold, iron, coal, diamond, sand, and gravel), but can also include fluid resources such as oil and natural gas. Modern mining has a long relationship with modern society. The oldest evidence of mining, with a concentrated area of digging into the Earth for materials, has a history that may go back 40,000 years to the hematite (used as red dye) of the Lion Cave in Swaziland [4]. Resources extracted by mining are generally considered to be nonrenewable. After a potentially profitable mineral deposit is located, it is mined by one of several techniques. Which technique is used depends upon the type of deposit and whether the deposit is shallow and thus suitable for surface mining or deep and thus requiring sub-surface mining. The style of mining is a function of technology, social license, and economics. It is in the best interest of the company extracting the resources to do so in a cost-effective way.

It is implicit that the technology to mine is available, economic conditions are suitable, and political, social and environmental considerations are satisfied in order to classify a natural resource deposit as ore. Depending on the substance, it can be concentrated in a narrow vein or distributed over a large area as a low-concentration ore. Some materials are mined directly from bodies of water (e.g. sylvite for potassium; water through desalination) and the atmosphere (e.g. nitrogen for fertilizers). These differences lead to various methods of mining, and differences in terminology depending on the certainty. **Ore mineral resource** is used for an indication of ore that is potentially extractable, and the term **ore mineral reserve** is used for a well defined (proven), profitable amount of extractable ore.

## Surface Mining Techniques

Surface mining techniques include: open-pit mining, area strip mining, contour strip mining and hydraulic mining. **Open-pit mining** involves digging a large, terraced hole in the ground in order to remove a near-surface ore body (Figure **6.5.2.2**). This technique is used in copper ore mines in Arizona and Utah and iron ore mines in Minnesota. Typically, the pit progressively deepens through additional mining cuts to extract the ore, and the walls of the pit are as steep as can safely be managed. A steep wall means there is less waste (non-valuable) rock or overburden to remove and is an engineering balance between efficient mining and mass wasting. Occasionally landslides do occur, including a very large landslide that occurred in the Bingham Canyon mine in 2013. These events are costly and dangerous, though careful monitoring gave the Bingham Canyon mine ample warning time.



Figure **6.5.2.2**: **Open Pit Mine** Bingham Canyon copper mine in Utah, USA. At 4 km wide and 1.2 km deep, it is the world's deepest open-pit mine. It began operations in 1906. *Source: Tim Jarrett via Wikimedia Commons*

**Area strip mining** is used in relatively flat areas. The overburden of soil and rock is removed from a large trench in order to expose the ore body. After the minerals are removed, the old trench is filled and a new trench is dug. This process is repeated until the available ore is exhausted. **Contour strip mining** is a similar technique except that it is used on hilly or mountainous terrains. A series of terraces are cut into the side of a slope, with the overburden from each new terrace being dumped into the old one below. In **mountaintop removal mining**, a specific type of strip mining, an entire mountaintop or rock layer is removed to gain access to the ore below. The environmental impacts of surface mining are usually greater due to the larger surface disturbance footprint [7].



Figure **6.5.2.3**: A surface coal mine in Wyoming.

**Hydraulic mining** is used in places such as the Amazon in order to extract gold from hillsides. Powerful, high-pressure streams of water are used to blast away soil and rock containing gold, which is then separated from the runoff. This process is very damaging to the environment, as entire hills are eroded away and streams become clogged with sediment. If land subjected to any of these surface mining techniques is not properly restored after its use, then it leaves an unsightly scar on the land and is highly susceptible to erosion.

## Sub-Surface Mining Techniques

Some mineral deposits are too deep to be surface mined and therefore require a **sub-surface mining** method. **Underground mining** is often used for higher-grade, more localized, or very concentrated resources. In the traditional sub surface method a deep vertical shaft is dug and tunnels are dug horizontally outward from the shaft into the ore body. The ore is removed and transported

to the surface. The deepest such subsurface mines (deeper than 3500 m) in the world are located in the Witwatersrand basin of South Africa, where gold is mined. This type of mining is less disturbing to the land surface than surface mining. It also usually produces fewer waste materials. However, it is more expensive and more dangerous than surface mining methods.

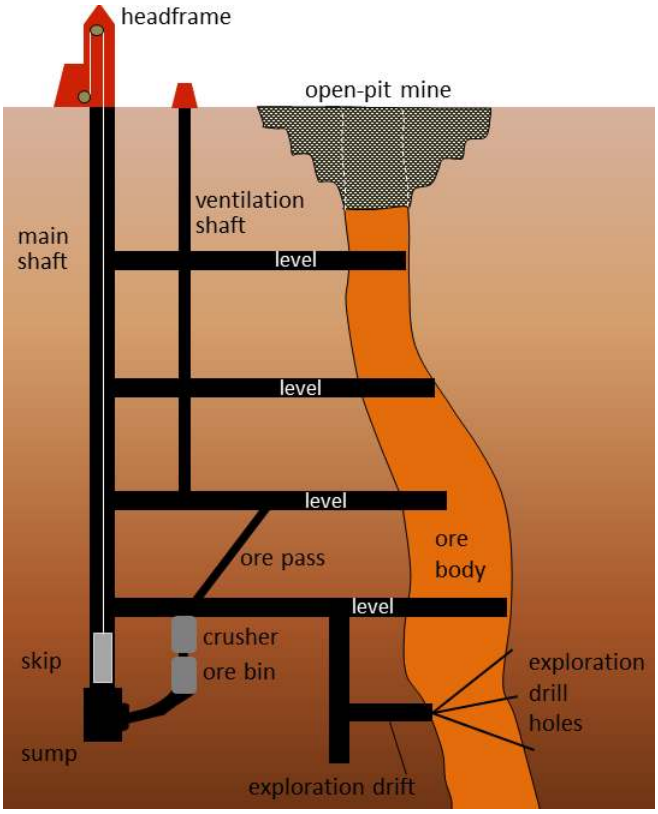



Figure 6.5.2.4: *Schematic cross-section of a typical underground mine.* © Steven Earle. CC BY



Figure 6.5.2.5: Underground mining in Estonia of Oil Shale.

A newer form of subsurface mining known as **in-situ mining** is designed to co-exist with other land uses, such as agriculture. An in-situ mine typically consists of a series of injection wells and recovery wells built with acid-resistant concrete and polyvinyl chloride casing. A weak acid solution is pumped into the ore body in order to dissolve the minerals. Then, the metal-rich solution is drawn up through the recovery wells for processing at a surface refining facility where the mineral of interest is precipitated. This method is used for the in-situ mining of copper ore.

## Ore Processing

Once an ore has been mined, it must be processed to extract pure metal. All ore minerals are mixed with less desirable components called **gangue**. The process of physically separating gangue minerals from ore-bearing minerals is called **concentrating**. When ore

is processed (typically very close to the mine), it is ground to a fine powder and the ore minerals are physically separated from the rest of the rock to make a **concentrate**. At a molybdenum mine, for example, this concentrate may be almost pure molybdenite ($MoS_2$). The rest of the rock is known as **tailings**. It comes out of the concentrator as a wet slurry and must be stored near the mine, in most cases, in a tailings pond. The tailings pond at the Myra Falls Mine on Vancouver Island is shown in Figure **6.5.2.6**, and the settling ponds for waste water from the concentrator are shown in Figure **6.5.2.7**. The tailings are contained by an embankment. Also visible in the foreground of Figure **6.5.2.6** is a pile of waste rock, which is non-ore rock that was mined in order to access the ore. Although this waste rock contains little or no ore minerals, at many mines it contains up to a few percent pyrite.

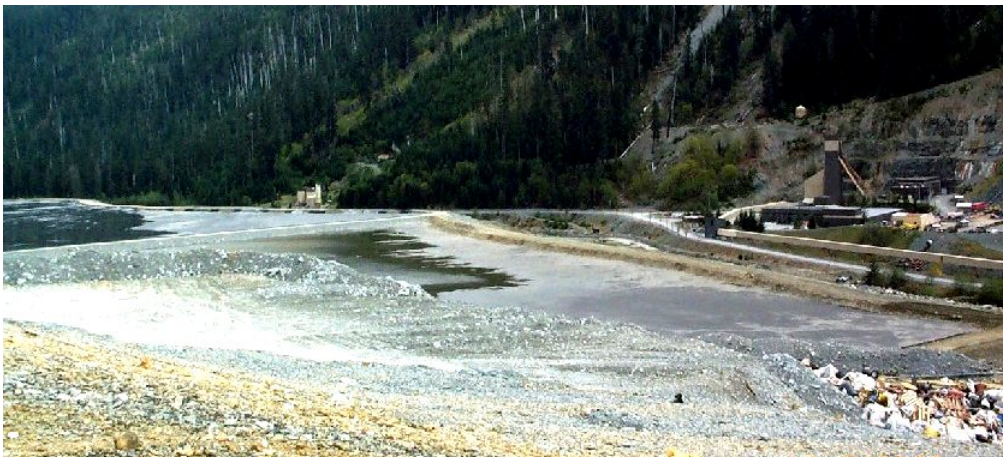


Figure **6.5.2.6**: *The tailings pond at the Myra Falls Mine on Vancouver Island. The dry rock in the middle of the image is waste rock. The structure on the right is the headframe for the mine shaft. Myra Creek flows between the tailings pond and the headframe.* © Steven Earle. CC BY

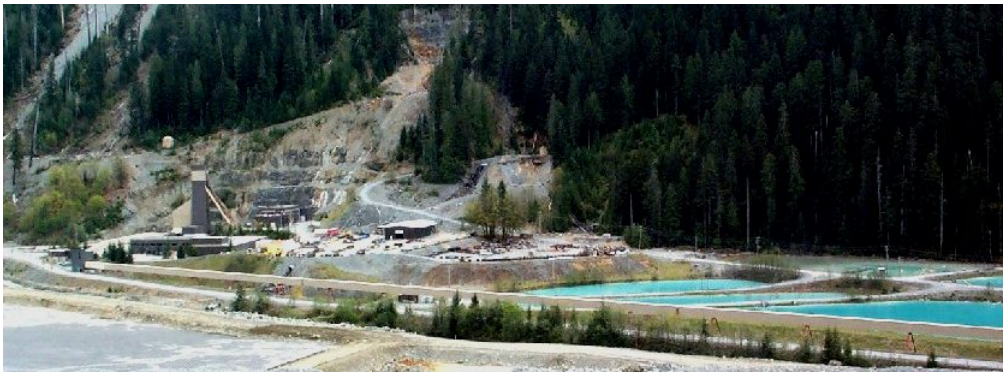


Figure **6.5.2.7**: *The tailings pond (lower left) at Myra Falls Mine with settling ponds (right) for processing water from the concentrator.* © Steven Earle. CC BY

Processes for extracting metal include smelting, electrowinning and heap leaching. In preparation for the **smelting** process, the ore is crushed and concentrated by a flotation method. The concentrated ore is melted in a smelting furnace where impurities are either burned-off as gas or separated as molten slag (Figure **6.5.2.8**). **Slag** is the glassy unwanted by-product of smelting ore and is sometimes mistaken for metorites. The smelting step is usually repeated several times to increase the purity of the metal.

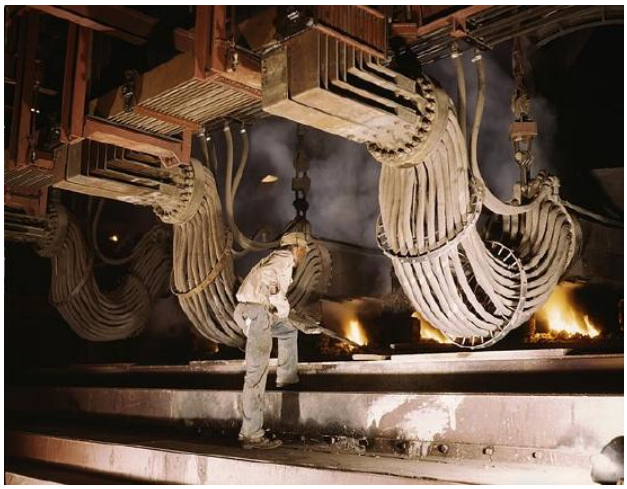


Figure 6.5.2.8: A phosphate smelting operation in Alabama, 1942.

For the **electrowinning** method ore or mine tailings are first leached with a weak acid solution to remove the desired metal. An electric current is passed through the solution and pure metal is electroplated onto a starter cathode made of the same metal. Copper can be refined from oxide ore by this method. In addition, copper metal initially produced by the smelting method can be purified further by using a similar electrolytic procedure.

Gold is sometimes extracted from ore by the **heap leaching** process. A large pile of crushed ore is sprayed with a cyanide solution. As the solution percolates through the ore it dissolves the gold. The solution is then collected and the gold extracted from it. All of the refining methods can damage the environment. Smelters produce large amounts of air pollution in the form of sulfur dioxide which leads to acid rain. Leaching methods can pollute streams with toxic chemicals that kill wildlife.

## Environmental Impacts of Metallic Mineral Mining

The primary impact of metallic mineral mining comes from the mining itself, including disturbance of the land surface, covering of landscapes by tailings impoundments, and increased mass wasting by accelerated erosion [48]. Surface mines can create enormous pits (Figure 6.5.2.2) in the ground as well as large piles of overburden and tailings that need to be **reclaimed**, i.e., restored to a useful landscape. Since 1977 surface mines in U.S. are required to be reclaimed, and commonly reclamation is relatively well done in this country. Unfortunately, surface mine reclamation is not done everywhere, especially in underdeveloped countries, due to lack of regulations or lax enforcement of regulations. Unreclaimed surface mines and active surface mines can be major sources of water and sediment pollution.

In addition, many metal deposits contain pyrite, an uneconomic sulfide mineral placed on waste dumps, which may generate **acid rock drainage** (ARD) during weathering (Figure 6.5.2.9). In the presence of oxygenated water, sulfides such as pyrite react undergo complex reactions to release metal ions and hydrogen ions, lowering pH to highly acidic levels. Mining and processing of mined materials typically increase the surface area to volume ratio in the material, causing reactions to occur even faster than what would occur naturally. If not managed properly, these reactions may lead to acidification of streams and groundwater plumes that can carry dissolved toxic metals. In mines where limestone is a waste rock of carbonate minerals like calcite or dolomite are present, their acid-neutralizing potential helps reduce the likelihood of generating ARD. Although this is a natural process too, it is very important to isolate mine dumps and tailings from oxygenated water, both to prevent the dissolution of sulfides and subsequent percolation of the sulfate-rich water into waterways. Industry has taken great strides in preventing contamination in recent decades, but earlier mining projects are still causing problems with local ecosystems.



Figure 6.5.2.9: **Acid Mine Drainage** The water in Rio Tinto River, Spain is highly acidic (pH = ~2) and the orange color is from iron in the water. A location along this river has been mined beginning some 5,000 years ago primarily for copper and more recently for silver and gold. *Source: Sean Mack of NASA via Wikimedia Commons*

Tailings ponds and waste-rock storage piles must be carefully maintained to ensure their integrity and monitored to ensure that acidic and metal-rich water is not leaking out. In August 2014, the tailings pond at the Mt. Polley Mine in central B.C. failed and 10 million cubic metres of waste water along with 4.5 million cubic metres of tailings slurry was released into Polley Lake, Hazeltine Creek, and Quesnel Lake (Figure 6.5.2.10 – 11). As of July 2015, the environmental implications of this event are still not fully understood.

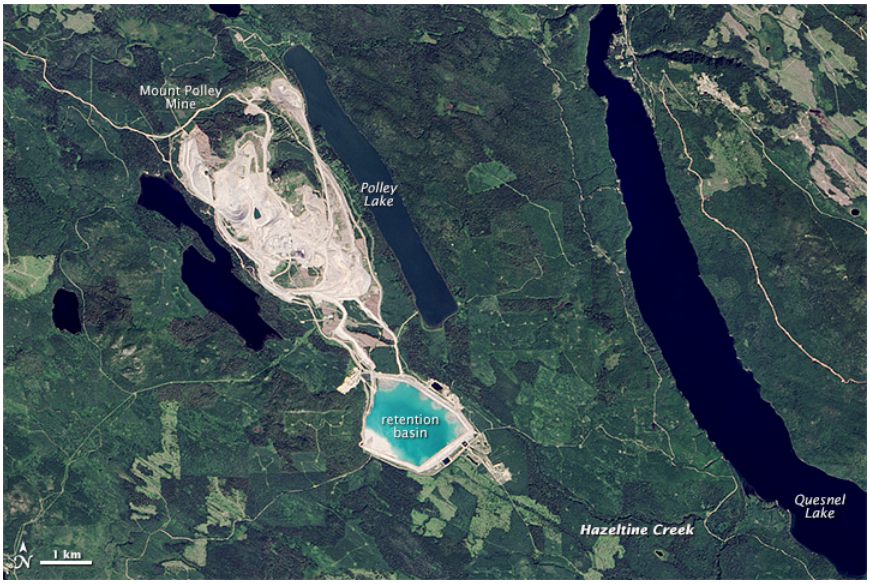



Figure 6.5.2.10: *The Mt. Polley Mine area prior to the dam breach of August 2014. The tailings were stored in the area labelled "retention basin." [https://en.Wikipedia.org/wiki/Mount_...mine_disaster]*




Figure 6.5.2.11: *The Mt. Polley Mine area after the tailings dam breach of August 2014. The water and tailings released flowed into Hazeltine Creek, and Polley and Quesnel Lakes. [https://en.Wikipedia.org/wiki/Mount_...mine_disaster]*

## MINERAL SUFFICIENCY AND THE FUTURE

Mineral resources are essential to life as we know it. A nation cannot be prosperous without a reliable source of minerals, and no country has all the mineral resources it requires. The United States has about 5 percent of the world's population and 7 percent of the world's land area, but uses about 30 percent of the world's mineral resources. It imports a large percentage of its minerals; in some cases sufficient quantities are unavailable in the U.S., and in others they are cheaper to buy from other countries. Certain minerals, particularly those that are primarily imported and considered of vital importance, are stockpiled by the United States in order to protect against embargoes or other political crises. These strategic minerals include: bauxite, chromium, cobalt, manganese and platinum.

Because minerals are produced slowly over geologic time scales, they are considered non-renewable resources. The estimated mineral deposits that are economically feasible to mine are known as mineral reserves. The growing use of mineral resources throughout the world raises the question of how long these reserves will last. Most minerals are in sufficient supply to last for many years, but a few (e.g. gold, silver, lead, tungsten and zinc) are expected to fall short of demand in the near future. Currently, reserves for a particular mineral usually increase as the price for that mineral increases. This is because the higher price makes it economically feasible to mine some previously unprofitable deposits, which then shifts these deposits to the reserves. However, in the long term this will not be the case because mineral deposits are ultimately finite.

There are ways to help prolong the life of known mineral reserves. Conservation is an obvious method for stretching reserves. If you use less, you need less. Recycling helps increase the amount of time a mineral or metal remains in use, which decreases the demand for new production. It also saves considerable energy, because manufacturing products from recycled metals (e.g. aluminum, copper) uses less energy than manufacturing them from raw materials. Government legislation that encourages conservation and recycling is also helpful. The current "General Mining Act of 1872," however, does just the opposite. It allows mining companies to purchase government land very inexpensively and not pay any royalties for minerals extracted from that land. As a result, mineral prices are kept artificially low which discourages conservation and recycling.

## References

4. Vogel, J. C. Groningen radiocarbon dates IX. *Radiocarbon* **12**, 444–471 (1970).

5. Schröder, K.-P. & Connon Smith, R. Distant future of the Sun and Earth revisited. *Mon. Not. R. Astron. Soc.* **386**, 155–163 (2008).

6. Duffield, W. A. Volcanoes, geothermal energy, and the environment. *Volcanoes and the Environment* 304 (2005).

7. Palmer, M. A. *et al.* Science and regulation. Mountaintop mining consequences. *Science* **327**, 148–149 (2010).

48. Nuss, P. & Eckelman, M. J. Life cycle assessment of metals: a scientific synthesis. *PLoS One* **9**, e101298 (2014).

Clugston, C. (2010) Increasing Global Nonrenewable Natural Resource Scarcity - An Analysis, The Oil Drum. Retrieved from http://www.theoildrum.com/node/6345

Craig J, Vaughan D, and Skinner B (2011) Earth Resources and the Environment (4th ed.). Pearson Prentice Hall, p. 92

## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- Metal Deposits from Physical Geology by Steven Earle (licensed under a Creative Commons Attribution 4.0 International License)
- Minerals from *AP Environmental Science* by University of California College Prep
- Non-Renewable Resources from *Environmental Science: A Canadian Perspective* by Bill Freedman (Creative Commons Attribution NonCommercial)
- Mineral Resources and Mining from *An Introduction to Geology* by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, & Cam Mosher (Geology Faculty at Salt Lake Community College), download free from OpenGeology
- Mineral Resources: Formation, Mining, Environmental Impact from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin

---

This page titled 6.5.2: Mining is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, & Cam Mosher (OpenGeology) .

# 6.5.3: Fossil Fuels- Formation and Mining

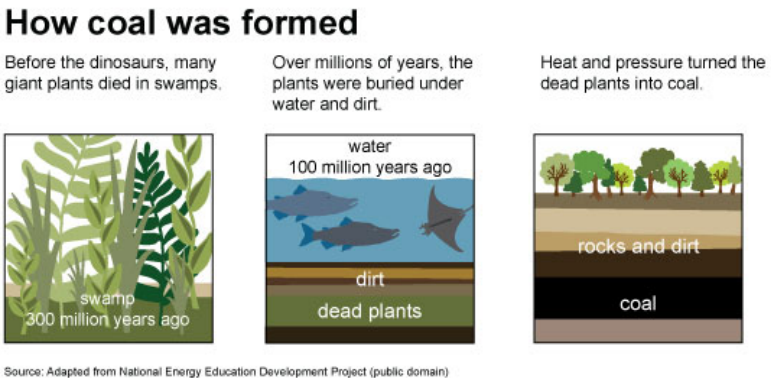
**Fossil fuels** are nonrenewable sources of energy formed from the organic matter of plants and microorganisms that lived millions of years ago. The natural resources that typically fall under this category are coal, oil (petroleum), and natural gas. This energy (and $CO_2$) was originally captured via photosynthesis by living organisms such as plants, algae, and photosynthetic bacteria. Sometimes this is known as fossil solar energy since the energy of the sun in the past has been converted into the chemical energy within a fossil fuel. Of course, as the energy is used, just like respiration from photosynthesis that occurs today, carbon can enter the atmosphere, causing climate consequences. Fossil fuels are nonrenewable because their formation took millions of years. Furthermore, higher productivity in the ancient environment allowed for more fossil fuel accumulation, meaning that the fossil fuel reserves available now could not necessarily be regenerated millions of years in the future.

The conversion of living organisms into fossil fuels is a complex process. As organisms die, most organic matter is oxidized back to $CO_2$ relatively quickly (within weeks or years in most cases), but any of it that gets isolated from the oxygen of the atmosphere (for example, deep in the ocean or in a stagnant bog) may last long enough to be buried by sediments and, if so, may be preserved for tens to hundreds of millions of years and the chemical energy within the organisms' tissues is added to surrounding geologic materials. Higher productivity in the ancient environment leads to a higher potential for fossil fuel accumulation, and there is some evidence of higher global biomass and productivity over geologic time [8]. Lack of oxygen and moderate temperatures seem to enhance the preservation of these organic substances [9; 10]. Heat and pressure that is applied after burial also can cause transformation into higher quality materials (brown coal to anthracite, oil to gas) and/or migration of mobile materials [11].

Fossil fuels are composed primarily of **hydrocarbons** (molecules of just carbon and hydrogen), but they contain lesser amounts nitrogen, sulfur, oxygen, and other elements as well. The precise chemical structures vary depending on the type of fossil fuel (coal, oil, or natural gas). The molecules in coal tend to be larger than those in oil and natural gas. Coal is thus solid at room temperature, oil is liquid, and natural gas is in a gaseous phase. Specifically, **coal** is a black or dark brown solid fossil fuel found as coal seams in rock layers formed from ancient swamp vegetation. Both oil and natural gas are fossil fuels found underground that formed from marine microorganisms. **Oil** (petroleum) is a liquid fossil fuel and consists of a variety of hydrocarbons while **natural gas** is a gaseous fossil fuel that consists of mostly methane and other small hydrocarbons.

## Coal

Coal, the first fossil fuel to be widely used, forms mostly on land in swampy areas adjacent to rivers and deltas in areas with humid tropical to temperate climates. Although some older coal deposits that predate terrestrial plants are presumed to come from algal buildups. Coal was formed when plant material is buried, heated, and compressed in oxygen-poor conditions over a long period of time (figure **6.5.3.1**). Millions of years ago, continents were in different locations with different climates, and swamp-like vegetation covered many regions. When the vegetation died, it could not fully decompose due to oxygen-poor conditions. Instead, it formed peat (a brown substance high in organic content). This situation, where the dead organic matter is submerged in oxygen-poor water, must be maintained for centuries to millennia in order for enough material to accumulate to form a thick layer. At some point, the peat was buried — typically because a river changes its course or sea level rises — and formed coal after millions of years of high pressure and temperature.



Source: Adapted from National Energy Education Development Project (public domain)

**Figure 6.5.3.1:** Coal was formed when large plants in swamps died 300 million years ago (before the dinosaurs). Over millions of years, this vegetation was buried under water and dirt (100 million years ago). Eventually, heat and pressure turned the dead plants into coal, which is found under layers of rock and dirt. Image from U.S. Energy Information Administration/National Energy Education Development Project (public domain).

There are several different types of coal ranging in quality (figure **6.5.3.2**). The more heat and pressure that coal undergoes during formation, the greater is its fuel value and the more desirable is the coal. The general sequence of a swamp turning into the various stages of coal are as follows:

Swamp → Peat → Lignite → Subbituminous coal → Bituminous coal → Anthracitic coal → Graphite

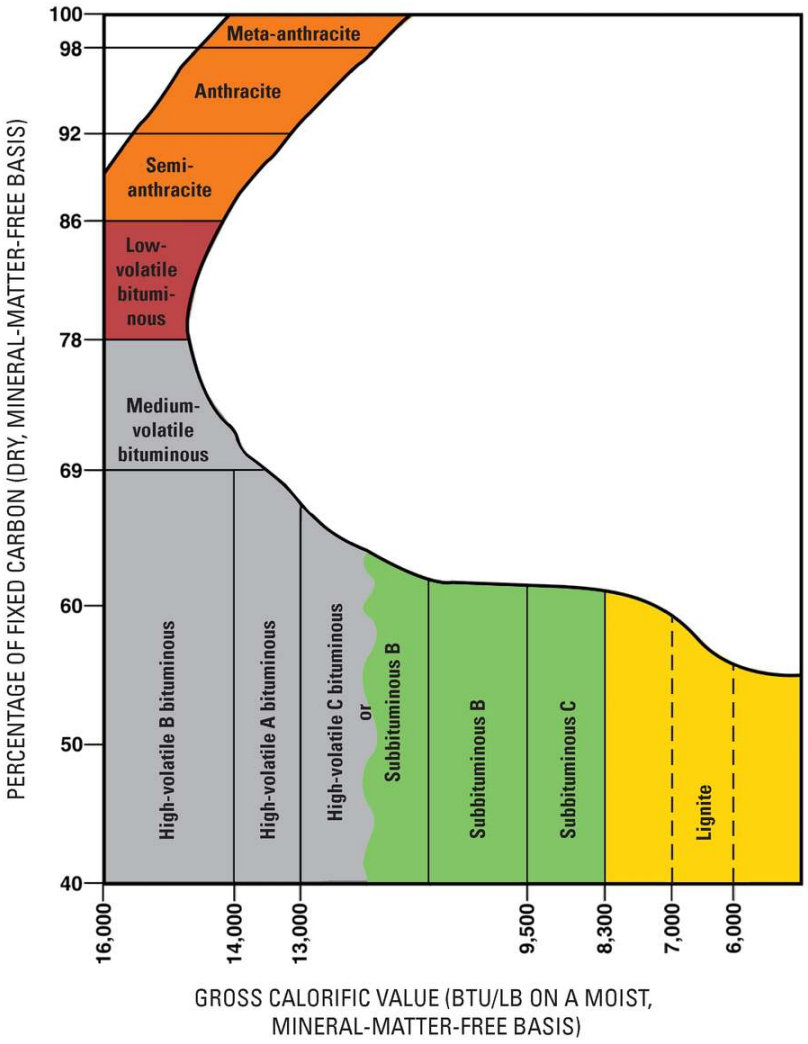



**Figure 6.5.3.2:** Coal rankings depend on energy content, measured as gross calorific value (how much energy is released from combustion) and carbon content that can be burned (percentage of fixed carbon). Anthracitic coal (orange) is the highest quality coal, with high energy and carbon content. Next in quality is bituminous coal (gray), subbituminous coal (green), and lignite (yellow). All three have less carbon content than anthracitic coal. Bituminous coal retains high energy content, but subbitminous coal and lignite have lower energy content. Image by USGS (public domain).

Specifically, peat compacts to form solid rock through a process called **lithification**, producing **lignite** (brown coal, a low-quality form of coal). With increasing heat and pressure, lignite turns to **subbituminous coal** and **bituminous coal**. Lignite, subbituminous coal, and bituminous coal are considered sedimentary rocks because they from from compacted sediments. At very high heat and pressure, bituminous coal is transformed to **anthracite**, a high-grade coal that is the most desirable coal since it provides the

highest energy output (figure 6.5.3.3). Anthracite is considered a metamorphic rock because it has been compacted and transformed to the extent that it is denser than the other forms of coal and no longer contains sheet-like layers of sediments. With even more heat and pressure driving out all the components that evaporate easily and leaving pure carbon, anthracite can turn to **graphite**.



**Figure 6.5.3.3:** Anthracitic coal, the highest grade of coal.

## Mining Coal

Coal is extracted by two principal methods, of which there are many variants: surface mining or subsurface mining. **Surface mining** uses large machines to remove the soil and layers of rock known as **overburden** to expose coal seams that are close to the Earth's surface (figure 6.5.3.4). **Strip mining** is a type of surface mining in which overburden is sequentially removed from each stretch (strip) of land. Once overburden is removed from the first strip, coal is removed. Overburden from the second strip is then deposited into the first strip, and coal is removed from the second strip. Overburden from the third strip is then placed in the first strip, and so on.



**Figure 6.5.3.4:** Surface mining of coal in Wyoming. Image by Bureau of Land Management (public domain).

**Mountaintop removal** is a more destructive type of surface mining in which all of the overburden is removed with explosives, revealing the entire coal seam at once (figure 6.5.3.5). The large mass of overburden (the mountaintop) is dumped into a nearby valley, and the coal is them removed.




**Figure 6.5.3.5:** Left: Mountaintop removal uses explosives to loosen overburden at the top of a mountain, which is then displaced into a nearby valley (valley fill). The underlying coal seam is then extracted. A sediment pond collects soil that erodes from the valley fill. Image by EPA (public domain). Right: A mountaintop removal site. Image by JW Randolph (public domain)

**Subsurface mining** (deep mining) employs underground tunnels to access deeper deposits (figure 6.5.3.6). Some underground mines are thousands of feet deep, and extend for miles. Miners ride elevators down deep mine shafts and travel on small trains in long tunnels to get to the coal. The miners use large machines that dig out the coal. In **drift mines**, a tunnel is dug horizontally into the side of a mountain. In **slope mines**, this tunnel is diagonal. In **shaft mines**, elevators are used to move coal through vertical tunnels.

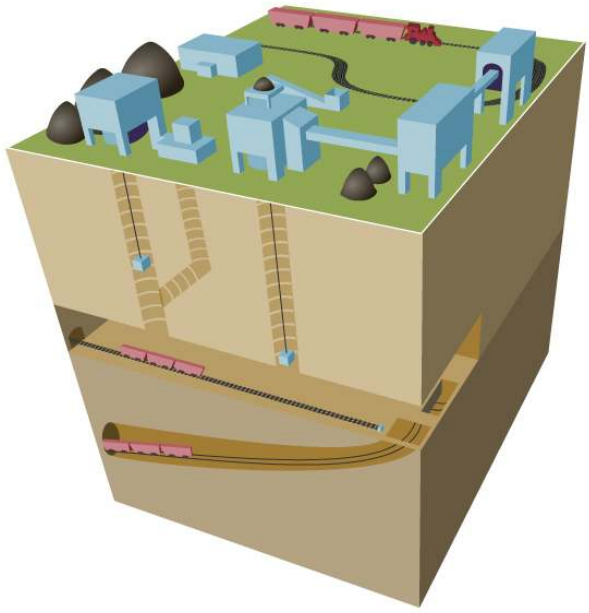


**Figure 6.5.3.6:** Subsurface mining involves digging tunnels to access coal deposits that are deep underground. Image by National Energy Education Development Project/EIA (public domain)

There are also significant health effects and risks to coal miners and those living in the vicinity of coal mines. Traditional underground mining is risky to mine workers due to the risk of entrapment or death. Over the last 15 years, the U.S. Mine Safety and Health Administration has published the number of mine worker fatalities and it has varied from 18-48 per year. Twenty-nine miners died on April 6, 2010 in an explosion at the Upper Big Branch coal mine in West Virginia, contributing to the uptick in

deaths between 2009 and 2010. In other countries, with less safety regulations, accidents occur more frequently. In May 2011, for example, three people died and 11 were trapped in a coalmine in Mexico for several days. There is also risk of getting **black lung disease** (pneumoconiosis). This is a disease of the lungs caused by the inhalation of coal dust over a long period of time**.** It causes coughing and shortness of breath. If exposure is stopped the outcome is good. However, the complicated form may cause shortness of breath that gets increasingly worse. **Mountaintop mining** (MTM), while less hazardous to workers, has particularly detrimental effects on land resources. MTM is a surface mining practice involving the removal of mountaintops to expose coal seams, and disposing of the associated mining waste in adjacent valleys. This form of mining is very damaging to the environment because it literally removes the tops of mountains, destroying the existing habitat. Additionally, the debris from MTM is dumped into valleys burying streams and other important habitat.



**Figure 6.5.3.7:** Mountaintop Removal Coal Mining in Martin County, Kentucky Photograph shows mountaintop coal removal mining in Martin County, Kentucky. Source: Flashdark.

## Oil and Gas

**Petroleum**, with the liquid component commonly called **oil** and gas component called **natural gas** (mostly made up of methane), is principally derived from organic-rich shallow marine sedimentary deposits [12] formed from ancient marine microorganisms (**plankton**) (Figure **6.5.3.8**). When plankton died, they were buried in sediments. As with coal, oxygen-poor conditions limited decomposition. As sediments continued to accumulate, the dead organisms were further buried. As the depth of burial increases, so does the temperature—due to the geothermal gradient—and gradually the organic matter within the sediments is converted to hydrocarbons. The first stage is the biological production (involving anaerobic bacteria) of methane. Most of this escapes back to the surface, but some is trapped in methane hydrates near the sea floor. At depths beyond about 2 km, and at temperatures ranging from 60° to 120°C, the organic matter is converted by chemical processes to oil. This depth and temperature range is known as the **oil window**. Beyond 120°C most of the organic matter is chemically converted to methane. (Some natural gas is also found associated with coal deposits (**coalbed methane**), consisting of methane produced during coal formation.)

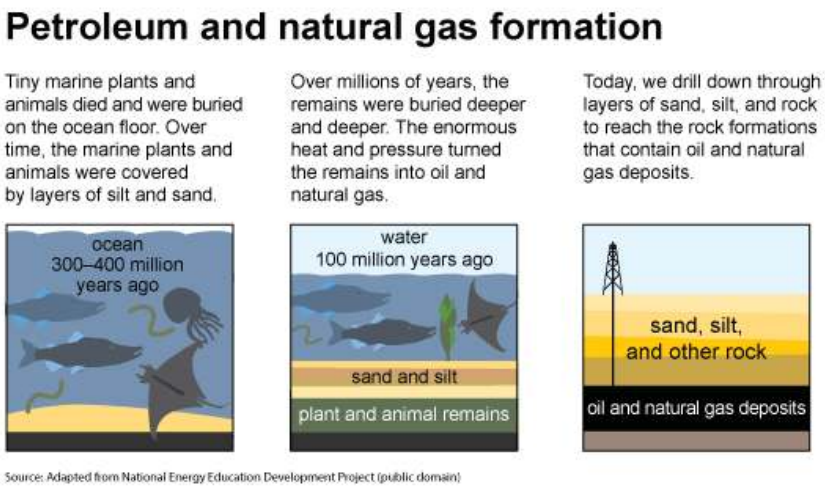



**Figure 6.5.3.8:** Petroleum (oil) and natural gas were formed from marine microorganisms. (The image text mentions tiny marine plants, but they were primarily algae and photosynthetic bacteria rather than plants.) These were covered by layers of silt and sand 300-400 million years ago. Over millions of years, the remains were buried deeper and deeper. They are pictured 100 million years ago. The enormous heat and pressure turned the remains into oil and natural gas. Now, oil and natural gas deposits are found underground and can be extracted via drilling through the layers of sand, silt, and rock. Image from U.S. Energy Information Administration/National Energy Education Development Project (public domain).

As the rock forms from the sediments that originally trapped the plankton, the oil and gas leak out of the **source rock** due to the increased pressure and temperature, and migrate to a different rock unit higher in the rock column. If the rock is porous and permeable rock, then that rock can act as a **reservoir** for the oil and gas. Petroleum is usually found one to two miles (1.6 – 3.2 km) below the Earth's surface, whether that is on land or ocean.

A **trap** is a combination of a subsurface geologic structure and an impervious layer that helps block the movement of oil and gas and concentrates it for later human extraction. Traps pool the fluid fossil fuels into a configuration in which extraction is more likely to be profitable, and such fossil fuels are called **conventional oil and natural gas** (figure **6.5.3.9**). Extraction of oil or gas outside of a trap (**unconventional oil and natural gas**) is less efficient and more expensive; sometimes it is not economically viable at all (does not produce a profit). Examples of unconventional fossil fuels include oil shale, tight oil and gas, tar sands (oil sands), and coalbed methane.

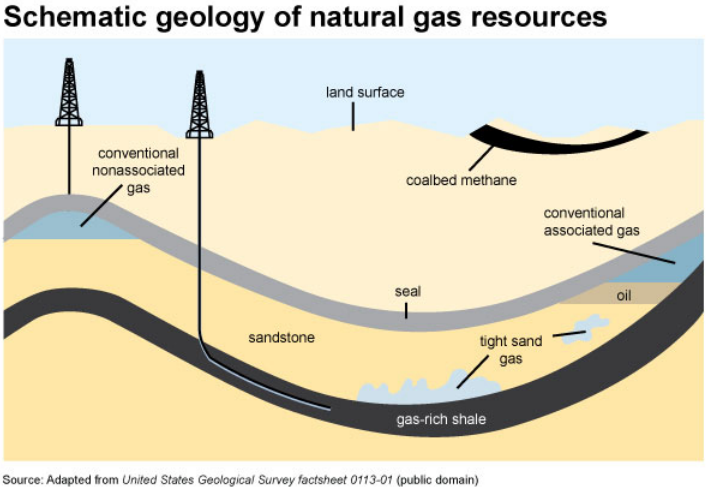



**Figure 6.5.3.9:** Conventional oil and natural gas deposits are trapped beneath impervious rock (gray). Conventional natural gas may be associated with oil or nonassociated. Coalbed methane and tight gas found in shale and sandstone are examples of unconventional fossil fuels. Image from USGS/EIA (public domain)

## Extraction of Conventional Oil and Natural Gas

Conventional oil and natural gas are contained under a trap (cap rock). Because natural gas consists of lighter molecules that are in a gaseous form at moderate temperatures, it is found on top of the oil, which may be floating on groundwater. To access conventional oil and natural gas, the trap is first pierced. Initially, they are under high enough pressure, and this drives them out of the well (**primary recovery**). Next, water (or gas) is injected to force more fossil fuels out (**secondary recovery**). Finally, enhanced oil recovery (**tertiary recovery**) may be used to extract further oil by applying heat (injecting steam) or injecting carbon dioxide, other gases, or larger molecules. For example, carbon dioxide causes the oil to thin and expand, making it easier to remove from the rocks. Note that secondary recovery simply increases pressure inside the reservoir whereas tertiary recovery changes the properties of the oil, making it easier to extract (figure **6.5.3.10**). Each stage of recovery is increasingly expensive, and extraction from a well continues as long as it remains profitable.

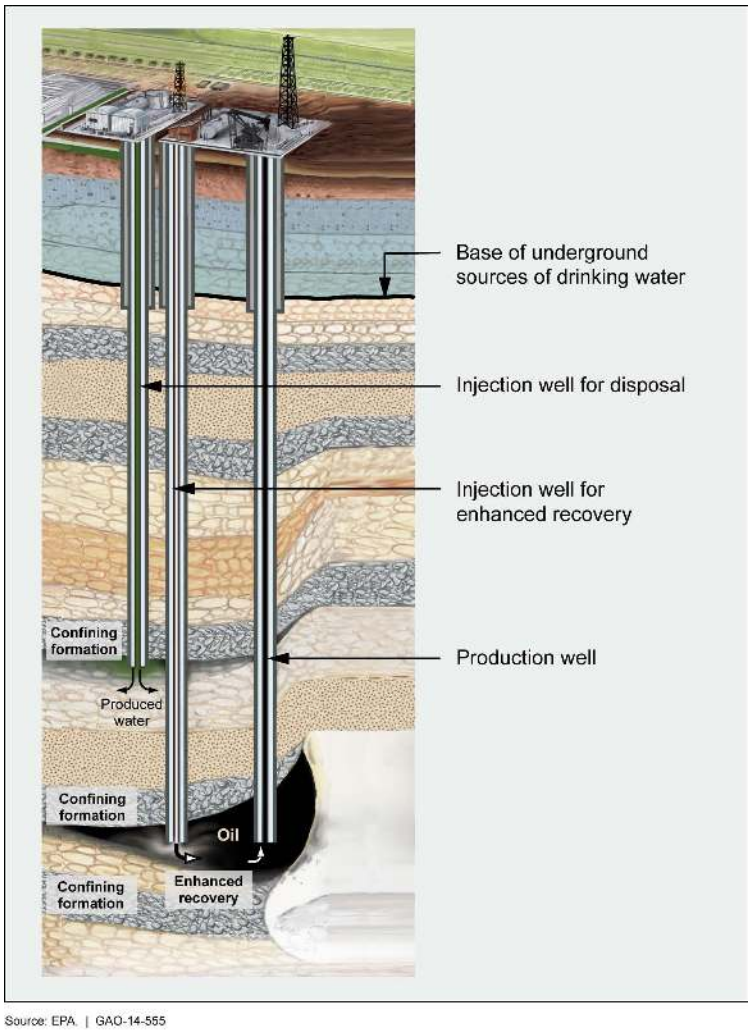



**Figure 6.5.3.10:** Injection wells transfer water, carbon dioxide, or other substances to an oil deposit increase pressure or change the oil's properties, facilitating extraction. On the right is the production well, through which extracted oil flows. This is facilitated by an injection well for enhanced recovery. On the left is an injection well for disposal through with waste water (produced water) is stored underground. Several confining formations trap substances underground. Closer to the top is the base of underground sources of drinking water, meaning that all drinking water is extracted from above this point. Image by Government Accountability Office (public domain).

Oil is mainly obtained by drilling either on land (onshore) or in the ocean (offshore). Early offshore drilling was generally limited to areas where the water was less than 300 feet deep. Oil and natural gas drilling rigs now operate in water as deep as two miles. Floating platforms are used for drilling in deeper waters (figure **6.5.3.11**). These self-propelled vessels are attached to the ocean

floor using large cables and anchors. Wells are drilled from these platforms which are also used to lower production equipment to the ocean floor. Some drilling platforms stand on stilt-like legs that are embedded in the ocean floor. These platforms hold all required drilling equipment as well as housing and storage areas for the work crews. Offshore production is much more expensive than land-based production.



**Figure 6.5.3.11:** A platform for offshore oil drilling. Image by Pixabay/keridjackson (Pixabay license).

## Extraction of Unconventional Oil and Natural Gas

Tight oil and natural gas trapped in shale (fine-grained sedimentary rocks with relatively high porosity and low permeability) as well as natural gas in tight sands (gas-bearing, fine-grained sandstones or carbonates with a low permeability) are extracted via **hydraulic fracturing**, informally referred to as “fracking”. This process uses explosives to create new fractures in these low-permeability rocks as well as increase the size, extent, and connectivity of existing fractures and then applies high-pressure fluid. First, a drill permeates the rock layers and then proceeds horizontally. Explosives then fracture rocks, freeing oil and natural gas. Finally, water, sand, and chemicals and injected, which flush out oil and natural gas (figure 6.5.3.12).

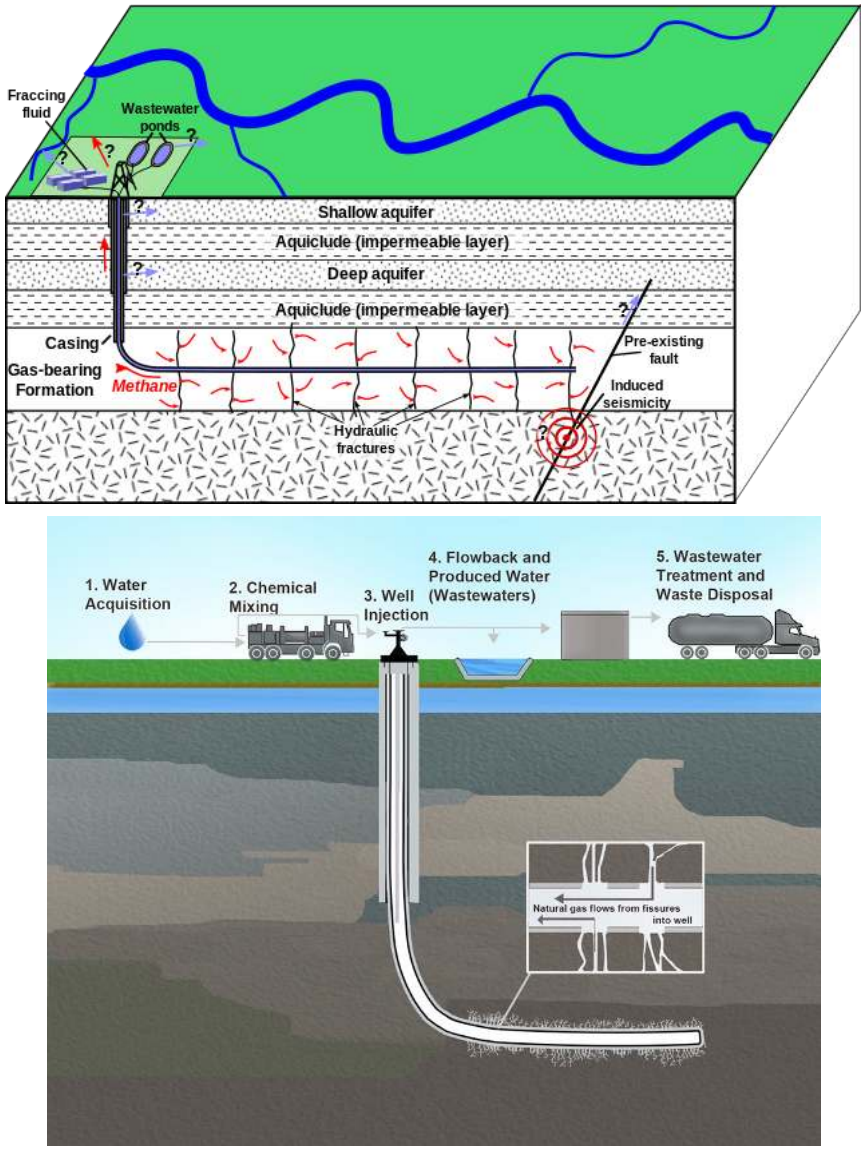



**Figure 6.5.3.12:** Two diagrams of fracking. Water is mixed with sand and chemicals and then injected into shale trapping tight oil and natural gas or tight sand. This flushes fossil fuels from the fissures that were previously created by explosives. Top: This section shows four underground layers. From top to bottom, they are the shallow aquifer, aquiclude (impermeable layer), deep aquifer, and another aquiclude. Fracking fluid is injected underground through a well surrounded by casing into the gas-bearing formation. Methane (red arrows) might escape from hydraulic fractures in this formation. Additionally, when the fractures intersect with a pre-existing fault, induced seismicity (earthquakes) is possible. Aboveground, fracking fluid is stored in wastewater ponds. Blue arrows and question marks indicate places where toxic wastewater may escape and contaminate groundwater such as from storage containers, wastewater ponds, casing, or faults. Bottom: Proper treatment and disposal of the wastewater is needed to limit environmental impact. 1: Water is acquired. 2: Chemicals are mixed. 3: Fracking fluid is injected into the well. Here, natural gas flows from fissures into the well. 4: Fracking results in flowback and produced water (wastewater). 5: Wastewater undergoes treatment and disposal. Bottom image by USGS (public domain).

Fracking causes more environmental damage than conventional extraction. The considerable use of water (figure **6.5.3.13**) may affect the availability of water for other uses in some regions, and this can affect aquatic habitats. In fact, fracking consumes more water than the use of nuclear energy, coal, or conventional oil and natural gas. If mismanaged, hydraulic fracturing fluid can be released by spills, leaks, or various other exposure pathways that contaminate land and groundwater (figure **6.5.3.13**). Fracking fluid flowback – the fluid pumped out of the well and separated from oil and gas – not only contains the chemical additives used in the drilling process but also contains heavy metals, radioactive materials (which release radiation), volatile organic compounds,

benzene (a carcinogen), toluene, ethylbenze, xylene, and other toxic air pollutants. Volatile organic compounds (VOCs) can react with the atmosphere to form ground-level ozone, which is associated with respiratory disease. Toulene can cause dizziness, confusion, headaches, and miscarriages. Ethylbenzene is a possible carcinogen that also causes dizziness, eye irritation, and hearing loss. Xylene also causes dizziness and headaches and furthermore can be fatal at high concentrations. In some cases, this contaminated water is sent to water treatment plants that are not equipped to deal with some of these classes of contamination. Finally, injecting wastewater for disposal can even induce earthquakes.

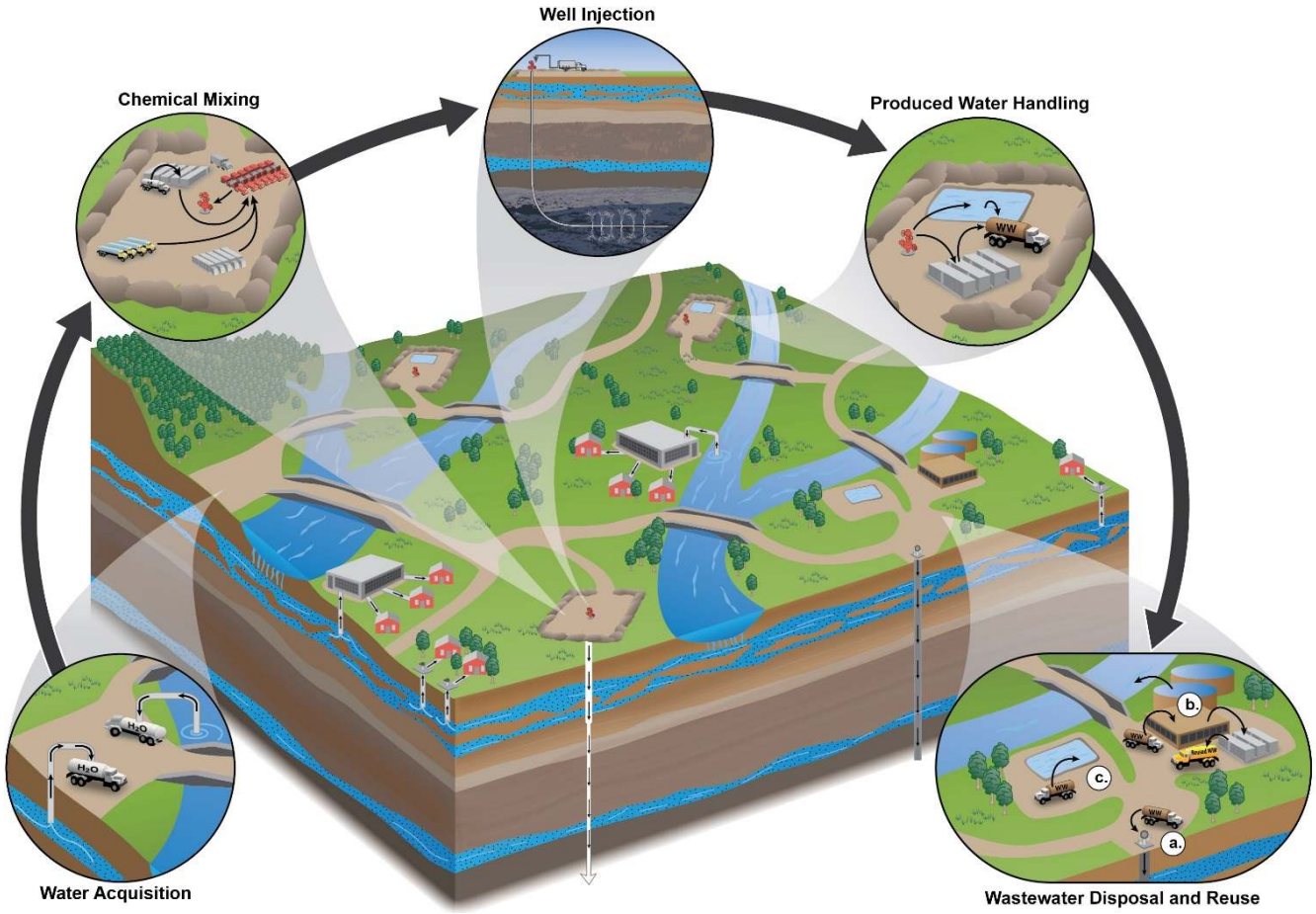



**Figure 6.5.3.13:** Fracking influences the water cycle. Water acquisition is the withdrawal of groundwater or surface water to make hydraulic fracturing fluids. Chemical mixing involves combining a base fluid, sand, and additives at the well site to create hydraulic fracturing fluids. During well injection, hydraulic fracturing fluids move through the oil and gas production well and in the targeted rock formation. Produced water handling refers to the onsite collection and handling of water that returns to the surface after hydraulic fracturing and the transportation of that water for disposal or reuse. Finally, the disposal and reuse of wastewater occurs. If this is not done properly, wastewater can pollute surrounding areas. Image and caption (modified) by EPA (public domain).

**Tar sands**, or oil sands, are sandstones that contain petroleum products that are highly viscous (like tar), and thus, can not be drilled and pumped out of the ground, unlike conventional oil (figure **6.5.3.14**). They are "unconventional" because the oil is exposed near the surface and is highly viscous because of microbial changes that have taken place at the surface. The fossil fuel in question is bitumen, which can be pumped as a fluid only at very low rates of recovery and only when heated or mixed with solvents. Thus, injections of steam and solvent or mining of the tar sands for later processing can be used to extract the tar from the sands. Tar sands can mined through strip mining or **open-pit mining**, a type of surface mining that involves forming a progressively deeper hole. The walls of the pit are as steep as can safely be managed. A steep wall means there is less waste overburden to remove and is an engineering balance between efficient mining and mass wasting. Alberta, Canada is known to have the largest reserves of tar sands in the world.



**Figure 6.5.3.14:** Tar sandstone from the Miocene Monterrey Formation of California.

The oil sands are controversial primarily because of the environmental cost of their extraction. Since the oil is so viscous, it requires heat to make it sufficiently liquid to process. This energy comes from gas; approximately 25 $m^3$ of gas is used to produce 0.16 $m^3$ (one barrel) of oil. (That's not quite as bad as it sounds, as the energy equivalent of the required gas is about 20% of the energy embodied in the produced oil.)

**Oil shale** (or tight oil) is a fine-grained sedimentary rock that sometimes contains **kerogen**, a solid material from which petroleum products can ultimately be manufactured. Shale is a common source of fossil fuels with high porosity but it has very low permeability. In order extract the fossil fuels, the material has to be mined and heated, which is expensive and typically has a negative impact on the environment. Oil shale is extracted by strip mining, creating subsurface mines, or open-pit mining. Oil shale can be burned directly like coal or baked in the presence of hydrogen to extract liquid petroleum (figure **6.5.3.15**). In order to get the oil out, the material has to be mined and heated, which, like with tar sands, is expensive and typically has a negative impact on the environment [17].




Figure **6.5.3.15**: Underground mining of oil shale in Estonia.

## References

8. Tappan, H. & Loeblich, A. R. Geobiologic Implications of Fossil Phytoplankton Evolution and Time-Space Distribution. *Geological Society of America Special Papers* **127**, 247–340 (1970).

9. Gordon, M., Jr, Tracey, J. I., Jr & Ellis, M. W. *Geology of the Arkansas bauxite region*. (1958).

10. Demaison, G. J. & Moore, G. T. Anoxic environments and oil source bed genesis. *Org. Geochem.* **2**, 9–31 (1980).

11. Tissot, B. Effects on prolific petroleum source rocks and major coal deposits caused by sea-level changes. *Nature* **277**, 463–465 (1979).

12. Pratt, W. E. *Oil in the Earth*. (University of Kansas Press, 1942).

13. Orton, E. *The Trenton Limestone as a Source of Petroleum and Inflammable Gas in Ohio and Indiana*. (U.S. Government Printing Office, 1889).

14. Dott, R. H. & Reynolds, M. J. Sourcebook for petroleum geology. (1969).

15. Vail, P. R. *et al.* Seismic stratigraphy and global sea level changes. *Seismic stratigraphy-applications to hydrocarbon exploration, edited by Payton, CE, Tulsa, American Association of Petroleum Geologists Memoir* **26**, 49–212 (1977).

16. Bauquis, P.-R. What future for extra heavy oil and bitumen: the Orinoco case. in **13**, 18 (1998).

17. Youngquist, W. Shale oil--The elusive energy. *Hubbert Center Newsletter* **4**, (1998).

18. Gandossi, L. An overview of hydraulic fracturing and other formation stimulation technologies for shale gas production. *Eur. Commisison Jt. Res. Cent. Tech. Reports* (2013).

19. Brown, V. J. Industry issues: Putting the heat on gas. *Environ. Health Perspect.* **115**, A76 (2007).

20. Kim, W.-Y. Induced seismicity associated with fluid injection into a deep well in Youngstown, Ohio. *J. Geophys. Res. [Solid Earth]* **118**, 3506–3518 (2013).

21. Taylor, E. L., Taylor, T. N. & Krings, M. *Paleobotany: The biology and evolution of fossil plants*. (Elsevier Science, 2009).

22. Mancuso, J. J. & Seavoy, R. E. Precambrian coal or anthraxolite; a source for graphite in high-grade schists and gneisses. *Econ. Geol.* **76**, 951–954 (1981).

23. Blander, M., Sinha, S., Pelton, A. & Eriksson, G. Calculations of the influence of additives on coal combustion deposits. *Argonne National Laboratory, Lemont, Illinois* 315 (2011).

24. Belloc, H. *The Servile State*. (T.N. Foulis, 1913).

25. McKenzie, H. & Moore, B. Social Origins of Dictatorship and Democracy. (1970).

26. Wrigley, E. A. *Continuity, Chance and Change: The Character of the Industrial Revolution in England*. (Cambridge University Press, 1990).

27. Quéré, C. L. *et al.* The global carbon budget 1959--2011. *Earth System Science Data* **5**, 165–185 (2013).

28. Barrie, L. A. & Hoff, R. M. The oxidation rate and residence time of sulphur dioxide in the Arctic atmosphere. *Atmos. Environ.* **18**, 2711–2722 (1984).

29. Crutzen, P. J. & Lelieveld, J. Human impacts on atmospheric chemistry. *Annu. Rev. Earth Planet. Sci.* **29**, 17–45 (2001).

## Contributors and Attributions

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

- Challenges and Impacts of Energy Use, Water Pollution, and Non-Renewable Energy Sources from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Shale Gas 101. Office of Fossil Energy. U.S. Department of Energy. Accessed 01-12-2021. (public domain)
- The Process of Unconventional Natural Gas Production. 2021. United States Environmental Protection Agency. Accessed 01-12-2021. (public domain)
- Non-Renewable Energy Sources from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Fossil Fuels and Mining from *An Introduction to Geology* by Chris Johnson et al. (licensed under CC-BY-NC-SA)
- Chapter 4: Non-Renewable Energy from *Introduction to Environmental Science: 2nd Edition* (2018) Biological Sciences Open Textbooks by Zehnder, Caralyn; Manoylov, Kalina; Mutiti, Samuel; Mutiti, Christine; VandeVoort, Allison; and Bennett, Donna (licensed under CC-BY-NC-SA).
- Fossil Energy Study Guide: Natural Gas. 2014. U.S. Department of Energy. Accessed 01-12-2021.
- Fossil Energy Study Guide: Oil. 2013. U.S. Department of Energy. Accessed 01-12-2021.
- Fossil Fuels from Physical Geology by Steven Earle (licensed under a Creative Commons Attribution 4.0 International License)
- Environmental Challenges in Energy, Carbon Dioxide, Air, Water and Land Use and Fossil Fuels from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors. Textbook content produced by Tom Theis and Jonathan (CC-BY). Download for free at CNX.

This page titled 6.5.3: Fossil Fuels- Formation and Mining is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 6.6: Soils

Soil is the outer loose layer that covers the surface of Earth. Soil quality is a major determinant, along with climate, of plant distribution and growth. Soil quality depends not only on the chemical composition of the soil, but also the topography (regional surface features) and the presence of living organisms. In agriculture, the history of the soil, such as the cultivating practices and previous crops, modify the characteristics and fertility of that soil.

Soil develops very slowly over long periods of time, and its formation results from natural and environmental forces acting on mineral, rock, and organic compounds. Soils can be divided into two groups: organic soils are those that are formed from sedimentation and primarily composed of organic matter, while those that are formed from the weathering of rocks and are primarily composed of inorganic material are called mineral soils. Mineral soils are predominant in terrestrial ecosystems, where soils may be covered by water for part of the year or exposed to the atmosphere.

## Importance of Soil

Soil is important to our society primarily because it provides the foundation of agriculture and forestry. Soil plays a key role in plant growth. Beneficial aspects to plants include providing physical support, water, heat, nutrients, and oxygen. Mineral nutrients from the soil can dissolve in water and then become available to plants. Soil plays a role in nearly all biogeochemical cycles on the Earth's surface. Global cycling of key elements such as carbon (C), nitrogen (N), phosphorous (P), and sulfur (S) all pass through soil. In the hydrologic (water) cycle, soil helps to mediate **infiltration** (percolating) from the surface into the groundwater. Microorganisms living in soil can also be important components of biogeochemical cycles through the action of decomposition and other processes such as nitrogen fixation.

Through their roots, plants absorb water and minerals (e.g., nitrates, phosphates, potassium, copper, zinc). Of the mineral nutrients absorbed from the soil, **macronutrients**, including nitrogen (N), potassium (K), calcium (Ca), magnesium (Mg), phosphorus (P), sulfur (S), and silicon (Si) are needed by plants in significant quantities. **Micronutrients** are essential elements that are needed only in small quantities, but can still be limiting to plant growth since these nutrients are not so abundant in nature. Micronutrients include chlorine (Cl), iron (Fe), boron (B), manganese (Mn), sodium (Na), zinc (Zn), copper (Cu), nickel (Ni), and molybdenum (Mo). There are some other elements that tend to aid plant growth but are not absolutely essential.

Micronutrients and macronutrients are desirable in particular concentrations and can be detrimental to plant growth when concentrations in soil solution are either too low (limiting) or too high (toxicity). Mineral nutrients are useful to plants only if they are in an extractable form in soil solutions, such as a dissolved ion rather than in solid mineral. Many nutrients move through the soil and into the root system as a result of concentration gradients, moving by diffusion from high to low concentrations. However, some nutrients are selectively absorbed by the root membranes, enabling concentrations to become higher inside the plant than in the soil.

Although many aspects of soil are beneficial to plants, excessively high levels of trace metals (either naturally occurring or added by humans) or applied herbicides can be toxic to some plants (figure **6.6.1**).



Figure **6.6.1**: These native plants are affected by herbicides that were used to kill nearby weeds. Image by Matt Lavin (CC-BY-SA).

An important factor affecting soil fertility is soil **pH** (the negative log of the hydrogen ion concentration). Soil pH is a measure of the acidity or alkalinity of the soil solution. On the pH scale (0 to 14) a value of seven represents a neutral solution; a value less than seven represents an **acidic** solution and a value greater than seven represents an **alkaline** solution (figure **6.6.2**). Soil pH affects the health of microorganisms in the soil and controls the availability of nutrients in the soil solution. Strongly acidic soils (less than 5.5) hinder the growth of bacteria that decompose organic matter in the soil. This results in a buildup of organic matter that has yet to be decomposed, which leaves important nutrients such as nitrogen in forms that are unusable by plants. Soil pH also affects the solubility of nutrient-bearing minerals. This is important because the nutrients must be dissolved in solution for plants to assimilate them through their roots. Most minerals are more soluble in slightly acidic soils than in neutral or slightly alkaline soils. Strongly acid soils (pH four to five), though, can result in high concentrations of aluminum, iron and manganese in the soil solution, which may inhibit the growth of some plants.

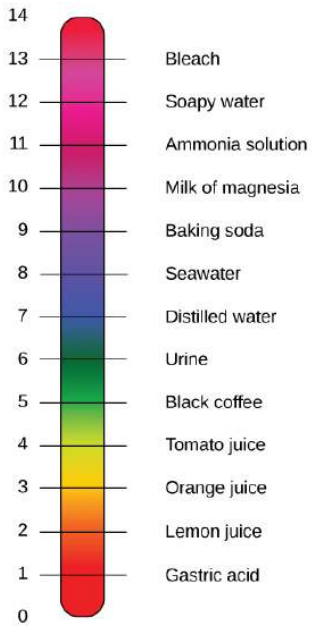



Figure **6.6.2**: The pH scale measures the amount of hydrogen ions ($H^+$) in a substance. This shows the pH of common household solutions. Those with a pH greater than 7 are basic (alkaline), and those with a pH less than 7 are acidic. The solutions from high to

low pH are bleach (13), soapy water (12), ammonia solution (11), milk of magnesia (10), baking soda (9), seawater (8), distilled water (7), urine (6), black coffee (5), tomato juice (4), orange juice (3), lemon juice (2), and gastric acid (1). Note that pH values are approximate. Image and caption (modified) from OpenStax/Edward Stevens (CC-BY). Access for free at openstax.org.

Several factors determine soil pH. Organic material in soil decreases pH to an extent, but it also acts as a **buffer**, limiting changes in pH. Climate is also important, with high amounts of rainfall increasing leaching and lowering pH. Some types of parent material, such as those high in silicon, decrease pH, while others, such as limestone increase pH.

# Soil Composition

Soil consists of organic matter (about 5%), inorganic mineral matter (40-45% of soil volume), water (about 25%) and air (about 25%) (Figure **6.6.3**). The amount of each of the four major components of soil depends on the amount of vegetation, soil compaction, and water present in the soil.

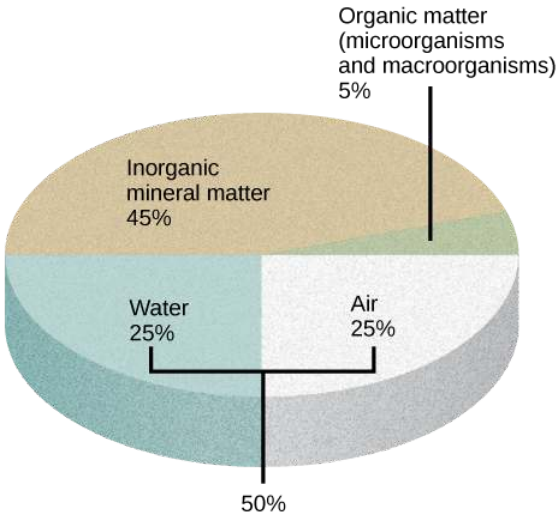



Figure **6.6.3**: The four major components of soil are shown: inorganic minerals, organic matter, water, and air.

The organic material consists of dead organisms in various stages of decomposition. It is dark-colored because it contains **humus**, partially decayed matter containing organic acids. Humus enriches the soil with nutrients, gives the soil a loose texture that holds water, and allows air to diffuse through it. Oxygen is important for plant roots and many inhabitants of the soil. The organic component of soil serves as a cementing agent, returns nutrients to the plant, allows soil to store moisture, makes soil tillable for farming, and provides energy for soil microorganisms. Most soil microorganisms—bacteria, algae, or fungi—are dormant in dry soil, but become active once moisture is available.

The inorganic material of soil consists of rock, slowly broken down into smaller particles that vary in size. Soil particles that are 100 µm to 2 mm in diameter are **sand**. (A **micrometer**, µm, $10^{-6}$ m, or a millionth of a meter.) Soil particles between 2 and 100 µm are called **silt**, and even smaller particles, less than 2 µm in diameter, are called **clay**.

Soil should ideally contain 50 percent solid material and 50 percent pore space (figure **6.6.4**). **Pore space** refers to the gaps in between soil particles. The larger the soil particles, the larger the pore spaces. Water can quickly pass through large pore spaces, so soils high in sand drain easily. Smaller soil particles have more surface area relative to volume and produce narrow pore spaces. Water clings to these surfaces, and soils high in clay thus retain water. (Clay is also negatively charged, which attracts water.) About one-half of the pore space should contain water, and the other half should contain air.

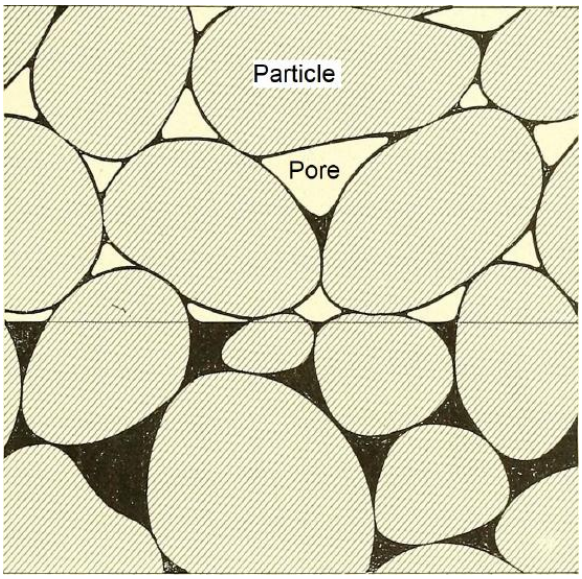



Figure **6.6.4**: Soil particles and the pore spaces in between them. The black shading represents water, which fills the pore spaces in the lower portion of the soil and clings to narrow portions of pore space in the top portion of the soil. Image from "Forest physiography; physiography of the United States and principles of soils in relation to forestry" (public domain).

Soil texture is based on percentages of sand, silt, and clay (figure **6.6.5**). Soils that have a high percentage of one particle size are named after that particle (a clay soil has a high percentage of clay). Other soils have a mixture of two particle sizes and very little of the third size. For example, silty clay has roughly 50% clay and 50% silt while sandy clay has 50-60% sand and 35-50% clay. Some soils have no dominant particle size and contain a mixture of sand, silt, and humus. These soils are called **loams**, and they are optimal for agriculture. A medium loam has roughly 40% sand, 40% silt, and 20% clay. Larger particles (sand) facilitate drainage, and small particles (clay) facilitate water retention, so loam soils both have good drainage and can remain moist. Soils that deviate slightly from a medium loam include loamy sand, sandy loam, sandy clay loam, clay loam, silty clay loam, and silty loam.

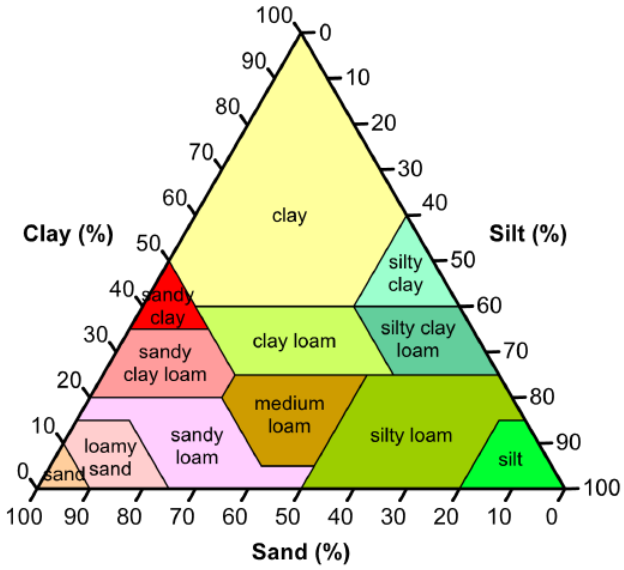



**Figure 6.6.5:** A soil texture triangle is used to determine the soil texture based on the percentages of sand, silt, and clay. The left axis represents the percentage of clay. The right axis represents the percentage of silt. The bottom axis represents the percentage of sand. Most soils that contain more than 40% are classified as clays. Exceptions are silty clays, which contain 40-60% clay and 40-60% silt. Another exception are sandy clays, which contain 35-50% clay and 50-70% sand. Clay loams and silty clay loams contain 25-40% clay. Clay loams contain 20-60% silt, and silty clay loams contain 60-75% silt. Sandy clay loams contain 20-35% clay and 50-80% sand. Medium loams contain 5-25% clay, 20-50% silt, and 30-55% sand. Soils with more than 50% silt and no more than

25% clay are either silty loams or silts. Silts have 80-100% silt and no more than 15% clay. Sand have 90-100% sand. Loamy sands have 75-90% sand and no more than 15% clay. A soil with a composition that does not fit into the already described categories would be a sandy loam. To determine the soil texture, first find the percentage of clay and trace a line horizontally to the right. At the same time, find the percentage of silt and trace a line diagonally (down and to the left, parallel to the clay axis). The two lines will meet at the correct soil type. For example, traced lines from 20% clay and 40% silt intersect at the medium loam category. To confirm, trace a diagonal line (down and to the right, parallel to the sit axis) from this point to the sand axis. This line cross the axis at 40%. In summary, a line can be traced from each axis that is parallel to axis that is counterclockwise of it. Image by Richard Wheeler/Zephyris at English Wikipedia (CC-BY-SA).

## Organic Versus Mineral Soils

Soils can be divided into two groups based on how they form. **Organic soils** are those that are formed from sedimentation and often contain more than 30% organic matter. They form when organic matter, such as leaf litter, is deposited more quickly than it can be decomposed (figure 6.6.6). **Mineral soils** are formed from the weathering of rocks, typically contain no more than 30% organic matter, and are primarily composed of inorganic material. **Weathering** occurs when biological, physical, and chemical processes, such as erosion, leaching, or high temperatures, break down rocks.



Figure 6.6.6): Decomposition of organic material occurs very slowly in this acidic bog. The organic material accumulates, which is characteristic of organic soils. Image by William L. Farr (CC-BY-SA).

## Soil Horizons

Soil distribution is not uniform because its formation results in the production of layers; together, the vertical section of a soil is called the **soil profile**. Within the soil profile, soil scientists define zones called horizons. A **horizon** is a soil layer with distinct physical and chemical properties that differ from those of other layers.

The soil profile has four distinct layers: 1) O horizon; 2) A horizon; 3) B horizon and 4) C horizon (figure 6.6.7 – 8). Upper horizons (labeled as the A and O horizons) are richer in organic material and so are important in plant growth, while deeper layers (such as the B and C horizons) retain more of the original features of the bedrock below. Some soils may have additional layers (like the E horizon, figure 6.6.7), or lack one of these layers. The thickness of the layers is also variable, and depends on the factors that influence soil formation. In general, immature soils may have O, A, and C horizons, whereas mature soils may display all of these, plus additional layers.

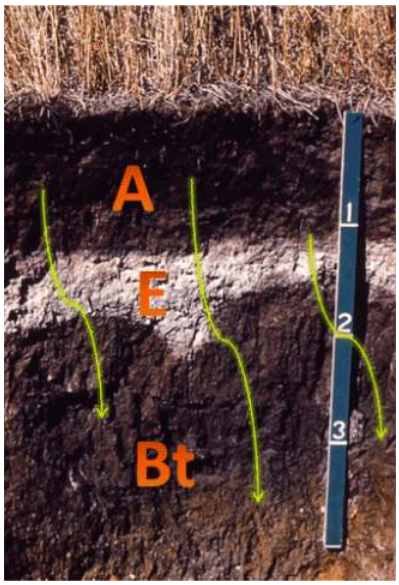



Figure 6.6.7): Soil Profile. Photograph shows a soil profile from South Dakota revealing three horizons. The A horizon (topsoil) is dark brown and extend about 1 foot. Below it is the white E (eluviated) horizon, which is about 6 inches thick. The B horizon (subsoil, labeled Bt) is the medium brown and the lowest visible layer The yellow arrows symbolize translocation of fine clays to the B horizon. The scale is in feet. Source: University of Idaho and modified by D. Grimley.

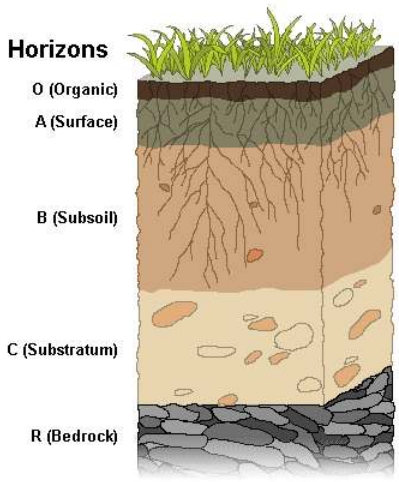



Figure 6.6.8): This image shows the different horizons, or layers, in soil. Vegetation grows at the surface. The top, deep brown layer is the O (organic) horizon. Below this is the gray A horizon (topsoil, or surface). Next is the light brown B horizon (subsoil). The C horizon (substratum, or soil base) is tan and includes larger pieces of rocks. The gray bedrock is the deepest layer. The E (eluviated) horizon that is sometimes found between the A and B horizons is absent here. This work by Wilsonbiggs is licensed under CC BY-SA 4.0

## O horizon

The very top of the **O horizon** consists of partially decayed organic debris like leaves (Figure 6.6.8). This horizon is usually dark in color because of humus.

## A horizon

The **A horizon** (topsoil) consists of a mixture of organic material with inorganic products of weathering, and it is therefore the beginning of true mineral soil (Figure 6.6.8). In this area, rainwater percolates through the soil and carries materials from the surface. The A horizon may be only 5 cm (2 in.), or it may over a meter. For instance, river deltas like the Mississippi River delta

have deep layers of topsoil. Microbial processes occur in the top soil, and this horizon supports plant growth. Many organisms, such as earthworms and insects live among the plant roots in this horizon.

### B horizon

The **B horizon** (subsoil) consists of small particles that have moved downward, resulting in a dense layer in the soil (Figure **6.6.8**). In some soils, the B horizon contains nodules or a layer of calcium carbonate. The subsoil is usually lighter in color than topsoil and often contains an accumulation of minerals.

### C horizon

The **C horizon** (soil base), includes the **parent material**, the organic and inorganic substances from which soils form (Figure **6.6.8**). Weathering parent material represents the first steps in the chemical breakdown of rock into soil. Often the weathered parent material is underlain by the parent material itself, although in some places it has been carried from another location by wind, water, or glaciers. Beneath the C horizon lies bedrock. The chemical nature of the parent material, whether granite, limestone, or sandstone, for example, has a great influence on the fertility of the soil derived from it.

## Factors Affecting Soil Formation and Composition

The fundamental factors that affect soil genesis can be categorized into five elements: climate, organisms, topography, parent material, and time. One could say that the relief, climate, and organisms dictate the local soil environment and act together to cause weathering and mixing of the soil parent material over time.

### Climate

The role of climate in soil development includes aspects of temperature and precipitation. Soils in very cold areas with permafrost conditions (such as the arctic tundra) tend to be shallow and weakly developed due to the short growing season. In warm, tropical climates, soils tend to be thicker (but lacking in organic matter), with extensive leaching and mineral alteration. In such climates, organic matter decomposition and chemical weathering occur at an accelerated rate. The presence of moisture and nutrients from weathering will also promote biological activity: a key component of a quality soil. See the Biomes chapter for more details about the effect of climate on soils.

Ancient soils, sometimes buried and preserved in the subsurface, are referred to as **paleosols** (figure **6.6.9**) and reflect past climatic and environmental conditions.

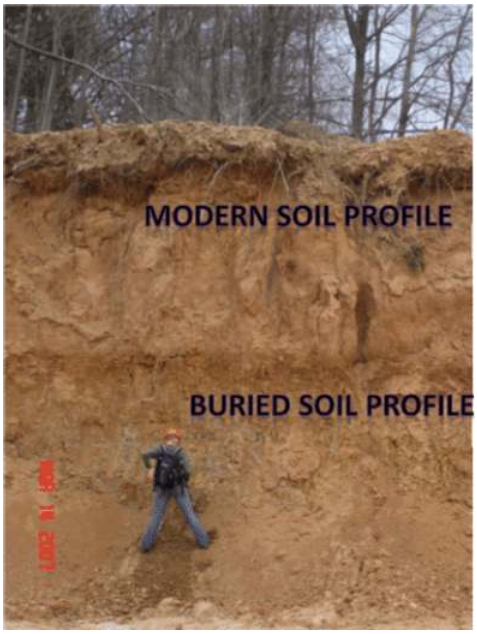



Figure **6.6.9**: Modern versus Buried Soil Profiles. A buried soil profile, or paleosol (above geologist ‘s head), represents soil development during the last interglacial period. A modern soil profile (Alfisol) occurs near the land surface. Source: D. Grimley.

## Organisms

The presence of living organisms in the soil (**soil biota**) greatly affects soil formation and structure. A diversity of animals are found in the soil such as nematodes, spiders, insects, centipedes, millipedes, pillbugs, slugs, and earthworms (figure **6.6.10**). The soil also contains microorganisms like bacteria, archaea, fungi, and "protists". Animals and microorganisms can produce pores and crevices, and plant roots can penetrate into crevices to produce more fragmentation. Additionally, leaves and other material that fall from plants decompose and contribute to soil composition. Microorganisms not only decompose organic matter, but contribute to other processes in nutrient cycles, such as nitrogen fixation.

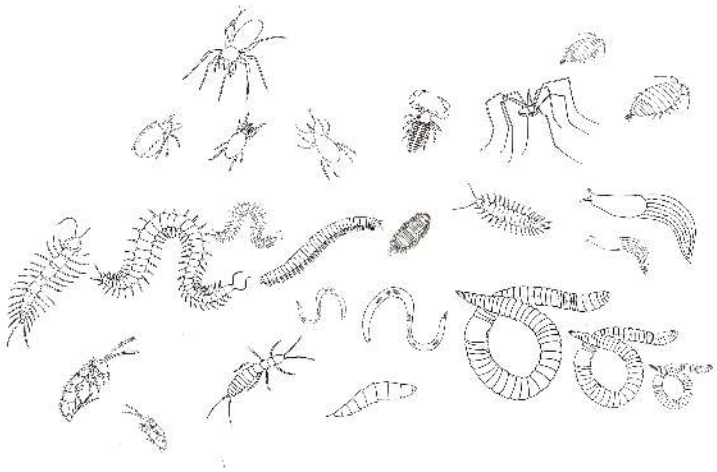


Figure **6.6.10**: Illustration of animals commonly found in the soil. Image by Arne Hendriks (CC-BY)

## Parent Material

Mineral soils form directly from the weathering of bedrock, the solid rock that lies beneath the soil, and therefore, they have a similar composition to the original rock. Other soils form in materials that came from elsewhere, such as sand and glacial drift. Materials located in the depth of the soil are relatively unchanged compared with the deposited material. Sediments in rivers may have different characteristics, depending on whether the stream moves quickly or slowly. A fast-moving river could have sediments of rocks and sand, whereas a slow-moving river could have fine-textured material, such as clay.

The type of parent material may also affect the rapidity of soil development. Parent materials that are highly weatherable (such as volcanic ash) will transform more quickly into highly developed soils, whereas parent materials that are quartz-rich, for example, will take longer to develop. Parent materials also provide nutrients to plants and can affect soil internal drainage.

## Topography

Regional surface features (familiarly called “the lay of the land”) can have a major influence on the characteristics and fertility of a soil. Topography affects water runoff, which strips away parent material and affects plant growth. Soils on steep slopes are more prone to erosion and may be thinner than soils that are on relatively level ground. **Infiltration**, the percolating of water through the soil, is limited in steep soils.

The local topography can have important microclimatic effects. In the northern hemisphere, south-facing slopes are exposed to more direct sunlight angles and are thus warmer and drier than north-facing slopes. The cooler, moister north-facing slopes have a more dynamic plant community and thicker soils because extensive root systems stabilize the soil and reduce erosion (figure **6.6.11**).



Figure **6.6.11**: A north-facing slope (left) and south-facing slope (right) in the Mediterranean climate (chaparral) of the Santa Monica Mountains, California. Vegetation on the north-facing slope is denser due to cool, moist conditions relative to the south-facing slope. Image by Noah Elhardt (CC-BY).

### Time

Time is an important factor in soil formation because soils develop over long periods. Soil formation is a dynamic process. Materials are deposited over time, decompose, and transform into other materials that can be used by living organisms or deposited onto the surface of the soil.

In general, soil profiles tend to become thicker (deeper), more developed, and more altered over time. However, the rate of change is greater for soils in youthful stages of development. The degree of soil alteration and deepening slows with time and at some point, after tens or hundreds of thousands of years, may approach an equilibrium condition where erosion and deepening (removals and additions) become balanced. **Young soils** (< 10,000 years old) are strongly influenced by parent material and typically develop horizons and character rapidly. Over time, as weathering processes deepen, mix, and alter the soil, the parent material becomes less recognizable as chemical, physical, and biological processes take effect. **Moderate age soils** (roughly 10,000 to 500,000 years old) are slowing in profile development and deepening, and may begin to approach equilibrium conditions. **Old soils** (>500,000 years old) have generally reached their limit as far as soil horizonation and physical structure, but may continue to alter chemically or mineralogically.

Soil development is not always continual. Geologic events such as landslides, glacier advance, or the rising of shorelines can rapidly bury soils. Erosion in rivers and shorelines can cause removal or truncation of soils, and wind or flooding slowly deposit sediment that adds to the soil. Animals can mix the soil and sometimes cause **soil regression**, a reversal or "bump in the road" for the normal path of development, and this increases development over time.

## Soil Taxonomy

Soils are classified into one of 12 soil orders based on soil horizons, how they form, and their chemical compositions. For example, Mollisols (figure **6.6.*f***), which are found in temperate grasslands, have a thick topsoil rich in organic content. Aridisols, on the other hand, are dry soils that contain calcium carbonate and are found in deserts. Each soil order is further divided into suborders. See USDA's The Twelve Orders of Soil Taxonomy and The Twelve Soil Orders from the University of Idaho for more details.

## Contributiors and Attributions

Modified by Kyle Whittinghill (University of Pittsburgh) and Melissa Ha from the following sources:

- The Soil from *General Biology* by OpenStax (CC-BY)
- Soil Profiles and Processes and Soil-Plant Interactions 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.

- Soil from *Biology* by John W. Kimball (licensed under CC-BY)

---

This page titled 6.6: Soils is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **31.2: The Soil** by OpenStax is licensed CC BY 4.0.

# CHAPTER OVERVIEW

## 7: Energy

Energy for lighting, heating and cooling our buildings, manufacturing products, and powering our transportation systems comes from a variety of natural sources. The earth's core provides geothermal energy. The gravitational pull of moon and sun create tides. The sun emits light (electromagnetic radiation), which creates wind, powers the water (hydrologic) cycle, and enables photosynthesis. Plants, algae, and cyanobacteria utilize solar energy to grow and create biomass that can be burned and used for **biofuels**, such as wood, biodiesel, bioethanol. Over the course of millions of years, biomass from photosynthetic organisms can create energy-rich **fossil fuels** through the geologic process of burial and transformation through heat and pressure.

Each of these types of energy can be defined as **renewable** or **non-renewable. Renewable energy** sources can be replenished within human lifespans. Examples include solar, wind, and biomass energy. **Nonrenewable energy** is finite and cannot be replenished within a human timescale. Examples include nuclear energy and fossil fuels, which take millions of years to form. All energy sources have some environmental and health costs, and the distribution of energy is not equally distributed among all nations. The environmental impacts of energy use on humans and the planet can happen anywhere during the life cycle of the energy source. The impacts begin with the extraction of the resource. They continue with the processing, purification or manufacture of the source; its transportation to place of energy generation, and ends with the disposal of waste generated during use.

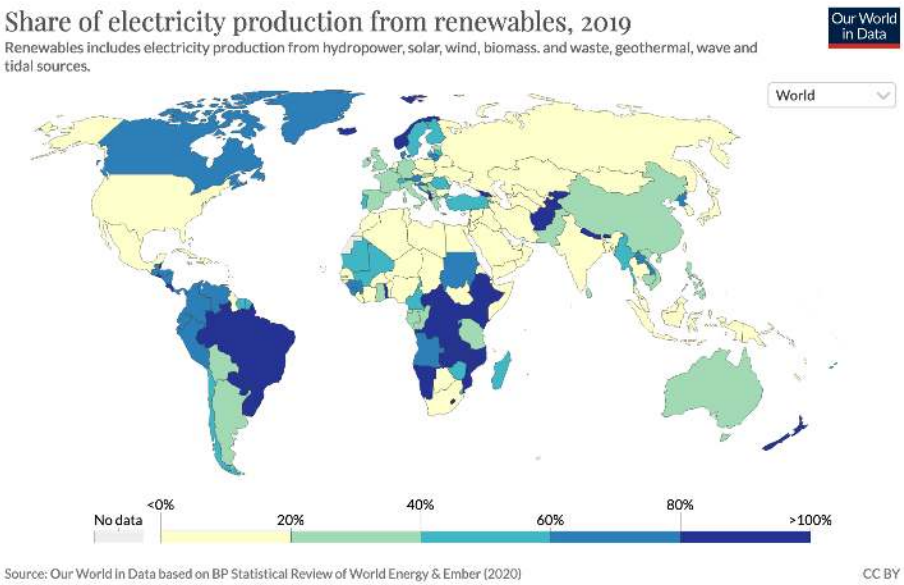



Figure **7.*a***: How much of the worlds energy comes from renewable sources. Graph by Our World in Data (CC-BY)

### Contributors and Attribution

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

- Challenges and Impacts of Energy Use 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


7.1: Energy from Fossil Fuels
7.2: Nuclear Energy
7.3: Solar Energy
7.4: Hydropower
7.5: Geothermal Energy

7.6: Wind Energy
7.7: Biofuels (Biomass Energy)
7.8: The Electrical Grid, Battery Challenges, and Energy Conservation

Thumbnail image - "Renewable energy on the grid" is in the Public Domain

---

This page titled 7: Energy is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

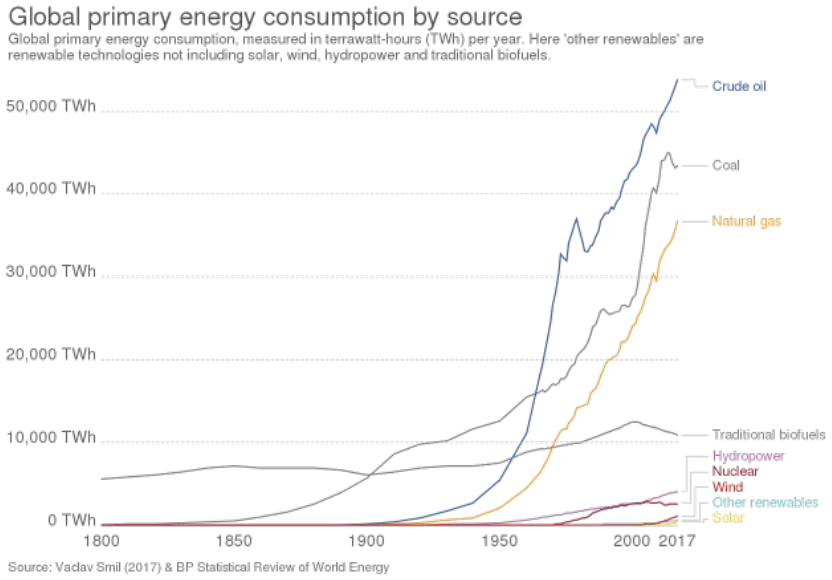
# 7.1: Energy from Fossil Fuels

The U.S. and the world overall heavily depend on fossil fuels. In 2019, fossil fuels contributed to 62.6% of electricity generation in the U.S. with coal contributing 23.4% and natural gas contributing 38.4% (table **7.1.1**). Note that oil (petroleum) is primary used for transportation and thus only contributes a fraction of a percentage to electricity generation. With respect to total energy consumption, the world continues to rely on crude oil more than any other energy source (33.1%) followed by coal (27%) and natural gas (24.3%; figure**7.1.1**).

**Table 7.1.1:** Contribution of Each Energy Source to Electricity Generation in the U.S. in 2019

| Energy source | Contribution to Electricity Generation (%) |
|---|---|
| **Fossil fuels (total)** | **62.6%** |
| Natural gas | 38.4% |
| Coal | 23.4% |
| Petroleum (total) | 0.4% |
| Other gases | 0.3% |
| **Nuclear** | **19.6%** |
| **Renewables (total)** | **17.6%** |
| Hydropower | 7.0% |
| Wind | 7.1% |
| Biomass (total) | 1.4% |
| Solar (total) | 1.7% |
| Geothermal | 0.4% |

Table modified from U.S. Energy Information Administration (public domain).



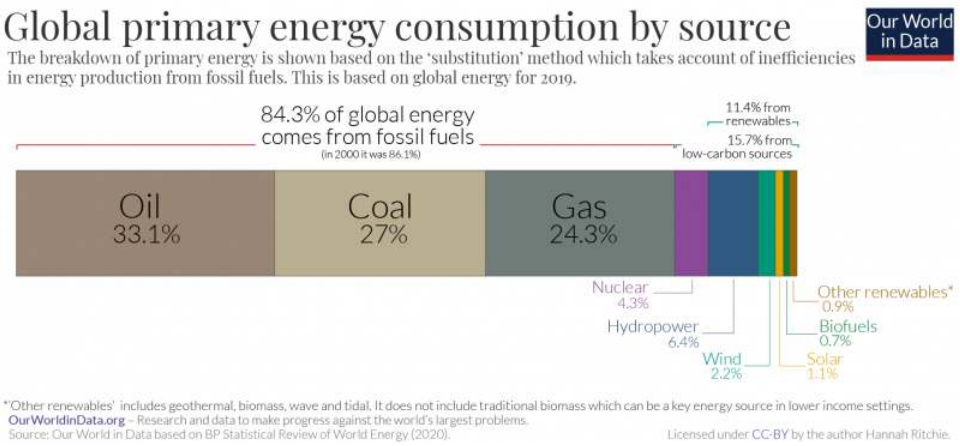



**Figure 7.1.1:** The top graph shows global energy consumption by source shows energy consumption in terrawatt-hours (TWh) on the y-axis and time in years (1800-2017) on the X-axis. The bottom graph shows the percentage that each energy source contributes to global energy consumption in 2019. Global consumption of most energy sources has increased over time. Consumption of crude oil is the highest (33.1% of global energy consumption in 2019), and it increased rapidly in the 1950s, surpassing coal consumption in the early 1960s. Coal is now the second most-consumed energy source (27%), but its global consumption has declined in recent years. Natural gas is the third most-consumed energy source (24.3%) and has also been increasing. Coal and natural gas consumption surpassed that of traditional biofuels in the early 1900s and 1970, respectively. Nuclear energy consumption has decreased recently, accounting for 4.3% of global energy consumption in 2019. Consumption of renewable energy overall (11.4%) is low compared to fossil fuels (84.3% combined), but it has generally increased in recent years. Images by Our World in Data (CC-BY).

## Generating Electricity from Coal

Coal has been used by humans for at least 6000 years, mainly as a fuel source. Coal resources in Wales are often cited as a primary reason for the rise of Britain (and later, the United States) in the Industrial Revolution. Coal electricity traces its origins to the early 20th Century, when it was the natural fuel for steam engines given its abundance, high energy density and low cost. Coal is the largest domestically produced source of energy. At the end of 2018, BP estimated coal reserves at 734,903 million tonnes, with nearly 23.7% of that in the United States. It is a major fuel resource that the United States controls domestically.

Once mined, coal may go to a preparation plant located near the mining site where it is cleaned and processed to remove impurities such as rocks and dirt, ash, sulfur, and other unwanted materials. This process increases the amount of energy that can be obtained from a unit of coal, known as its heating value.  Finally, the mined and processed coal must be transported. Transportation can be more expensive than mining the coal. Nearly 70% of coal delivered in the United States is transported, for at least part of its trip, by train (figure **7.1.2**). Coal can also be transported by barge, ship, or truck. Coal can also be crushed, mixed with water, and sent through a slurry pipeline. Sometimes, coal-fired electric power plants are built near coal mines to lower transportation costs.



**Figure 7.1.2:** A freight train loaded with coal briquettes, formed from compressed coal dust, in Morwell, Victoria, Australia. Image and caption (modified) by CSIRO (CC-BY).

Once at the power plant, coal is first pulverized into a fine powder and then mixed with hot air and blown into a furnace (figure **7.1.3**). This allows for the most complete combustion (burning) and maximum heat release. Purified water, pumped through pipes inside a boiler, is turned into steam by the heat from the combustion of coal. The high pressure of the steam pushing against a series of giant turbine blades turns the turbine shaft. The turbine shaft is connected to the shaft of the generator, where magnets spin within wire coils to produce electricity. After doing its work in the turbine, the steam is drawn into a condenser, a large chamber in the basement of the power plant. In this important step, millions of gallons of cool water from a nearby source (such as a river or lake) are pumped through a network of tubes running through the condenser. The cool water in the tubes converts the steam back into water that can be used over and over again in the plant. The cooling water is returned to its source without any contamination except at a higher temperature than when first extracted from the river or lake.

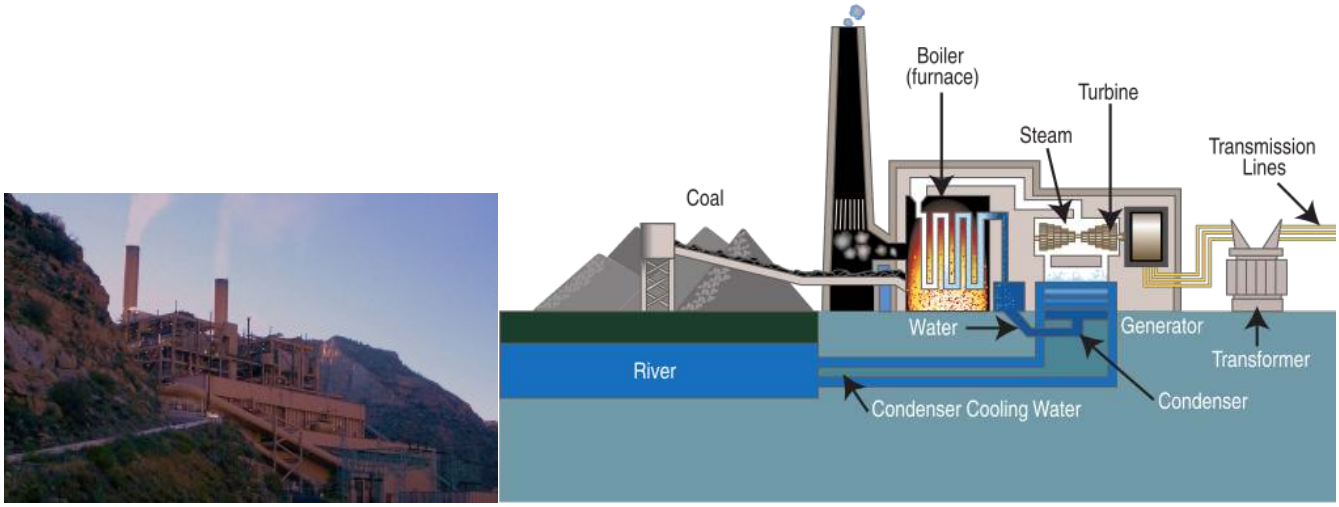



**Figure 7.1.3:** Left: Coal power plant in Helper, Utah. Right: Diagram of a typical steam-cycle coal power plant (proceeding from left to right). Coal enters the boiler (furnace), where it is burned. This boils water, producing steam. The steam turns a turbine, powering a generator. The resultant electricity is passed through a transformer (which changes the voltage) and is then sent through transmission lines. The steam cools and condenses back to liquid water in the condenser. This is facilitator by cooling water from a nearby river. Right image by Tennessee Valley Authority (public domain).

Coal is plentiful and inexpensive, when looking only at the market cost relative to the cost of other sources of electricity, but its extraction, transportation, and use produces a multitude of environmental impacts that the market cost does not truly represent. Coal emits sulfur dioxide, nitrogen oxide, and mercury, which have been linked to acid rain, smog, and health issues. Burning of

coal emits higher amounts of carbon dioxide per unit of energy than the use of oil or natural gas. Coal accounted for 35% of the total United States emissions of carbon dioxide released into the Earth's atmosphere in 2010. Ash generated from combustion contributes to water contamination.

This video shows how heat energy can be used to generate electricity.



## Generating Electricity from Oil and Natural Gas

**Proven oil reserves** (or natural gas reserves) refers to the amount of oil or natural gas that can be extracted economically with current methods (such as conventional wells or fracking). The U.S. Energy Administration estimates that there are enough liquid fuels to last through 2050 (and they include biofuels in this projection). In 2016, BP projected that proven reserves of oil and natural gas can support global demands for another 50 years. Additionally, they estimate that coal reserves can last another 115 years.

Scientists and policy-makers often discuss the question of when the world will reach **peak oil production**, the point at which oil production is at its greatest and then declines. It was initially predicted that global peak oil would be reached in 2000, but oil production and consumption continue to rise. The United States, however, already passed peak oil production in 1970.

The concentration of oil reserves in a few regions of the world makes much of the world dependent on imported energy for transportation (figure**7.1.4**). The rise in the price of oil in the last decade makes dependence on imported energy for transportation an economic as well as an energy issue. The United States spent $304.9 billion on oil imports in 2019. The United States has become more and more dependent on foreign oil since 1970 when our own oil production peaked.

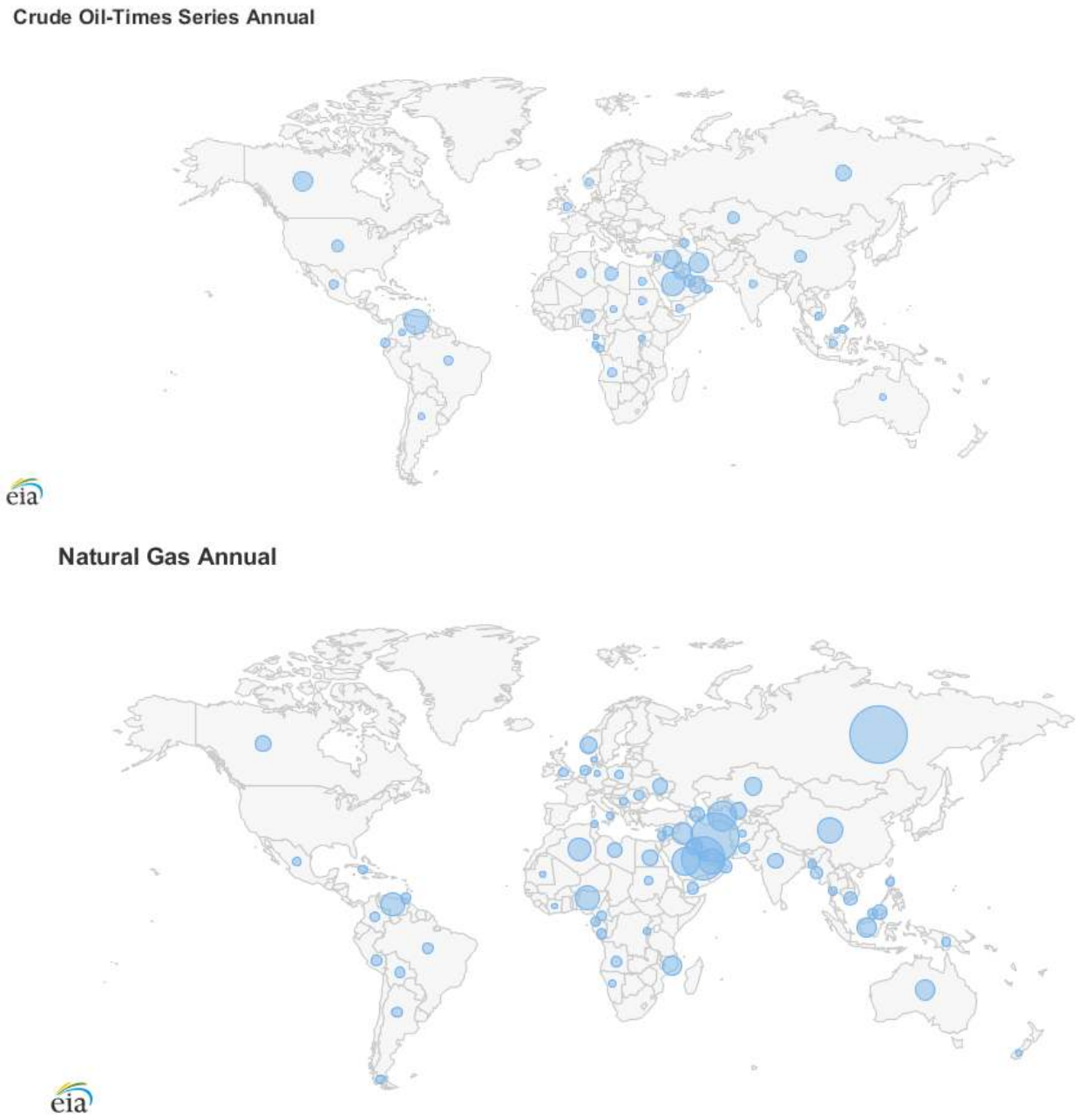



**Figure 7.1.4:** Global reserves of oil (top) and natural gas (bottom) by country in 2017 and 2020, respectively. The blue circle in each country is scaled to the size of the reserve. Both fossil fuels are abundant in the Middle East. Russia also has a rich supply of natural gas. Images from EIA (public domain).

The major holder of oil reserves is the **Organization of Petroleum Exporting Countries,** (OPEC). As of 2018, there were 15 member countries in OPEC: Algeria, Angola, Congo, Ecuador, Equatorial Guinea, Gabon, Iran, Iraq, Kuwait, Libya, Nigeria, Qatar, Saudi Arabia, the United Arab Emirates, and Venezuela. OPEC attempts to influence the amount of oil available to the world by assigning a production quota to each member except Iraq, for which no quota is presently set.

Overall compliance with these quotas is mixed since the individual countries make the actual production decisions. All of these countries have a national oil company but also allow international oil companies to operate within their borders. They can restrict the amounts of production by those oil companies. Therefore, the OPEC countries have a large influence on how much of world demand is met by OPEC and non-OPEC supply.

This pressure has lead the United States to developing policies that would reduce reliance on foreign oil such as developing additional domestic sources and obtaining it from non-Middle Eastern countries such as Canada, Mexico, Venezuela, and Nigeria. However, since fossil fuel reserves create jobs and provide dividends to investors, a lot is at stake in a nation that has oil reserves. Oil wealth may be shared with the country's inhabitants or retained by the oil companies and dictatorships, such as in Nigeria prior to the 1990s.

### Natural Gas Power Plants

Natural gas is burned to produce electricity following the same general process used in a coal power plant (figure **7.1.5**). Oil is occasionally used to generate electricity as well.

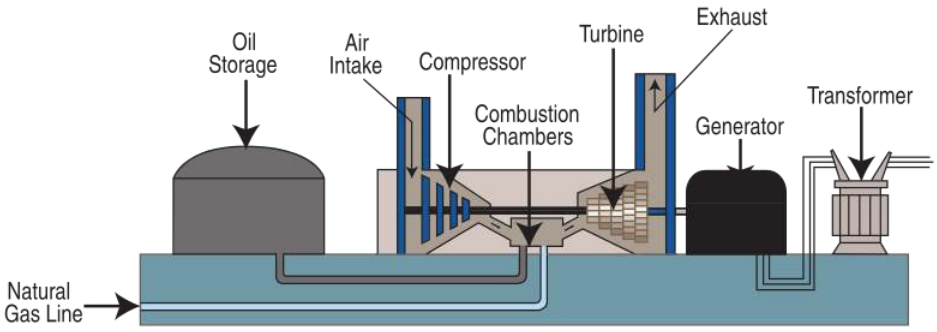



**Figure 7.1.5:** This combustion chamber burns either natural gas or oil. Fuel flows through a natural gas line or from oil storage into the combustion chamber. Air passes through the air intake and is compressed in the compressor. Natural gas and compressed air are mixed with compressed air in the combustion chamber and burned. High-pressure combustion gases spin the turbine, which drives the generator. The resultant electric current is then passed through a transformer, which alters voltage. Image and caption (modified) by Tennessee Valley Authority (public domain).

## The Global Dependence of Transportation on Oil

Two-thirds of oil consumption is devoted to transportation, providing fuel for cars, trucks, trains and airplanes. For the United States and most developed societies, transportation is woven into the fabric of our lives, a necessity as central to daily operations as food or shelter. The concentration of oil reserves in a few regions or the world makes much of the world dependent on imported energy for transportation. The rise in the price of oil in the last decade makes dependence on imported energy for transportation an economic as well as an energy issue. The United States, for example, now spends upwards of $350 billion annually on imported oil, a drain of economic resources that could be used to stimulate growth, create jobs, build infrastructure and promote social advances at home.

### Refining Crude Oil

The result of oil recovery is crude oil (petroleum), which contains many types of hydrocarbons as well as some unwanted substances such as sulfur, nitrogen, oxygen, dissolved metals, and water all mixed together. Unprocessed crude oil is therefore, not generally useful in industrial applications and must first be separated into different useable products (**petrochemicals**) at a refinery. Gasoline (petrol), diesel, tar, and asphalt are examples of petrochemicals.

**Fractional distillation** is the key process used in oil refineries to separate the components of crude oil. During fractional distillation, crude oil is heated and then allowed to cool. The heaviest compounds sink to the bottom as residues. Components of vaporized crude oil condense at different levels in the distillation column depending on their boiling points, which is primarily due to their molecular sizes. The heaviest compounds (condense near the bottom of the column, where the temperature is still high. Lighter compounds condense at cooler temperatures higher up in the column. Some compounds remain as gases at the top of the column (figure **7.1.6**).

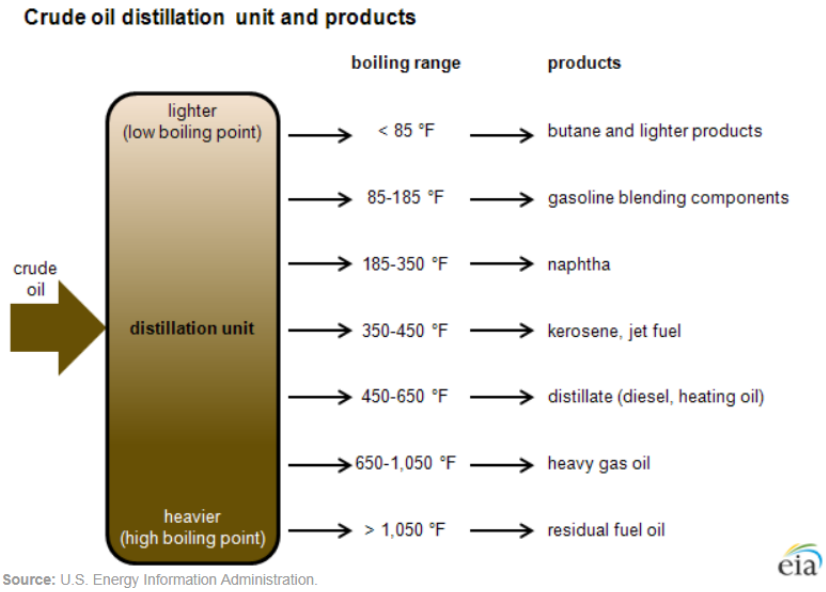



**Figure 7.1.6:** The process of fractional distillation involves heating crude oil and allowing the components to cool. As they move up the distillation column, they condense at different levels based on their boiling points. The heaviest compounds have the highest boiling points and condense at the bottom of the column while lighter compounds (low boiling point) condense at the top. From lowest to highest boiling point the petrochemicals produced are residual fuel oil (1050 °F), heavy gas oil (650-1050 °F), diesel and heating oil (450-650 °F), kerosene and jet fuel (350-450 °F), naphtha (used to make gasoline, solvents, cleaning solutions, etc.; 185-350 °F), gasoline blending components (85-185 °F), and butane and lighter products (< 85 °F). Image by U.S. Energy Information Administration (public domain).

The video below explains the process of fractional distillation. The labeled distillation column at 3:00 shows heated crude oil (400 °C) separating into various petrochemicals. From bottom to top, they are bitumen (> 350 °C), diesel (250-350 °C), kerosene (160-250 °C), naphtha (70-160 °C), petrol (20-70 °C), and gas (< 20 °C).



**Conversion** is the chemical processing in which some of the fractions (produced from fractional distillation) are transformed in to other products. For example, a refinery can turn diesel fuel into gasoline depending on the demand for gasoline. Conversion can involve breaking larger hydrocarbon chains into smaller ones (cracking), combining smaller chains into larger ones (unification), or rearranging the molecules to created desired products (alteration).

**Treatment** is done to the fractions to remove impurities such as sulfur, nitrogen and water among others. Refineries also combine the various fractions (processed and unprocessed) into mixtures to make desired products. For example, different mixtures of hydrocarbon chains can create gasolines with different octane ratings, with and without additives, lubricating oils of various

weights and grades (WD-40, 10W-40, 5W-30, etc.), heating oil, and many others. The products are stored onsite until they can be delivered to various markets such as gas stations, airports and chemical plants.

A 42 U.S. gallon barrel of crude oil yields about 45 gallons of petroleum products because of refinery processing gain (figure **7.1.7**). This increase in volume is similar to what happens to popcorn when it is popped. Gasoline makes up the largest fraction of all petroleum products obtained. Other products include diesel fuel and heating oil, jet fuel, petrochemical feedstocks (to manufacture plastics, synthetic rubber, or other chemicals), waxes, lubricating oils, and asphalt.

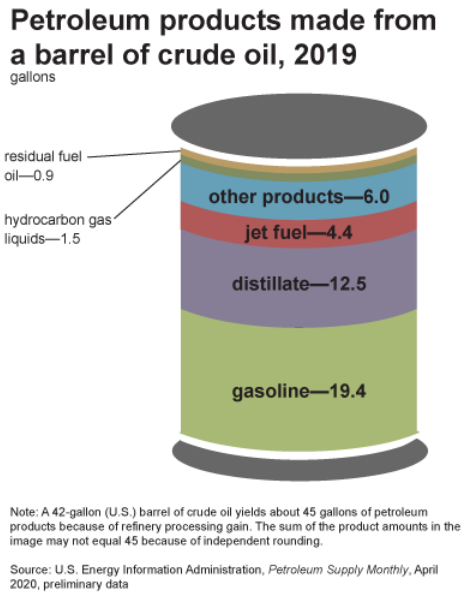



**Figure 7.1.7:** Main products (measured in gallons) made from a barrel of crude oil in 2019. These include residual fuel oil (0.9 gallons), hydrocarbon gas (1.5), other products (6.0), jet fuel (4.4), distillate (12.5), and gasoline (19.4). Note: A 42-gallon (U.S.) barrel of crude oil yields about 45 gallons of petroleum products because of refinery processing gain. The sum of the product amounts in the image may not equal 45 because of independent rounding. Image by EIA (public domain).

Oil refining is one of top sources of air pollution in the United States for volatile organic hydrocarbons and toxic emissions, and the single largest source of carcinogenic benzene. When petroleum is burned as gasoline or diesel, or to make electricity or to power boilers for heat, it produces a number of emissions that have a detrimental effect on the environment and human health:

- Carbon dioxide ($CO_2$) is a greenhouse gas and a source of climate change.
- Sulfur dioxide ($SO_2$) causes acid rain, which damages plants and animals that live in water, and it increases or causes respiratory illnesses and heart diseases, particularly in vulnerable populations like children and the elderly.
- Nitrous oxides ($NO_x$) and Volatile Organic Carbons (VOCs) contribute to ozone at ground level, which is an irritant and causes damage to the lungs.
- Particulate Matter (PM) produces hazy conditions in cities and scenic areas, and combines with ozone to contribute to asthma and chronic bronchitis, especially in children and the elderly. Very small, or “fine PM,” is also thought to penetrate the respiratory system more deeply and cause emphysema and lung cancer.
- Lead can have severe health impacts, especially for children.

## Transportation of Oil and Natural Gas

Most of the production of oil and natural gas occurs in remote areas, while most consumption is in densely populated areas. Consequently, enormous quantities of petroleum and its refined products are transported over great distances, mostly by overland pipelines, railroads, and trucks. A system of pipelines runs throughout the United States to transport oil and fuels from one location to another. There are pipelines that transport crude oil from the oil well to the refinery. At the refinery, there are additional pipelines that transport the finished product to various storage terminals where it can then be loaded onto trucks for delivery, such as to a gas station. On the global scale, most petroleum and its refined products are transported by oceanic tankers and overland pipelines. Local distribution systems involve smaller tankers, barges, pipelines, railroads, and trucks. There is a risk of accidental spillage

from all of these means of transportation. Some of the spills have been spectacular in their volume and environmental damage. In addition, petroleum is discharged into the environment by many smaller sources, which sum to a large cumulative amount.

Once natural gas is produced from underground rock formations, it is sent by pipelines to storage facilities and then on to the end user. The United States has a vast pipeline network that transports gas to and from nearly any location in the lower 48 states. There are more than 210 natural gas pipeline systems, using more than 300,000 miles of interstate and intrastate transmission pipelines (figure 7.1.8). Compressor stations that maintain pressure on the natural gas to keep it moving through the system. There are more than 400 underground natural gas storage facilities that can hold the gas until it is needed back in the system for delivery.



**Figure 7.1.8:** Construction of the controversial Dakota Access Pipeline, which stretches from North Dakota to Illinois. The pictured portion of the pipeline is in Central Iowa. You can learn more about the 2016 protests against this pipeline here. Image by Dakota Access Pipe Line (CC-BY).

Natural gas is released into the atmosphere from coal mines, oil and gas wells, and natural gas storage tanks, pipelines, and processing plants. These leaks are the source of about 25% of total U.S. methane emissions, which translates to three percent of total U.S. greenhouse gas emissions, contributing to climate change. When natural gas is produced but cannot be captured and transported economically, it is "flared," or burned at well sites, which converts it to carbon dioxide. This is considered to be safer and better than releasing methane into the atmosphere because carbon dioxide is a less potent greenhouse gas than methane. However, when natural gas with high concentrations of the toxic gas hydrogen sulfide is flared, it produces carbon dioxide, carbon monoxide, sulfur dioxide, nitrogen oxides, and many other compounds (see Air Pollution for more details).

Leaks also happen when we use petrochemicals on land. For example, gasoline sometimes drips onto the ground when people are filling their gas tanks, when motor oil gets thrown away after an oil change, or when fuel escapes from a leaky storage tank. When it rains, the spilled petrochemicals get washed into the gutter and eventually flow to rivers and into the ocean. Another way that oil sometimes gets into water is when fuel is leaked from motorboats and jet skis. When a leak in a storage tank or pipeline occurs, petrochemicals can also get into the ground, and the ground must be cleaned up. To prevent leaks from underground storage tanks, all buried tanks are supposed to be replaced by tanks with a double lining.

### Oil Spills

Human-caused oil spills in rivers and oceans harm ecosystems. From an economic perspective, oil spills disrupt the fishing industry and tourism. Oil spills at sea are generally much more damaging than those on land, since they can spread for hundreds of nautical miles in a thin oil slick which can cover beaches with a thin coating of oil. This can kill sea birds, mammals, shellfish and other organisms it coats. Oil spills on land are more readily containable if a makeshift earth dam can be rapidly bulldozed around the spill site before most of the oil escapes, and land animals can avoid the oil more easily.

Oil spills onto land are relatively common. Between 1989 and 1995, about 3,500 spills per year were reported in Canada – most all were relatively small, although by law they must be reported (Environment Canada, 1998). About 42% of the spills occurred in the vicinity of production wells, while 29% were from pipelines, and 16% from tanker trucks. During that period, up to 140-thousand t

of oil was spilled per year in petroleum-producing areas, due to accidental losses and well blowouts. In another study of the period 2000 to 2011, a total of 1,047 spills were reported from oil or gas pipelines in Canada (Kheraj, 2013).

During an oil spill on water, oil floats to the surface because it is less dense than water, and the lightest hydrocarbons evaporate, decreasing the size of the spill but polluting the air. Then, bacteria begin to decompose the remaining oil, in a process that can take many years. After several months only about 15% of the original volume may remain, but it is in thick asphalt lumps, a form that is particularly harmful to birds, fish, and shellfish. Cleanup operations can include a variety of components, but each has its procs and cons. Skimmer ships that vacuum oil from the water surface, but these are effective only for small spills. Controlled burning works only in early stages before the light, ignitable part evaporates, but this also pollutes the air. **Dispersants** are detergents that break up oil to accelerate its decomposition, but some dispersants may be toxic to the ecosystem. **Bioremediation** refers to adding microorganisms that specialize in quickly decomposing oil, but this can disrupt the natural ecosystem.

Oil spills can result from supertanker accidents such as the accidental grounding of the Exxon Valdez in 1989, which spilled 10 million gallons of oil into the rich ecosystem of coastal Alaska and killed massive numbers of animals. A significant amount of the petroleum extracted in the United States is mined in northern Alaska, from where a 1,280 km pipeline carries it south to the port of Valdez. The oil is then transported to markets in the western U.S. by a fleet of supertankers. The first part of the oceanic passage runs through a narrow shipping channel in Prince William Sound. Before the Exxon Valdez accident in March, 1989, tankers had navigated that passage about 16-thousand times. However, the Exxon Valdez, the newest tanker in the Exxon fleet, was steered onto a submerged reef, resulting in a spill of 36-thousand tonnes of its 176-thousand tonnes load of petroleum. About 40% of the spill washed onto shoreline habitat of Prince William Sound, while 25% was carried out of the sound by currents, and 35% evaporated at sea. Less than 10% was recovered or burned at sea.

The environmental damage was compounded by a lack of preparedness by industry and government for dealing with an oil-spill emergency. Essential equipment for containment and oil recovery was not immediately available, and it took too long to mobilize trained personnel. Consequently, despite favorable sea conditions during the first critical days after the grounding, few effective oil-spill countermeasures were mounted. Not until the second day of the spill was it possible to off-load unspilled petroleum from the Exxon Valdez to another tanker, and not until the third day were floating booms deployed to contain part of the spill. Unfortunately, a gale developed on the fourth day, making it impossible to contain or recover spilled petroleum, which then became widely dispersed. The region around Prince William Sound is famous for its spectacular scenery and large populations of wildlife. Some ecological communities and species were severely damaged by the oil spill. However, controversies have arisen about both the poor understanding of some ecological effects, and the role of science and scientists in sorting out legal and political aspects of the disaster (Holloway, 1996; Weins, 1996). For a long time, some scientists were prohibited from sharing their information because of legal needs for confidentiality. Controversies arose among scientists, environmental advocates, and other interest groups about the scale and intensity of some of the reported damages. About 1,900 km of shoreline habitat was oiled to some degree. A survey found that 140 km was “heavily oiled,” meaning there was at least a 6 m wide oiled zone. Another 93 km were “moderately oiled” (3-6 m wide zone), 323 km were “lightly oiled” (3 m wide), and the rest “very lightly oiled” (< 10% cover of oil). Overall, about 20% of the shoreline of the Sound, plus 14% of beaches on the nearby Kenai Peninsula and Kodiak Island, suffered some degree of oiling. A heroic and extremely expensive effort was undertaken to clean up some of the pollution from oiled beaches. About 11-thousand people were involved, costing the Exxon corporation about US$2.5 billion. The U.S. government spent an additional US$154 million. Residues were removed from heavily oiled beaches by machines and people wielding shovels and bags. Other places were cleaned by pressurized streams of hot or cold water. On some beaches, people actually wiped oiled rocks with absorbent cloths, a procedure that was ironically referred to as “rock polishing”.

These cleanup efforts helped greatly, and they were aided by natural processes, especially winter storms and microbial degradation of residues. Consequently, the amount of residues on beaches declined rapidly in the years following the spill. One survey of 28 polluted sites found an average of 37% surface oil cover in the first post-spill summer of 1989, but less than 2% in 1990. Another survey in 1991, after two post-spill winters and three summers, found that fewer than 2% of the beaches still had visible surface residues, compared with 20% in the first summer after the spill. However, subsurface residues still existed in many places. Initially, severe damage was caused to the seaweed-dominated intertidal zone of affected coastlines. These effects were made worse by certain cleanup methods, particularly washing with pressurized hot water. Fortunately, much of this damage proved to be short term, and by the end of 1991 a substantial recovery of seaweeds and invertebrates had begun. However, there were lingering effects on community structure, and vestiges of oil were still present 15 years later at some sites.



**Figure 7.1.9:** The top photo shows a heavily oiled beach on Green Island, Prince William Sound, soon after the Exxon Valdez disaster in 1989. The site was cleaned with warm-water washing in 1989, and then manually in 1990. In 1990 and 1991, it was treated with fertilizer to enhance microbial breakdown of the petroleum residues. The bottom photo shows the improved condition of the same beach in 1992, as a result of the natural and managed cleanups. Although little visible damage occurs on the surface, there are hydrocarbon residues deeper in the substrate. Source: Exxon Corporation.

Sea otters (*Enhydra lutris*) were the hardest-hit marine mammals. More than 3,500 otters were killed by oiling, out of a population of 5-10-thousand. A total of 357 oiled sea otters were captured and treated, of which 223 survived and were released or placed in zoos. Seabirds are very abundant in the region, particularly so in the autumn when certain species aggregate there during their southern migration. At that time, about 10-million seabirds may inhabit the Sound. Fortunately, the Exxon Valdez disaster happened in late winter, but there were still about 600-thousand seabirds present. About 36-thousand dead birds were found, but many additional corpses sank or drifted out to sea, and the total mortality may have been 375-435-thousand seabirds. About 400 people, 140 boats, and 5 aircraft were hired by Exxon to capture and rehabilitate oiled birds. They managed to treat 1,600 birds of 71 species, but half of them died of their injuries. The rest were treated and released to the wild, but the lingering effects of hydrocarbon poisoning likely prevented most of them from surviving for long. Even bird and mammal populations that suffered large mortality recovered to their natural abundance within a decade or less. From a strictly environmental perspective, the habitats affected by the disaster showed an impressive amount of resilience.

On the night of April 20, 2010, the Deepwater Horizon oil rig, one of hundreds operating in the Gulf of Mexico, exploded, killing eleven men, and placing one of the most rich and diverse coastal regions on earth in imminent danger of petroleum poisoning (figure 7.1.10). BP had been drilling in waters a mile deep, and in the next two days, as the rig slowly sank, it tore a gash in the pipe leading to the oil well on the ocean floor. The blowout was apparently caused by a failure of the casing of the borehole and also of the fail-safe blowout preventer. These likely occurred because of an unwise engineering decision to use an insufficient cementing regime for the borehole despite encountering extremely high geological pressure during the drilling. This resulted in a fire and explosion on the drilling platform, which sank during the fire-fighting action because of the enormous amounts of water poured

into it. The blowout lasted for 87 days and two hundred million gallons of crude oil poured into the Gulf before the technological means could be found to seal the undersea well. It was the worst environmental disaster in American history, and the largest peacetime oil spill ever.



**Figure 7.1.10:** Platform supply vessels battle the blazing remnants of the off shore oil rig Deepwater Horizon. A Coast Guard MH-65C dolphin rescue helicopter and crew document the fire aboard the mobile offshore drilling unit Deepwater Horizon, while searching for survivors. Multiple Coast Guard helicopters, planes and cutters responded to rescue the Deepwater Horizon's 126 person crew. Image and caption (modified) from US Coast Guard (public domain).

Wildlife, ecosystems, and people's livelihood were adversely affected as crude oil spread over as much as 176-thousand $km^2$ of water and affected beaches from western Florida to Texas, and caused tens of billions of dollars of economic damage. A lot of money and huge amounts of energy were expended on immediate clean-up efforts. About 7,000 $m^3$ of dispersant was used to help protect coastal infrastructure and habitats, and also to disperse the petroleum as it issued from the subsea blowout itself. Despite the enormous effort, considerable damage was done to natural habitats, recreational beaches, the commercial fishery, and harbors. The long-term impacts are still not known. Subsequent legal actions found that British Petroleum (BP), the operator of the drilling project, bore primary responsibility for the disaster. Eventually, BP paid more than US$42-billion in criminal and civil settlements. The very fact that BP was drilling under such risky conditions—a mile underwater, in quest of oil another three miles under the ocean floor—is an expression of the global demand for oil, the world's most valuable energy resource.

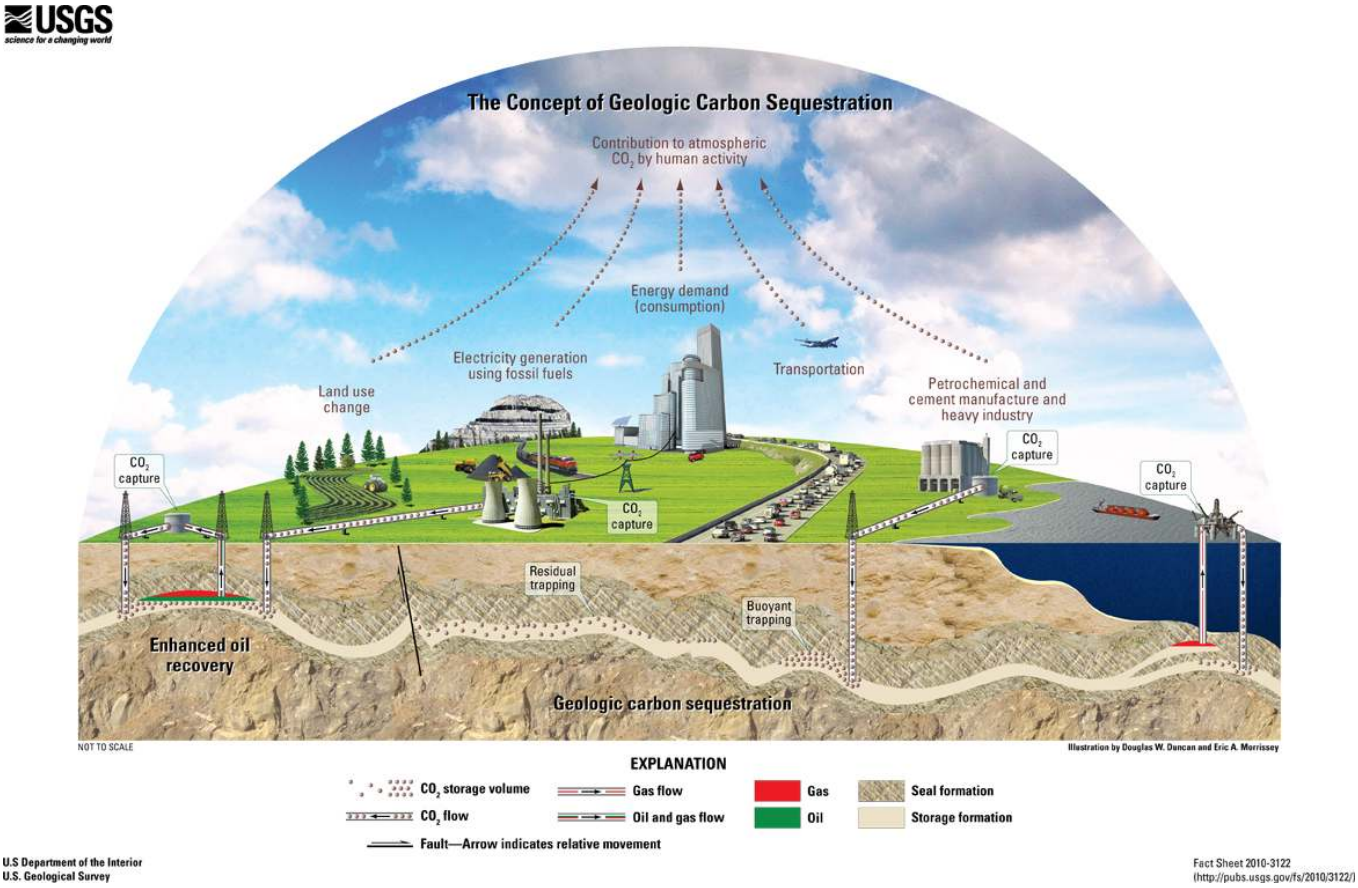
The world's largest-ever marine spill occurred during the brief Persian Gulf War of 1991. Iraqi forces deliberately released huge quantities of petroleum (about 0.8-million tonnes or approximately 200 million gallons) into the Gulf of Arabia from a Kuwaiti coastal loading facility. In part, this spill was a tactic of warfare – an attempt to make it difficult for Allied forces to execute an amphibious landing during the liberation of Kuwait. Mostly, however, the spill was an act of economic and ecological terrorism. The Iraqis also caused an enormous spillage on land during that war by igniting the more than 700 production wells in Kuwait that released enormous clouds of smoke and acid rain for over nine months.. An estimated 2-6 million tonnes of petroleum per day were emitted from the burning wells. After the Gulf War was over, it took 11 months to control and cap the blowouts. By that time, an immense 42-126-million tonnes of petroleum had spilled. Most of the crude oil burned in the atmosphere or evaporated, but 5-21 million tonnes accumulated as vast crude-oil lakes in the desert around the blowouts.

### The Future of Natural Gas and Coal

The future development of coal and natural gas depend on the degree of public and regulatory concern for carbon emissions, and the relative price and supply of the two fuels. Supplies of coal are abundant in the United States, and the transportation chain from mines to power plants is well established. The primary unknown factor is the degree of public and regulatory pressure that will be placed on carbon emissions. Strong regulatory pressure on carbon emissions would favor retirement of coal and addition of natural gas power plants. This trend is reinforced by the recent dramatic expansion of shale gas reserves in the United States due to advances in drilling technology. Shale natural gas production has increased 48% annually in the years 2006 – 2010, with more

increases expected. Greater United States production of shale gas will gradually reduce imports and could eventually make the United States a net exporter of natural gas.

**Clean coal technologies** can limit the air pollution released when burning coal. Some of these technologies remove toxins from coal before burning it while others capture toxins that are released while burning coal. For example, **smokestack scrubbers** in power plants clean sulfur dioxide, nitrous oxide, particulate matter, and mercury from the smoke before it is released. **Carbon capture and sequestration** involves capturing carbon dioxide released and storing it, but it requires 25-40% more energy, reducing the efficiency of coal (figure **7.1.10**). In this process, smoke from a coal power plant is passed through a solvent to trap carbon dioxide, but other waste gases are still released in the smoke. Carbon dioxide is then separated from the solvent. Some can be used in industry (such as for carbonated beverages or to tertiary recovery of oil), and the rest is sequestered (stored) underground. Note that clean coal technologies can reduce coal's contribution to climate change and reduce the amount of toxins that are released, but it does not fully prevent coal-generated air pollution (figure **7.1.10**).




## References


Baker, B., B. Campbell, R. Gist, L. Lowry, S. Nickerson, C. Schwartz, and L. Stratton. 1989. Exxon Valdez oil spill: the first eight weeks. Alaska Fish & Game, 21 (4): 2-37.

Davidson, A. 1990. In the Wake of the Exxon Valdez. Douglas & McIntyre, Toronto, ON.

Earle, S. 1992. Assessing the damage one year later (after the Gulf oil spill). National Geographic, 179 (2): 122-134.

GESAMP. 1991. Carcinogens: Their Significance as Marine Pollutants. Report 46, Joint Group of Experts on the Scientific Aspects of Marine Pollution (GESAMP), International Marine Organization, London, UK.

GESAMP. 1993. Impact of Oil and Related Chemicals and Wastes in the Marine Environment. Report 50, Joint Group of Experts on the Scientific Aspects of Marine Pollution (GESAMP), International Marine Organization, London, UK.London.

GESAMP. 2007. Estimates of Oil Entering the Marine Environment from Sea-Based Activities. Joint Group of Experts on the Scientific Aspects of Marine Pollution (GESAMP), International Marine Organization, London, UK.

Holloway, M. 1996. Sounding out science. Scientific American, 275 (4): 106-112.

Holloway, M. and J. Horgan. 1991. Soiled shores. Scientific American, 265 (4): 103-116.

Keeble, J. 1999. Out of the Channel: The Exxon Valdez Oil Spill in Prince William Sound. Eastern Washington University Press, Seattle, WA.

Kheraj, S. 2013. Tracking Canada's History of Oil Pipeline Spills, 1949-2012. ActiveHistory.ca http://activehistory.ca/2013/11/tracking-canadas-history-of-oil-pipeline-spills/

Koons, C.B. and H.O. Jahns. 1992. The fate of oil from the Exxon Valdez; A perspective. Marine Technical Society Journal, 26: 61-69.

National Academy of Sciences (U.S.). 2003. Oil in the Sea: Impacts, Fates, and Effects. National Academy of Sciences, Washington, DC.

National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling. 2011. Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling: Report to the President. Washington, DC.

Paine, R.T., J.L. Ruesink, A. Sun, E.L. Soulanille, M.L. Wonham, C.D.G. Harley, D.R. Brumbaugh, and D.L. Secord. 1996. Trouble on oiled waters: lessons from the Exxon Valdez oil spill. Annual Reviews of Ecology and Systematics, 27: 197-235.

Warner, F. 1991. The environmental consequences of the Gulf War. Environment, 33 (5): 7–26.

Weins, J.A. 1996. Oil, seabirds, and science: the effects of the Exxon Valdez oil spill. BioScience, 46: 587-597.

Wells, P.G., J.N. Butler, and J.S. Hughes. 1995. Exxon Valdez Oil Spill: Fate and Effects in Alaskan Waters. American Society for Testing and Materials, Philadelphia, PA.

## Contributors and Attribution

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

- Challenges and Impacts of Energy Use, and Water Pollution, and Non-Renewable Energy Sources from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Hannah Ritchie (2017) - "Fossil Fuels". Published online at OurWorldInData.org. Accessed 01-16-2021. (licensed under CC-BY)
- Non-Renewable Energy Sources from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Fossil Fuels and Mining from *An Introduction to Geology* by Chris Johnson et al. (licensed under CC-BY-NC-SA)
- Chapter 4: Non-Renewable Energy from *Introduction to Environmental Science: 2nd Edition* (2018) Biological Sciences Open Textbooks by Zehnder, Caralyn; Manoylov, Kalina; Mutiti, Samuel; Mutiti, Christine; VandeVoort, Allison; and Bennett, Donna (licensed under CC-BY-NC-SA).
- Fossil Energy Study Guide: Natural Gas. 2014. U.S. Department of Energy. Accessed 01-12-2021.
- Fossil Energy Study Guide: Oil. 2013. U.S. Department of Energy. Accessed 01-12-2021.
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.
- Environmental Challenges in Energy, Carbon Dioxide, Air, Water and Land Use, Case Study: Energy and the BP Oil Disaster, and Fossil Fuels from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors. Textbook content produced by Tom Theis and Jonathan (CC-BY). Download for free at CNX.
- Oil Spills in E*nvironmental Science: A Canadian Perspective* by Bill Freedman (CC-Attribution-NC)

This page titled 7.1: Energy from Fossil Fuels is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 7.2: Nuclear Energy

## Nuclear Energy Consumption

The use of nuclear energy presents an interesting dilemma. On the one hand, nuclear electricity produces no carbon emissions, a major sustainable advantage in a world facing climate change. On the other hand, there is environmental risk of storing spent fuel for thousands or hundreds of thousands of years, societal risk of nuclear proliferation, and the impact of accidental releases of radiation from operating reactors. Thoughtful scientists, policy makers, and citizens must weigh these advantages and disadvantages.

Nuclear electricity came on the energy scene remarkably quickly. Following the development of nuclear technology at the end of World War II for military ends, nuclear energy quickly acquired a new peacetime path for inexpensive production of electricity. Eleven years after the end of World War II, a very short time in energy terms, the first commercial nuclear reactor produced electricity at Calder Hall in Sellafield, England. The number of nuclear reactors grew steadily to more than 400 by 1990 (figure **7.2.1**), four years after the Chernobyl disaster in 1986 and eleven years following Three Mile Island in 1979 (see Consequences of Nuclear Energy). The number of operating reactors remained approximately flat for two decades, and the United States has not built a new nuclear facility since 1996. The number of operating nuclear power plants decreased in 2011, when the meltdown at the Fukushima Daiichi Nuclear Power Station caused Japan to shut down all of its nuclear power plants. Japan has since resumed use of some of its nuclear reactors.



Figure **7.2.1**: Arkansas Nuclear One, a power plant in Russellville, Arkansas. Image by Edibobb (CC-BY).

World production of electricity from nuclear power was about 2795.96 terawatt-hours (TWh) in 2019, comprising 10.4% of electricity production and 4.3% of total energy consumption globally (figure **7.2.2**). (For reference, the U.S. generated about 4100 TWh of electricity total in 2019.) The United States produced and consumed about 30.5% of the world's nuclear power in 2019, where nuclear power provided about 19.6% of the electricity and 8.0% of the total energy consumption (figure **7.2.3**).

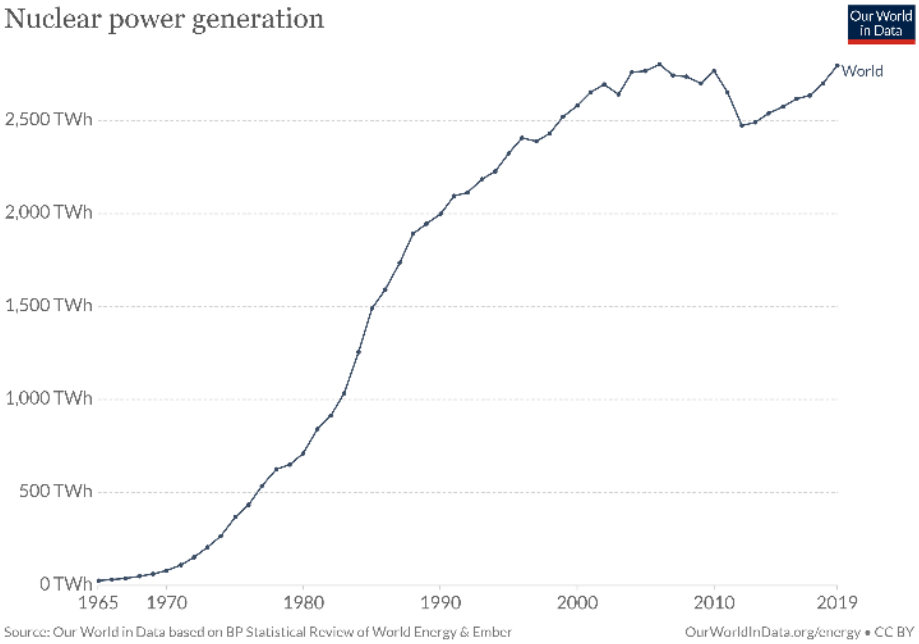



Figure **7.2.2**: Global nuclear power generation in TWh over time. Nuclear power production initially increased but remained relative flat in the early 2000s. The dip in 2011 is due to the temporary shutdown of nuclear reactors in Japan following the Fukushima Daiichi Nuclear Power Station meltdown. Image from Hannah Ritchie/Our World in Data (CC-BY).

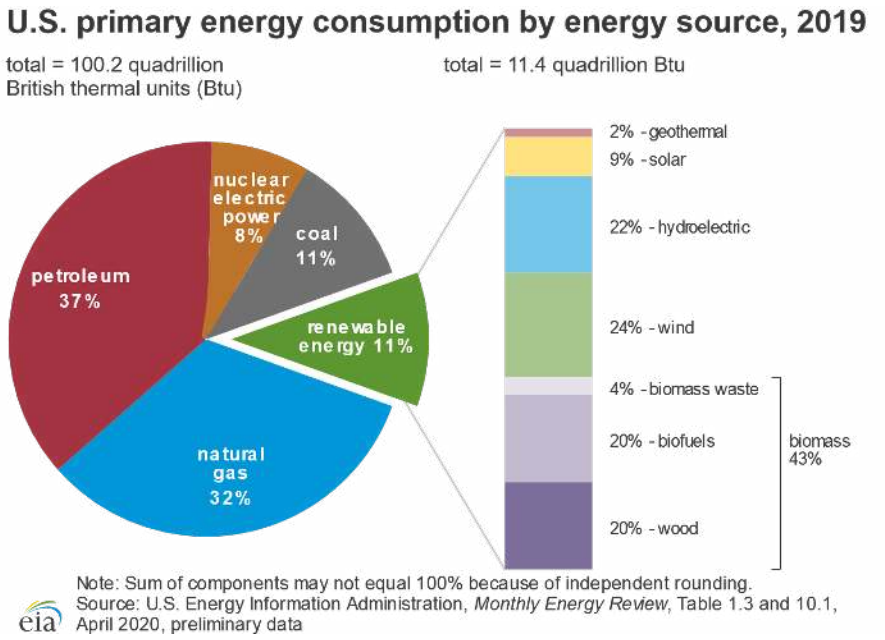



Figure **7.2.3**: U.S. primary energy consumption by energy source in 2019. Nuclear power comprised 8% of total energy consumption in the U.S. Total energy consumption was 100.2 quadrillion British thermal units (Btu). Other source of energy were petroleum (37%), natural gas (32%), coal (11%), and renewable energy (11%).Total renewable energy consumption was 11.4 quadrillion Btu. This includes biomass (43%), wind (24%), hydroelectric (22%), solar (9%), and geothermal (2%). Biomass includes wood (20%), biofuels other than wood and waste (20%), and biomass waste (4%). Note: Sum of components may not equal 100% because of independent rounding. Image by EIA (public domain).

## The Nuclear Fuel Cycle

Uranium ore must be mined, milled, and enriched to produce nuclear fuel. The **nuclear fuel cycle** represents the progression of nuclear fuel from creation to disposal (figure **7.2.4**). The first stage of the nuclear fuel cycle is uranium recovery, in which uranium ore is **mined**. It is then milled to produce yellowcake (uranium ore concentrate/uranium oxide/$U_3O_8$). **Milling** separates the uranium from the other parts of the ore. Each ton of mined uranium ore typically yields 1-4 pounds of yellowcake (0.05% to 0.20% yellowcake). Next, the uranium ore concentrate is converted into uranium hexafluoride ($UF_6$). It is then **enriched** to increase the concentration of uranium-235 ($^{235}U$) relative to $^{238}U$. During **fuel fabrication,** natural and enriched $UF_6$ is converted into into uranium dioxide ($UO_2$) or uranium metal alloys for use as fuel for nuclear power plants. Disposal of spent fuel rods and other hazardous waste generated in this process are discussed in Consequences of Nuclear Energy.

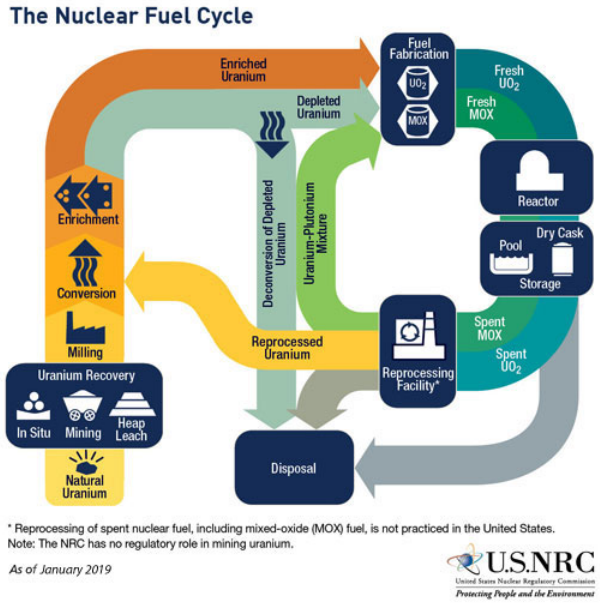



Figure **7.2.4**: The nuclear fuel cycle begins with recover (mining) of natural uranium, followed by milling, conversion, and enrichment. The enriched uranium then undergoes fuel fabrication, producing uranium dioxide ($UO_2$) or uranium metal alloys (MOX). The fuel then goes to a nuclear reactor at the power plant. Spent fuel rods are stored in a pool or dry cask, potentially reprocessed, and sent to disposal. Reprocessed uranium can reenter earlier steps of the fuel cycle. Wastes produced during enrichment undergo deconversion before disposal. Image by U.S. NRC (public domain)

## Nuclear Reactors

The fuel, which is now in the form of cylindrical ceramic pellets are then sealed into long metal tubes called **fuel rods**, which are assembled in reactor cores along with control rods. Each fuel pellet, which is about 1 cm in length, stores the same amount of energy as a ton of coal. Thousands of fuel rods form the **reactor core**, the site of nuclear fission in a nuclear power plant (figure **7.2.5**).

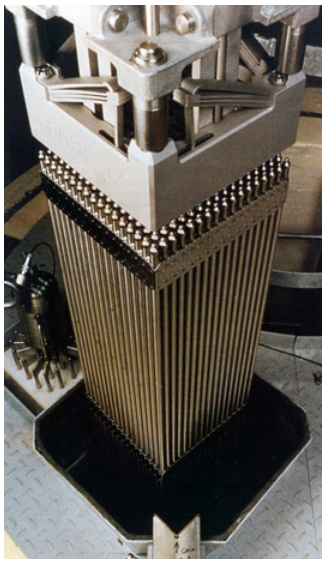


Figure 7.2.5: An assembly of fuel rods containing pellets of nuclear fuel. Image by Alternative Energies and Atomic Energy Commission, France (public domain)

Heat is produced in a nuclear reactor when neutrons strike uranium atoms, causing them to split in a continuous chain reaction that releases heat energy (figure 7.2.6). Specifically, fission of $^{235}U$, releases additional neutrons, which then cause the fission of nearby $^{235}U$ nuclei. However, if fission occurs in too many atoms simultaneously, too much energy is released, which can cause an explosion or meltdown. This is prevented by **control rods**, which are made of a material such as boron that absorbs excess neutrons released in nuclear fission. When the neutron-absorbing control rods are pulled out of the core, more neutrons become available for fission, and the chain reaction speeds up, producing more heat. When they are inserted into the core, fewer neutrons are available for fission, and the chain reaction slows or stops, reducing the heat generated.

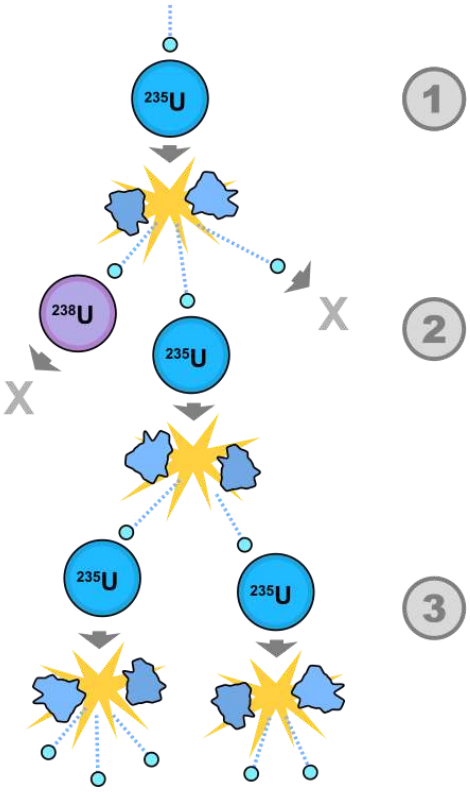


Figure 7.2.6: Schematic diagram of a fission chain reaction. (1) A uranium-235 atom absorbs a neutron, and fissions into two new atoms (fission fragments), releasing three new neutrons and energy. (2) One of those neutrons is absorbed by an atom of uranium-238, and does not continue the reaction. Another neutron is simply lost and does not collide with anything, also not continuing the reaction. However one neutron does collide with an atom of uranium-235, which then fissions and releases two neutrons and some binding energy. (3) Both of those neutrons collide with uranium-235 atoms, each of which fission and release between one and three neutrons, and so on. Image and caption (modified) by Fastfission (public domain).

**Nuclear reactors** (figure 7.2.7) contain the reactor core and the machinery needed to generate electricity from the heat released. The reactor core is submerged in water. In addition to transferring heat energy, the water also serves to slow down, or "moderate" the neutrons which is necessary for sustaining the fission reactions. Ultimately, the heat energy is used to generate high-pressure steam, which turns a turbine to generate electricity. The mechanism is similar to that of coal- or natural gas-generated electricity, but nuclear fission rather than combustion of coal is the source of heat energy.

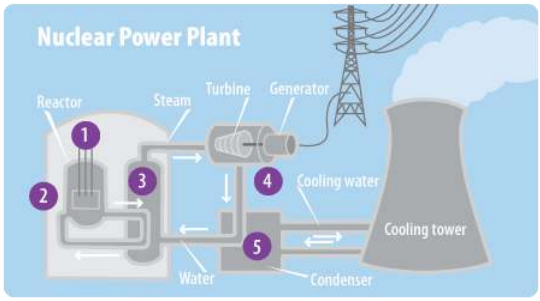



Figure 7.2.7: (1) In a nuclear reactor, fuel rods full of uranium pellets are placed in water. (2) Inside the fuel rods, uranium atoms split, releasing energy. (3) This energy heats water, creating steam. (4) The steam moves through a turbine, which turns a generator to create electricity. (5) In the condensor, the steam cools back into water, which can then be used over again. At some nuclear power plants, extra heat is released from a cooling tower. Image and caption (modified) by EPA (public domain).

There are two main types of nuclear reactors: pressurized water reactors and boiling water reactors.

## Pressurized Water Reactor

In a **pressurized water reactor**, there are three separate streams of water: the water in contact with the reactor core, the water that produces steam, and the cooling water (figure 7.2.8). The reactor core is submerged in water, which is held by a steel vessel. This is surrounded by a containment structure. As the nuclear fission reaction heats the water surrounding it, the water is pumped in a cyclical stream. It transfers heat to the second stream of water, which is in a separate vessel. This second stream is kept at a lower pressure, allowing the water to boil and create steam The steam turns a turbine, generating electricity. The steam then is cooled in the **condenser** by a separate stream of cooling water. Because water from the reactor core does not mix with other parts of the reactor, not all of the reactor is radioactive.

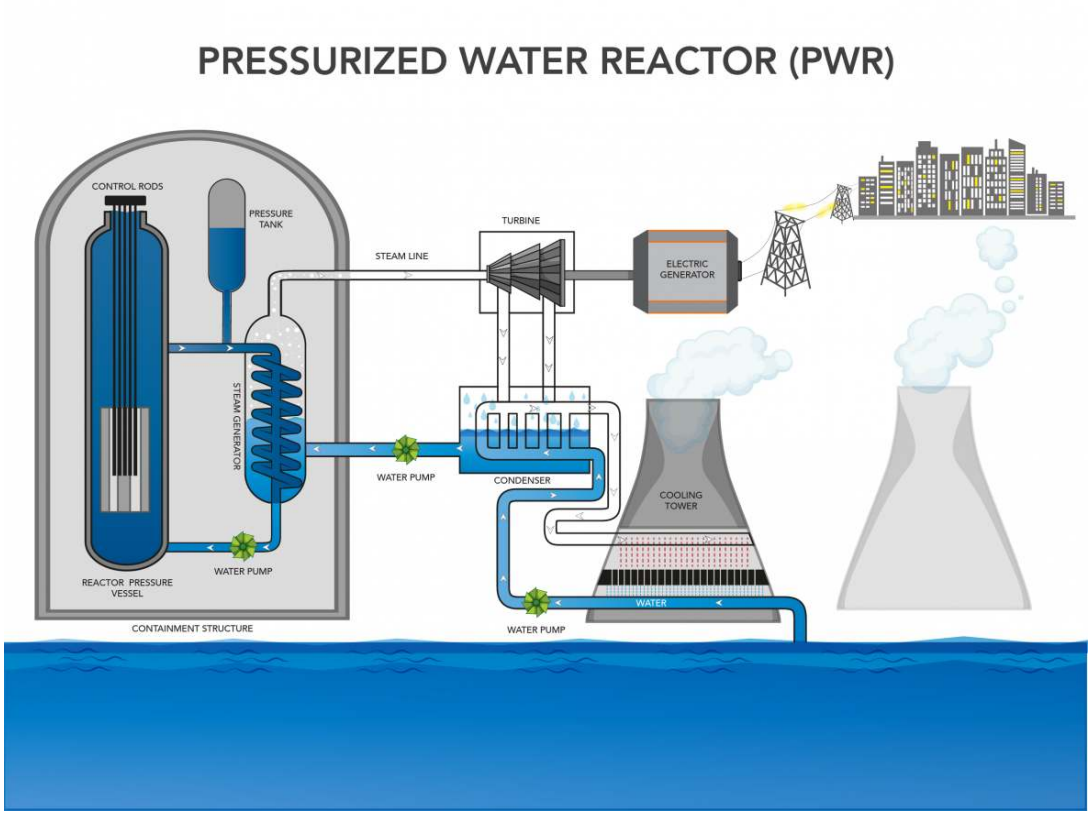



Figure **7.2.8**: There are three separate streams of water in a pressurized water reactor. The first stream is associated with the reactor core in the reactor pressure vessel. The second stream is associated with the steam generator. The third stream is from an external source and is used to condense the steam. The reactor pressure vessel contains control rods and fuel rods (not labeled). The pressure tank regulates water pressure. Hot water from the reactor moves to the steam generator, where it heats water in the second stream to produce steam. Steam from the steam generator moves through the steam line and turns a turbine, powering the electric generator. Steam cools in the condenser, and the cooled water is then pumped back into the steam generator. The reactor pressure vessel, pressure tank, and steam generator are all in a containment structure. Water from an external source is used to cool steam in the condenser. The water from the external source cools in the cone-shaped cooling tower. Image by Office of Nuclear Energy/ U.S. Department of Energy (public domain).

## Boiling Water Reactor

In a **boiling water reactor**, there are two separate streams of water: the water in contact with the reactor core and the cooling water (figure **7.2.9**). The reactor core heats the water in which it is submerged. This water is held by a steel vessel that is surrounded by a containment structure. The steam produced as the reactor core heats water turns a turbine, which generates electricity. The steam then is cooled in the condenser by a separate stream of cooling water. Because water from the reactor core comes in contact with all parts of the reactor, the entire thing is radioactive.

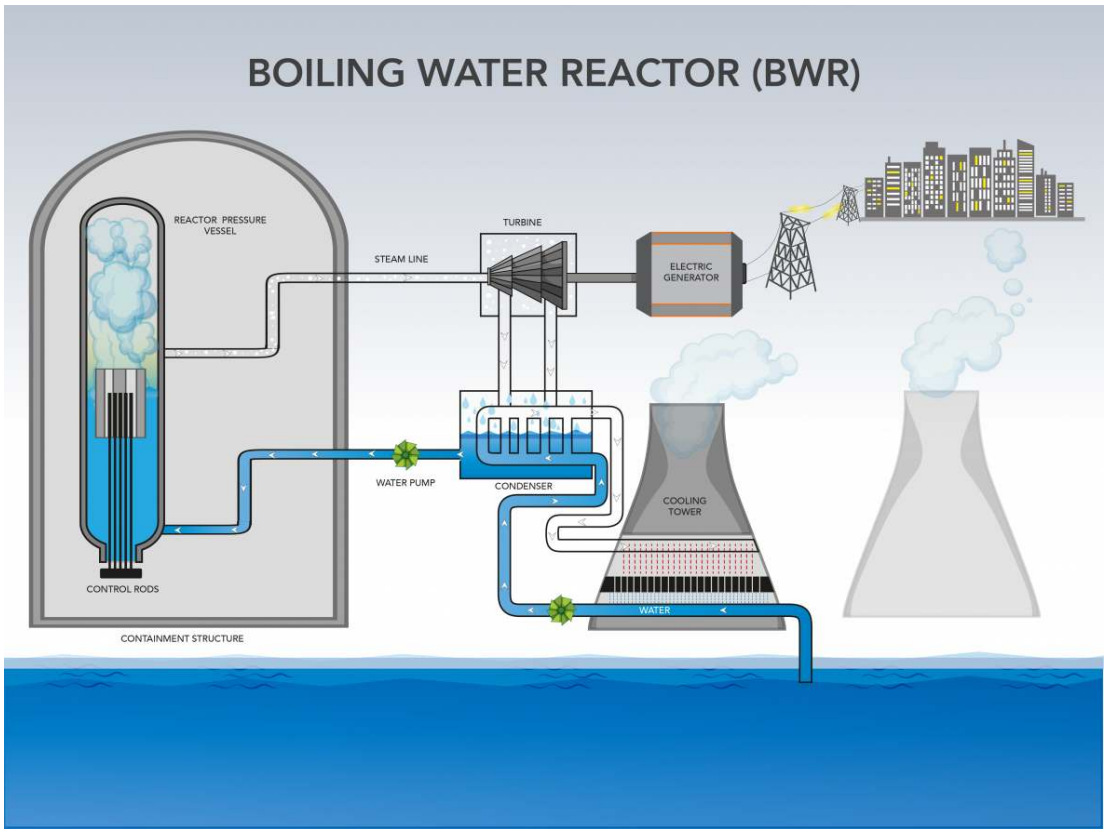



Figure **7.2.9**: A boiling water reactor has two streams of water. In the first stream, the same water that bathes the reactor core in the reactor pressure vessel and turns to steam and turns a turbine. The third stream is from an external source and is used to condense the steam. The reactor pressure vessel contains control rods and fuel rods (not labeled). It is within a larger containment structure. When water is heated by the reactor, it evaporates to steam. This moves through the steam line and turns a turbine, powering the electric generator. Steam cools in the condenser, and the cooled water is then pumped back into the reactor pressure vessel. Water from an external source is used to cool steam in the condenser. The water from the external source cools in the cone-shaped cooling tower. Image by Office of Nuclear Energy/ U.S. Department of Energy (public domain).

## Advantages of Nuclear Energy

In contrast to fossil fuels, generating electricity from nuclear energy does not pollute the air or significantly contribute to climate change (figure **7.2.10**). As we continue to deplete global reserves of fossil fuels, supplies of nuclear fuel are abundant. It is

estimated that uranium supplies will last over 200 years, and there is potential to use other radioactive isotopes as well.

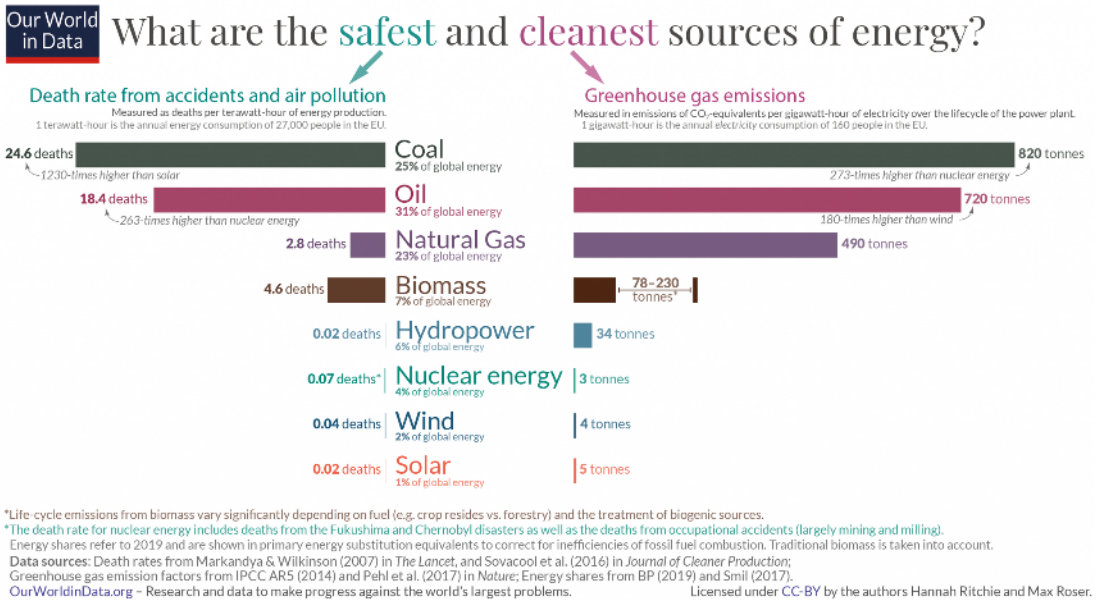



**Figure 7.2.10:** Nuclear energy causes fewer deaths and releases miniscule greenhouse gases compared to fossil fuels. These bar graphs explore the question, "What are the safest and cleanest sources of energy?". On the left is the death rate from accidents and air pollution measured as deaths per terawatt-hour (deaths/TWh) of energy production. 1 terawatt-hour is the annual energy consumption of 27,000 people in the European Union (EU). On the right are greenhouse gas emissions measured in tonnes of CO2-equivalents emitted per gigawatt-hour ($tCO_2$-eq/gWh) of electricity over the lifecycle of the power plant. 1 gigawatt-hour is the annual electricity consumption of 160 people in the EU. Coal accounted for 24.6 deaths/TWh, 25% of global energy, and 820 $tCO_2$-eq/gWh. Oil accounted for 18.4 deaths/TWh, 31% of global energy, and 720 $tCO_2$-eq/gWh. Coal accounted for 2.8 deaths/TWh, 23% of global energy, and 490 $tCO_2$-eq/gWh. Biomass accounted for 4.6 deaths/TWh, 7% of global energy, and 78-230 $tCO_2$-eq/gWh. Hydropower accounted for 0.02 deaths/TWh, 6% of global energy, and 34 $tCO_2$-eq/gWh. Nuclear energy accounted for 0.07 deaths/TWh, 4% of global energy, and 4 $tCO_2$-eq/gWh. Wind accounted for 0.04 deaths/TWh, 2% of global energy, and 4 $tCO_2$-eq/gWh. Coal accounted for 0.02 deaths/TWh, 1% of global energy, and 5 $tCO_2$-eq/gWh. Coal caused 1230 times more deaths than solar, and oil caused 263 times more deaths than nuclear energy. Coal emissions were 273 times higher than nuclear energy, and oil emissions were 180 times higher than wind. Image by Hannah Ritchie and Max Roser/Our World in Data (CC-BY).

Furthermore, nuclear power plants are more reliable than any other source, with a capacity factor of 93.5% (figure 7.2.11). **Capacity** is the amount of electricity a generator can produce when it's running at full blast, and the **capacity factor** is a measure of how often a plant is running at maximum power. (A power plant with a capacity factor of 100% means that it's producing power all of the time.)

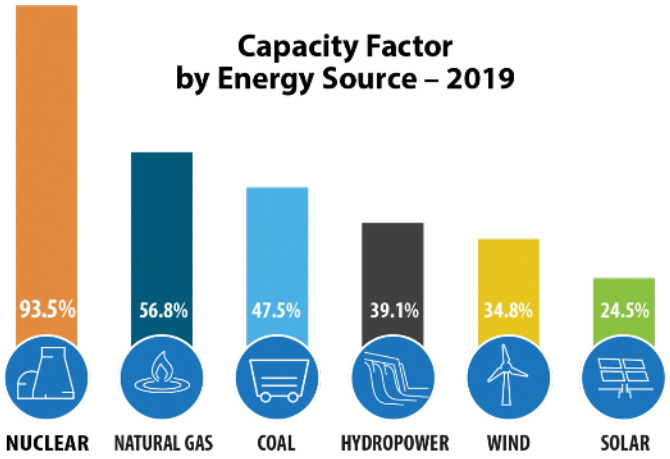



**Figure 7.2.11:** The capacity factor of six energy sources in 2019. Nuclear has the highest capacity factor at 93.5% followed by natural gas (56.8%), coal (47.5%), hydropower (39.1%), wind (34.8%), and solar (24.5%). Image from Office of Nuclear Energy/ U.S. Department of Energy (public domain).

## Negative Impacts of Nuclear Energy

Despite its benefits, nuclear power has downsides. It requires more water than any other energy source. Water used for cooling is released back into the environment, and while it does not contain radioactive materials or other harmful chemicals, it is warmer than before. This is called **thermal pollution**, and it can harm aquatic life, which are adapted to cooler temperatures. Surface mining for uranium ore degrades habitat and releases toxins from underground (similar to surface mining for coal). Nuclear power plants are expensive to build and maintain, and they require large amounts of metal and concrete. Enriched uranium for nuclear fuel if in the wrong hands can be used to make nuclear weapons (figure **7.2.12**). While nuclear accidents are rare, they can cause great harm, and their impacts are long-lasting. Furthermore, the problem of safely disposing spent nuclear fuel remains unresolved. The latter two concerns are discussed in more detail below.

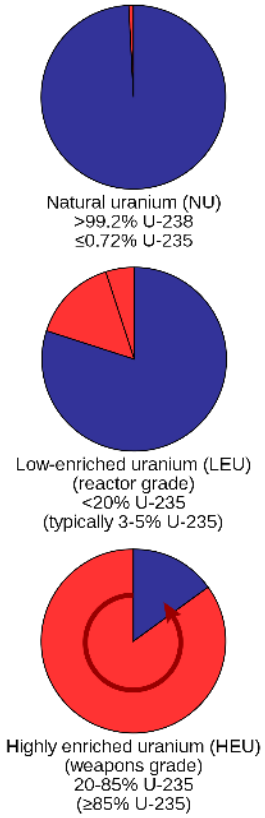



**Figure 7.2.12:** Pie charts showing the relative proportions of uranium-238 (blue) and uranium-235 (red) at different levels of enrichment. Nuclear weapons require more highly enriched uranium than is needed for nuclear fuel. Natural uranium (NU) consists of more than 99.2% U-238 and 0.72% or less of U-235. Low-enriched uranium (LEU) consists of less than 20% U-235, with reactor-grade uranium typically 2-5% U-235. High-enriched uranium (HEU) consists of 20-85% U-235 with weapons grade at least 85% U-235. Image and caption (modified) by Fastfission (public domain).

## Nuclear Waste

The main environmental challenge for nuclear power is the wastes including high-level radioactive waste, low-level radioactive waste, and uranium mill tailings. These materials have long radioactive half-lives and thus remain a threat to human health for thousands of years.

**High-level radioactive waste** (HLRW) consists of used nuclear reactor fuel (spent nuclear fuel rods). These contain the products of nuclear fission, which are radioactive themselves. This HLRW is temporarily stored in a pool at the nuclear power plant or a dry cask, steel cylinders within another container, made of steel or concrete (figure **7.2.13**). Dry casks contain inert (nonreactive) gas and may be located at the power plant, a decommissioned power plant, or a separate storage site. High-level radioactive waste may only be moved to a dry cask after one year of cooling in a pool. The U.S. has no long-term storage for HLRW, and spent fuel thus remains interim storage.

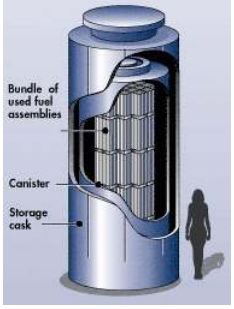


**Figure 7.2.13:** Dry casks seal spent fuel rods in a canister, which is surrounded by a larger storage cask. Bundles of used fuel assemblies are inside. Image by NRC (public domain).

Yucca Mountain in Nevada was proposed as a long-term geologic storage site, where HLRW could be buried for hundreds of thousands of years. The storage facility was constructed, but it has not been used due to opposition from local residents and concern about the safety of transporting HLNW (figure **7.2.14**)



**Figure 7.2.14:** Yucca Mountain is the proposed site for long-term storage of high-level radioactive waste in the U.S., but it is not in use due to political controversy. Image by whitehouse.gov (public domain).

Some countries reprocess (recycle) spent nuclear fuel, but no recycling or reprocessing facility or a federal waste repository is currently licensed in the United States. Reprocessing separates the useable fraction of spent fuel and recycles it through the reactor, using a greater fraction of its energy content for electricity production, and sends the remaining high-level waste to permanent geologic storage.

The primary motivation for reprocessing is greater use of fuel resources, extracting about 25 percent more energy than the once through cycle. A secondary motivation for recycling is a significant reduction of the permanent geologic storage space (to 20% or less of what would otherwise be needed) and time (from hundreds of thousands of years to thousands of years). While these advantages seem natural and appealing from a sustainability perspective, they are complicated by the risk of theft of nuclear material from the reprocessing cycle for use in illicit weapons production or other non-sustainable ends. At present, France, the United Kingdom, Russia, Japan, and China engage in some form of reprocessing; the United States, Sweden, and Finland do not reprocess.

**Low-level radioactive waste** (LLRW) refers to items that were exposed to radiation includes clothing, filters, and gloves. These can be contained with concrete or lead (through which radiation cannot pass; figure **7.2.15**).

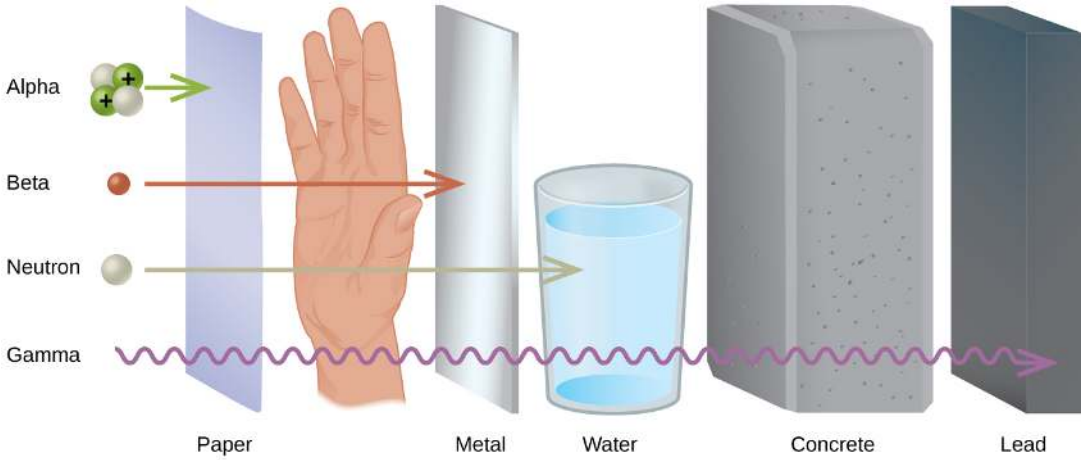



**Figure 7.2.15:** Water, thick concrete, lead, and steel (not shown) can stop several types of radiation released from radioactive wastes. Note that gamma rays can somewhat penetrate all of these substances, but lead, concrete, and steel provide a partial shield. The ability of different types of radiation to pass through material is shown. From least to most penetrating, they are alpha < beta < neutron < gamma. The top particle listed is made up of two white spheres and two green spheres that are labeled with positive signs and is labeled "Alpha." A right-facing arrow leads from this to the paper. The second particle is a red sphere labeled "Beta" and is followed by a right-facing arrow that passes through the paper and stops at the hand. The third particle is a white sphere labeled "Neutron" and is followed by a right-facing arrow that passes through the paper, hand and metal but is stopped at the glass of water. The fourth particle is shown by a squiggly arrow and it passes through all of the substances but stops at the lead. Terms at the bottom read, from left to right, "Paper," "Metal," "Water," "Concrete" and "Lead." Image and caption (modified) from Flowers, Theopold, and Langley/OpenStax (CC-BY). Download for free at CNX.

Low-level waste is typically stored at the nuclear power plant, either until it has decayed away and can be disposed of as ordinary trash, or until amounts are large enough for shipment to one of the five LLRW disposal sites in the U.S. (figure **7.2.16**).

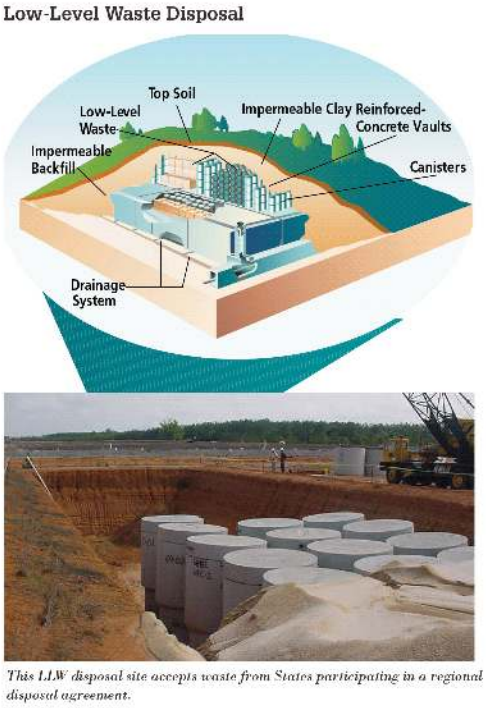



**Figure 7.2.16:** Diagram (top) and photo (bottom) of a low-level radioactive waste (LLRW or LLW) disposal site. In the diagram, low-level waste is contained in canisters within concrete vaults. This is surrounded by impermeable clay and backfill. The drainage

system prevents the wastes from contaminating groundwater. The entire disposal facility is underground, and a layer of top soil is above it. The disposal site in the photo accepts waste from States participating in a regional disposal agreement.The waste is sealed in canisters and shallowly buried. Image by NRC (public domain).

Enrichment of uranium produces depleted uranium hexafluoride ($DUF_6$), or **uranium mill tailings**, as a byproduct, which does not have high enough concentrations of $^{235}U$ to use as nuclear fuel but is still hazardous. Tailings represent the greatest percentage of nuclear waste by volume, and there are more than 200 million tons of radioactive mill-tailings in the United States. Tailings contain several radioactive elements including radium, which decays to produce radon, a radioactive gas. They are stored in **impoundments**, lined pits in the ground that are flooded with water, in remote areas. **Deconversion** involves chemically treating the tailings to reduce their hazards so that they can be stored as LLRW.

## Nuclear Disasters

There are many other regulatory precautions governing permitting, construction, operation, and decommissioning of nuclear power plants due to risks from an uncontrolled nuclear reaction. The potential for contamination of air, water and food is high should an uncontrolled reaction occur. Even when planning for worst-case scenarios, there are always risks of unexpected events. The nuclear accidents at Three Mile Island, Chernobyl, and Fukushima raised concerns about the safety of nuclear power.

The Three Mile Island accident occurred in Pennsylvania in 1979. It was a partial meltdown that resulted from an electrical failure and errors in operation. There were no direct deaths. Studies investigated the possibility of exposure to radiation from the accident indirectly causing deaths through increased rates of cancer or other disease, but there has not been evidence of this. In contrast, the 1986 meltdown at Chernobyl Nuclear Power Plant in what is now the Ukraine was responsible for 50 direct deaths. This disaster occurred from a test of the emergency systems gone wrong. Estimates of indirect deaths from radiation exposure range from 4,000 to 60,000.

The global discussion regarding nuclear energy has been strongly impacted by March 2011 earthquake and subsequent tsunami that hit Japan resulted in reactor meltdowns at the Fukushima Daiichi Nuclear Power Station causing massive damage to the surrounding area. The disaster disabled the cooling system for a nuclear energy complex, ultimately causing a partial meltdown of some of the reactor cores and release of significant radiation. The design of the reactors (boiling water reactors) made it more difficult to vent the system without releasing radiation. Cooling the radioactive fuel generated a large volume of contaminated water, and the disaster costed at least $300 billion dollars. While there were no immediate deaths, one person later died from cancer attributed to radiation exposure. Thousands died as a result of stress associated with the evacuation, and about 20% of the over 160,000 evacuees had not yet returned home as of 2019.

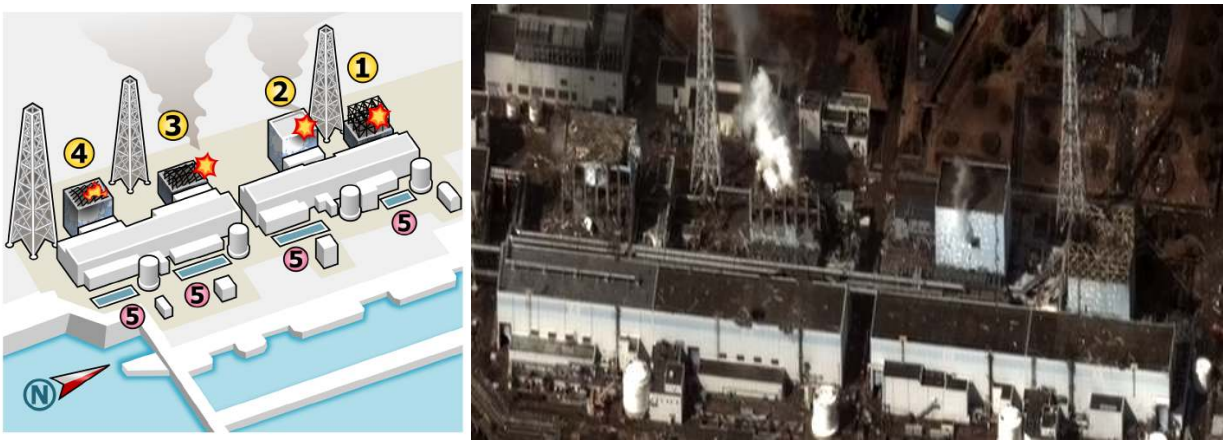



**Figure 7.2.17:** Left: Diagram (approximate) of the Fukushima I Nuclear Power plant accidents. (1) Unit 1: Explosion, roof blown off on 12 March. (2) Unit 2: Explosion on 15 March; contaminated water in underground trench, possible leak from suppression chamber. (3) Unit 3: Explosion, most of concrete building destroyed on 14 March, possible plutonium leak. (4) Fire on 15 March; water level in spent fuel pools partly restored. (5) Multiple trenches: probable source of contaminated water, partly underground, leaked stopped on 6 April. Right: The Fukushima I Nuclear Power Plant after the 2011 Tōhoku earthquake and tsunami. Reactor 1

to 4 from right to left. Left image and caption (modified) by Sodacan (CC-BY). Right image and caption by Digital Globe (CC-BY-SA).

To learn more about the Fukushima you can watch this three-minute segment, What Recovery Looks Like In Japan Almost A Decade After Fukushima Nuclear Disaster, which provides on update on evacuees from the Fukushima nuclear disaster.

## Attribution

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

- Non-renewable Energy Sources from *AP Environmental Science* by University of California College Prep (CC-BY). Download for free at CNX.
- Stages of the Nuclear Fuel Cycle. 2020. U.S. NRC. Accessed 01-16-2021 (public domain).
- Nuclear explained: The nuclear fuel cycle. 2020. U.S. Energy Information Administration. Accessed 01-16-2021 (public domain).
- Non-Renewable Energy Sources from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Nuclear Energy, Environmental Challenges in Energy, Carbon Dioxide, Air, Water and Land Use, and Systems of Waste Management from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors. Textbook content produced by Tom Theis and Jonathan (CC-BY). Download for free at CNX.
- What is Generation Capacity? 2020. Mike Mueller. Office of Nuclear Energy. U.S. Department of Energy. Accessed 01-16-2021 (public domain)

---

This page titled 7.2: Nuclear Energy is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 7.3: Solar Energy

**Solar energy** refers to heat or light energy from the sun. Solar energy is by far the most plentiful type of renewable energy, delivered to the surface of the Earth at a rate of 120,000 Terawatts (TW) per hour, compared to the global human use of 19.8 TW in the entire year of 2019. To put this in perspective, covering 1.2% of the Sahara desert with solar panels could meet Earths energy needs. Of course, this does not consider limitations on storage capacity and the ability to distribute that energy.

Technologies to harness solar energy may be passive or active. **Passive solar technologies** do not require complex equipment and can be as simple as using natural light from a window or skylight to illuminate a room (figure **7.3.1**). Similarly, solar tubes are lined with reflective material and can concentrate light energy to better illuminate a room (figure **7.3.1**). These are embedded in the ceiling as regular light fixtures would be.




**Figure 7.3.1:** Skylights and solar tubes (tubular daylighting devices, TDD). While skylights resemble windows that are installed in the ceiling, solar tubes use lenses and a reflective lining to concentrate sunlight, providing bright illumination for darker parts of the home. What makes a skylight energy efficient? Traditional skylights are skylights that use the same technologies as windows, but these technologies are even more valuable for skylights, which receive direct sun in summer and greater outside/inside temperature differentials in winter. Tubular daylighting devices gather sunlight at the roof and transmit it down to a diffusing lens mounted in an interior surface, usually a ceiling. The natural light from a TDD can illuminate closets, bathrooms, hallways, or other spaces that typically would not have access to sunlight, decreasing the need for electric lighting. Left image by U.S. Department of Energy (public domain), and right image by Energy Star (public domain).

Solar energy can also be used as heat, which can be maximized through careful architecture (figure **7.3.2**). Firstly, the building requires south-facing windows (or glass doors). As sunlight passes through these areas, the energy is stored in the thermal mass of the building. This refers to heat trapping materials such as rock or tiles. The building is also designed such that heat is then distributed throughout the building. Finally, roof eaves or similar structures block sunlight from entering the home during the summer.

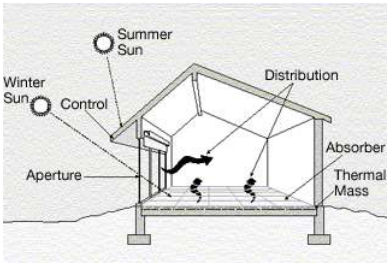



**Figure 7.3.2:** This building design uses the sun for heating during the winter and is thus a passive solar technology. The roof overhang serves as a control, blocking the higher summer sun while allowing the lower winter sunlight to pass through the window (aperture). Heat energy from the sun is then distributed in the building. Thermal mass (such as a tile or stone floor) absorbs and later releases heat energy. Image by U.S. Department of Energy (public domain).

A simple solar water heater is a passive technology consisting of a network of tubes that are heated by the sun (figure **7.3.3**). The hot water is then transferred through the plumping of a home. (Some solar water heaters are more complex, using pumps, and are thus considered active solar technologies.)




**Figure 7.3.3:** A solar water heater. Image by Vijayanarasimha/Pixabay (public domain).

**Active solar technologies** are more complex. For example, solar panels use light energy to generate electricity (figure**7.3.4**). This occurs in the units of the solar panel, which are called **photovoltaic cells** (PV cells; figure**7.3.4 – 5**). Each photovoltaic cell consists of two layers of **semiconductors**, substances that only conduct electricity under certain circumstances. (In contrast, conducts always conduct electricity, and insulators do not.) One semiconductor has extra electrons, but the other has extra spaces for electrons. When light shines on the photovoltaic cell, it causes electrons to move from between semiconductor layers through the conductor that connects them (such as metal wires or plates). This movement results in an electric current.



**Figure 7.3.4:** Rooftop solar installation on Douglas Hall at the University of Illinois at Chicago has no effect on land resources, while producing electricity with zero emissions. Source: Office of Sustainability, UIC

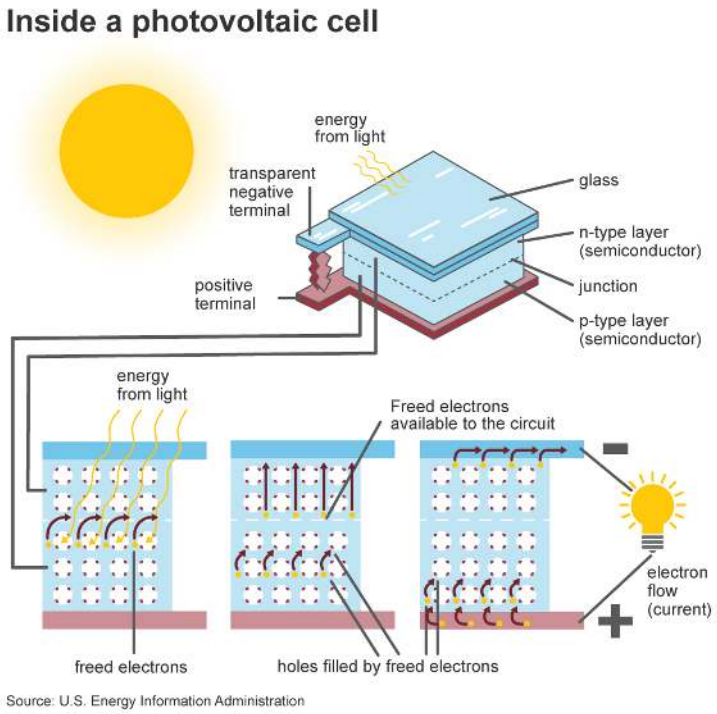



**Figure 7.3.5:** Diagram of a **photovoltaic cell** consisting of two layers of semiconductors. The top (n-type) semiconductor has extra electrons and the bottom (p-type) semiconductor has extra spots (holes) for electrons. A junction divides these two layers. Light frees the electrons and allows them to move through a wire from the n-type to fill "electron holes" in the p-type semiconductor. The movement of electrons through the wire results in an electrical current. Image by U.S. Energy Information Administration (public domain).

This video explains how the photovoltaic cells within solar panels generate electricity.



Another example of active solar technology is **solar thermal technology**. This involves using a series of mirror to concentrate solar energy, ultimately generating steam. From there, the steam turns a turbine and powers a generator (figure**7.3.6**).

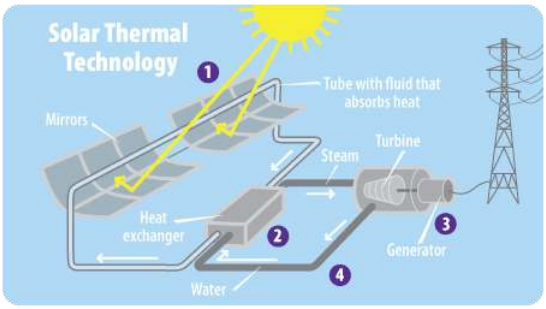



**Figure 7.3.6:** In solar thermal technology, (1) mirrors or reflectors concentrate the sun's rays to heat a special kind of liquid, which absorbs heat. (2) The heat from this liquid boils water to create steam, facilitated by the heat exchanger. (3) Steam spins a turbine that is connected to a generator, which creates electricity. (4) The steam cools and condenses back to water, which is recycled, reheated, and converted into steam again. Image and caption (modified) from EPA (public domain).

Solar energy is by far the most plentiful renewable source of energy, delivered to the surface of the earth at a rate of 120,000 Terawatts (TW), compared to the global human use of 15 TW. To put this in perspective, covering 100×100 km2 of desert with 10% efficient solar cells would produce 0.29 TW of power, about 12% of the global human demand for electricity. To supply all of the earth's electricity needs (2.4 TW in 2007) would require 7.5 such squares, an area about the size of Panama (0.05% of the earth's total land area). The world's conventional oil reserves are estimated at three trillion barrels, including all the oil that has already been recovered and that remain for future recovery. The solar energy equivalent of these oil reserves is delivered to the earth by the sun in 1.5 days.

Not only is solar energy abundant, but the use of solar panels for electricity does not generate air pollution or contribute to climate change. (The manufacture of solar panels can generate some pollution, including greenhouse gas emissions, but this is minimal compared to that of fossil fuels.) Like wind energy, expansions on solar energy can create jobs and boost economies. Also like wind, sunlight is intermittent and storage of solar energy is limited by battery capacity. Some locations do not receive consistently direct sunlight and are ill-suited for solar panels. While solar energy has historically been the most expensive form of renewable energy, new technologies have lowered its cost.

The geographical distribution of useable renewable solar energy is quite uneven, concentrated in deserts where cloud cover is rare. The solar resources of the United States, Germany and Spain are compared in Figure **7.3.7**. The solar irradiation in the southwestern United States is exceptional, equivalent to that of Africa and Australia, which contain the best solar resources in the world. Much of the United States has solar irradiation as good or better than Spain, considered the best in Europe, and much higher than Germany.

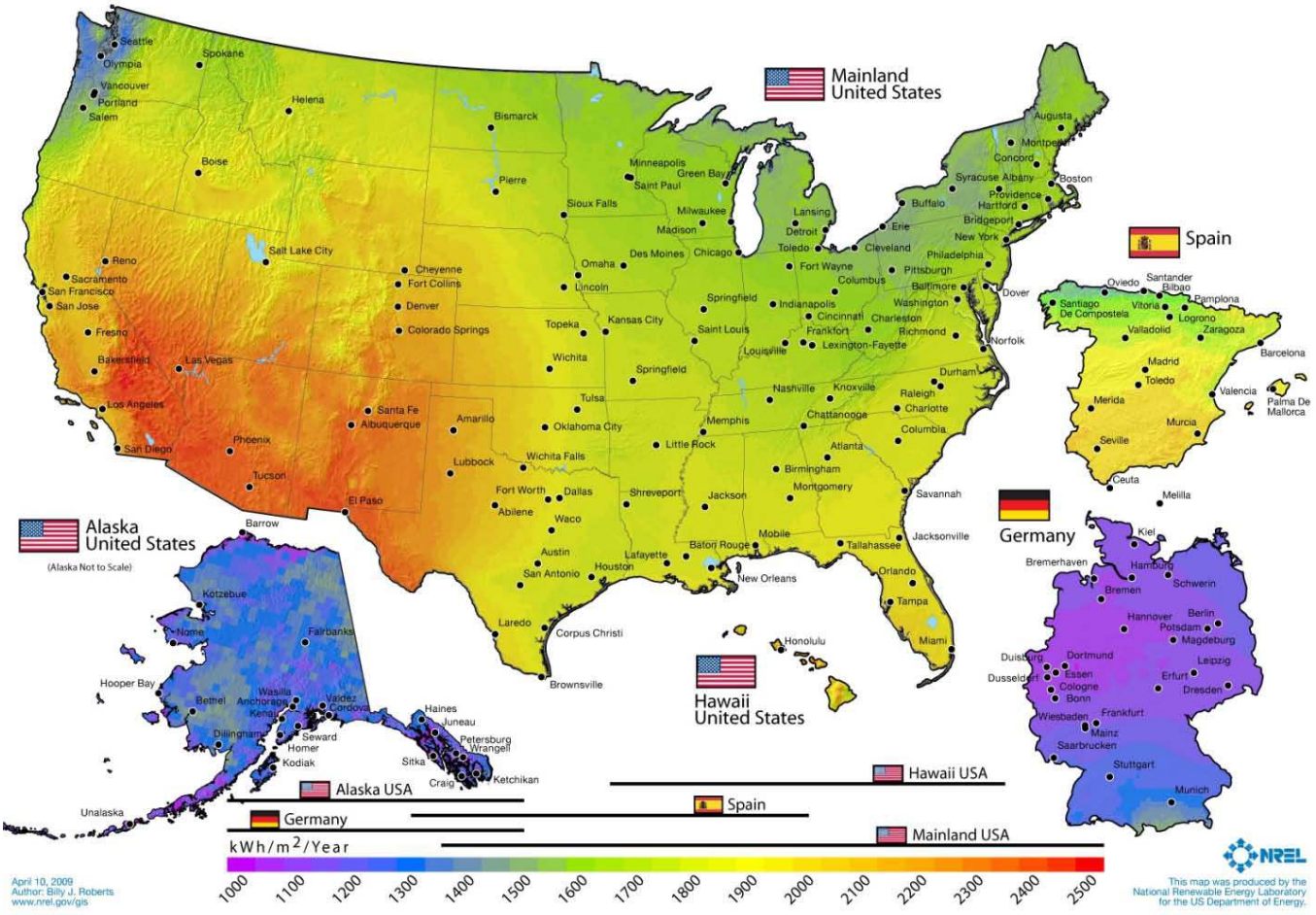



**Figure 7.3.7: Figure Solar Resources of the United States, Spain and Germany** The solar resources of the United States, Spain and Germany, expressed as solar insolation averaged over the year. The geographic variation of solar irradiation in the United States is only a factor of two, much less than for other renewable energy sources. *Source: U.S. Department of Energy, Energy Efficiency and Renewable Energy, 2008 Solar Technologies Market Report, DOE/GO-102010-2867 (January, 2010), p52.*

The placement of solar panels determines how their environmental impact. Solar arrays are often placed on roofs of buildings or over parking lots or integrated into construction in other ways. However, large systems may be placed on land and particularly in deserts where those fragile ecosystems could be damaged if care is not taken. Additionally, solar farms can compete for agricultural space.

Other downsides of solar energy are water consumption (for some uses) and generation of hazardous wastes. Large networks of mirrors and lenses that concentrate solar energy for electricity generation in thermal solar systems or for heating may need to be cleaned regularly with water. Water is also needed for cooling the turbine-generator. Using water from underground wells may affect the ecosystem in some arid locations. The manufacturing of photovoltaic cells generates some hazardous waste from the chemicals and solvents used in processing. Some solar thermal systems use potentially hazardous fluids (to transfer heat) that require proper handling and disposal. Nuclear power exceeds solar energy in water consumption and hazardous waste generation, however.

## Attribution

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

- Renewable Energy and Challenges and Impacts of Energy Use 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.

- Renewable Energy: Solar, Wind, Hydro and Biomass in *Sustainability a Comprehensive Foundation* by Tom Theis and Jonathan Tomkin

---

This page titled 7.3: Solar Energy is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **11.3: Renewable Energy Sources** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# 7.4: Hydropower

**Hydropower** (hydroelectric energy) refers to the energy in moving water, which can be used to generate electricity. It is ultimately driven by solar energy. Water at high elevations is replenished through evaporation (induced by solar heat), condensation, and precipitation. The moving water turns a turbine (figure **7.4.1**) and this moves magnets within a wire coil, causing the movement of electrons in a generator. The result is an electric current (electricity). Combustion of fossil fuels or biofuels, nuclear energy, wind energy, and thermal solar systems all follow this same general mechanism; the difference is what causes the turbine to spin.

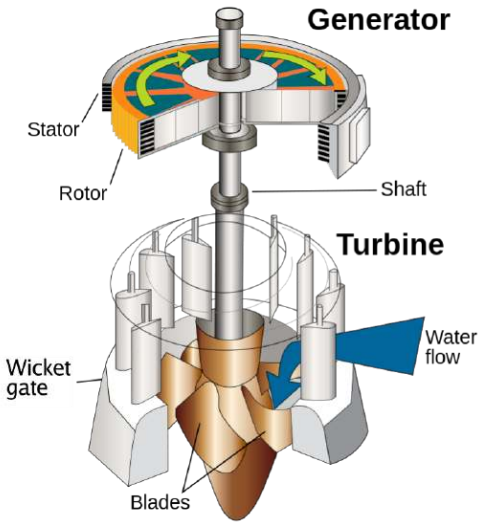



**Figure 7.4.1:** A water turbine. Moving water turns the blades of the turbine. The wicket gate controls the water flow. The moving water causes magnets in the rotor to spin and move electrons within the loops of wire in the stator, generating an electric current. Image by U.S. Army Corps of Engineers (public domain).

There are several types of hydropower, but each generates electricity through the same general mechanism described above. The most well-known use of hydropower involves dams. **Dams** block the flow of rivers, producing an artificial lake (reservoir) upstream. The release of water is then controlled. Some of this water moves through a channel (penstock), and this kinetic energy (energy of motion) is harnessed into electricity (figure **7.4.2**).

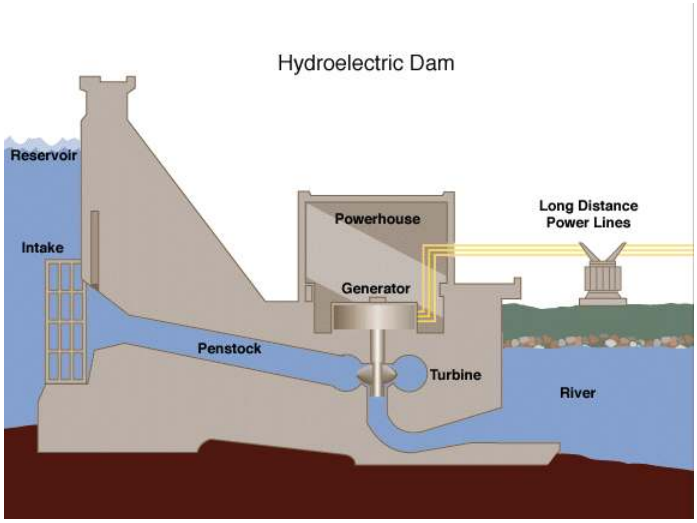



**Figure 7.4.2:** A section of a hydroelectric dam. Water in the reservoir passes through the intake into the penstock, which is a narrow channel. The moving water turns a turbine. This powers the generator in the powerhouse. The water continues down the river. Meanwhile electricity is distributed through long distance power lines. Image by Tennessee Valley Authority (public domain).

Because water is stored in a reservoir for large-scale hydropower systems with dams, these facilities have a large effect on river flow below the dam. This kind of facility has a large effect on the seasonal variation of below-dam riverflow, because it greatly reduces the spring peak flow while increasing the flow during the summer, fall, and winter (Figure **7.4.3**).

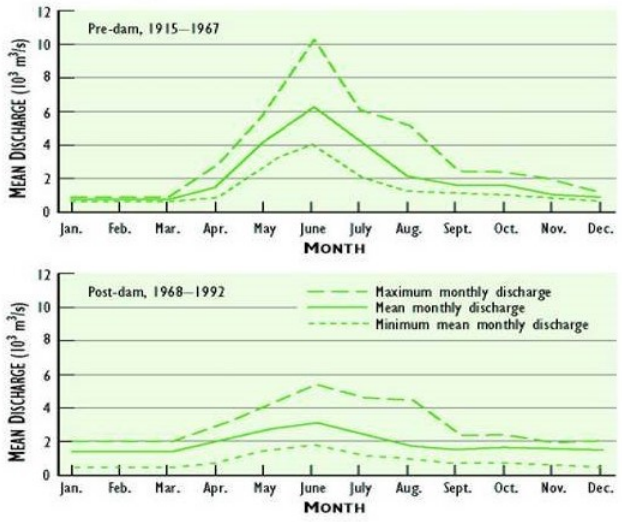



**Figure 7.4.3: Effects of Reservoir Development on Seasonal Flow of the Peace River**. The Bennett Dam and its associated reservoir were built in 1968, so the upper curves (for 1915–1967) represent pre-impoundment conditions. The flows were measured at the town of Peace River, Alberta. The most notable changes are (1) an overall reduction in the annual variation of riverflow; (2) a large reduction in peak flow during mid-April to mid-July; and (3) an increase during the low-flow period of September to March. Source: Modified from Rosenberg et al. (1997) by Bill Freedman.

In so-called peaking or **hydropeaking** systems water flow below the dam is altered throughout the day to increase flow of water (and therefore electricity generation) at times when demand for electricity is high and decrease flow of water at times when demand for electricity is lower (Figure **7.4.4**). If the peak-flow system causes large flow variations downriver, important ecological damage may be caused.




**Figure 7.4.4:** Effects of Peaking Discharges on Hourly Mean Discharge to the Nelson River. The data are for a one-week period in July 1984. This figure illustrates the changes in hydrology that occur downstream if the generation of electricity is timed to meet the peak daily demand. Note the relatively low demand during the weekend. Kelsey Dam is not subject to these peaking discharges, so it represents the background pattern of low flow in the summer. Source: Modified from Rosenberg et al. (1997) by Bill Freedman

Hydropower dams and the reservoirs they create degrades habitat and negatively impacts native species. For example, migration of fish to their upstream spawning areas can be obstructed by dams. In areas where salmon must travel upstream to spawn, such as along the Columbia River in Washington and Oregon, the dams block their way (figure **7.4.5**). This problem can be partially alleviated by using "fish ladders" that help salmon get around the dams (figure **7.4.6**). Fish traveling downstream, however, can get

killed or injured as water moves through turbines in the dam. Reservoirs and operation of dams can also affect aquatic habitats due to changes in water temperatures, water depth, chemistry, flow characteristics, and sediment loads, all of which can lead to significant changes in the ecology and physical characteristics of the river both upstream and downstream. When reservoirs first fill with water, they inundate terrestrial (land) habitats, farms, cities, and archeological and cultural sites, sometimes forcing populations to relocate. The terrestrial vegetation slowly decomposes in oxygen-poor conditions, releasing methane into the atmosphere, a potent greenhouse gas. In this sense, hydropower generates few air pollutants during operation, but construction does contribute to climate change.

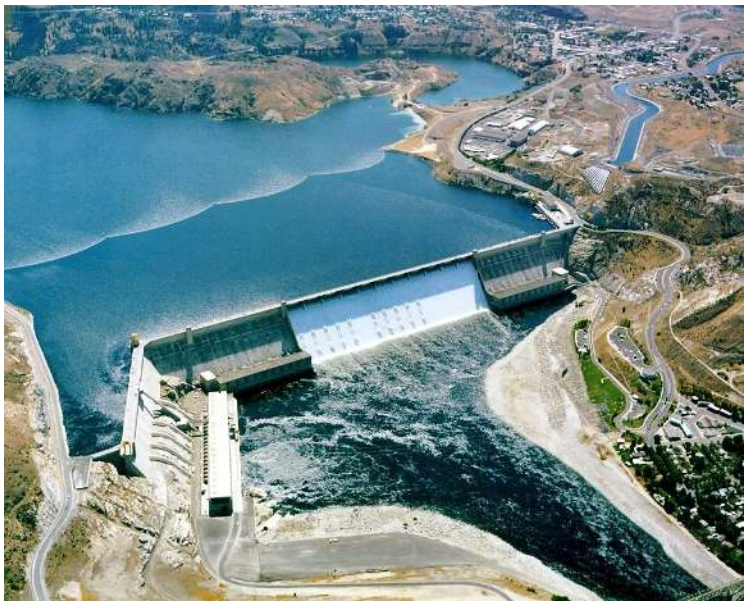


**Figure 7.4.5:** The Grand Coulee Dam on the Columbia River in Washington state created Lake Roosevelt, which provides outdoor recreation such as fishing and swimming. However, it compromised salmon populations and flooded sacred sites for people of the Spokane Tribe. Image by U.S. Bureau of Reclamation (public domain).

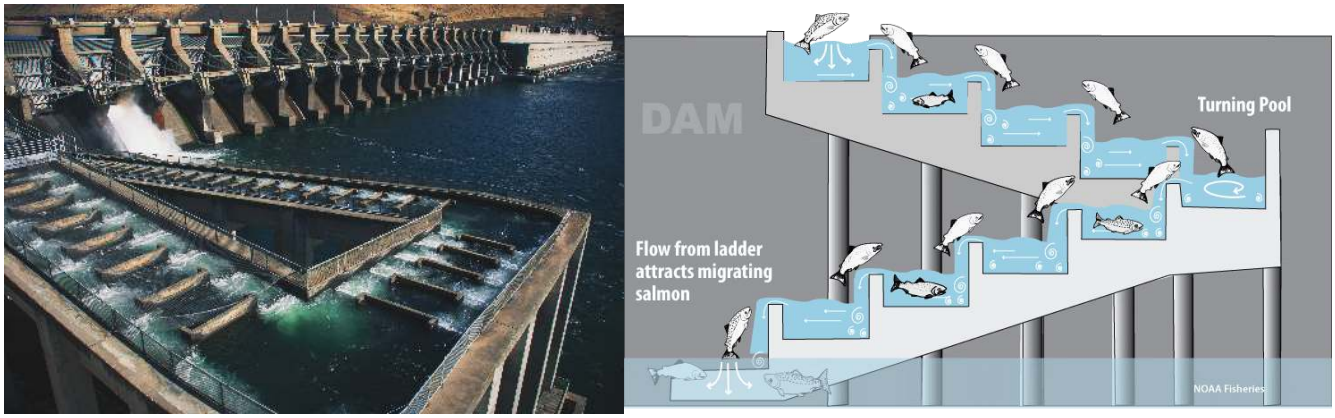



**Figure 7.4.6:** Left: The fish ladder at John Day Lock and Dam along the Columbia River in Oregon allows adult fish to migrate upstream of the dam on their own. Right: Diagram of a fish ladder, which allows the passage of salmon over a dam. Salmon swim upstream to spawn (reproduce) and are able to swim up the steps of the ladder. The flow from the ladder attracts migrating salmon. At the turning pool, they rotate and then proceed up the next flight of the ladder until they are over the dam. Left image and caption (modified) from U.S. Army Corps of Engineers, and right image by NOAA Fisheries (public domain).

Large-scale dam hydropower projects are often criticized for their impacts on wildlife habitat, fish migration, and water flow and quality. However, small **run-of- the-river projects** are free from many of the environmental problems associated with their large-scale relatives because they use the natural flow of the river, and thus produce relatively little change in the stream channel and flow. In **run-of-the-river hydroelectricity**, the turbine is placed directly into a river, and the natural flow of water through the river

spins the turbine. The dams built for some run-of-the-river projects are very small and impound little water, and many projects do not require a dam at all. Thus, effects such as oxygen depletion, increased temperature, decreased flow, and impeded upstream migration are not problems for many run-of-the-river projects.

Small hydropower projects offer power solutions for many remote communities throughout the world, such as those in Nepal, India, China, and Peru, as well as for highly industrialized countries like the United States. **Small hydropower systems** are those that have a capacity between 0.01 to 30 megawatt (MW) of power. For reference, a 1-MW generator operating for an hour (producing 1 MWh) generates enough electricity to powers the average home in the U.S. for a little over a month. Small hydropower systems that generate less than 0.1 MW are more specifically called **microhydropower** systems (figure **7.4.7**). Most of the systems used by home and small business owners would qualify as microhydropower systems. In fact, a 10 kW system generally can provide enough power for a large home, a small resort, or a hobby farm.

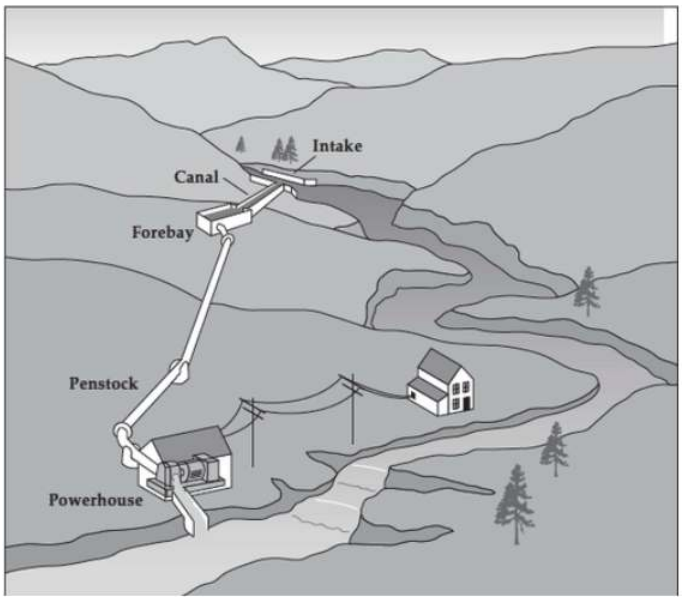



**Figure 7.4.7:** A microhydropower system. Although there are several ways to harness the moving water to produce energy, run-of-the-river systems, which do not require large storage reservoirs, are often used for microhydropower projects. For run-of-the-river hydro projects, a portion of a river's water is diverted to a channel, pipeline, or pressurized pipeline (penstock) that delivers it to a waterwheel or turbine. In this image, water from the river moves through the intake then to the canal, forebay, penstock, and powerhouse, which contains the turbine. The moving water rotates the wheel or turbine, which spins a shaft. The motion of the shaft can be used for mechanical processes, such as pumping water, or it can be used to produce electricity. In this image, electrical lines connect the powerhouse to a nearby house.

One benefit of using hydropower is that dams and reservoirs can additionally be used for recreation, flood control, and storing freshwater. While we will never run out of moving water, the potential to use hydropower fluctuates based on precipitation. For example, during a drought, water levels in a reservoir decreases, and there is less water released to generate electricity. As with most sources of renewable energy, hydropower is not practical everywhere, and it is most effective in mountainous regions with high rainfall and snowmelt.

**Tidal energy** and **energy from ocean waves** (figure **7.4.8**) are also considered forms of hydropower. However, tidal energy arises from the gravitational attraction of the moon and the more distant sun on the earth's oceans, combined with rotation of the earth. In other words, most forms of hydropower are indirect forms of solar energy, with the exception of tidal energy.

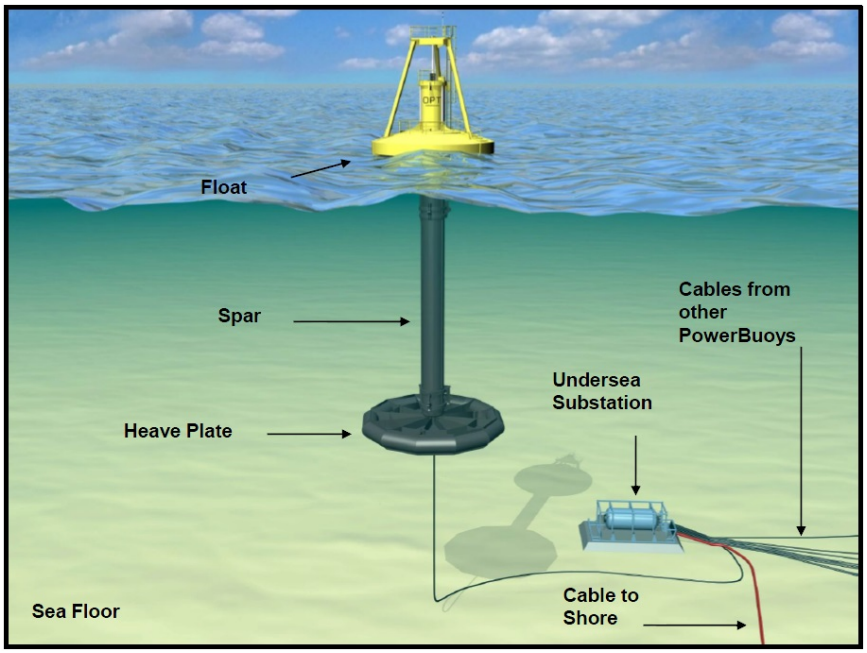



**Figure 7.4.8:** One way to harness wave energy is with an absorber. Absorbers extract energy from the rise and fall of the waves with a buoy. The buoy consists of the float, which is above the water and a heavy plate. The narrow spar connects the two. A cable connects the heavy plate of each buoy to an undersea substation, and another cable runs from the substation to the shore. Image and caption (modified) by OpenEI (public domain).

## Contributors and Attributions

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

- Renewable Energy and Challenges and Impacts of Energy Use 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.
- Additional Problems of Surface Waters from *Environmental Science: A Canadian Perspective* by Bill Freedman.

This page titled 7.4: Hydropower is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 7.5: Geothermal Energy

**Geothermal energy** originates from heat rising to the surface from Earth's molten iron core created during the formation and compression of the early Earth as well as from heat produced continuously by radioactive decay of uranium, thorium, and potassium in the Earth's crust. **Geothermal power plants** harness this heat energy to produce electricity much in the same way that heat from burning coal generates energy (figure **7.5.1 – 3**). Water is injected underground and heated. The steam that emerges can be used directly, the heat can be transferred to closed system of another fluid, which then boils (figure **7.5.3**). Either way, the steam (or other high-pressure gas) ultimately turns a turbine and powers a generator.



**Figure 7.5.1:** Steam is released from a geothermal power plant. Image by Open Access Government (CC-BY).

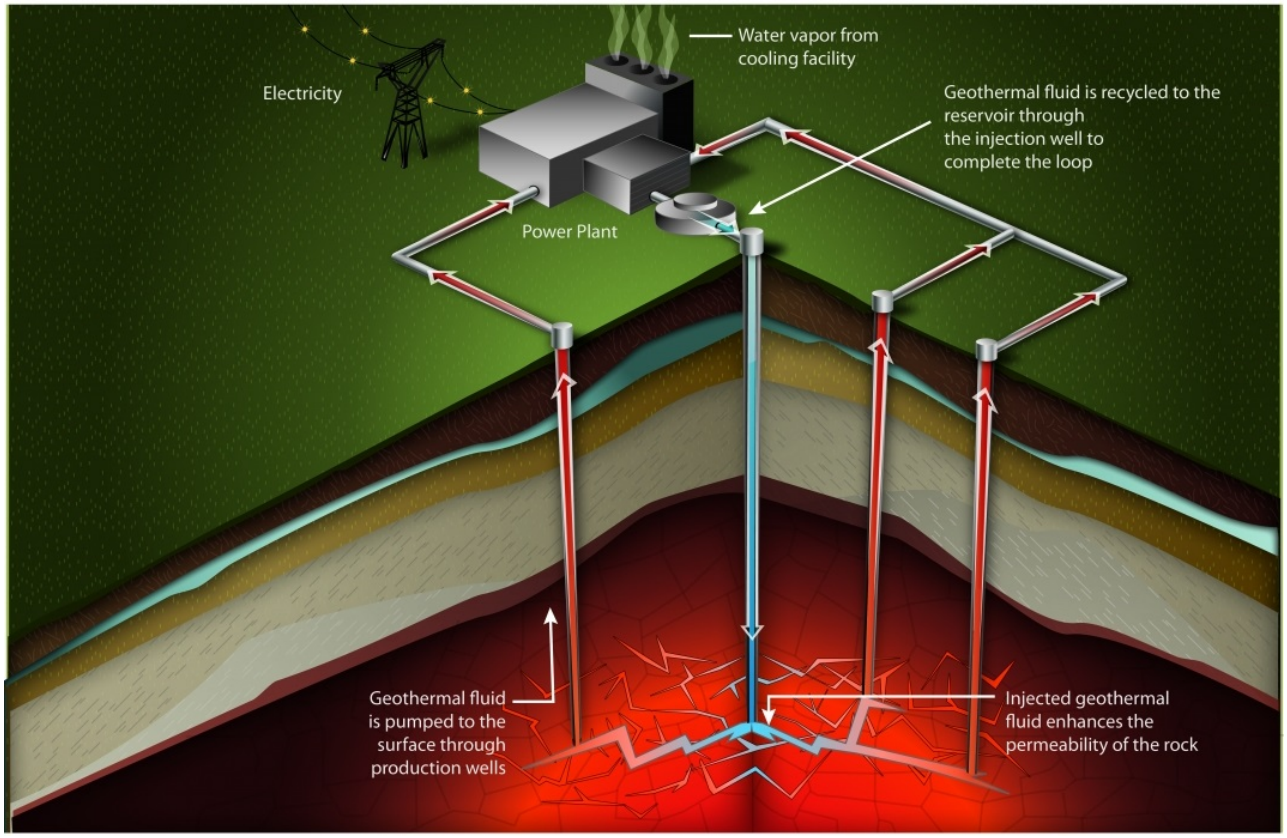


**Figure 7.5.2:** In a geothermal power plant, geothermal fluid is injected underground. (Geothermal water refers to the hot water/steam used in this process.) It enhances the permeability of the rocks and is heated underground. It exits to the surface through production wells, where it is used to generate electricity in the geothermal power plant. Water vapor is released from the cooling facility of the power plant. To recycle the geothermal fluid, it is returned to the reservoir and injected back into the ground. Image from Office of Energy Efficiency and Renewable Energy/U.S. Department of Energy (public domain).

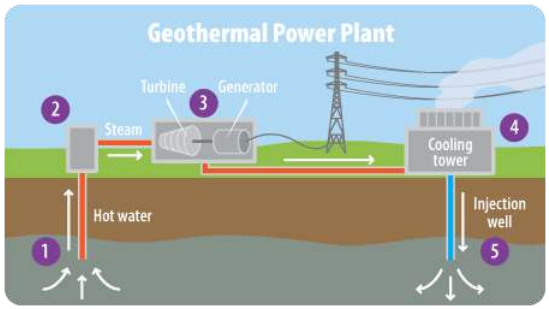



**Figure 7.5.3:** (1) In a geothermal power plant, hot water is pumped from deep underground through a well under high pressure. (2) When the water reaches the surface, the pressure is dropped, which causes the water to turn into steam. (3) The steam spins a turbine, which is connected to a generator that produces electricity. (4) The steam cools off in a cooling tower and condenses back to water. (5) The cooled water is pumped back into the Earth to begin the process again. Image and caption (modified) from EPA (public domain).

**Geothermal heat pumps** (**ground-source heat pumps**) rely on cool temperatures underground to cool or heat homes (figure **7.5.4**). They are sometimes considered a second type of geothermal energy, but they are also a means of energy conservation. Geothermal heat pumps use a heat-exchange system that runs in the subsurface about 20 feet (5 meters) below the surface, which is consistently cool (around 55°F, or 12.5° C). Fluid is pumped underground and then along ducts in the home. This cools the house during the summer, acting as a heat sink. During a cold winter, it warms the house to 55° F (acting as a heat source), and traditional heating systems do the rest. This reduces the energy consumption required to generate heat from gas, steam, hot water, and conventional electric air-conditioning systems.

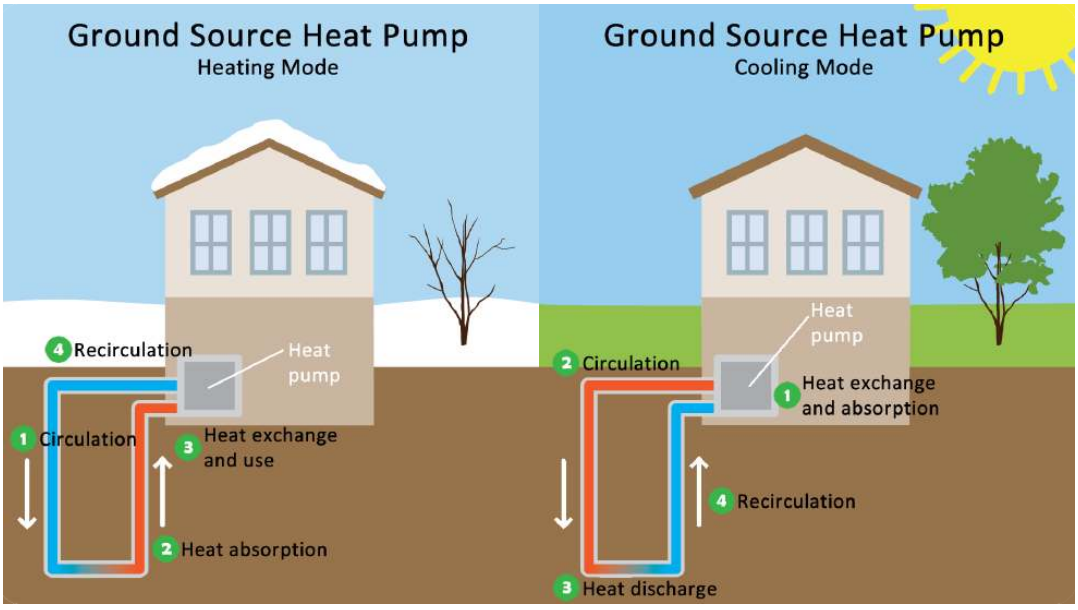



**Figure 7.5.4:** A ground source heat pump (geothermal heat pump) in heating mode (during the winter; left) and in cooling mode (during the summer; right). There are four steps for heating mode (left). (1) Circulation: The above-ground heat pump moves water or another fluid through a series of buried pipes or ground loops. (2) Heat absorption: As the fluid passes through the ground loop, it absorbs heat from the warmer soil, rock, or ground water around it. (3) Heat exchange and use: The heated fluid returns to the building where it used for space or water heating. The system uses a heat exchanger to transfer heat into the building's existing air

handling, distribution, and ventilation system. (4) Recirculation: Once the fluid transfers its heat to the building, it returns at a lower temperature to the ground loop to be heated again. This process is repeated, moving heat from one point to another. The four steps for cooling mode (right) are similar: (1) heat exchange and absorption, (2) circulation, (3) heat discharge, and (4) recirculation. Images and caption (modified) from EPA (public domain).

This video explains the construction and mechanism of geothermal heat pumps.



Not only does geothermal energy have multiple applications (generating electricity, heating, and cooling), but it is reliable. While solar and wind energy are intermittent, heat is consistently radiated from deep underground. Additionally, the cool temperatures closer to the surface needed for geothermal heat pumps are present year round and at all locations. Geothermal power plants for electricity generation, however, can only be built in specific locations where hot magma is close enough to the Earth's surface. These locations are typically associated with geysers, hot springs, or volcanoes (figure 7.5.5). Additionally, geothermal power plants are costly to build.



**Figure 7.5.5:** Locations of the world's volcanoes. They are concentrated along the west coasts of North American and South America as well as just east of Asia and Australia. These regions, which wrap around the Pacific Ocean, form what is known as the Ring of Fire. Image by Salazar, S.S., Muñoz, Y. & Ospino, A. Analysis of geothermal energy as an alternative source for electricity in Colombia. *Geotherm Energy* **5,** 27 (2017). (CC-BY)

The environmental impact of geothermal energy depends on how it is being used. The use of geothermal heat pumps has almost no negative impact on the environment. Geothermal power plants do not burn fuel to generate electricity, so they generate minimal air pollution. They release less than 1% of the carbon dioxide emissions of a fossil fuel plant. Geothermal plants plants use scrubber systems to clean the air of hydrogen sulfide that is naturally found in the steam and hot water. They emit 97% less sulfur compounds (one cause of acid deposition/acid rain) than are emitted by fossil fuel plants. After the steam and water from a

geothermal reservoir have been used, they are injected back into the Earth. One environmental concern associated with geothermal power plants is geothermal drilling during their construction has caused earthquakes, similar to the effects of injection wells for fracking.

## Attribution

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

- Renewable Energy and Challenges and Impacts of Energy Use 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 7.5: Geothermal Energy is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 7.6: Wind Energy

**Wind energy** arises from the motion of the air. It is driven by solar energy (differences in air temperature causes air currents). The wind turns a turbine, which powers a generator (figure **7.6.1**). The rotor blades of a wind turbine work like an airplane wing or helicopter rotor blade. When wind flows across the blade, the air pressure on one side of the blade decreases, and this causes the rotor to spin. The rotor connects to the generator, either directly or through a series of gears that speed up the rotation and allow for a physically smaller generator. Similar to the electricity generation from coal, natural gas, or nuclear energy, the rotating motion causes magnets spin within wire coils to produce electricity. According to the American Wind Energy Association, 39% of all new electrical generating capacity in the United States in 2019 was due to wind.

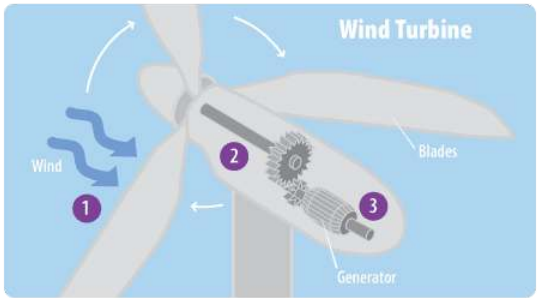



**Figure 7.6.1:** (1) As the wind blows over the blades of a wind turbine, it causes the blades to lift and rotate. (2) The rotating blades turn a shaft that is connected to a generator. (3) The generator creates electricity as it turns. Image and caption (modified) from EPA (public domain).

## Advantages of Wind Energy

Wind is among the lowest-cost sources of renewable energy, and its expansion creates jobs (figure **7.6.2**). Like many renewable energy sources, wind turbines do not release air pollutants or contribute to climate change, and they do not require water for cooling. Because a wind turbine has a small physical footprint relative to the amount of electricity it produces, many wind farms are located on crop and pasture land. They contribute to economic sustainability by providing extra income to farmers and ranchers, allowing them to stay in business and keep their property from being developed for other uses. For example, energy can be produced by installing wind turbines in the Appalachian mountains of the United States instead of engaging in mountain top removal for coal mining. Offshore wind turbines on lakes or the ocean may have smaller environmental impacts than turbines on land, and winds are up to 50% stronger and steadier offshore than on land (figure **7.6.3**).

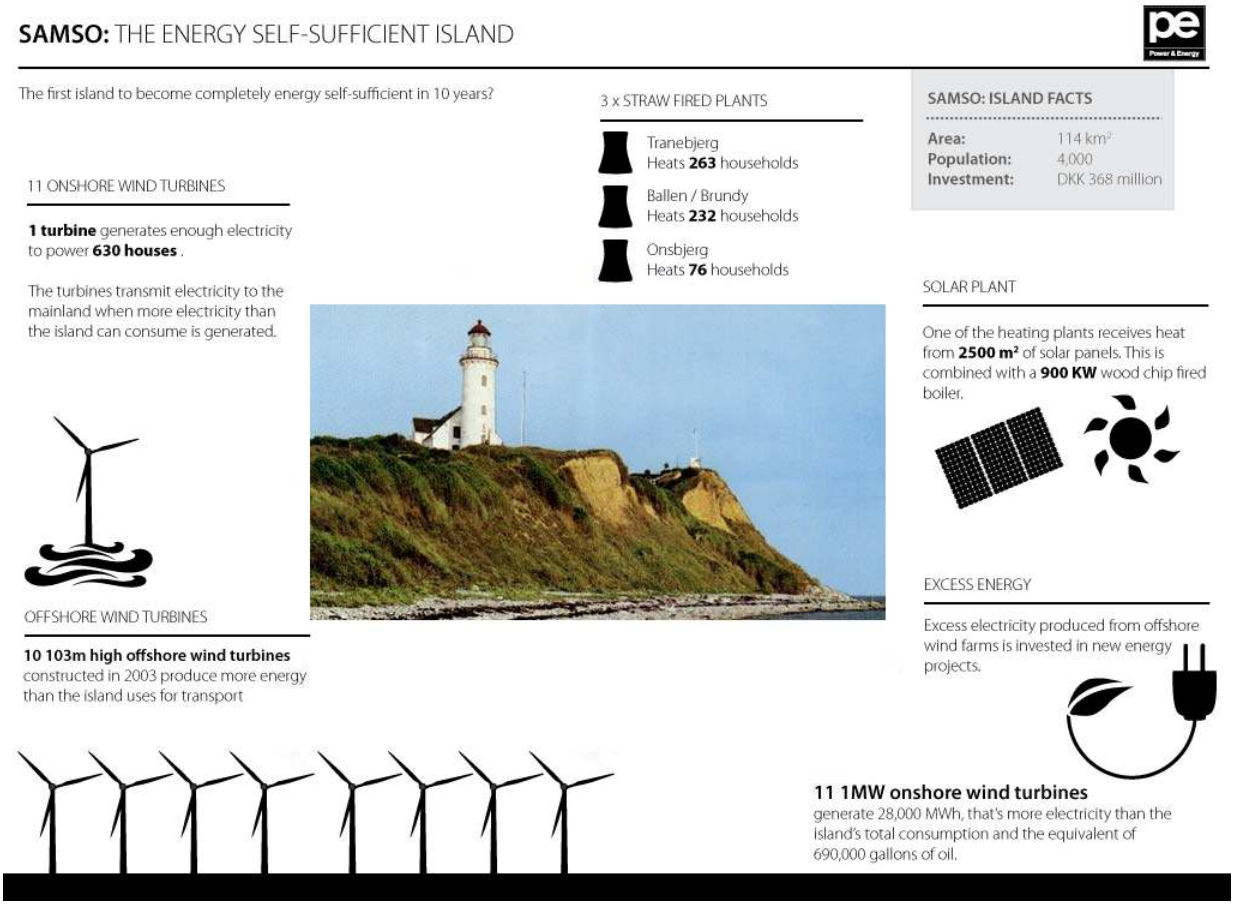



**Figure 7.6.2:** The Danish island of Samsø was the first region in the country to transition to 100% renewable energy, relying on offshore wind, onshore wind, straw (a biofuel), and solar. The text in the image is, "Samsø : The Energy Self-sufficient Island. The first island to become complete energy self-sufficient in 10 years. Samsø: Island Facts— Area: 114 $km^2$, Population: 4,000, Investment: DKK 368 million. 11 Onshore Wind Turbines— 1 turbine generates enough electricity to power 630 houses. The turbines transmit electricity to the mainland when more electricity than the island can consume is generated. Offshore Wind Turbines- 10 103-m-high offshore wind turbines constructed in 2003 produce more energy than the island uses for transport. 11 1-MW Onshore Wind Turbines generate 28,000 MWh. That's more electricity than the island's total consumption and the equivalent of 690,000 gallons of oil. 3X Straw-fired Plants— Traneberg- heats 263 households, Ballen/Brundy- heats 232 households, Onsbjerg- heats 76 households. Solar Plant- One of the heating plants receives heat from 2500 $m^2$ of solar panels. This is combined with a 900 KW wood chip-fired boiler. Excess Energy- Excess electricity produced from offshore wind farms is invested in new energy projects. Images by GDS Infographics (CC-BY) and National Geospatial-Intelligence Agency (public domain).



**Figure 7.6.3:** Wind turbines may be on land (onshore, left) or in the ocean (offshore, right). Left image by Joshua Winchell/USFWS (public domain). Right image by CGP Grey (CC-BY).

## Disadvantages of Wind Energy

Wind energy does have a few challenges. Wind turbines are only effective in regions with strong enough winds to generate sufficient electricity. Even in regions with strong winds, wind availability is intermittent. This can be mitigated by the use batteries to store energy, but battery capacity, despite continuous technological advances, is still limited. There are aesthetic concerns to some people when they see them on the landscape, and some people do not like the sound that wind turbine blades make. A few wind turbines have caught on fire, and some have leaked lubricating fluids, though this is relatively rare. Turbines have been found to cause bird and bat deaths particularly if they are located along their migratory path, although communication towers and domestic cats are bigger threats. There are some small impacts from the construction of wind projects or farms, such as the construction of service roads, the production of the turbines themselves, and the concrete for the foundations.

The wind resource of the United States, while abundant, is less homogeneous. Strong winds require steady gradients of temperature and pressure to drive and sustain them, and these are frequently associated with topological features such as mountain ranges or coastlines. The onshore wind map of the United States shows this pattern, with the best wind along a north-south corridor roughly at mid-continent (Figure **7.6.4**). Offshore winds over the Great Lakes and the east and west coasts are stronger and steadier though they cover smaller areas. The technical potential for onshore wind is over 8000 GW of capacity (Lu, 2009; Black & Veatch, 2007) and offshore is 800 – 3000 GW (Lu, 2009; Schwartz, Heimiller, Haymes, & Musial, 2010). For comparison, the United States used electricity in 2009 at the rate of 450 GW averaged over the day-night and summer-winter peaks and valleys.

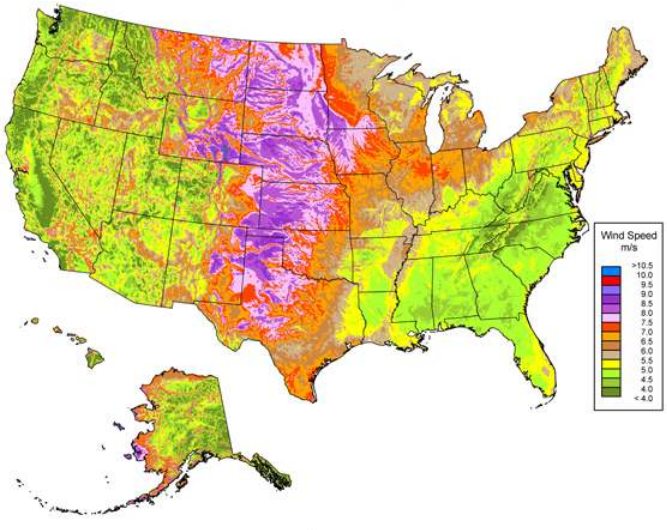



**Figure 7.6.4: 80 Meter Wind Resource Map** Figure shows the average wind speeds in the United States at 80 meters. Also see offshore wind resource maps. *Source: U.S. Department of Energy, National Renewable Energy Laboratory and AWS Truepower LLC*

## Attribution

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

- Renewable Energy and Challenges and Impacts of Energy Use from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- The Power of Wind. Office of Energy Efficiency and Renewable Energy. U.S. Department of Energy. Accessed 01-17-2021 (public domain).
- Renewable Energy: Solar, Wind, Hydro and Biomass in *Sustainability a Comprehensive Foundation* by Tom Theis and Jonathan Tomkin
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.

This page titled 7.6: Wind Energy is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 7.7: Biofuels (Biomass Energy)

**Biofuels** (biomass energy) contain energy produced from organisms, such as animal waste, plants, or algae. It is another indirect form of solar energy. Biofuels have many uses. They be burned directly or first converted to ethanol (often with the help of bacteria and fungi) to generate electricity. The heat from combustion produces steam and turns a turbine to power a generator. Biodiesel offers an alternative to petrochemicals for fueling vehicles. Biofuels have even been used to power small planes (figure **7.7.*a***). Furthermore, burning wood or straw provides heating.



**Figure 7.7.*a*:** Test flight of the "Green Hornet", a Navy fighter jet partially powered by biofuels. Image by U.S. Navy (public domain).

Unlike fossil fuels, biofuels are **carbon neutral** (figure **7.7.*b***). Fossil fuels store carbon that was captured by organisms millions of years ago. When we burn them, carbon dioxide is released much more rapidly than it was removed. Biofuels removed carbon dioxide from the atmosphere more recently, and they form over shorter time spans. When biofuels are burned, this carbon dioxide that was recently removed is released back into the atmosphere.

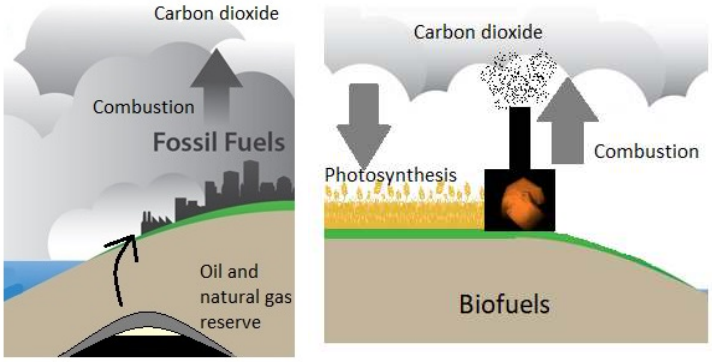



**Figure 7.7.*b*:** Fossil fuels, such as oil and natural gas, contain carbon that has been stored underground for millions of years. Combustion (burning) of these fossil fuels releases carbon dioxide into the atmosphere, contributing to climate change. When biofuels are grown, they capture carbon dioxide through photosynthesis. Burning them or their products for heat, transportation, or electricity generation release this carbon dioxide, which was only recently captured. For this reason, biofuels are considered carbon neutral. Images created by Melissa Ha from NETL/DOE (public domain) and publicdomainvectors.org (public domain).

Another advantage of biofuels is that they can be produced locally and can be cultivated in many different locations. On the other hand, they occupy space that could otherwise be used for food production. To further complicate matters, the characteristics that

make a plant species ideal for biofuel (such as being resistant to pests and fast-growing) are also characteristics that help invasive species thrive. Care must taken to contain these species if grown outside of their native ranges.

Combustion of solid municipal waste (see below) or animal wastes as biofuels, reduces waste and generates electricity simultaneously. Unlike most forms of renewable energy, however, combustion of biofuels pollutes the air. (The carbon dioxide released is not an issue since biofuels are carbon neutral, but other air pollutants are also released.) In fact, indoor air pollution from fires used for cooking inside the home is a major cause of death in developing countries.

Each type of biomass must be evaluated for its environmental and social impact in order to determine if it is really advancing sustainability and reducing environmental impacts. For example, cutting down large swaths of forests just for energy production is not a sustainable option because our energy demands are so great that we would quickly deforest the world, destroying critical habitat. For biomass to be a sustainable option, it usually needs to come from waste material, such as lumber mill sawdust, paper mill sludge, yard waste, or oat hulls from an oatmeal processing plant, livestock manure, or trash. These materials would otherwise just accumulate or decompose. Several examples of biofuel use are discussed below in more detail, including the specific advantages and disadvantages of type of use.

## Burning Wood

Using wood, and charcoal made from wood, for heating and cooking can replace fossil fuels and may result in lower carbon dioxide emissions. If wood is harvested from forests or woodlots that have to be thinned or from urban trees that fall down or needed be cut down anyway, then using it for biomass does not impact those ecosystems. However, wood smoke contains harmful pollutants like carbon monoxide and particulate matter (see Air Pollution).

For home heating, it is most efficient and least polluting when using a modern wood stove or fireplace insert that are designed to release small amounts of particulates. However, in places where wood and charcoal are major cooking and heating fuels such as in developing countries, the wood may be harvested faster than trees can grow resulting in deforestation (figure **7.7.*c***). The largest share of biofuel use comes from traditional biomass, mostly fuel wood gathered for household cooking and heating, often without regard for sustainable replacement.




**Figure 7.7.*c*:** Charcoal production in Uganda. (A) Trees cut down for charcoal. (B) Pieces of firewood. (C) Wood logs read for charcoal burning. (D) Charcoal packed and ready for market. Image and caption (modified) from Bamwesigye et al. *Sustainability* **2020**, 12(20), 8337. (CC-BY)

Biomass can be used in small power plants. For instance, Colgate College has had a wood-burning boiler since the mid-1980s (figure **7.7.*d***). In one year it processed approximately 20,000 tons of locally and sustainably harvested wood chips, the

equivalent of 1.17 million gallons (4.43 million liters) of fuel oil, avoiding 13,757 tons of emissions and saving the university over $1.8 million in heating costs. The University's steam-generating wood-burning facility now satisfies more than 75% of the campus's heat and domestic hot water needs.



**Figure 7.7.*d*:** Wood chips are one example of a biofuel that can be burned to generate electricity. Source: Ulrichulrich

## Municipal Solid Waste

**Municipal solid waste** (MSW) is commonly known as garbage and can create electricity by burning it directly or by burning the methane produced as it decays. Waste-to-energy processes are gaining renewed interest as they can solve two problems at once: disposal of waste and production of energy from a renewable resource. Many of the environmental impacts are similar to those of a coal plant: air pollution, ash generation, etc. Because the fuel source is less standardized than coal and hazardous materials may be present in MSW, incinerators and waste-to-energy power plants need to clean the gases of harmful materials. The Environmental Protection Agency in the U.S. regulates these plants very strictly and requires anti-pollution devices to be installed. Also, while incinerating at high temperature, many of the toxic chemicals may break down into less harmful compounds. The ash from these plants may contain high concentrations of various metals that were present in the original waste. If ash is clean enough it can be "recycled" as an MSW landfill cover or to build roads, cement blocks, and artificial reefs (similar to coral reefs, but built by humans).

## Landfill Gas (Biogas)

Landfill gas (**biogas**) is a sort of human-made "biogenic" gas as discussed above (figure **7.7.*e***). Methane is formed as a result of biological processes in sewage treatment plants, waste landfills, anaerobic composting, and livestock manure management systems. This gas is captured and burned to produce heat or electricity. The electricity may replace electricity produced by burning fossil fuels, reducing carbon dioxide emissions. The only environmental impacts are from the construction of the plant itself, similar to that of a natural gas plant.

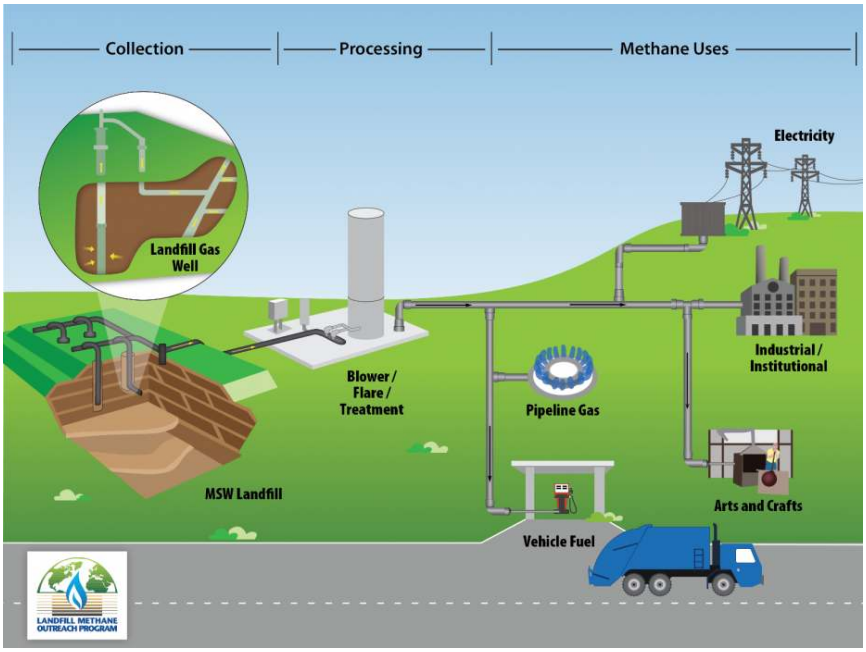



**Figure 7.7.*e*:** The collection and processing of landfill gas to produce methane for multiple uses. First, landfill gas is collected from the landfill gas well. A system of vertical and horizontal piping buried in an municipal solid waste landfill. The landfill gas is then processed and treated for use (blower/flare/treatment). Potential end uses include industrial/institutional uses, arts and crafts, pipeline gas, and vehicle fuel. Image and caption (modified) from Landfill Methane Outreach Program/EPA (public domain).

## Bioethanol and Biodiesel

**Bioethanol** and **biodiesel** are liquid biofuels manufactured from plants, typically crops. Bioethanol can be easily fermented from sugar cane juice, as is done in Brazil. Additionally, it can be fermented from broken down corn starch, as is mainly done in the United States.

The economic and social effects of growing plants for fuels need to be considered, since the land, fertilizers, and energy used to grow biofuel crops could be used to grow food crops instead. The competition of land for fuel versus food can increase the price of food, which has a negative effect on society. It could also decrease the food supply increasing malnutrition and starvation globally. Also, in some parts of the world, large areas of natural vegetation and forests have been cut down to grow sugar cane for bioethanol and soybeans and palm-oil trees to make biodiesel. This is not sustainable land use. Derived biofuels from parts of plants not used for food, such as stalks, reduces their environmental impact. Biodiesel can be made from used vegetable oil and has been produced on a very local basis. Compared to diesel, a petrochemical derived from crude oil, biodiesel combustion produces less sulfur oxides, particulate matter, carbon monoxide, and unburned and other hydrocarbons, but it produces more nitrogen oxide (see Air Pollution).

Liquid biofuels typically replace petroleum and are used to power vehicles (figure 7.7.*f*). Although ethanol-gasoline mixtures burn cleaner than pure gasoline, they also are more volatile and thus have higher “evaporative emissions” from fuel tanks and dispensing equipment. These emissions contribute to the formation of harmful, ground level ozone and smog (see Air Pollution). Gasoline requires extra processing to reduce evaporative emissions before it is blended with ethanol.




**Figure 7.7.*f*:** A bus running on soybean biodiesel. Image and caption (modified) from U.S Department of Energy (public domain).

## Attribution

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

- Renewable Energy and Challenges and Impacts of Energy Use 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 7.7: Biofuels (Biomass Energy) is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 7.8: The Electrical Grid, Battery Challenges, and Energy Conservation

## Electricity Grid

Over the past century and a half **electricity** has emerged as a popular and versatile energy carrier. Today, electricity is exploited not only for its diverse end uses such as lighting, motion, refrigeration, communication and computation, but also as a primary carrier of energy. Electricity is one of two backbones of the modern energy system (liquid transportation fuels are the other), carrying high density energy over short and long distances for diverse uses. In 2009, electricity consumed the largest share of the United States' primary energy, 38 percent, with transportation a close second at 37 percent. These two sectors also accounted for the largest shares of U.S. carbon emissions, 38 percent for electricity and 33 percent for transportation.

By far most electricity is generated by combustion of fossil fuels to turn steam or gas turbines. This is the least efficient step in the energy chain, converting only 36 percent of the chemical energy in the fuel to electric energy, when averaged over the present gas and coal generation mix. It also produces all the carbon emissions of the electricity chain. Beyond production, electricity is a remarkably clean and efficient carrier. Conversion from rotary motion of the turbine and generator to electricity, the delivery of electricity through the power grid, and the conversion to motion in motors for use in industry, transportation and refrigeration can be more than 90 percent efficient. None of these steps produces greenhouse gas emissions. It is the post-production versatility, cleanliness, and efficiency of electricity that make it a prime energy carrier for the future. Electricity generation, based on relatively plentiful domestic coal and gas, is free of immediate fuel security concerns. The advent of electric cars promises to increase electricity demand and reduce dependency on foreign oil, while the growth of renewable wind and solar generation reduces carbon emissions. The primary sustainability challenges for electricity as an energy carrier are at the production step: efficiency and emission of carbon dioxide and toxins.

### The Electricity Grid: Capacity and Reliability

Beyond production, electricity faces challenges of capacity, reliability, and implementing storage and transmission required to accommodate the remoteness and variability of renewables. The largest capacity challenges are in urban areas, where 79 percent of the United States and 50 percent of the world population live. The high population density of urban areas requires a correspondingly high energy and electric power density. In the United States, 33 percent of electric power is used in the top 22 metro areas, and electricity demand is projected to grow 31 percent by 2035. This creates an "urban power bottleneck" where underground cables become saturated, hampering economic growth and the efficiencies of scale in transportation, energy use and greenhouse gas emission that come with high population density. Saturation of existing cable infrastructure requires installation of substantial new capacity, an expensive proposition for digging new underground cable tunnels.

The **reliability** of the electricity grid presents a second challenge. The United States' grid has grown continuously from origins in the early 20th Century; much of its infrastructure is based on technology and design philosophy dating from the 1950s and 1960s, when the major challenge was extending electrification to new rural and urban areas. Outside urban areas, the grid is mainly above ground, exposing it to weather and temperature extremes that cause most power outages. The response to outages is frustratingly slow and traditional – utilities are often first alerted to outages by telephoned customer complaints, and response requires sending crews to identify and repair damage, much the same as we did 50 years ago. The United States' grid reliability is significantly lower than for newer grids in Europe and Japan, where the typical customer experiences ten to 20 times less outage time than in the United States. Reliability is especially important in the digital age, when an interruption of even a fraction of a cycle can shut down a digitally controlled data center or fabrication line, requiring hours or days to restart.

Reliability issues can be addressed by implementing a **smart grid** with two-way communication between utility companies and customers that continuously monitors power delivery, the operational state of the delivery system, and implements demand response measures adjusting power delivered to individual customers in accordance with a previously established unique customer protocol. Such a system requires installing digital sensors that monitor power flows in the delivery system, digital decision and control technology and digital communication capability like that already standard for communication via the Internet. For customers with on-site solar generation capability, the smart grid would monitor and control selling excess power from the customer to the utility.

Figure 7.8.1 illustrates the two-way communication features of the smart grid. The conventional grid in the upper panel sends power one way, from the generating station to the customer, recording how much power leaves the generator and arrives at the

customer. In the smart grid, the power flow is continuously monitored, not only at the generator and the customer, but also at each connection point in between. Information on the real time power flow is sent over the Internet or another special network to the utility and to the customer, allowing real time decisions on adding generation to meet changes in load, opening circuit breakers to reroute power in case of an outage, reducing power delivered to the customer during peak periods to avoid outages (often called "demand response"), and tracking reverse power flows for customers with their own solar or other generation capacity. The conventional power grid was designed in the middle of the last century to meet the simple need of delivering power in one direction. Incorporating modern Internet-style communications and control features could bring the electricity grid to a qualitatively new level of capability and performance required to accommodate local generation and deliver higher reliability.




**Figure 7.8.1: Smart Grid** The addition of real-time monitoring and communicating capability like that used on the Internet would add 'smart' operation of the electricity grid. *Source:* *National Institute of Standards and Technology*

Smart components incorporated throughout the grid would be able to detect overload currents and open breakers to interrupt them quickly and automatically to avoid unnecessary damage and triggering a domino effect cascade of outages over wide areas as happened in the Northeast Blackout of 2003. For maximum effectiveness, such smart systems require fast automatic response on millisecond time scales commensurate with the cycle time of the grid. Even simple digital communication meets this requirement, but many of the grid components themselves cannot respond so quickly. Conventional mechanical circuit breakers, for example, take many seconds to open and much longer to close. Such long times increase the risk of dangerous overload currents damaging the grid or propagating cascades. Along with digital communications, new breaker technology, such as that based on fast, self-healing superconducting fault current limiters, is needed to bring power grid operation into the modern era.

## Integrating Renewable Electricity on the Grid

Accommodating renewable electricity generation by wind and solar plants is among the most urgent challenges facing the grid. Leadership in promoting renewable electricity has moved from the federal to the state governments, many of which have legislated Renewable Portfolio Standards (RPS) that require 20 percent of state electricity generation to be renewable by 2020. 30 states and the District of Columbia have such requirements, the most aggressive being California with 33 percent renewable electricity required by 2020 and New York with 30 percent by 2015. To put this legal requirement in perspective, wind and solar now account for about 1.6 percent of U.S. electricity production; approximately a factor of ten short of the RPS requirements. (Crabtree & Misewich, 2010).

The grid faces major challenges to accommodate the variability of wind and solar electricity. Without significant storage capacity, the grid must precisely balance generation to demand in real time. At present, the variability of demand controls the balancing process: demand varies by as much as a factor of two from night to day as people go through their daily routines. This predictable variability is accommodated by switching reserve generation sources in and out in response to demand variations. With **renewable generation**, variation can be up to 70 percent for solar electricity due to passing clouds and 100 percent for wind due to calm days, much larger than the variability of demand. At the present level of 1.6 percent wind and solar penetration, the relatively small variation in generation can be accommodated by switching in and out conventional resources to make up for wind and solar fluctuations. At the 20 percent penetration required by state Renewable Portfolio Standards, accommodating the variation in generation requires a significant increase in the conventional reserve capacity. At high penetration levels, each addition of wind or solar capacity requires a nearly equal addition of conventional capacity to provide generation when the renewables are quiescent.

This double installation to insure reliability increases the cost of renewable electricity and reduces its effectiveness in lowering greenhouse gas emissions.

A major complication of renewable variation is its unpredictability. Unlike demand variability, which is reliably high in the afternoon and low at night, renewable generation depends on weather and does not follow any pattern. Anticipating weather-driven wind and solar generation variability requires more sophisticated forecasts with higher accuracy and greater confidence levels than are now available. Because today's forecasts often miss the actual performance target, additional conventional reserves must be held at the ready to cover the risk of inaccuracies, adding another increase to the cost of renewable electricity. Storage of renewable electricity offers a viable route to meeting the variable generation challenge.

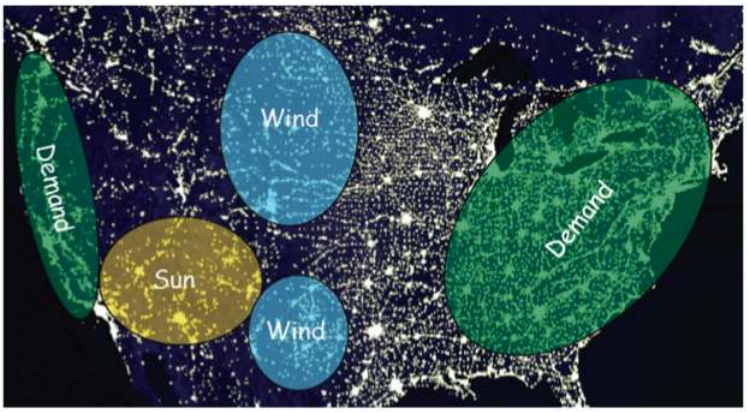



**Figure 7.8.2: Renewable Resource Location vs. Demand Location** Wind and solar electricity resources are located far from population centers, requiring a dramatic improvement in long-distance electricity transmission – an "interstate highway system for electricity." *Source:* *Integrating Renewable Electricity on the Grid, Report of the Panel on Pubic Affairs, American Physical Society (2010)*.

The final challenge for accommodating renewables is long distance transmission (Figure **7.8.2**). Although long distance delivery is possible where special high voltage transmission lines have been located, the capacity and number of such lines is limited. The situation is much like automobile transportation before the interstate highway system was built in the 1950s. It was possible to drive coast to coast, but the driving time was long and uncertain and the route indirect. To use renewable electricity resources effectively, we must create a kind of interstate highway system for electricity.

## Combined Heat and Power

Electricity in the United States is generated, for the most part, from central station power plants at a conversion efficiency of roughly 30 to 35 percent. Meaning, for every 100 units of fuel energy into a simple cycle central station electric power plant, we get only 30 to 35 units of electricity. The remainder of the energy in the fuel is lost to the atmosphere in the form of heat.

The thermal requirements of our buildings and facilities are generally provided **on-site** through the use of a boiler or furnace. The efficiencies of this equipment have improved over the years and now it is common to have boilers and furnaces in commercial and industrial facilities with efficiencies of 80 percent and higher. Meaning, for every 100 units of fuel energy into the boiler/furnace, we get about 80 units of useful thermal energy.

Commercial and industrial facilities that utilize the conventional energy system found in the United States (electricity supplied from the electric grid and thermal energy produced on-site through the use of a boiler/furnace) will often times experience overall fuel efficiencies of between 40 to 55 percent (actual efficiency depends on the facilities heat to power ratio).

Combined Heat and Power (known also as CHP or “cogeneration”) is an integrated system located at or near the building/facility that generates utility grade electricity which satisfies at least a portion of the electrical load of the facility, and captures and recycles the waste heat from the electric generating equipment to provide useful thermal energy to the facility.

CHP implies that heat and electricity are produced simultaneously in one process. Use of combined heat and power helps to improve the overall efficiency of electricity and heat production as these systems combine electricity production technologies with heat recovery equipment. Increasing the conversion efficiency of power generation through the use of CHP helps to reduce the

environmental impact of power generation. These systems can reach fuel use efficiencies of as high as 75 to 85 percent (versus the conventional energy system at approximately 40 to 55 percent).

A well designed, installed and operated CHP system provides benefits for the facility owner (end user), the electric utility, and society in general. The high efficiency attained by the CHP system provides the end user with lower overall energy costs, improved electric reliability, improved electric power quality, and improved energy security. In areas where the electric utility distribution grid is in need of expansion and/or upgrades, CHP systems can provide the electric utility with a means of deferring costly modifications to the grid.

Although the electricity generated on-site by the end user displaces the electricity purchased from the local electric utility and is seen as lost revenue by many utilities, energy efficiency and lower utility costs are in the best interest of the utility customer and should be considered as a reasonable customer option by forward-looking customer oriented utilities. Finally, society in general benefits from the high efficiencies realized by CHP systems. The high efficiencies translate to less air pollutants (lower greenhouse gas and NOx emissions) than produced from central station electric power plants.

## Electricity as an Alternative Fuel Source

Electric cars represent an alternative to oil for transportation. Electric vehicles are run by an electric motor, as in a fuel cell car, up to four times as efficient as a gasoline engine. The electric motor is far simpler than a gasoline engine, having only one moving part, a shaft rotating inside a stationary housing and surrounded by a coil of copper wire. Electricity comes from a battery, whose storage capacity, like that of hydrogen materials, is too small to enable long distance driving. Developing higher energy density batteries for vehicles is a major challenge for the electric car industry. The battery must be charged before driving, which can be done from the grid using excess capacity available at night, or during the day from special solar charging stations that do not add additional load to the grid. Because charging typically takes hours, a potentially attractive alternative is switching the battery out in a matter of minutes for a freshly charged one at special swapping stations. A large fleet of electric cars in the United States would require significant additional electricity, as much as 130 GW if the entire passenger and light truck fleet were converted to electricity, or 30 percent of average United States electricity usage in 2008.

The energy usage of electric cars is about a factor of four less than for gasoline cars, consistent with the higher efficiency of electric motors over internal combustion engines. Although gasoline cars vary significantly in their energy efficiency, a "typical" middle of the road value for a five-passenger car is 80kWh/100km. A typical electric car (such as the Think Ox from Norway, the Chevy Volt operating in its electric mode, or the Nissan Leaf) uses ~ 20 kWh/100km. While the energy cost of electric cars at the point of use is significantly less, one must consider the cost at the point of production, the electricity generating plant. If the vehicle's electricity comes from coal with a conversion efficiency of 33 percent, the primary energy cost is 60 kWh/100km, approaching but still smaller than that of the gasoline car. If electricity is generated by combined cycle natural gas turbines with 60 percent efficiency, the primary energy cost is 33 kWh/100km, less than half the primary energy cost for gasoline cars. These comparisons are presented in Table below.

Comparison of energy use for gasoline driven and battery driven cars, for the cases of inefficient coal generation (33%) and efficient combined cycle natural gas generation (60%) of electricity. *Source:* George Crabtree.

| | **Gasoline Engine 5 passenger car** | **Battery Electric Nissan Leaf, Chevy Volt (battery mode), Think Ox** |
|---|---|---|
| **Energy use at point of use** | 80 kWh/100km | 20 kWh/100km |
| **Energy use at point of production: Coal at 33% efficiency** | | 60 kWh/100km |
| **Combined Cycle Natural Gas at 60% efficiency** | | 33 kWh/100km |

Comparison of carbon emissions from gasoline driven and battery driven cars, for the cases of high emission coal generation (2.1 lb CO2/kWh), lower emission natural gas (1.3 lbCO2/kWh) and

very low emission nuclear, hydro, wind or solar electricity.*Source:* George Crabtree.

| | Gasoline Engine 5 passenger car | Battery Electric Nissan Leaf, Chevy Volt (battery mode), Think Ox |
|---|---|---|
| **CO2 Emissions at point of use** | 41 lbs | ~ 0 |
| **CO2 Emissions at point of production Coal@2.1 lb CO2/kWh** | | 42 lbs |
| **Gas@1.3 lb CO2/kWh** | | 25 lbs |
| **Nuclear, hydro, wind or solar** | | < 1 lb |

The carbon footprint of electric cars requires a similar calculation. For coal-fired electricity producing 2.1 lb $CO_2$/kWh, driving 100km produces 42 lbs (19 kgs) of carbon dioxide; for gas-fired electricity producing 1.3 lb $CO_2$/kWh, 100km of driving produces 26 lbs (11.7 kgs) of carbon dioxide. If electricity is produced by nuclear or renewable energy such as wind, solar or hydroelectric, no carbon dioxide is produced. For a "typical" gasoline car, 100km of driving produces 41 lbs (18.5 kgs) of carbon dioxide. Thus the carbon footprint of a "typical" electric car is, at worst equal, to that of a gasoline car and, at best, zero. Table Comparisons of Carbon Emissions summarizes the carbon footprint comparisons.

## Hybrid Transportation

Unlike electric cars, hybrid vehicles rely only on gasoline for their power. Hybrids do, however, have a supplemental electric motor and drive system that operates only when the gasoline engine performance is weak or needs a boost: on starting from a stop, passing, or climbing hills. Conventional gasoline cars have only a single engine that must propel the car under all conditions; it must, therefore, be sized to the largest task. Under normal driving conditions the engine is larger and less efficient than it needs to be. The hybrid solves this dilemma by providing two drive trains, a gasoline engine for normal driving and an electric motor for high power needs when starting, climbing hills and passing. The engine and motor are tailored to their respective tasks, enabling each to be designed for maximum efficiency. As the electric motor is overall much more efficient, its use can raise fuel economy significantly.

The battery in hybrid cars has two functions: it drives the electric motor and also collects electrical energy from regenerative braking, converted from kinetic energy at the wheels by small generators. Regenerative braking is effective in start-stop driving, increasing efficiency up to 20 percent. Unlike gasoline engines, electric motors use no energy while standing still; hybrids therefore shut off the gasoline engine when the car comes to a stop to save the idling energy. Gasoline engines are notoriously inefficient at low speeds (hence the need for low gear ratios), so the electric motor accelerates the hybrid to ~15 mph (24 kph) before the gasoline engine restarts. Shutting the gasoline engine off while stopped increases efficiency as much as 17 percent. The energy saving features of hybrids typically lower their energy requirements from 80 kWh/100km to 50-60 kWh/100km, a significant savings. It is important to note, however, that despite a supplementary electric motor drive system, all of a hybrid's energy comes from gasoline and none from the electricity grid.

The **plug-in hybrid** differs from conventional hybrids in tapping both gasoline and the electricity grid for its energy. Most plug-in hybrids are designed to run on electricity first and on gasoline second; the gasoline engine kicks in only when the battery runs out. The plug-in hybrid is thus an electric car with a supplemental gasoline engine, the opposite of the conventional hybrid cars described above. The value of the plug-in hybrid is that it solves the "driving range anxiety" of the consumer: there are no worries about getting home safely from a trip that turns out to be longer than expected. The disadvantage of the plug-in hybrid is the additional supplemental gasoline engine technology, which adds cost and complexity to the automobile.

## The Battery Challenge

To achieve reasonable driving range, electric cars and plug-in hybrids need large batteries, one of their greatest design challenges and a potentially significant consumer barrier to widespread sales. Even with the largest practical batteries, driving range on electricity is limited, perhaps to ~100km. Designing higher energy density batteries is currently a major focus of energy research, with advances in Li-ion battery technology expected to bring significant improvements. The second potential barrier to public

acceptance of electric vehicles is charging time, up to eight hours from a standard household outlet. This may suit overnight charging at home, but could be a problem for trips beyond the battery's range – with a gasoline car the driver simply fills up in a few minutes and is on his way. Novel infrastructure solutions such as battery swapping stations for long trips are under consideration.

From a sustainability perspective, the comparison of gasoline, electric, hybrid and plug-in hybrid cars is interesting. Hybrid cars take all their energy from gasoline and represent the least difference from gasoline cars. Their supplementary electric drive systems reduce gasoline usage by 30-40 percent, thus promoting conservation of a finite resource and reducing reliance on foreign oil. Electric cars, however, get all of their energy from grid electricity, a domestic energy source, completely eliminating reliance on foreign oil and use of finite oil resources. Their sustainability value is therefore higher than hybrids. Plug-in hybrids have the same potential as all electric vehicles, provided their gasoline engines are used sparingly. In terms of carbon emissions, the sustainability value of electric vehicles depends entirely on the electricity source: neutral for coal, positive for gas and highly positive for nuclear or renewable hydro, wind or solar. From an energy perspective, electric cars use a factor of four less energy than gasoline cars at the point of use, but this advantage is partially compromised by inefficiencies at the point of electricity generation. Even inefficient coal-fired electricity leaves an advantage for electric cars, and efficient gas-fired combined cycle electricity leaves electric cars more than a factor of two more energy efficient than gasoline cars. However, none of these comparisons take into account differences in the emissions and environmental impact of mining materials for construction of the different types of cars and their batteries.

## Energy Conservation

**Energy conservation** refers to reducing energy waste and increasing efficiency. Energy conservation can involve behavior changes as well as technologies. Some examples of energy conservation have no financial impact. These include turn off and unplugging electronics when not in use, turning down the water heater, and driving efficiently (figure **7.8.3**). Additionally, opening blinds on south-facing windows in the morning during the winter takes advantage of a passive solar technology. Relying on the sun for heating and lighting reduces the use of electricity.




**Figure 7.8.3:** A plant on the dashboard of this Ford Fusion (right) indicates if the driver's behavior promotes fuel efficiency. In this picture, there is only one leaf on the plant, indicating low efficiency. Some ways to drive efficiently include following the speed limit and adjusting your speed gradually. Image by ogilvyprworldwide (CC-BY).

Other examples of energy conservation require some financial investment, but they quickly pay for themselves with savings on an energy bill. An **energy audit** is a first step to investigate inefficiencies in one's home. This helps homeowners identify where their home is losing energy, and which problem areas and fixes they should prioritize to save energy and money. For example, an energy audit may reveal places in the home where hot escaping in the winter or entering in the summer. An energy auditor might recommend installing insulation to better seal the home as well as to insulate the hot water heater and pipes. Investing in most high-efficiency appliances also pays for itself relatively quickly (figure **7.8.4**).




**Figure 7.8.4:** Energy Star appliances use energy efficiently. Image by MoneyBlogNewz (CC-BY)

This video provides a walkthrough of an energy audit.

Finally, some strategies for energy conservation require sizeable financial investment. They can eventually pay for themselves over extended periods of time. Once example are double-paned, low emissivity (low e) windows (figure 7.8.5). The two layers of glass trap air between them, which serves as insulation. Additionally, the glass is coated with very small metal dots that allow light to pass through, but infrared (heat) energy is reflected back. If it is hotter outside, heat is emitted back outside; if warmer inside, heat will be emitted back inside. Energy-efficient air conditioners, geothermal heat pumps, and on-demand (tankless) water heaters (figure 7.8.6) are also examples of energy-conserving technologies that require sizeable investment.

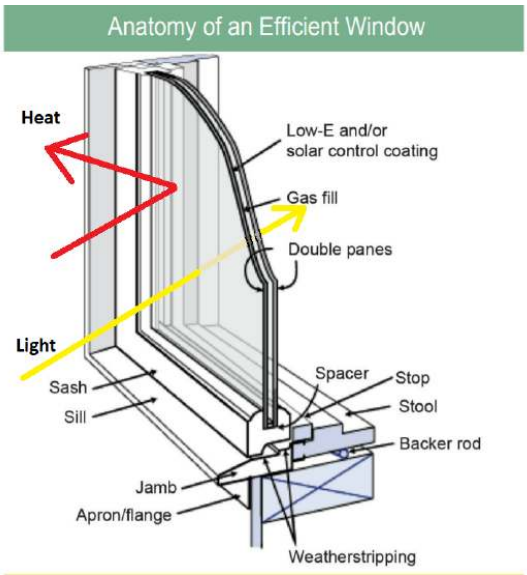



**Figure 7.8.5:** Anatomy of an efficient window. A double-paned, low emissivity window allows light from the sun to pass through, but tiny metal dots (not visible) reflect heat (labeled "Low-E and/or solar control coating"). The double panes trap air between them, an insulator (labeled "gas fill"). Spacers are between the glass and the sash, which surrounds the glass. The sash can move

out of the frame when the window is open, and weatherstipping prevents air from leaking between the sash and the frame. The stop on the outside prevents the sash swinging outward, and the stool extends perpendicular to the wall outside of the stop. Under it is a backer rod, filling the space between the stool and the wall beneath it. The frame consists of the horizontal sill and vertical jamb. The apron/flange is the trim just below the frame. Image modified from U.S. Department of Energy (public domain).

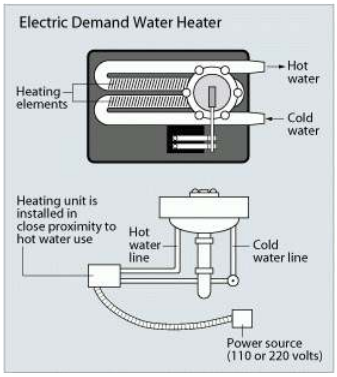



**Figure 7.8.6:** An tankless water heater (electric demand water heater). Because hot water is not stored, there is no opportunity for it to cool and need to be reheated. This conserves energy. The heating unit is installed in close proximity to hot water use. It is suppled by a power source (110 or 220 volts). Hot and cold water lines run from the heating unit (which contains heating elements) to the sink. Image by U.S. Department of Energy (public domain).

## References


Crabtree, G. & Misewich, J. (Co-Chairs). (2010). *Integrating Renewable Electricity on the Grid, American Physical Society*. American Physical Society, Washington D.C. Retrieved August 12, 2011 from http://www.aps.org/policy/reports/popa-reports/upload/integratingelec.pdf

Owen, D. (2009). *Green Metropolis*: *Why Living Smaller, Living Closer, And Driving Less Are the Keys to Sustainability*. New York: Riverhead Books.

U.S. Energy Information Administration. (2010). *Annual Energy Review 2009*. Retrieved August 12, 2011 from http://www.eia.gov/totalenergy/data/annual/pdf/aer.pdf


## Contributions and Attribution

Modified by Kyle Whittinghill and Melissa Ha from the following sources

- Energy Conservation by Melissa Ha (CC-BY-NC)
- Hydrogen and Electricity as Alternative Fuels, Electricity Grid and Sustainability Challenges, and Combined Heat and Power from *Essentials of Environmental Science* by Kamala Doršner is licensed under CC BY 4.0.
- Electricity and Combined Heat and Power in *Sustainability a Comprehensive Foundation* by Tom Theis and Jonathan Tomkin
- Home Energy Audits. U.S. Department of Energy. Accessed 01-18-2021. (public domain)


This page titled 7.8: The Electrical Grid, Battery Challenges, and Energy Conservation is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# CHAPTER OVERVIEW

## 8: Atmosphere and Air Pollution

All biomes are universally affected by global conditions, such as climate, that ultimately shape each biome's environment. Scientists who study climate have noted a series of marked changes that have gradually become increasingly evident during the last sixty years. Global climate change is the term used to describe altered global weather patterns, including a worldwide increase in temperature, due largely to rising levels of atmospheric carbon dioxide.

### Learning Objectives

- Define global climate change
- Summarize the effects of the Industrial Revolution on global atmospheric carbon dioxide concentration
- Describe three natural factors affecting long-term global climate
- List two or more greenhouse gases and describe their role in the greenhouse effect

### Global Climate Change

Climate change can be understood by approaching three areas of study:

- current and past global climate change
- causes of past and present-day global climate change
- ancient and current results of climate change

### Summary

The Earth has gone through periodic cycles of increases and decreases in temperature. During the past 2000 years, the Medieval Climate Anomaly was a warmer period, while the Little Ice Age was unusually cool. Both of these irregularities can be explained by natural causes of changes in climate, and, although the temperature changes were small, they had significant effects. Natural drivers of climate change include Milankovitch cycles, changes in solar activity, and volcanic eruptions. None of these factors, however, leads to rapid increases in global temperature or sustained increases in carbon dioxide. The burning of fossil fuels is an important source of greenhouse gases, which plays a major role in the greenhouse effect. Long ago, global warming resulted in the Permian extinction: a large-scale extinction event that is documented in the fossil record. Currently, modern-day climate change is associated with the increased melting of glaciers and polar ice sheets, resulting in a gradual increase in sea level. Plants and animals can also be affected by global climate change when the timing of seasonal events, such as flowering or pollination, is affected by global warming.


8.1: The Atmosphere, Climate, and Weather

8.2: Earth's Energy Balance

8.3: Atmospheric and Ocean Circulation

8.4: Air Pollution

8.5: Acid Deposition

8.6: Stratospheric Ozone Depletion


### Contributors and Attributions


- 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)

This page titled 8: Atmosphere and Air Pollution is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Matthew R. Fisher (OpenOregon) via source content that was edited to the style and standards of the LibreTexts platform.

# 8.1: The Atmosphere, Climate, and Weather

## 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 $1.0 \times 10^5$ 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 **8.1.1**). 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.

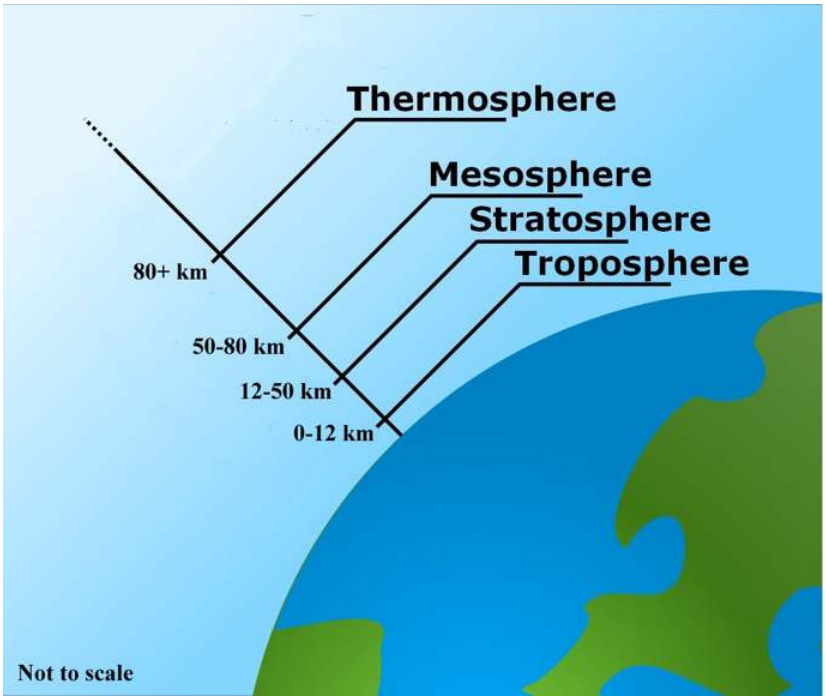


**Figure 8.1.1:** The layers of the atmosphere. the troposphere is the closest to the Earth's surface (0-12 km). Next, is the stratosphere (12-50 km), mesophere (50-80 km), and thermosphere (80+ km). The outermost layer (the exosphere) is not shown. **Ground-level ozone** in the troposphere is a form 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.

---

8.1: The Atmosphere, Climate, and Weather is shared under a CC BY-NC-SA 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.

# 8.2: Earth's Energy Balance

## Earth's Temperature is a Balancing Act

Earth's temperature depends on the balance between energy entering and leaving the planet. When incoming energy from the sun is absorbed, Earth warms. When the sun's energy is reflected back into space, Earth avoids warming. When energy is released from Earth into space, the planet cools. Many factors, both natural and human, can cause changes in Earth's energy balance, including:

- Changes in the greenhouse effect, which affects the amount of heat retained by Earth's atmosphere;
- Variations in the sun's energy reaching Earth;
- Changes in the reflectivity of Earth's atmosphere and surface.

The balance of incoming and outgoing heat on Earth is referred to as its **energy budget**. As with any budget, to maintain constant conditions the budget must be balanced so that the incoming heat equals the outgoing heat. The energy budget of Earth appears below (figure 8.2.1).

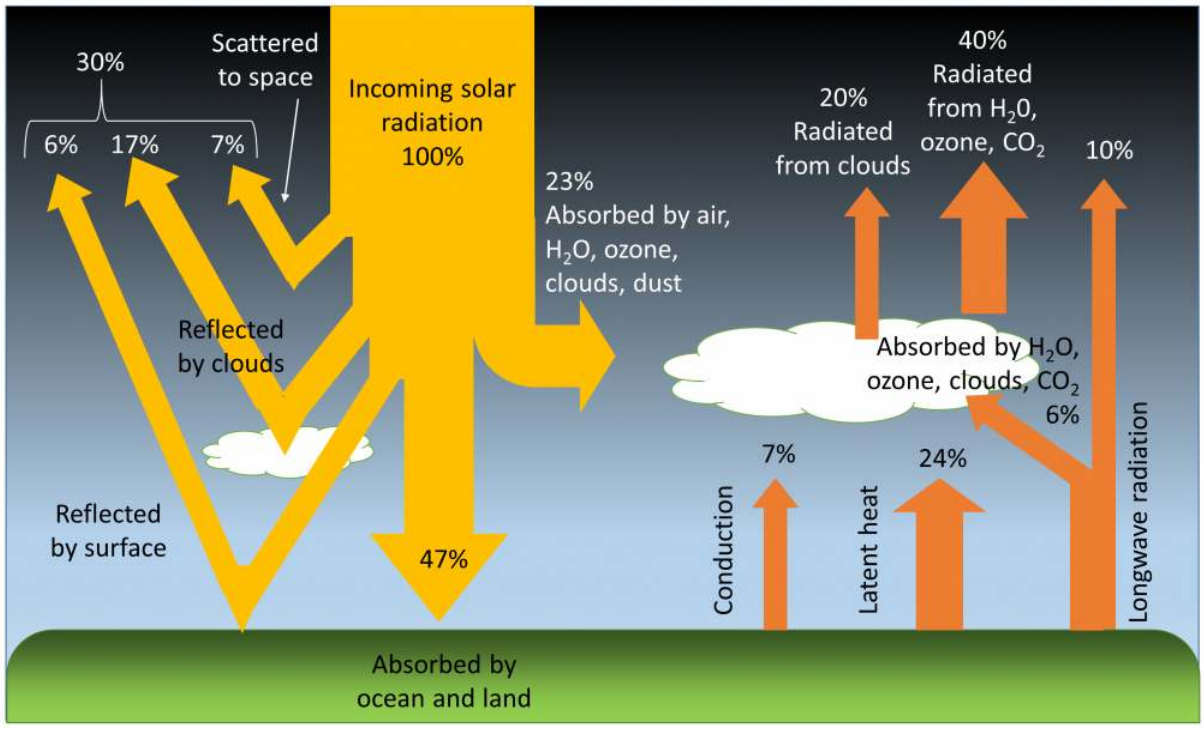



Figure 8.2.1: Earth's energy budget. Of all of the solar radiation reaching Earth, 30% is reflected back to space and 70% is absorbed by the Earth (47%) and atmosphere (23%). The heat absorbed by the land and oceans is exchanged with the atmosphere through conduction, radiation, and latent heat (phase change). The heat absorbed by the atmosphere is eventually radiated back into space (Paul Webb, Roger Williams University).

Solar radiation from the sun is composed of mostly ultraviolet (UV), visible light, and infrared (IR) radiation (figure 8.2.2). Components of solar radiation include parts with a shorter wavelength than visible light, like ultraviolet light, and parts of the spectrum with longer wavelengths, like IR and others (figure 8.2.2). Of all of the solar energy reaching the Earth, about 30% is reflected back into space from the atmosphere, clouds, and surface of the Earth (figure 8.2.1). Another 23% of the energy is absorbed by the water vapor, clouds, and dust in the atmosphere, where it is converted into heat.

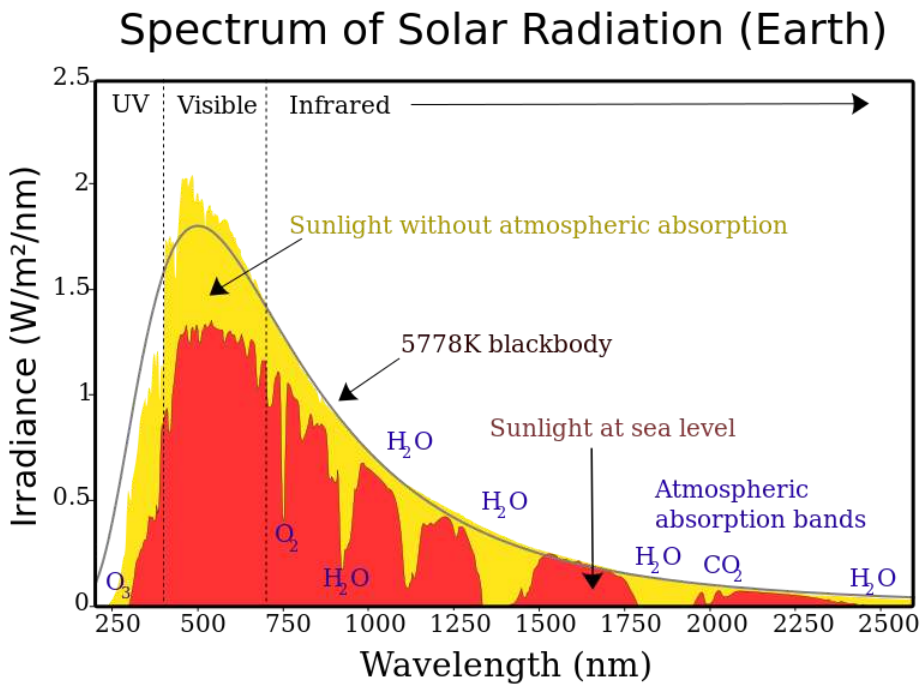



Figure 8.2.2: Incoming radiation absorbed, scattered, and reflected by atmospheric gases.

Just under half (47%) of the incoming solar radiation is absorbed by the land and ocean, and this energy heats up the Earth's surface. The energy absorbed by the Earth returns to the atmosphere through three processes; conduction, radiation, and latent heat (phase change) (figure 8.2.1).

**Conduction** is the transfer of heat through direct contact between the surface and the atmosphere. Air is a relatively poor thermal conductor (which means it is a good insulator), so conduction represents only a small part of the energy transfer between the Earth and the atmosphere; equal to about 7% of the incoming solar energy.

All bodies with a temperature above absolute zero (-273$^o$ C) radiate heat in the form of longwave, infrared **radiation** (figure 8.2.2). The warmed Earth is no exception, and about 16% of the original solar energy is radiated from the Earth to the atmosphere (figure 8.2.1). When sunlight warms a surface such as a paved surface, a patio, or deck, the warmer surface emits more thermal radiation, which is a type of IR radiation. So, there is a conversion from visible, UV, and IR to just thermal IR. This thermal IR is what we experience as heat. Some of this radiated energy will dissipate into space, but a significant amount of heat will be absorbed by the atmosphere. This is the basis for the **greenhouse effect** (figure 8.2.3). In the greenhouse effect, shortwave solar radiation passes through the atmosphere and reaches the Earth's surface where it gets absorbed. When the radiation is re-emitted by the Earth, it is now in the form of long wavelength, infrared radiation, which does not easily pass through the atmosphere. Instead, this infrared radiation is absorbed by the atmosphere, particularly by the greenhouse gases such as $CO_2$, methane, and water vapor. As a result, the atmosphere heats up. Without the greenhouse effect, the average temperature on Earth would be about -18$^o$ C, which is too cold for liquid water, and therefore life as we know it could not exist!

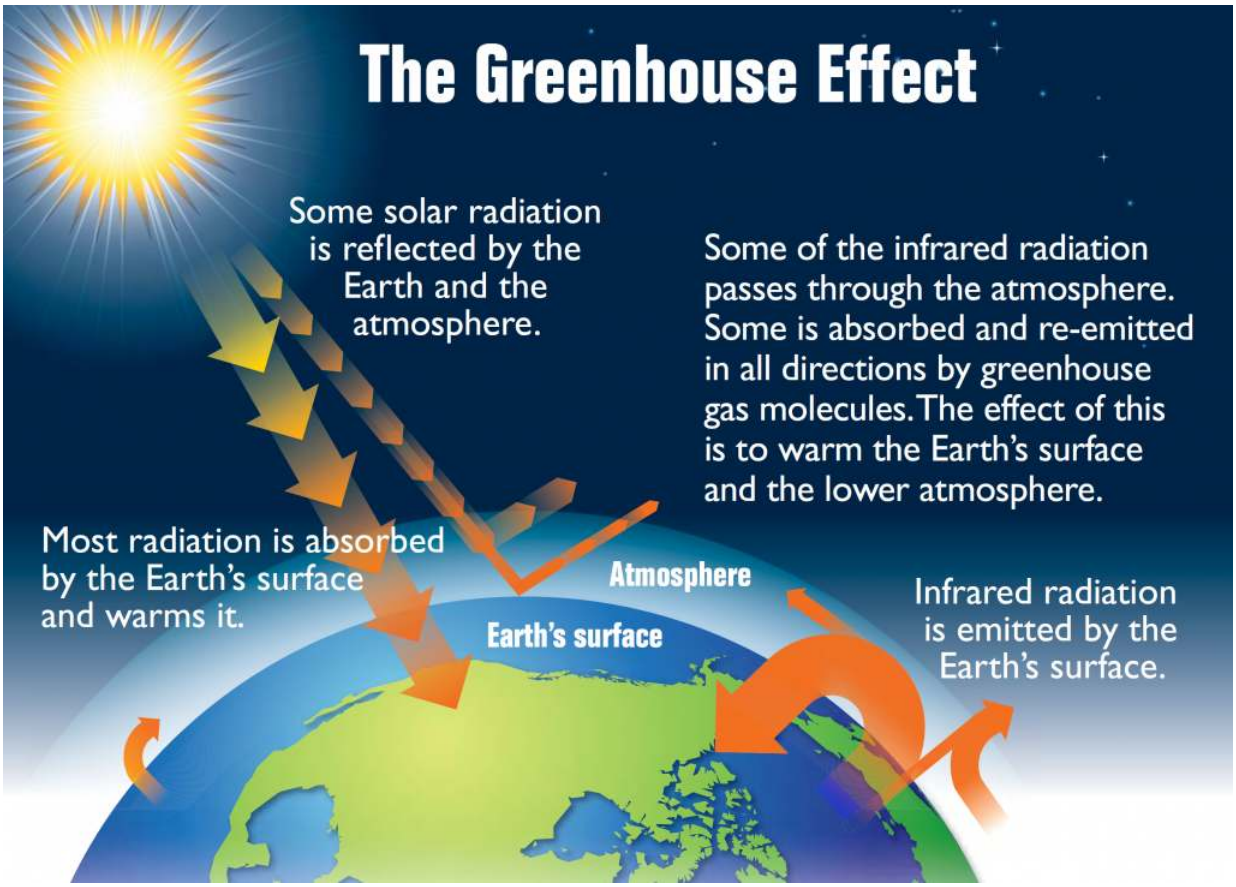



**Figure 8.2.3:** An explanation of the greenhouse effect (By US EPA [Public domain], via Wikimedia Commons).

The largest pathway for energy exchange between the land or oceans and the atmosphere is **latent heat** transferred through **phase changes**; heat released or absorbed when water moves between solid, liquid, and vapor forms. Heat must be added to liquid water to make it evaporate, and when water vapor is formed, that heat is removed from the ocean and transferred to the atmosphere along with the water vapor. When water vapor condenses into rain, that heat is then returned to the oceans. The same process happens with the formation and melting of ice. Heat is absorbed by ice when it melts, and heat is released when ice forms, and these phase changes transfer heat between the oceans and the atmosphere.

To complete the energy budget, the heat that is absorbed by the atmosphere either directly from solar radiation or as a result of conduction, radiation and latent heat, is eventually radiated back into space (figure 8.2.1).

## Greenhouse Gasses

Throughout this chapter we'll be talking about the role of **greenhouse gases** (GHGs) in controlling the climate, so it's important to understand what greenhouse gases are and how they work. As you know, the dominant gases of the atmosphere are nitrogen (as $N_2$) and oxygen (as $O_2$). These gas molecules have only two atoms each and are not GHGs. Some of the other important gases of the atmosphere are water vapor ($H_2O$), carbon dioxide ($CO_2$), and methane ($CH_4$). All of these have more than two atoms, and they are GHGs. All molecules vibrate at various frequencies and in various ways, and some of those vibrations take place at frequencies within the range of the infrared (IR) radiation that is emitted by Earth's surface. Gases with two atoms, such as $O_2$, can only vibrate by stretching (back and forth), and those vibrations are much faster than the IR radiation (Figure 8.2.4). Gases with three or more atoms (such as $CO_2$) vibrate by stretching as well, but they can also vibrate in other ways, such as by bending (Figure 8.2.4). Those vibrations are slower and match IR radiation frequencies. When IR radiation interacts with $CO_2$ or with one of the other GHGs, the molecular vibrations are enhanced because there is a match between the wavelength of the IR light and the vibrational frequency of the molecule. This makes the molecule vibrate more vigorously, heating the surrounding air in the process. These molecules also emit IR radiation in all directions, some of which reaches Earth's surface. The heating caused by the more vigorous vibrations of GHGs is the **greenhouse effect**.

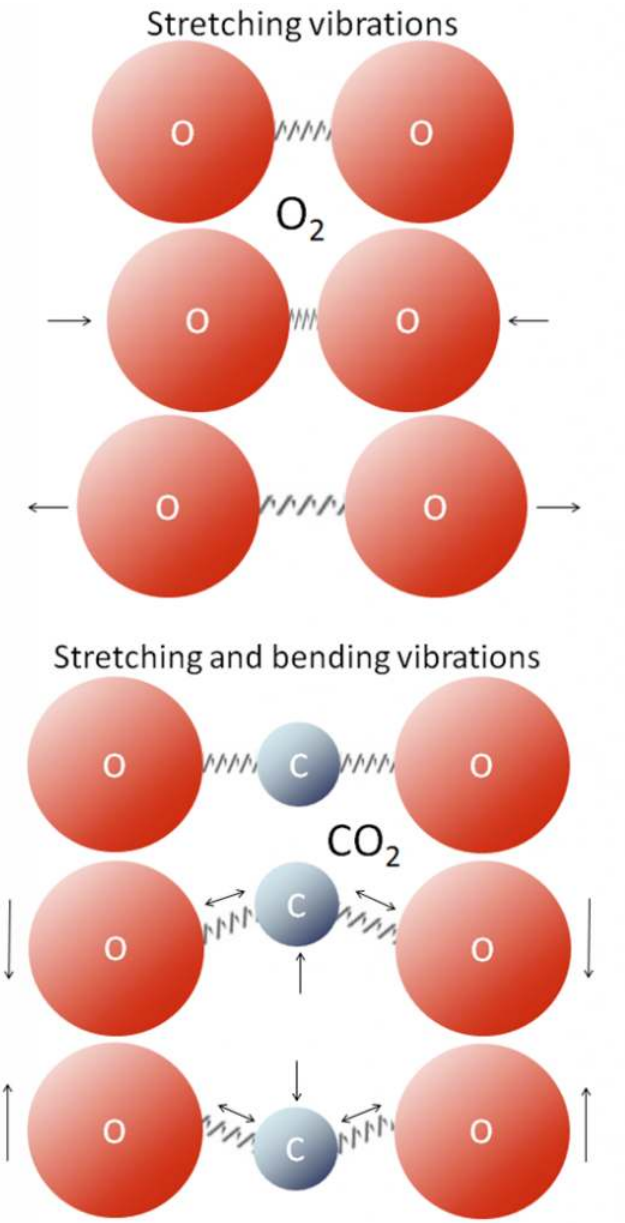



Figure 8.2.4: Stretching versus bending vibrations in atmospheric gases. © Steven Earle. CC BY.

**Differential Heating of Earth's Surface**

If the Earth was a flat surface facing the sun, every part of that surface would receive the same amount of incoming solar radiation. However, because the Earth is a sphere, sunlight is not equally distributed over the Earth's surface, so different regions of Earth will be heated to different degrees. This differential heating of Earth's surface occurs for a number of reasons. First, because of the curvature of Earth, sunlight only falls perpendicularly to the surface at the center of the sphere (equatorial regions). At any other point on Earth, the angle between the surface and the incoming solar radiation is less than $90^o$. Because of this, the same amount of incoming solar radiation will be concentrated in a smaller area at the equator, but will be spread over a much larger area at the poles (figure 8.2.5). Thus the tropics receive more intense sunlight and a greater amount of heating per unit of area than the polar regions.

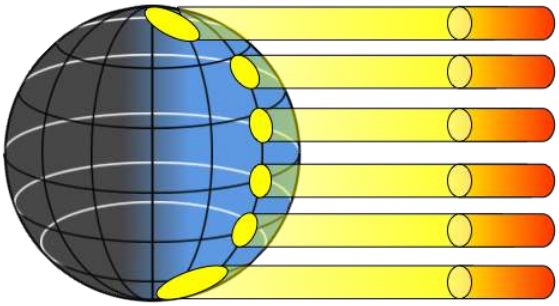


**Figure 8.2.5:** Because of the curvature of the Earth, the same amount of sunlight will be spread out over a larger area at the poles compared to the equator. The equator therefore receives more intense sunlight, and a greater amount of heat per unit of area (By Thebiologyprimer (Own work) [CC0], via Wikimedia Commons).

The angle at which sunlight strikes the Earth contributes to differential heating of the surface in an additional way. At the poles, because of the angle at which the solar energy strikes the surface, more of the light will glance off of the surface and the atmosphere and be reflected back into space. At the equator, the direct angle with which light reaches the surface results in more of the energy being absorbed rather than reflected.

Finally, the poles reflect more solar energy than other parts of the Earth because the poles have a higher **albedo**. The albedo refers to reflectivity of a surface, expressed as the percentage of light that reflects off a given material. Lighter surfaces are more reflective than darker surfaces (which absorb more energy), and therefore have a higher albedo. 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 (Figure 8.2.6).

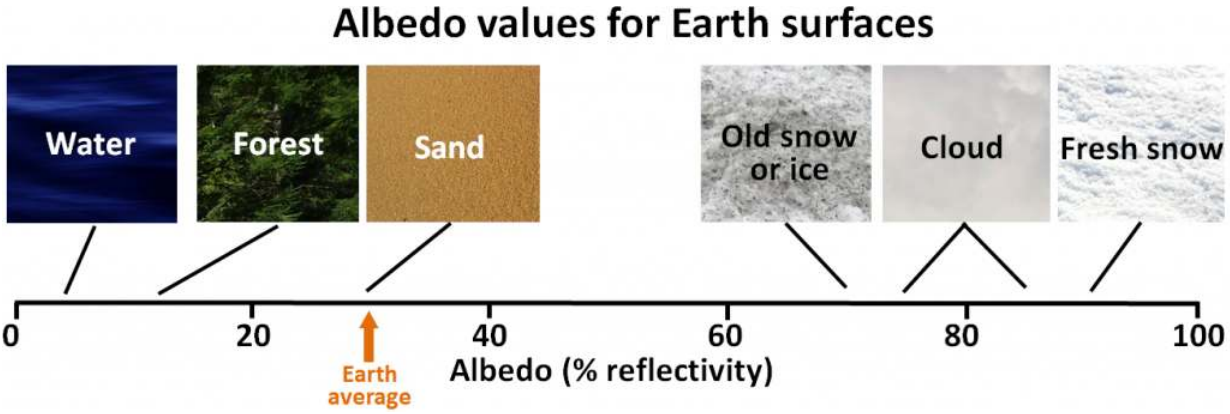



Figure 8.2.6: Typical albedo values for Earth surfaces. © Steven Earle. CC BY.

At the poles, the ice, snow and cloud cover create a much higher albedo, and the poles reflect more and absorb less solar energy than the lower latitudes (figure 8.2.7). Through all of these mechanisms, the poles absorb much less solar radiation than equatorial regions, which is why the poles are cold and the tropics are very warm.

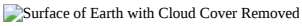



Figure 8.2.7: **Surface of Earth with Cloud Cover Removed** The surface of the Earth with cloud cover removed. The poles and deserts are much brighter than the oceans and forests. *Source:* *NASA Goddard Space Flight Center Image by Reto Stöckli. Courtesy of NASA's Earth Observatory.*

But there is an interesting twist to this global distribution of heat. The tropical regions actually receive more radiant heat than they emit, and the poles emit more heat than they receive (figure 8.2.8). We should therefore expect that the tropics will be getting continually warmer, while the poles become increasingly cold. Yet this is not the case; so what is happening? Rather than the heat

remaining isolated near the equator, about 20% of the heat from the tropics is transported to the poles before it is emitted. This large scale transport of energy moderates the climates at both extremes. The mechanisms for this heat transfer are ocean and atmospheric circulation.

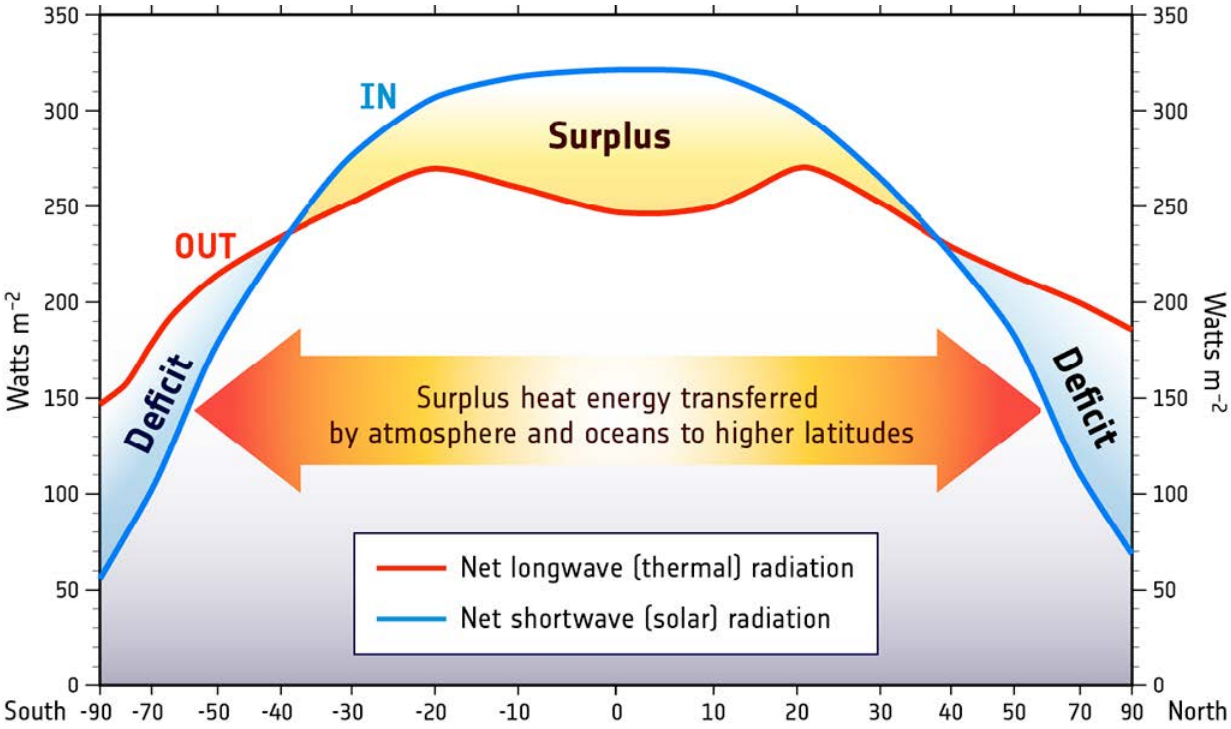



Figure 8.2.*h*: The balance between heat gain and heat loss as a function of latitude. Excess heat received near the equator is transferred towards the poles (National Oceanography Centre (NOC). Creative Commons 3.0 unported license).

The National Aeronautical and Space Administration (NASA) Earth Observatory website has an array of climate resources. For a more in-depth discussion of Earth's energy budget, go to http://earthobservatory.nasa.gov/Features/EnergyBalance/

## Contributors and Attributions

- 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
- Contributed by Paul Webb Professor (Biology) at Rodger Williams University
- Contributed by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, & Cam Mosher Faculty (Geology) at Salt Lake Community College
- Contributed by Steven Earle Faculty (Geology) at Vancover Island University
- 5.2 Climate Processes; External and Internal Controls from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin
- 19.1 What Makes Climate Change from *Physical Geology* by Steven Earle (licensed under CC-BY)
- Sourced from OpenGeology from *An Introduction to Geology* Section 15.2 Earth's Temperature
- 8.1 Earth's Heat Budget from *Introduction to Oceanography* by Paul Webb (licensed under CC-BY)
- 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)

8.2: Earth's Energy Balance is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

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

# 8.3: Atmospheric and Ocean Circulation

## Atmospheric Circulation

Differential heating of the Earth's surface results in equatorial regions receiving more heat than the poles. As air is warmed at the equator it becomes less dense and rises, while at the poles the cold air is denser and sinks. If the Earth was non-rotating, the warm air rising at the equator would reach the upper atmosphere and begin moving horizontally towards the poles. As the air reached the poles it would cool and sink, and would move over the surface of Earth back towards the equator. This would result in one large atmospheric convection cell in each hemisphere (Figure **8.3.1**), with air rising at the equator and sinking at the poles, and the movement of air over the Earth's surface creating the winds. On this non-rotating Earth, the prevailing winds would thus blow from the poles towards the equator in both hemispheres (Figure **8.3.1**).

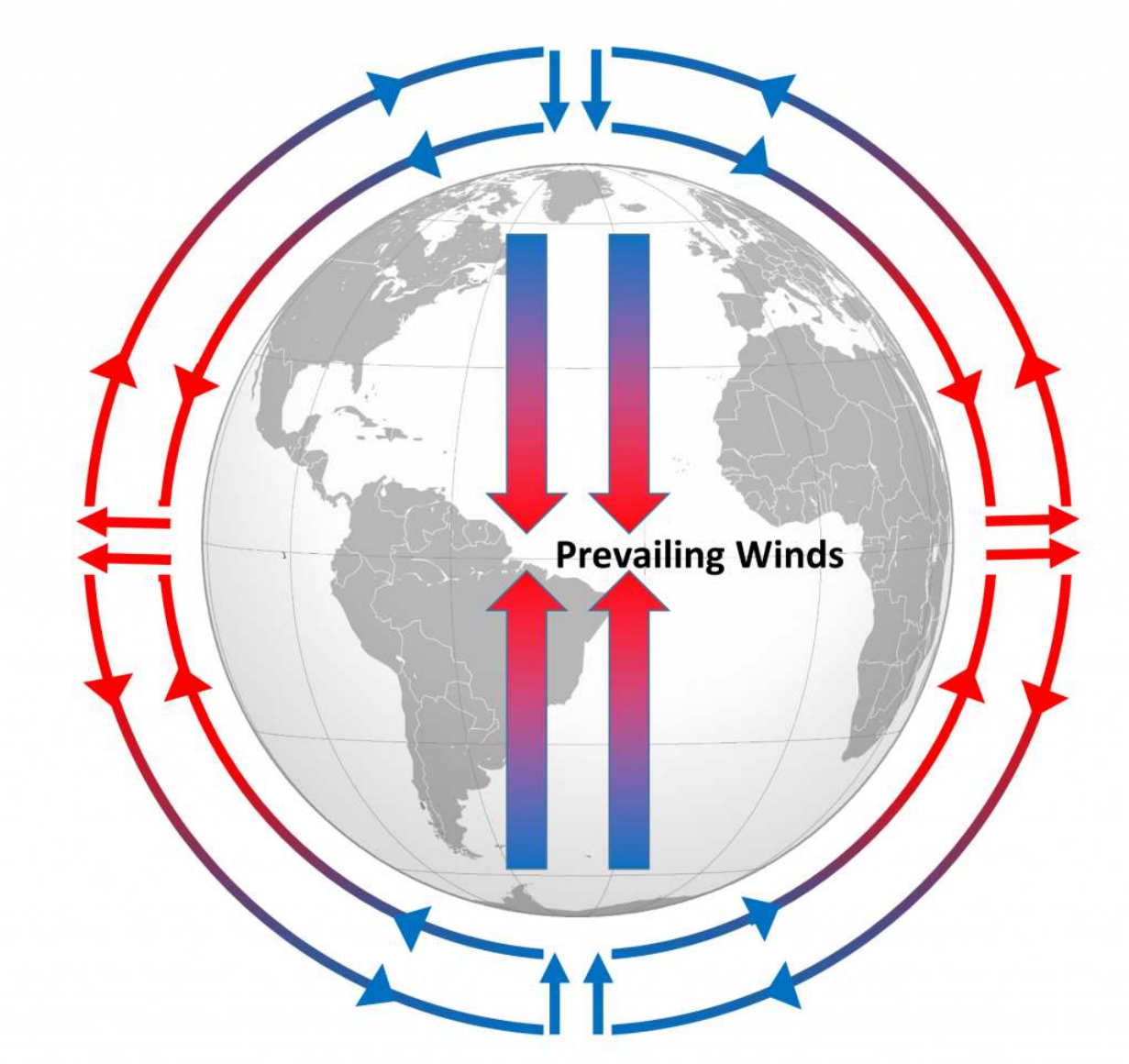



Figure **8.3.1**: Hypothetical atmospheric convection cells on a non-rotating Earth. Air rises at the equator and sinks at the poles, creating a single convection cell in each hemisphere. The prevailing winds moving over the Earth's surface blow from the poles towards the equator in both hemispheres (Modified by Paul Webb from globe image by Location_of_Cape_Verde_in_the_globe.svg: Eddo derivative work: Luan fala! [CC BY-SA 3.0], via Wikimedia Commons).

The non-rotating situation in Figure **8.3.1** is of course only hypothetical, and in reality the Earth's rotation makes this atmospheric circulation a bit more complex. The paths of the winds on a rotating Earth are deflected by the **Coriolis Effect**. The Coriolis Effect

is a result of the fact that different latitudes on Earth rotate at different speeds. This is because every point on Earth must make a complete rotation in 24 hours, but some points must travel farther, and therefore faster, to complete the rotation in the same amount of time. In 24 hours a point on the equator must complete a rotation distance equal to the circumference of the Earth, which is about 40,000 km. A point right on the poles covers no distance in that time; it just turns in a circle. So the speed of rotation at the equator is about 1600 km/hr, while at the poles the speed is 0 km/hr. Latitudes in between rotate at intermediate speeds; approximately 1400 km/hr at $30^o$ and 800 km/hr at $60^o$. As objects move over the surface of the Earth they encounter regions of varying speed, which causes their path to be deflected by the Coriolis Effect.

To explain the Coriolis Effect, imagine a cannon positioned at the equator and facing north. Even though the cannon appears stationary to someone on Earth, it is in fact moving east at about 1600 km/hr due to Earth's rotation. When the cannon fires the projectile travels north towards its target; but it also continues to move to the east at 1600 km/hr, the speed it had while it was still in the cannon. As the shell moves over higher latitudes, its momentum carries it eastward faster than the speed at which the ground beneath it is rotating. For example, by $30^o$ latitude the shell is moving east at 1600 km/hr while the ground is moving east at only 1400 km/hr. Therefore, the shell gets "ahead" of its target, and will land to the east of its intended destination. From the point of view of the cannon, the path of the projectile appears to have been deflected to the right (red arrow, Figure **8.3.2**). Similarly, a cannon located at $60^o$ and facing the equator will be moving east at 800 km/hr. When its shell is fired towards the equator, the shell will be moving east at 800 km/hr, but as it approaches the equator it will be moving over land that is traveling east *faster* than the projectile. So the projectile gets "behind" its target, and will land to the west of its destination. But from the point of view of the cannon facing the equator, the path of the shell still appears to have been deflected to the right (green arrow, Figure **8.3.2**). Therefore, in the Northern Hemisphere, the apparent Coriolis deflection will *always* be to the **right**.

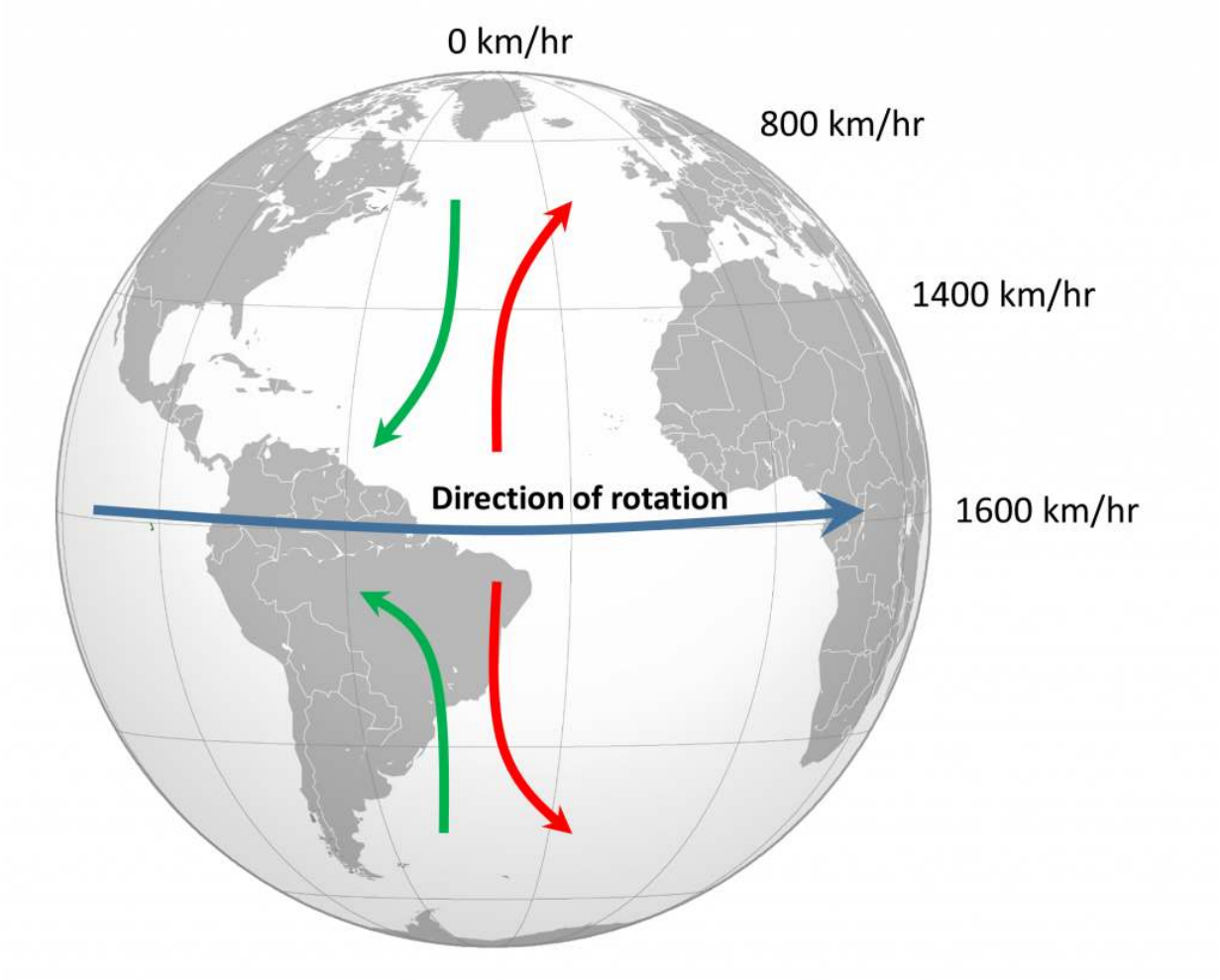



Figure **8.3.2**: The Coriolis Effect. Objects moving from the equator towards the poles (red arrows) move into a region of slower rotational speed and their paths are deflected "ahead" of their point of origin. Movement from high latitudes to low latitudes (green arrows) goes from a region of low speed to a region of higher rotation speed, and there is deflection "behind" their point of origin. In the Northern Hemisphere this deflection is always to the right from the point of origin, and in the Southern Hemisphere the deflection is always to the left
(Modified by Paul Webb from globe image by Location_of_Cape_Verde_in_the_globe.svg: Eddo derivative work: Luan fala! [CC BY-SA 3.0], via Wikimedia Commons).

In the Southern Hemisphere the situation is reversed (Figure **8.3.2**). Objects moving towards the equator from the south pole are moving from low speed to high speed, so are left behind and their path is deflected to the left. Movement from the equator towards the south pole also leads to deflection to the left. In the Southern Hemisphere, the Coriolis deflection is always to the **left** from the point of origin.

The magnitude of the Coriolis deflection is related to the difference in rotation speed between the start and end points. Between the poles and 60$^o$ latitude, the difference in rotation speed is 800 km/hr. Between the equator and 30$^o$ latitude, the difference is only 200 km/hr (Figure **8.3.2**). Therefore the strength of the Coriolis Effect is stronger near the poles, and weaker at the equator.

Because of the rotation of the Earth and the Coriolis Effect, rather than a single atmospheric convection cell in each hemisphere, there are three major cells per hemisphere. Warm air rising at the equator cools as it moves through the upper atmosphere, and it descends at around 30$^o$ latitude. The convection cells created by rising air at the equator and sinking air at 30$^o$ are referred to as **Hadley Cells**, of which there is one in each hemisphere. The cold air that descends at the poles moves over the Earth's surface towards the equator, and by about 60$^o$ latitude it begins to rise, creating a **Polar Cell** between 60$^o$ and 90$^o$. Between 30$^o$ and 60$^o$ lie the **Ferrel Cells**, composed of sinking air at 30$^o$ and rising air at 60$^o$ (Figure **8.3.3**). With three convection cells in each hemisphere that rotate in alternate directions, the surface winds no longer always blow from the poles towards the equator as in the non-rotating

Earth in Figure **8.3.1**. Instead, surface winds in both hemispheres blow towards the equator between 90° and 60° latitude, and between 0° and 30° latitude. Between 30° and 60° latitude, the surface winds blow towards the poles (Figure **8.3.3**).

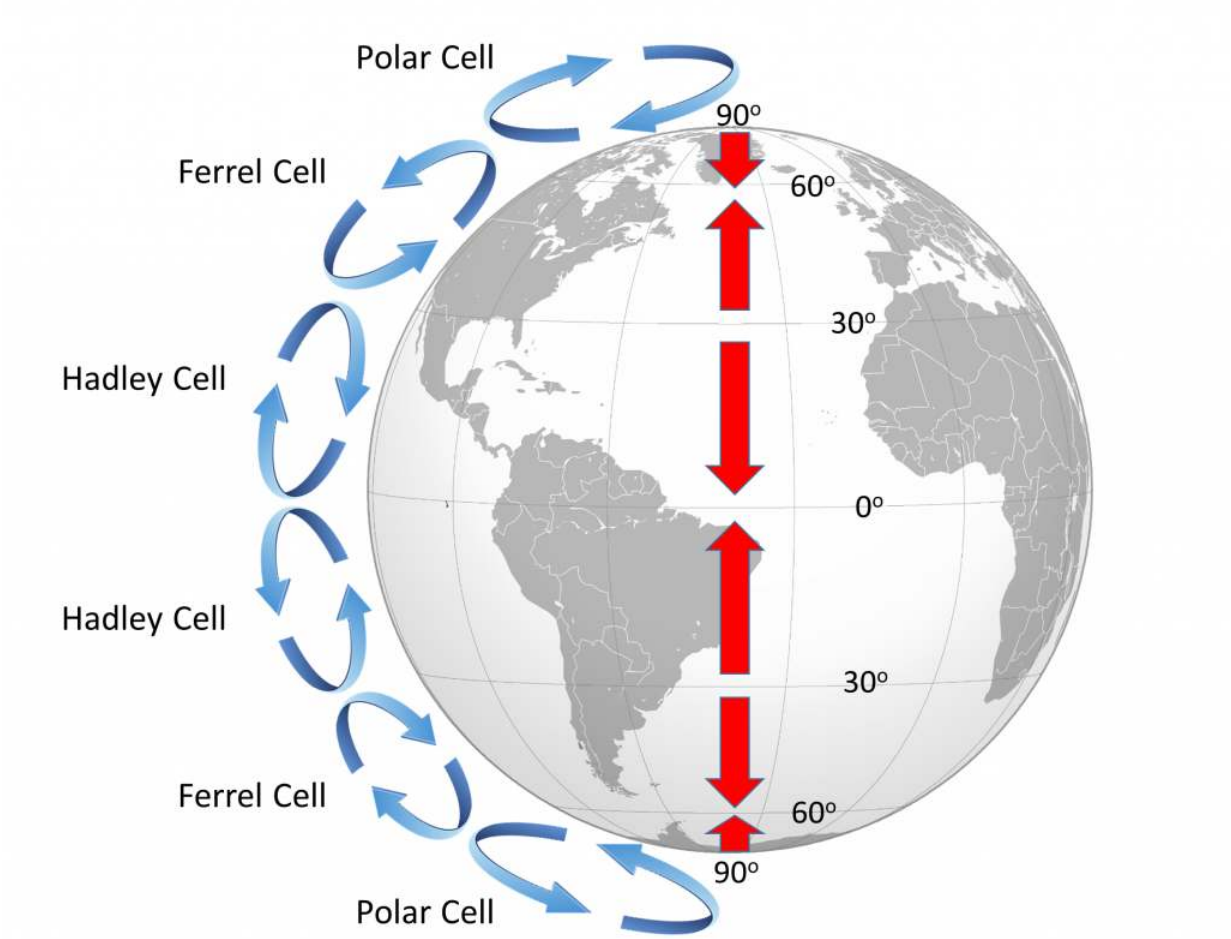



Figure **8.3.3**: On a rotating Earth, there are three atmospheric convection cells in each hemisphere, leading to alternating bands of surface winds (red arrows)
(Modified by Paul Webb from globe image by Location_of_Cape_Verde_in_the_globe.svg: Eddo derivative work: Luan fala! [CC BY-SA 3.0], via Wikimedia Commons).

The surface winds created by the atmospheric convection cells are also influenced by the Coriolis Effect as they change latitudes. The Coriolis Effect deflects the path of the winds to the right in the Northern Hemisphere and to the left in the Southern Hemisphere. Adding this deflection leads to the pattern of prevailing winds illustrated in Figure **8.3.4**. Between the equator and 30° latitude are the **trade winds**; the northeast trade winds in the Northern Hemisphere and the southeast trade winds in the Southern Hemisphere (note that winds are named based on the direction from which they originate, not where they are going). The **westerlies** are the dominant winds between 30° and 60° in both hemispheres, and the **polar easterlies** are found between 60° and the poles.

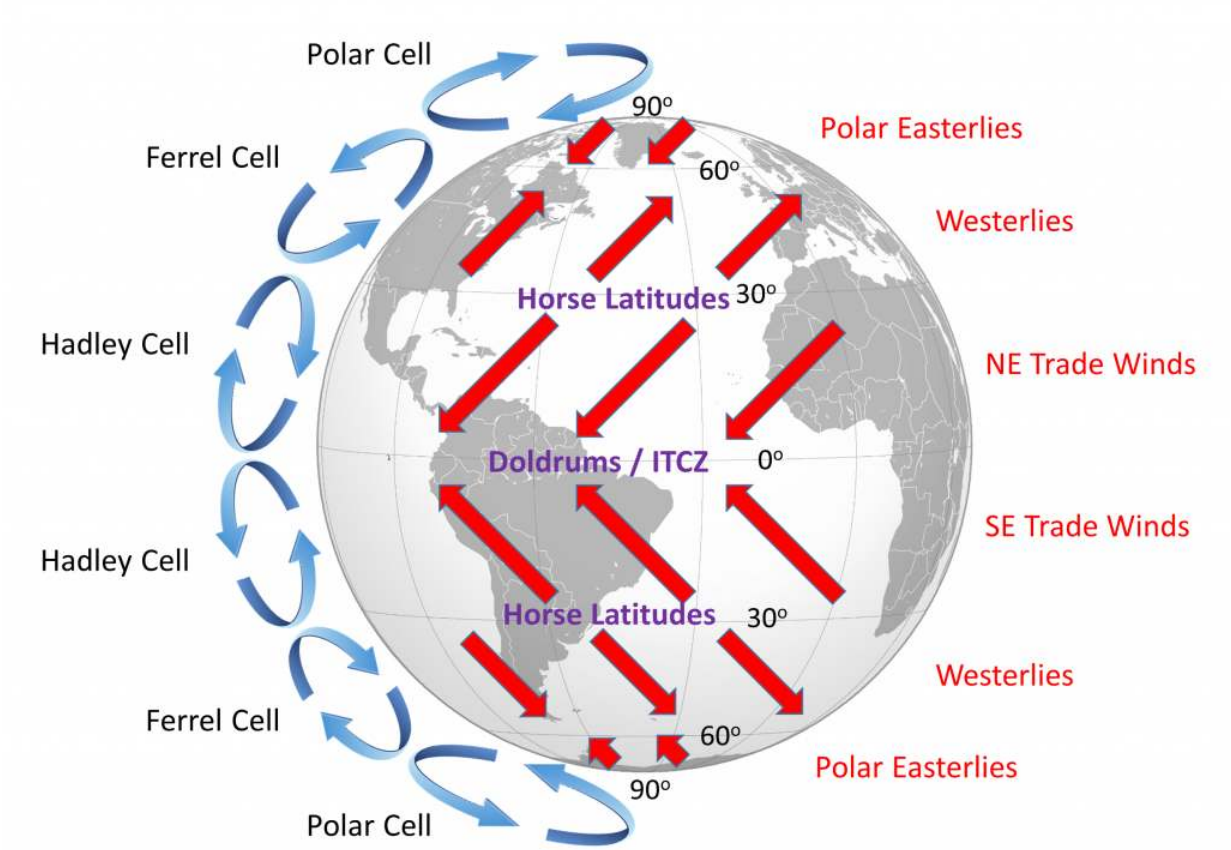



Figure **8.3.4**: The prevailing wind patterns of Earth (Modified by PW from globe image by Location_of_Cape_Verde_in_the_globe.svg: Eddo derivative work: Luan fala! [CC BY-SA 3.0], via Wikimedia Commons).

In between these wind bands lie regions of high and low pressure. High pressure zones occur where air is descending, while low pressure zones indicate rising air. Along the equator the rising air creates a low pressure region called the **doldrums**, or the **Intertropical Convergence Zone** (ITCZ) (convergence zone because this is where the trade winds converge). At 30º latitude there are high pressure zones of descending air known as the **horse latitudes**, or the subtropical highs. Finally, at 60º lies another low pressure region called the **polar front**. It should be noted that these high and low pressure zones are not fixed in place; their latitude fluctuates depending on the season, and these fluctuations have important implications for regional climates.

Doldrums, horse latitudes, and trade winds may seem like some odd names for these atmospheric phenomena, but many of them can be traced back to maritime traditions and lore. The **doldrums** refer to regions of low pressure around the equator. In these areas, air is rising rather than moving horizontally, so these regions commonly encounter very light winds. The lack of wind could leave sailing ships becalmed for days or weeks at a time, which was not good for the morale of the ship's crew. Like the doldrums the **horse latitudes** are also areas with light winds, this time due to descending air, which could leave ships becalmed. The term **trade winds** may have originally derived from the terms for "track" or "path", but the term may have become more common during European exploration and commercialization of the New World. Mariners sailing from Europe to the New World could sail south until they reached the trade winds, which would then propel their ships across the Atlantic to the Caribbean. To return to Europe, ships could sail to the northeast until they entered the westerlies, which would then steer them back to Europe.

In addition to their role in creating the surface winds, these high and low pressure systems also influence other climatic phenomena (Figure **8.3.5**. Along the equator air is rising as it is warmed by solar radiation. Warm air contains more water vapor than cold air, which is why we experience humidity during the summer and not during the winter. The water content of air roughly doubles with every 10º C increase in temperature. So the air rising at the equator is warm and full of water vapor; as it rises into the upper atmosphere it cools, and the cool air can no longer hold as much water vapor, so the water condenses and forms rain. Therefore, low pressure systems are associated with precipitation, and we see wet habitats like tropical rainforests near the equator (Figure **8.3.5**.

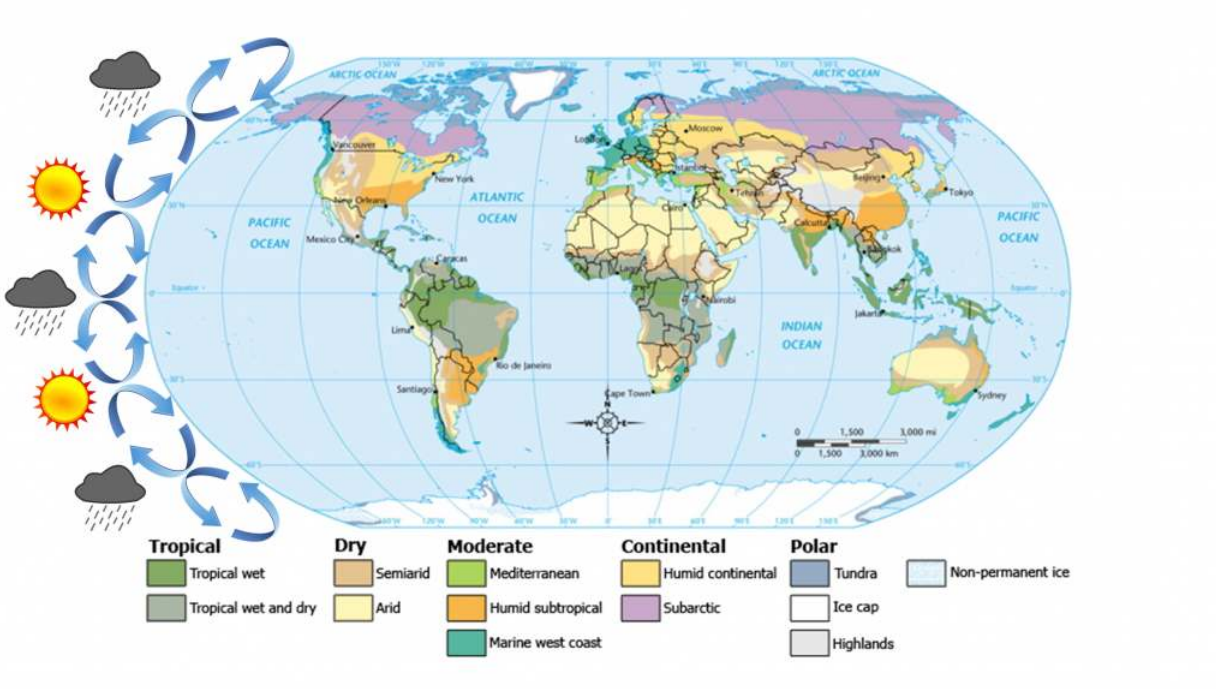



Figure **8.3.5**: Major global climatic regions in relation to atmospheric convection cells. Rising air and low pressure creates rain and wet environments at $0^o$ and $60^o$ latitudes, while high pressure, sinking air creates drier conditions at $30^o$ and $90^o$ latitudes.
(Modified by Paul Webb; Map by Waitak at en.Wikipedia Later version(s) were uploaded by Splette at en.Wikipedia; Sun by Inductiveload (Own work Based on File:Nuvola_apps_kweather.svg); Raincloud by Calusarul (Own work); all [CC BY-SA 3.0], via Wikimedia Commons).

After rising and producing rain near the equator, the air masses move towards $30^o$ latitude and sink back towards Earth as part of the Hadley convection cells. This air has lost most of its moisture after producing the equatorial rains, so the sinking air is dry, resulting in arid climates near $30^o$ latitude in both hemispheres. Many of the major desert regions on Earth are located near $30^o$ latitude, including much of Australia, the Middle East, and the Sahara Desert of Africa (Figure **8.3.5**. The air also becomes compressed and heats up as it sinks, absorbing any moisture from the clouds and creating clear skies. Thus high pressure systems are associated with dry weather and clear skies. This cycle of high and low pressure regions continues with the Ferrel and Polar convection cells, leading to rain and the boreal forests at $60^o$ latitude in the Northern Hemisphere (there are no corresponding large land masses at these latitudes in the Southern Hemisphere). At the poles, descending, dry air produces little precipitation, leading to the polar desert climate.

The elevation of the land also plays a role in precipitation and climactic characteristics. As moist air moves over land and encounters mountains it rises, expands, and cools because of the declining pressure and temperature. The cool air holds less water vapor, so condensation occurs and rain falls on the windward side of the mountains (Figure **8.3.6**. As the air passes over the mountains to the leeward side, it is now dry air, and as it sinks the pressure increases, it heats back up, any moisture revaporizes, and it creates dry, deserts regions behind the mountains (Figure **8.3.6**. This phenomenon is referred to as a **rain shadow**, and can be found in areas such as the Tibetan Plateau and Gobi Desert behind the Himalayas, Death Valley behind the Sierra Nevada mountains, and the dry San Joaquin Valley in California.

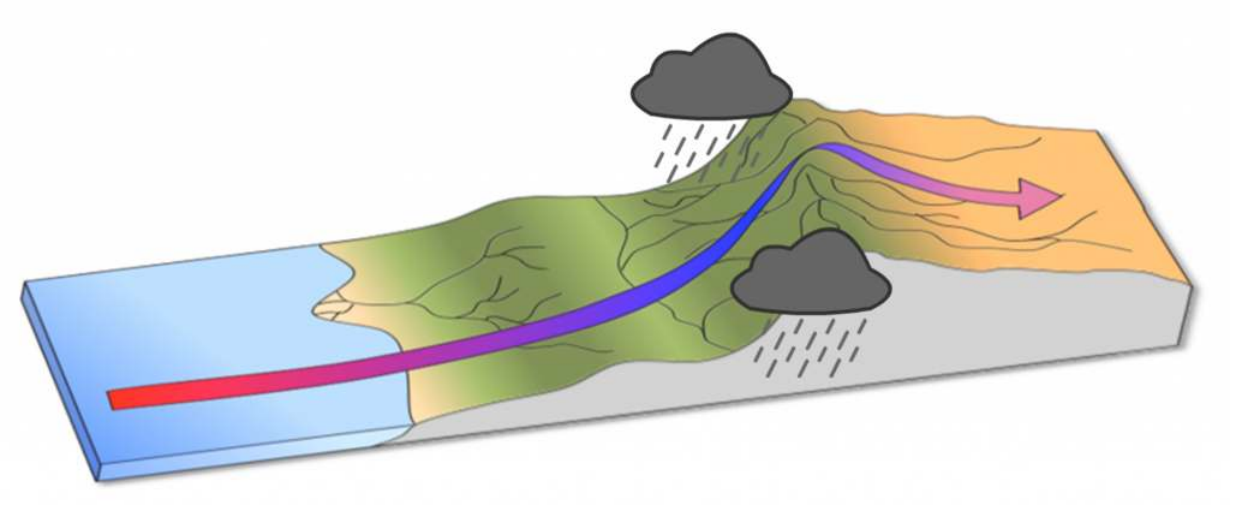


Figure **8.3.6**: A rain shadow. Air rising over mountains cools and condenses and forms rain, leaving dry descending air and arid conditions on the other side of the mountain (Modified by Paul Webb from Thebiologyprimer, Public domain via Wikimedia Commons).

## Ocean Circulation

Not unexpectedly, the oceans are warmest near the equator—typically 25° to 30°C—and coldest near the poles—around 0°C (Figure **8.3.7**). (Sea water will remain unfrozen down to about -2°C.) Variations in sea-surface temperatures (SST) are related to redistribution of water by ocean currents, as we'll see below. A good example of that is the plume of warm Gulf Stream water that extends across the northern Atlantic. St. John's, Newfoundland, and Brittany in France are at about the same latitude (47.5° N), but the average SST in St. John's is a frigid 3°C, while that in Brittany is a reasonably comfortable 15°C.

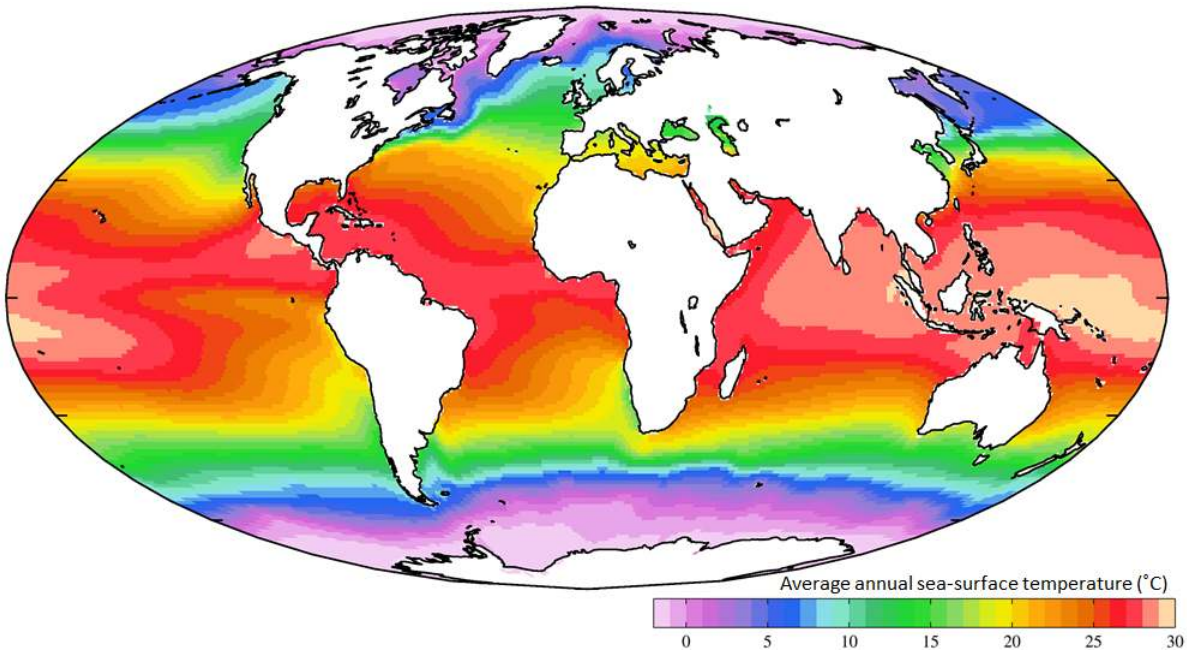



Figure **8.3.7**: The global distribution of average annual sea-surface temperatures.
[© Plumbago. CC BY-SA. *https://upload.wikimedia.org/wikiped..._TMP_AYool.png*]

Currents in the open ocean are created by wind moving across the water and by density differences related to temperature and salinity. An overview of the main ocean currents is shown in Figure **8.3.8**. As you can see, the northern hemisphere currents form circular patterns (**gyres**) that rotate clockwise, while the southern hemisphere gyres are counter-clockwise. This happens for the

same reason that the water in your northern hemisphere sink rotates in a clockwise direction as it flows down the drain; this is caused by the **Coriolis effect**.

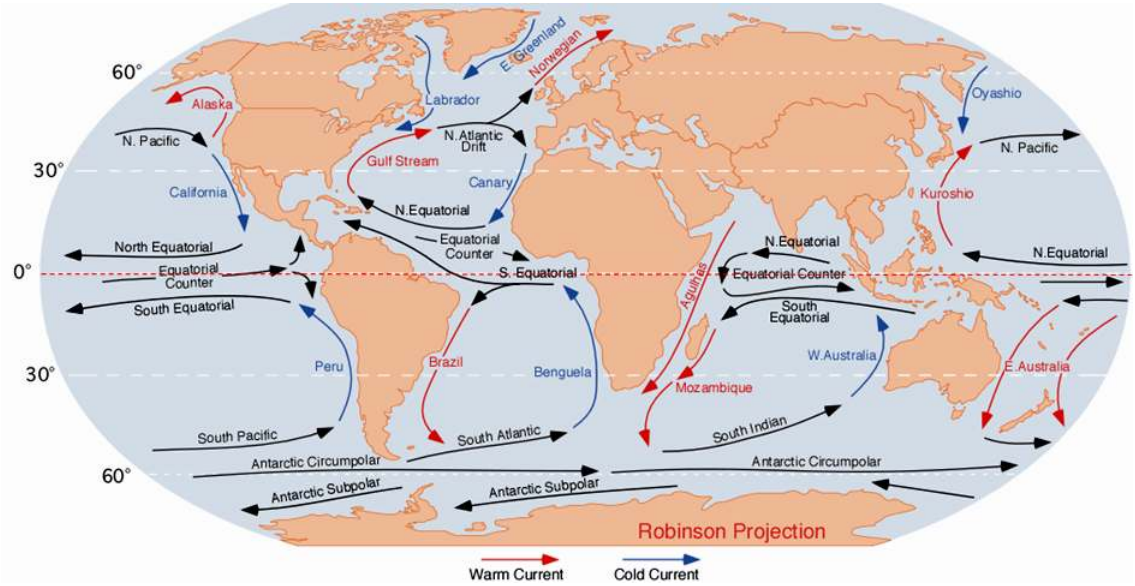



Figure **8.3.8**: Overview of the main open-ocean currents. Red arrows represent warm water moving toward colder regions. Blue arrows represent cold water moving toward warmer regions. Black arrows represent currents that don't involve significant temperature changes.
["Corrientes Oceanicas" by Dr. Michael Pidwirny. Public domain. From: *https://upload.wikimedia.org/wikiped...-oceanicas.png*]

Because the ocean basins aren't like bathroom basins, not all ocean currents behave the way we would expect. In the North Pacific, for example, the main current flows clockwise, but there is a secondary current in the area adjacent to our coast—the Alaska Current—that flows counter-clockwise, bringing relatively warm water from California, past Oregon, Washington, and B.C. to Alaska. On Canada's eastern coast, the cold Labrador Current flows south past Newfoundland, bringing a stream of icebergs past the harbour at St. John's (Figure **8.3.9**). This current helps to deflect the Gulf Stream toward the northeast, ensuring that Newfoundland stays cool, and western Europe stays warm.



Figure **8.3.9**: An iceberg floating past Exploits Island on the Labrador Current.
"Newfoundland Iceberg just off Exploits Island" © Shawn. CC BY-SA.

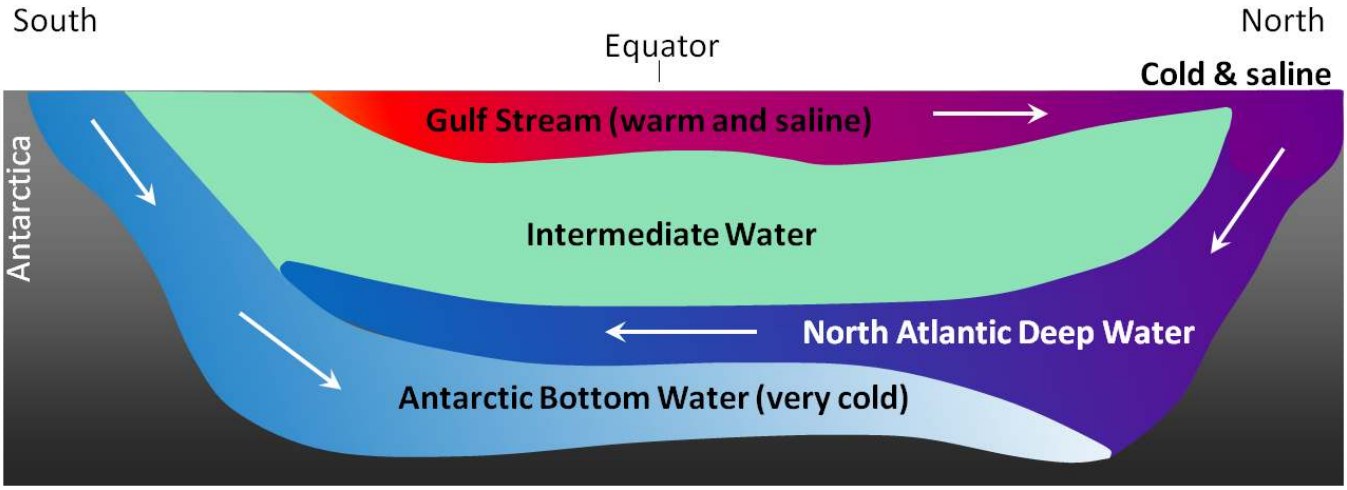



Figure 8.3.10: A depiction of the vertical movement of water along a north-south cross-section through the Atlantic basin.
© Steven Earle. CC BY.

The currents shown in Figure 8.3.8 are all surface currents, and they only involve the upper few hundred meters of the oceans. But there is much more going on underneath. The Gulf Stream, for example, which is warm and saline, flows past Britain and Iceland into the Norwegian Sea (where it becomes the Norwegian Current). As it cools down, it becomes denser, and because of its high salinity, which also contributes to its density, it starts to sink beneath the surrounding water (Figure 8.3.10). At this point, it is known as **North Atlantic Deep Water** (NADW), and it flows to significant depth in the Atlantic as it heads back south. Meanwhile, at the southern extreme of the Atlantic, very cold water adjacent to Antarctica also sinks to the bottom to become **Antarctic Bottom Water** (AABW) which flows to the north, underneath the NADW.

The descent of the dense NADW is just one part of a global system of seawater circulation, both at surface and at depth, as illustrated in Figure 8.3.11. The water that sinks in the areas of deep water formation in the Norwegian Sea and adjacent to Antarctica moves very slowly at depth. It eventually resurfaces in the Indian Ocean between Africa and India, and in the Pacific Ocean, north of the equator.

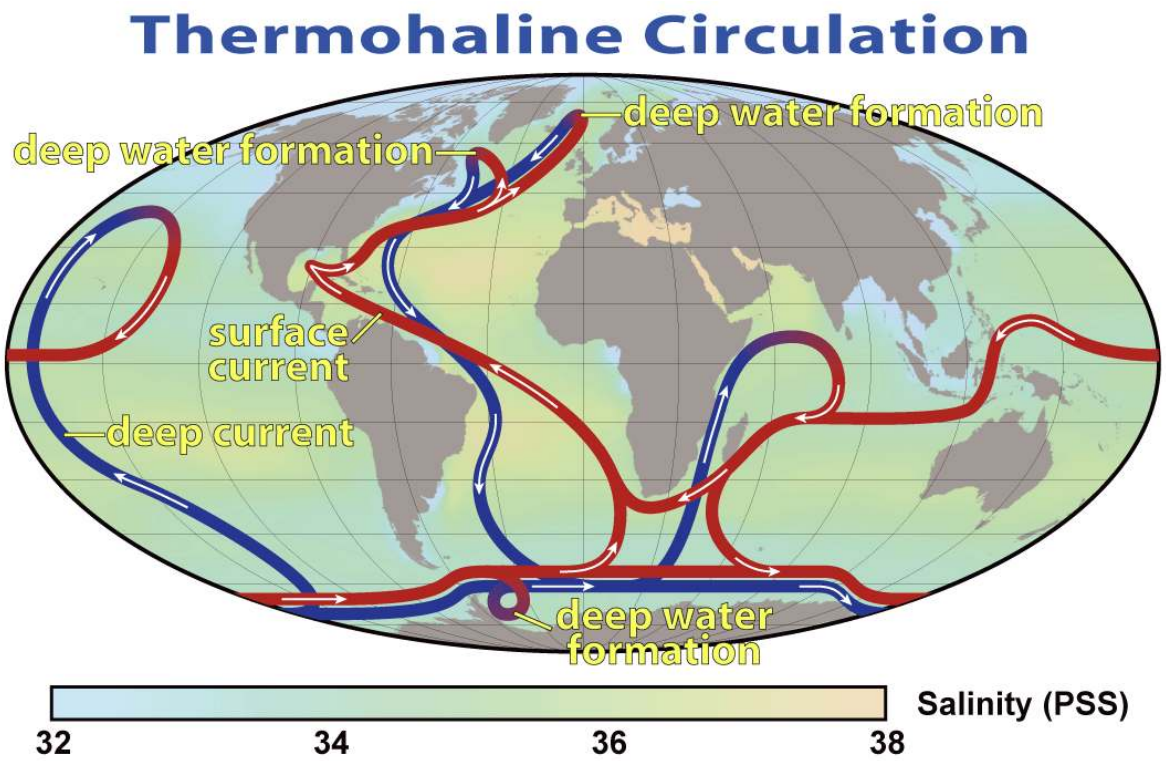



Figure **8.3.11**: The thermohaline circulation system, also known as the Global Ocean Conveyor.
["Thermohaline Circulation" by NASA. Public domain from: *https://en.Wikipedia.org/wiki/Thermo...culation_2.png*]

The thermohaline circulation is critically important to the transfer of heat on Earth. It brings warm water from the tropics to the poles, and cold water from the poles to the tropics, thus keeping polar regions from getting too cold and tropical regions from getting too hot. A reduction in the rate of thermohaline circulation would lead to colder conditions and enhanced formation of sea ice at the poles. This would start a positive feedback process that could result in significant global cooling. There is compelling evidence to indicate that there were major changes in thermohaline circulation, corresponding with climate changes, during the Pleistocene Glaciation.

## Contributors and Attributions

- Kyle Whittinghill (University of Pittsburgh)
- Contributed by Paul Webb Professor (Biology) at Rodger Williams University
- Contributed by Steven Earle Faculty (Geology) at Vancover Island University
- 18.4 Ocean Water from *Physical Geology* by Steven Earle (licensed under CC-BY)
- 8.2 Winds and the Coriolis Effect from *Introduction to Oceanography* by Paul Webb (licensed under CC-BY)
- 8.3 Winds and Climate from *Introduction to Oceanography* by Paul Webb (licensed under CC-BY)

This page titled 8.3: Atmospheric and Ocean Circulation is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Paul Webb.

# SECTION OVERVIEW

## 8.4: Air Pollution

**Air pollution** occurs in many forms but can generally be thought of as gaseous and particulate contaminants that are present in the earth's atmosphere. Chemicals discharged into the air that have a direct impact on the environment are called **primary pollutants**. These primary pollutants sometimes react with other chemicals in the air to produce **secondary pollutants**.

Air pollution is typically separated into two categories: outdoor air pollution and indoor air pollution. **Outdoor air pollution** involves exposures that take place outside of the built environment. Examples include fine particles produced by the burning of coal, noxious gases such as sulfur dioxide, nitrogen oxides and carbon monoxide; ground-level ozone and tobacco smoke. **Indoor air pollution** involves exposures to particulates, carbon oxides, and other pollutants carried by indoor air or dust. Examples include household products and chemicals, out-gassing of building materials, allergens (cockroach and mouse dropping, mold, pollen), and tobacco smoke.

This video summarizes the causes, consequences, and solutions to air pollution.



### Sources of Air Pollution

A **stationary source** of air pollution refers to an emission source that does not move, also known as a point source. Stationary sources include factories, power plants, and dry cleaners. The term **area source** is used to describe many small sources of air pollution located together whose individual emissions may be below thresholds of concern, but whose collective emissions can be significant. Residential wood burners are a good example of a small source, but when combined with many other small sources, they can contribute to local and regional air pollution levels. Area sources can also be thought of as non-point sources, such as construction of housing developments, dry lake beds, and landfills.

A **mobile source** of air pollution refers to a source that is capable of moving under its own power. In general, mobile sources imply "on-road" transportation, which includes vehicles such as cars, sport utility vehicles, and buses. In addition, there is also a "non-road" or "off-road" category that includes gas-powered lawn tools and mowers, farm and construction equipment, recreational vehicles, boats, planes, and trains.

**Agricultural sources** arise from operations that raise animals and grow crops, which can generate emissions of gases and particulate matter. For example, animals confined to a barn or restricted area produce large amounts of manure. Manure emits various gases, particularly ammonia into the air. This ammonia can be emitted from the animal houses, manure storage areas, or from the land after the manure is applied. In crop production, the misapplication of fertilizers, herbicides, and pesticides can potentially result in aerial drift of these materials and harm may be caused.

Unlike the above mentioned sources of air pollution, air pollution caused by **natural sources** is not caused by people or their activities. An erupting volcano emits particulate matter and gases, forest and prairie fires can emit large quantities of pollutants (figure 8.4.1), dust storms can create large amounts of particulate matter, and plants and trees naturally emit volatile organic

compounds which can form aerosols that can cause a natural blue haze. Wild animals in their natural habitat are also considered natural sources of “pollution”.




**Figure 8.4.1:** Smoke from the Camp fire in November of 2018 covers much of Northern California and extends over the Pacific Ocean. Image by NASA Earth Observatory/Lauren Dauphin (public domain).

## Criteria Air Pollutants

The Environmental Protection Agency has identified six common air pollutants called **criteria air pollutants**. They are particulate matter, ground-level ozone, carbon monoxide, sulfur oxides, nitrogen oxides, and lead. These pollutants can harm health and the environment, and cause property damage. Of the six pollutants, particle pollution and ground-level ozone are the most widespread health threats. The U.S. Environmental Protection Agency (EPA) regulates them by developing criteria based on considerations of human and environmental health.

1. **Ground-level ozone** is not emitted directly into the air, but is created by chemical reactions between nitrogen oxides ($NO_X$) and the gaseous oxygen ($O_2$) in the atmosphere react in the presence of sunlight. It also forms when nitrogen oxides and volatile organic compounds (VOCs) react. Emissions from industrial facilities and electric utilities, motor vehicle exhaust, gasoline vapors, and chemical solvents are some of the major sources of $NO_X$ and VOCs. Breathing ozone can trigger a variety of health problems, particularly for children, the elderly, and people of all ages who have lung diseases such as asthma. Ground level ozone can also have harmful effects on sensitive vegetation and ecosystems. (Ground-level ozone should not be confused with the ozone layer, which is high in the atmosphere and protects Earth from ultraviolet light; ground-level ozone provides no such protection).
2. **Particulate matter**, also known as particle pollution, is a complex mixture of extremely small particles and liquid droplets. Particle pollution is made up of a number of components, including acids (such as nitrates and sulfates), organic chemicals, metals, and soil or dust particles. Examples of human-generated particulate matter are from combustion (such as burning wood or coal) and mining (which releases dust). Wildfires or wind erosion are natural sources of particulate matter. Exposure to these particles can lead to respiratory irritation, reduction of lung capacity, lung cancer, and emphysema. The size of particles is directly linked to their potential for causing health problems. EPA is concerned about particles that are 10 micrometers in diameter or smaller because those are the particles that generally pass through the throat and nose and enter the lungs. Furthermore, particles that are 2.5 micrometers in diameter or smaller can pass from the lungs into the bloodstream, through which they are transported to various organs and cause serious health effects.
3. **Carbon monoxide** (CO) is a colorless, odorless gas emitted from combustion processes. Nationally and, particularly in urban areas, the majority of CO emissions to ambient air come from vehicles. However, volcanic eruptions and wildfires also release carbon monoxide. Carbon monoxide can cause harmful health effects by reducing oxygen delivery to the body’s organs (like the heart and brain) and tissues because carbon monoxide binds to hemoglobin. This causes dizziness, nausea, drowsiness, and headaches; at high concentrations CO can cause death.
4. **Nitrogen oxides** ($NO_X$) are a group of highly reactive gasses, which includes nitrogen monoxide (NO) and nitrogen dioxide ($NO_2$). EPA’s National Ambient Air Quality Standard uses $NO_2$ as the indicator for the larger group of nitrogen oxides. Nitrogen oxides forms quickly from emissions from cars, trucks and buses, power plants, and off-road equipment. Additionally, nitrogen

oxides are released naturally by certain bacteria. Nitric oxide is a clear, colorless gas formed during the combustion of fossil fuels. Nitrogen dioxide forms when nitric oxide reacts with atmospheric oxygen; the reddish-brown pungent gas is considered to be a secondary pollutant. Exposure to oxides of nitrogen can cause lung damage, aggravate asthma and bronchitis, and increase susceptibility to the flu and colds. In addition to contributing to the formation of ground-level ozone, and fine particle pollution, nitrogen oxides react in the atmosphere to form nitric acid ($HNO_3$) and nitrous acid ($HNO_2$), which are components of Acid Deposition. Nitrogen oxides is linked with a number of adverse effects on the respiratory system.

5. **Sulfur dioxide** ($SO_2$) is mainly released from fossil fuel combustion at power plants (73%) and other industrial facilities (20%). Smaller sources of $SO_2$ emissions include industrial processes such as extracting metal from ore, and the burning of high sulfur containing fuels by locomotives, large ships, and non-road equipment. Volcanoes are a natural source of $SO_2$. Oxides of sulfur have a characteristic rotten egg odor. Sulfur dioxide is linked with a number of adverse effects on the respiratory system. It reacts in the atmosphere to form sulfuric acid ($H_2SO_4$), a component of Acid Deposition.
6. **Lead** is a metal found naturally in the environment as well as in manufactured products. The major sources of lead emissions have historically been from fuels in on-road motor vehicles (such as cars and trucks) and industrial sources. As a result of regulatory efforts in the U.S. to remove lead from on-road motor vehicle gasoline, emissions of lead from the transportation sector dramatically declined by 95 percent between 1980 and 1999, and levels of lead in the air decreased by 94 percent between 1980 and 1999. Today, the highest levels of lead in air are usually found near lead smelters. The major sources of lead emissions to the air today are ore and metals processing and piston-engine aircraft operating on leaded aviation gasoline. Volcanoes and erosion are natural sources of lead.

## Other Air Pollutants

Carbon dioxide is produced by the complete combustion of fossil fuels. It is considered a greenhouse gas because it heats up the atmosphere by absorbing infrared radiation. As a result of this characteristic, excess amounts of carbon dioxide in the atmosphere may contribute to global warming. Carbon dioxide can also react with water in the atmosphere and produce slightly acidic rain. Carbon dioxide emissions can be reduced by limiting the amount of fossil fuels burned. Carbon dioxide is not currently regulated as an air pollutant by the EPA, but is included here because it is a greenhouse gas.

**Volatile organic compounds (VOCs)** include **hydrocarbons** such as methane (CH4), propane (C3H8), and octane (C8H18), and **chlorofluorocarbons (CFCs)** such as dichlorodifluoromethane (CCl2F2). Hydrocarbons are released into the atmosphere in automobile exhaust and from the evaporation of gasoline. They contribute to the formation of **photochemical smog**. Chlorofluorocarbons were used as propellants for aerosols and as refrigerants until it was discovered they can cause depletion of the protective ozone layer.

Radioactive substances include radon-222, iodine-131, and strontium-90. **Radon** is gas produced during the decay of uranium that is naturally present in rocks and building materials made with these rocks. It is known to cause lung cancer in humans. The other radioisotopes are produced by nuclear power plants (iodine-131) or are contained in the fallout from atmospheric nuclear testing (strontium-90). They can be introduced into the food chain through plants and become incorporated in the tissues of humans and other animals. Their ionizing radiation can produce cancers, especially those related to the thyroid and bone.

Hazardous air pollutants include **benzene** (C6H6) and **carbon tetrachloride** (CCl4). Benzene is a common organic solvent with numerous industrial uses. Carbon tetrachloride was formerly used as a solvent in the dry cleaning business. It is still used in industrial processes. Exposure to these compounds can cause cancer, birth defects and central nervous system problems.

## Indoor Air Pollution (Major concerns in developed countries)

Most people spend approximately 90 percent of their time indoors. However, the indoor air we breathe in homes and other buildings can be more polluted than outdoor air and can increase the risk of illness. There are many sources of indoor air pollution in homes. They include biological contaminants such as bacteria, molds and pollen, burning of fuels and environmental tobacco smoke, building materials and furnishings, household products, central heating and cooling systems, and outdoor sources. Outdoor air pollution can enter buildings and become a source of indoor air pollution.

**Sick building syndrome** is a term used to describe situations in which building occupants have health symptoms that are associated only with spending time in that building. Causes of sick building syndrome are believed to include inadequate ventilation, indoor air pollution, and biological contaminants. Usually indoor air quality problems only cause discomfort. Most

people feel better as soon as they remove the source of the pollution. Making sure that your building is well-ventilated and getting rid of pollutants can improve the quality of your indoor air.

### Secondhand Smoke (Environmental Tobacco Smoke)

Secondhand smoke is the combination of smoke that comes from a cigarette and smoke breathed out by a smoker. When a non-smoker is around someone smoking, they breathe in secondhand smoke.

Secondhand smoke is dangerous to anyone who breathes it in. There is no safe amount of secondhand smoke. It contains over 7,000 harmful chemicals, at least 250 of which are known to damage human health. It can also stay in the air for several hours after somebody smokes. Even breathing secondhand smoke for a short amount of time can hurt your body.

Over time, secondhand smoke can cause serious health issues in non-smokers. The only way to fully protect non-smokers from the dangers of secondhand smoke is to not allow smoking indoors. Separating smokers from nonsmokers (like “no smoking” sections in restaurants), cleaning the air, and airing out buildings does not completely get rid of secondhand smoke.

Source: Smokefree.gov

## Addressing Air Pollution

A combination of regulatory, economic, and technological strategies can reduce air pollution. The 1970 **Clean Air Act** is a key federal regulation that has successfully reduced emissions of air pollutants such as lead and sulfur dioxide. The Environmental Protection Agency sets an air quality standard for large emitters of criteria air pollutants, which is the maximum allowable amount of the pollutant that can be released. Each state then develops a plan to comply with these standards. Note that hundreds of air pollutants beyond the six criteria air pollutants are regulated under the Clean Air Act.

**Economic incentives** make it financially beneficial for individuals, institutions, or companies to use technologies or otherwise take action that limits air pollution. For examples, **green taxes** increase the cost of an environmentally harmful action. For example, Canada levies a green tax on fuel-inefficient vehicles. **Subsidies**, on the other hand reduce the cost of environmentally beneficial choices. For example, the federal solar tax credit in the United States allows homeowners to claim a percentage of the cost of installing solar panels when their filing taxes This amount would thus reduce taxes owed or be issued as a tax refund. (Increased usage of solar electricity ultimately reduces the need for coal and natural gas power plants, which generate air pollution.)

**Tradable permits** (**cap-and trade**) is another economic incentive. First, the total allowable amount of emissions for an entity (such as a state or country) is set. Next, emission credits are distributed to potential polluters. Industries that reduce emissions can sell their credits, while those with high emissions may be required. For example, large emitters of greenhouse gases in California must by carbon pollution permits. The purchase of these permits funds the California Climate Credit, which Californians receive as a reduction to their April and October electricity bills.

Finally, a variety of technologies reduce air pollution. For example, air filters and proper ventilation are essential for limiting indoor air pollution. Advances in public transportation (and policies that promote them), and energy-efficient vehicles further reduce air pollution. Electric vehicles also reduce air pollution (figure **8.4.2**), particularly if the the electricity that supplies them comes from a clean, renewable source, such as solar panels. **Catalytic converters** make emissions from standard vehicles less harmful by facilitating chemical reactions. For example, they catalyze reactions that convert nitrogen oxides ($NO_X$) to nitrogen gas ($N_2$) and carbon monoxide (CO) to carbon dioxide ($CO_2$). (While carbon dioxide is a greenhouse gas, it is not directly poisonous as carbon monoxide is.)




Figure 8.4.2: Mass transportation, especially those that use clean energy, reduces air pollution. Image by Mtattrain (CC-BY-SA).

Other technologies limit emissions from industry. For example, **smokestack scrubbers** remove some pollutants from power plant emissions before they are released and have reduced sulfur dioxide emissions from burning coal. **Electrostatic precipitators** negatively charge pollutants and remove charged pollutants with a positive electrode (figure 8.4.3). While electrostatic precipitators were originally developed for industry, smaller versions can be used as air purifiers in homes or businesses.

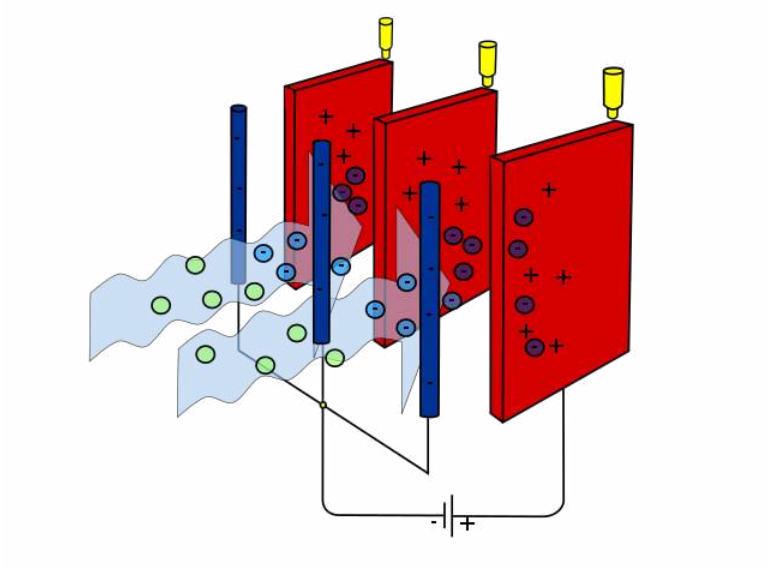


Figure 8.4.3: An electrostatic precipitator. Blue arrows represent the flow of air past metal rods, which negatively charge the pollutants. These are then drawn to positive electrodes (red, flat metal plates), removing the pollutants form the air. Image by Evan Mason (CC-BY-SA).

## Contributors and Attributions

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

- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.
- Atmospheric Pollution from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Basic Information about $NO_2$. Environmental Protection Agency. Accessed 11-04-2021 (public domain).
- Air, Water, and Soil from *AP Environmental Science* by The University of California College Prep

**Air Pollution**

8.4: Air Pollution is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 8.5: 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 **8.5.1**). 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.

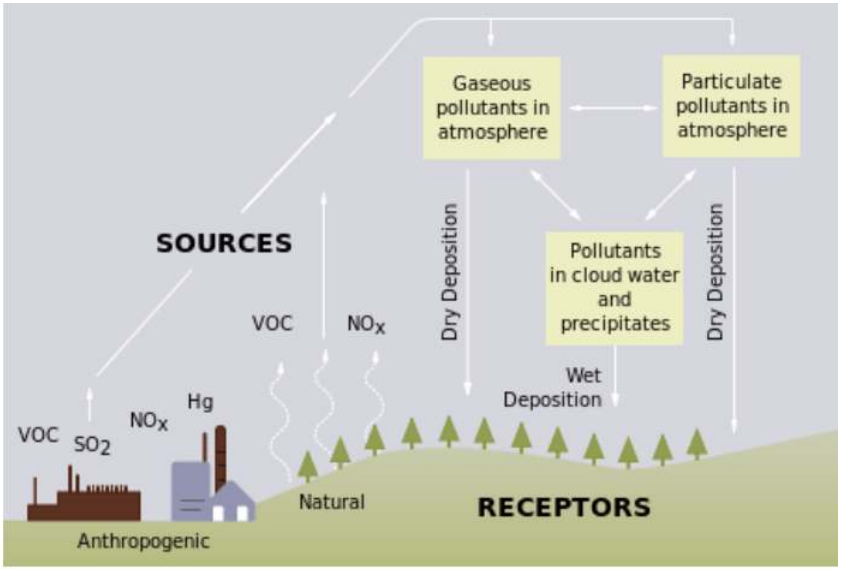



**Figure 8.5.1:** This diagram shows both anthropogenic and natural sources of acid deposition. Volatile organic compounds (VOCs), sulfur dioxide ($SO_2$), nitrogen oxides ($NO_X$), and mercury (Hg) are released, but only $SO_2$ and $NO_X$ react to form acids. Three categories of pollutants interact 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 **8.5.2**).

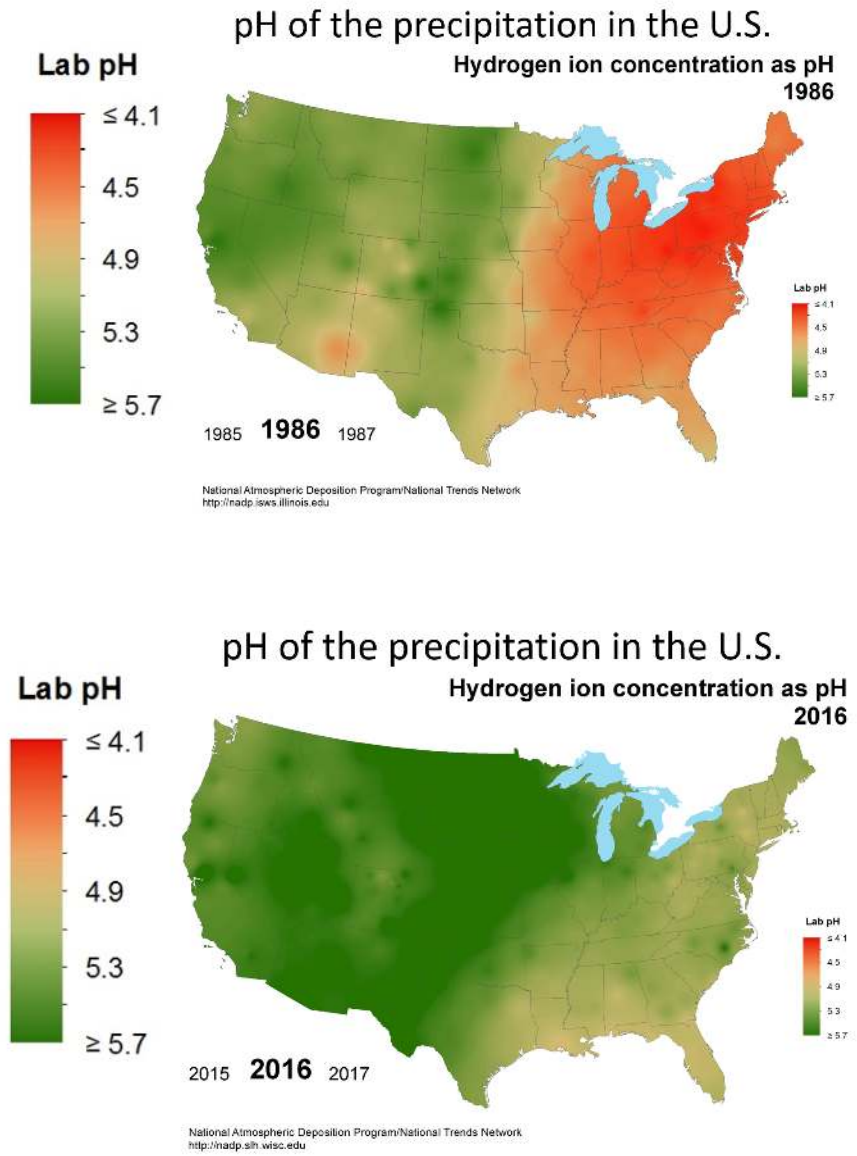



**Figure 8.5.2:** pH measures how acidic or basic a solution is, which depends on the concentration of hydrogen ions. Precipitation is normally slightly acidic, with a pH of around 5.6. (Anything below pH = 7 is considered acidic.) pH of precipitation was even more acidic in much of the East Coast of 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 8.5.3).



**Figure 8.5.3:** 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 8.5: Acid Deposition is shared under a CC BY-NC-SA 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.

# 8.6: Stratospheric Ozone Depletion

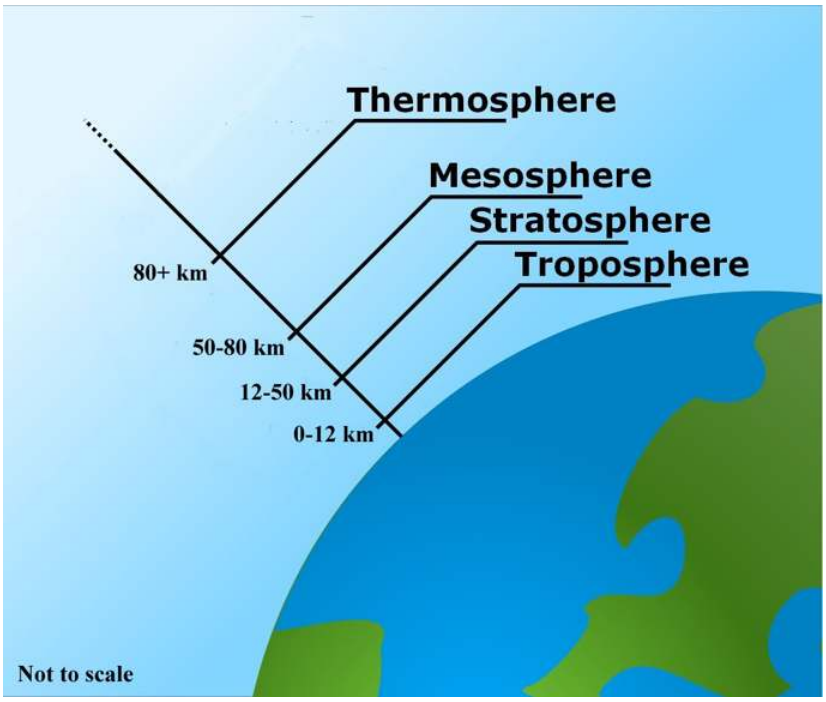



**Figure 8.6.*b*:** The layers of the atmosphere. the toposphere is the closest to the Earth's surface (0-12 km). Next, is the stratosphere (12-50 km), mesophere (50-80 km), and thermosphere (80+ km). The outermost layer (the exosphere) is not shown. **Ground-level ozone** in the troposphere is a form of air pollution, but the **ozone layer** in the stratosphere helps filter UV rays. Image by GFDL (CC-BY-SA).

In contrast to ground-level ozone, ozone in the stratosphere (which begins about 7 miles [11 km] up — where airliners cruise) protects life on Earth from the damaging effects of solar ultraviolet (UV) radiation (figure 8.6.1). The concentration of ozone high in the stratosphere has declined over the past two decades . Satellite monitoring of the stratosphere, which began in 1978, has revealed a marked decline. The most serious decline occurs over Antarctica in spring (October) when a precipitous drop in ozone causes an **ozone hole** (figure 8.6.2). The Dobson unit is a measure of the number of molecules of ozone in a vertical column of the atmosphere.

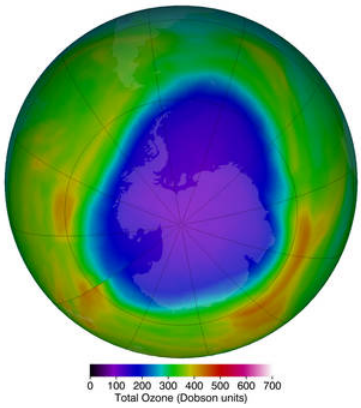



**Figure 8.6.2:** The Antarctic ozone hole that occurs annually in September and October during the Southern Hemisphere spring typically sees much lower ozone levels in than the Arctic. The purples and deep blues show the extent of low ozone levels on October 12, 2018, when they dropped to 104 Dobson units. Image and caption by NASA's Goddard Space Flight Center (public domain).

The ozone depletion process begins when **CFCs** (**chlorofluorocarbons**) and other ozone-depleting substances (ODS) are emitted into the atmosphere (figure **8.6.3**). After a period of several years, ODS molecules reach the stratosphere, about 10 kilometers above the Earth's surface (figure **8.6.1**). CFCs are synthetic gases in which the hydrogen atoms of methane are replaced by atoms of fluorine and chlorine (e.g., $CHF_2Cl$, $CFCl_3$, $CF_2Cl_2$). These gases are noninflammable, nontoxic, and very stable. They were widely used in industry as refrigerants (e.g., in refrigerators and air conditioners), solvents, propellants in aerosol cans (now banned in some countries), and in the manufacture of plastic foams. CFCs escape to the air from all of these uses (e.g., from leaky and discarded refrigeration units). Their chemical inertness, which makes CFCs so desirable for industry, also makes them a threat to the atmosphere. Once in the atmosphere, it may take 60–100 years for them to decompose and disappear.

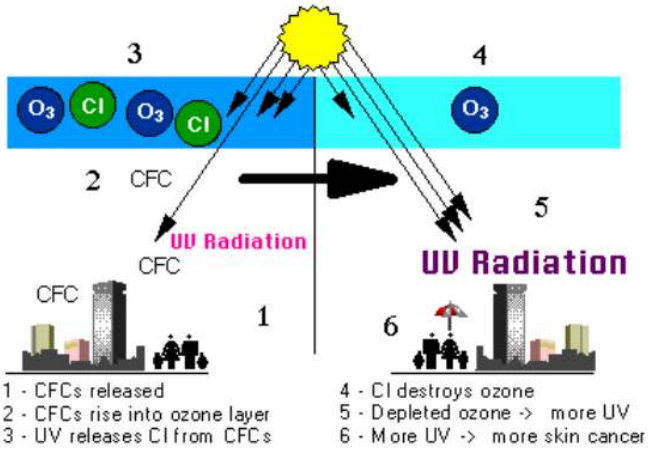



**Figure 8.6.3:** Strong ultraviolet (UV) light breaks apart the ozone-depleting substances (ODS). The process is as follows: (1) chlorofluorocarbons (CFCs) released, (2) CFCs rise into ozone layer in the stratosphere, which contains ozone ($O_3$), (3) UV radiation releases chlorine (Cl) from CFCs, (4) Cl destroys ozone, (5) depleted ozone allows more UV radiation to pass through the atmosphere, and (6) more UV radiation causes more skin cancer. Some ODS, including CFCs, hydrochlorofluorocarbons (HCFCs), carbon tetrachloride, and methyl chloroform release Cl atoms. Other ODS, including halons and methyl bromide, release bromine (Br) atoms. It is these atoms that actually destroy ozone, not the intact ODS molecule. It is estimated that one chlorine atom can destroy over 100,000 ozone molecules before it is removed from the stratosphere. Credit: NASA GSFC.

Ozone ($O_3$) is constantly produced and destroyed in a natural cycle called the **Chapman Cycle**, as shown in figure **8.6.4**. Ozone is produced in the stratosphere by natural photochemical reactions. They involve the absorption of solar UV radiation by oxygen molecules ($O_2$), which creates highly reactive oxygen atoms (O) that join with other $O_2$ molecules to form $O_3$ (figure **8.6.4**). Ozone ($O_3$) is then destroyed as it absorbs further UV radiation to produce an $O_2$ oxygen molecule and a oxygen atom (O). These reactions proceed relatively quickly in the stratosphere because high-energy UV radiation is abundant there. As a result, $O_3$ concentrations are typically 200-300 ppb in the stratosphere, about 10 times greater than in the ambient troposphere.

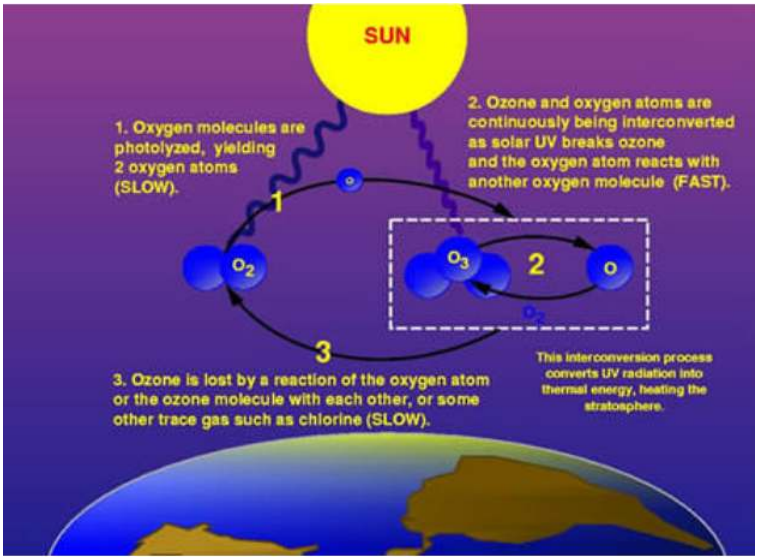



**Figure 8.6.4:** Because ozone ($O_3$) filters out a harmful type of ultraviolet radiation (UVB), less ozone means higher UVB levels at the surface. The more the depletion, the larger the increase in incoming UVB. UVB has been linked to skin cancer, cataracts, damage to materials like plastics, and harm to certain crops and marine organisms. Although some UVB reaches the surface even without ozone depletion, its harmful effects will increase as a result of this problem. The process of the ozone layer filtering UV radiation is as follows: (1) Gaseous oxygen molecules ($O_2$) are photolyzed (split), yielding two oxygen atoms. This is a slow process. (2) Ozone and oxygen atoms are continuously being interconverted as solar UV breaks $O_3$ into $O_2$ and a single oxygen atom (O). The oxygen atom (O) reacts with another $O_2$ molecule to form $O_3$. This interconversion is a fast process and converts UV radiation into thermal energy, heating the stratosphere. (3) Ozone is lost by a reaction of the oxygen atom or the ozone molecule with each other, or some other trace gas such as chlorine. This is a slow process.

While ozone production and destruction are balanced, ozone levels remain stable. Large increases in stratospheric ODS, however, have upset that balance. In effect, they are removing ozone faster than natural ozone creation reactions can keep up. Therefore, ozone levels fall. Stratospheric $O_3$ can be destroyed by various processes, including reactions with the trace gases $NO_x$and $N_2O$ and with reactive ions of bromine, chlorine, and fluorine. Because of anthropogenic emissions, the concentrations of some of these $O_3$-consuming chemicals have been increasing in the stratosphere, leading to concerns about the depletion of stratospheric $O_3$. It is widely believed that emissions of chlorofluorocarbons (CFCs), particularly the industrial gases known as freons, have been especially important in this regard. Because CFCs are extremely unreactive in the troposphere, they eventually migrate up to the stratosphere, where they are bombarded with UV radiation and slowly degrade (photodissociate) to release free chlorine. The chlorine efficiently reacts with and destroys $O_3$.

The $O_3$-destroying reactions proceed most effectively under extremely cold and stagnant conditions in the stratosphere, such as those occurring above polar latitudes at the end of the Antarctic and Arctic winters. These polar-focused $O_3$ depletions result in the development of so-called ozone holes during the early springtime. These phenomena have been observed regularly since the early 1980s. The $O_3$ holes over Antarctica are particularly extensive and typically involve decreases of the $O_3$ concentration of 30-50% during the spring. Smaller depletions of $O_3$ occur above the Arctic, including northern Canada. The affected areas in the Northern Hemisphere are much smaller than their counterparts in Antarctica. Although the seasonal losses of $O_3$ only take place over polar regions, lower latitudes are also affected when the $O_3$-depleted air becomes dispersed during the breakup of the holes. This temporarily reduces the stratospheric $O_3$ concentrations throughout the hemisphere, though not nearly to the same degree as occurs in the holes themselves.

## Policies to Reduce Ozone Destruction

One success story in reducing pollutants that harm the atmosphere concerns ozone-destroying chemicals. In 1973, scientists calculated that CFCs could reach the stratosphere and break apart. This would release chlorine atoms, which would then destroy ozone. Based only on their calculations, the United States and most Scandinavian countries banned CFCs in spray cans in 1978. More confirmation that CFCs break down ozone was needed before more was done to reduce production of ozone-destroying

chemicals. In 1985, members of the British Antarctic Survey reported that a 50% reduction in the ozone layer had been found over Antarctica in the previous three springs.

Two years after the British Antarctic Survey report, the "Montreal Protocol on Substances that Deplete the Ozone Layer" was ratified by nations all over the world. The **Montreal Protocol** controls the production and consumption of 96 chemicals that damage the ozone layer. CFCs have been mostly phased out since 1995, although they were used in developing nations until 2010. Some of the less hazardous substances will not be phased out until 2030. The Protocol also requires that wealthier nations donate money to develop technologies that will replace these chemicals. Because CFCs take many years to reach the stratosphere and can survive there a long time before they break down, the ozone hole did not immediately disappear after CFC emissions were reduced; however, it has been shrinking (figure **8.6.2**). The ozone layer will reach the same levels it had before 1980 around 2068 and 1950 levels in one or two centuries.

CFCs have largely been replaced by **hydrofluorocarbons** (HFCs) in air conditioners and refrigerators. HFCs do not destroy ozone but are potent greenhouse gases.

## Health and Environmental Effects of Ozone Layer Depletion

There are three types of UV light: UVA, UVB, and UVC. Reductions in stratospheric ozone levels will lead to higher levels of UVB reaching the Earth's surface. The sun's output of UVB does not change; rather, less ozone means less protection, and hence more UVB reaches the Earth. Studies have shown that in the Antarctic, the amount of UVB measured at the surface can double during the annual ozone hole.

Laboratory and epidemiological studies demonstrate that UVB causes non-melanoma skin cancer and plays a major role in malignant melanoma development. In addition, UVB has been linked to cataracts -- a clouding of the eye's lens. All sunlight contains some UVB, even with normal stratospheric ozone levels. It is always important to protect your skin and eyes from the sun (see this website from the EPA on sun safety for more information). Ozone layer depletion increases the amount of UVB and the risk of health effects.

UVB is generally harmful to cells, and therefore all organisms. UVB cannot penetrate into an organism very far and thus tends to only impact skin cells. Microbes like bacteria, however, are composed of only one cell and can therefore be harmed by UVB. Physiological and developmental processes of plants are affected by UVB radiation, even by the amount of UVB in present-day sunlight. Despite mechanisms to reduce or repair these effects and a limited ability to adapt to increased levels of UVB, plant growth can be directly affected by UVB radiation. Phytoplankton form the foundation of aquatic food webs. Phytoplankton productivity is limited to the euphotic zone, the upper layer of the water column in which there is sufficient sunlight to support net productivity. The position of the organisms in the euphotic zone is influenced by the action of wind and waves. In addition, many phytoplankton are capable of active movements that enhance their productivity and, therefore, their survival. Exposure to solar UVB radiation has been shown to affect both orientation mechanisms and motility in phytoplankton, resulting in reduced survival rates for these organisms.

## Contributors and Attributions

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

- Ozone Depletion 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
- **Ozone** in *Biology* by John W. Kimball, is licensed CC BY
- Gaseous Air Pollution in *Environmental Science: A Canadian Perspective* by Bill Freedman

This page titled 8.6: Stratospheric Ozone Depletion is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **10.2: Ozone Depletion** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# CHAPTER OVERVIEW

## 9: Climate Change


9.1: What Makes the Climate Change?

9.2: Prehistoric Climate Change and Proxys for Studying Past Climates

9.3: Anthropogenic Climate Change

9.4: Implications of Climate Change


---


9: Climate Change is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 9.1: What Makes the Climate Change?

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 feedbacks**), or suppress the change (**negative feedbacks**). In this section we will be discussing primarily 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.

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

### Life Cycle of the Sun

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 4.57 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 (Figure **9.1.1**). 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 GHGs—to one with only a few hundred parts per million of $CO_2$ and just under 1 part per million of $CH_4$ because, over geological time, life and its metabolic processes have evolved and changed the atmosphere.

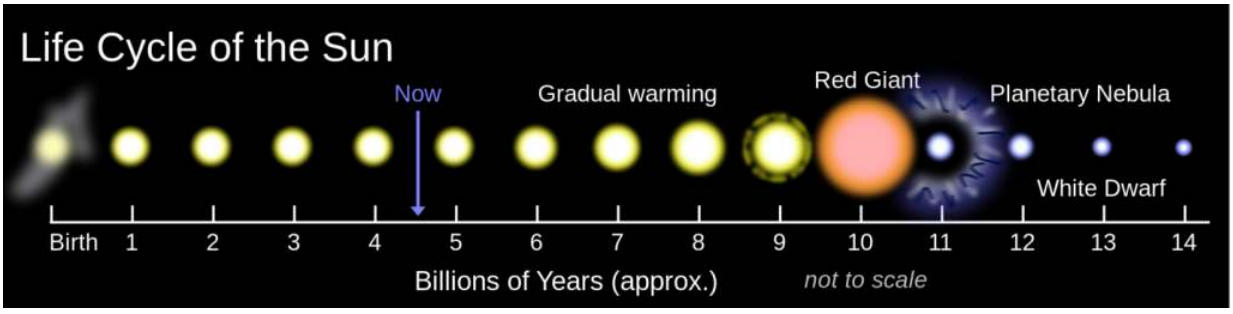



Figure **9.1.1**: The life cycle of our Sun and of other similar stars.
[from https://upload.wikimedia.org/wikiped..._Cycle.svg.png
"Solar Life Cycle" by Oliver Beatson. Public domain.]

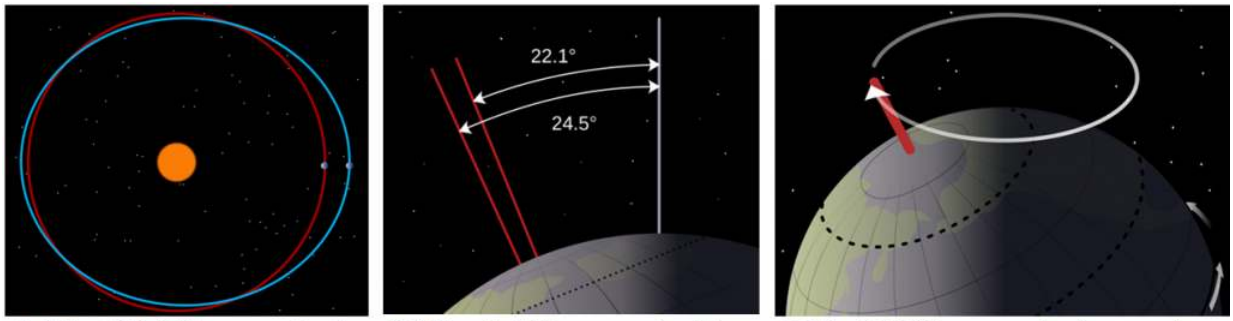
Changes occurring in the sun itself can affect the intensity of the sunlight that reaches Earth's surface. The intensity of the sunlight can cause either warming (during periods of stronger solar intensity) or cooling (during periods of weaker solar intensity). The sun follows a natural 11-year cycle of small ups and downs in intensity, but the effect of these 11-year cycles on Earth's climate is small.

### Milankovitch Cycles

Earth's orbit around the Sun is nearly circular, but like all physical systems, it has natural oscillations (Figure **9.1.2**). The importance of changes in the **eccentricity** (shape) of the Earth's orbit, **tilt** of the Earth, and **precession** (where the Earth's axis points) 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*. Glaciations are most sensitive to the insolation received at latitudes of around 65°, and with the current configuration of continents, it would have to be 65° north (because there is almost no land at 65° south).

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 **9.1.2**a). **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 **9.1.2**: The cycles of Earth's orbit and rotation. These are also called the Milankovitch Cycles.
[© Steven Earle. CC BY. after a: https://upload.wikimedia.org/wikiped..._zero.svg.png], b: https://upload.wikimedia.org/wikiped...range.svg.png] c: https://upload.wikimedia.org/wikiped...ns/thumb/4/43/Earth_precession.svg/2000px-Earth_precession.svg.png]

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 **9.1.2**b). 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 **9.1.2**c). 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 most important aspects are whether the northern hemisphere is pointing toward the Sun at its closest or farthest approach, and how eccentric the Sun's position is in Earth's orbit. Two opposing situations are illustrated in Figure **9.1.3**. In the upper panel, the northern hemisphere is at it farthest distance from the Sun during summer, which means cooler summers. In the lower panel, the northern hemisphere is at its closest distance to the Sun during summer, which means hotter summers. Cool summers—as opposed to cold winters—are the key factor in the accumulation of glacial ice, so the upper scenario in Figure **9.1.3** is the one that promotes glaciation. This factor is greatest when eccentricity is high.

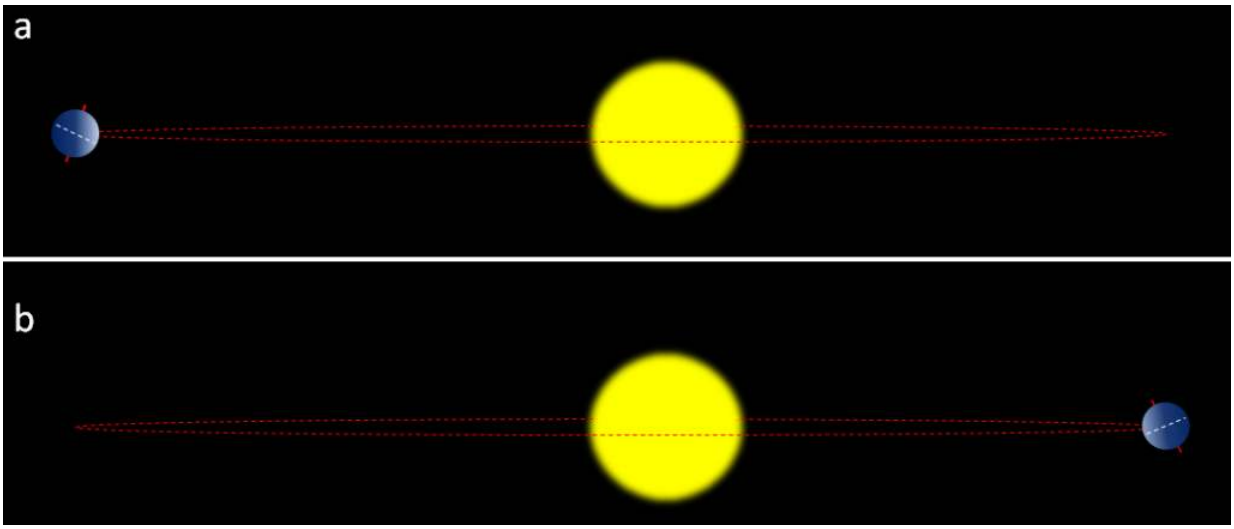



Figure 9.1.3: The effect of precession on insolation in the northern hemisphere summers. In (a) the northern hemisphere summer takes place at greatest Earth-Sun distance, so summers are cooler. In (b) (10,000 years or one-half precession cycle later) the opposite is the case, so summers are hotter. The red dashed line represents Earth's path around the Sun. © Steven Earle. CC BY.

As already stated, **climate feedbacks** are critically important in amplifying weak climate forcings into full-blown climate changes. When Milankovitch published his theory in 1924, it was widely ignored, partly because it was evident to climate scientists that the forcing produced by the orbital variations was not strong enough to drive the significant climate changes of the glacial cycles. Those scientists did not recognize the power of positive feedbacks. It wasn't until 1973, 15 years after Milankovitch's death, that sufficiently high-resolution data were available to show that the Pleistocene glaciations were indeed driven by the orbital cycles, and it became evident that the orbital cycles were just the forcing that initiated a range of feedback mechanisms that made the climate change.

Data for tilt, eccentricity, and precession over the past 400,000 years have been used to determine the insolation levels at 65° north, as shown in Figure 9.1.4. Also shown in Figure 9.1.4 are Antarctic ice-core temperatures from the same time period. The correlation between the two is clear, and it shows up in the Antarctic record because when insolation changes lead to growth of glaciers in the northern hemisphere, southern-hemisphere temperatures are also affected.

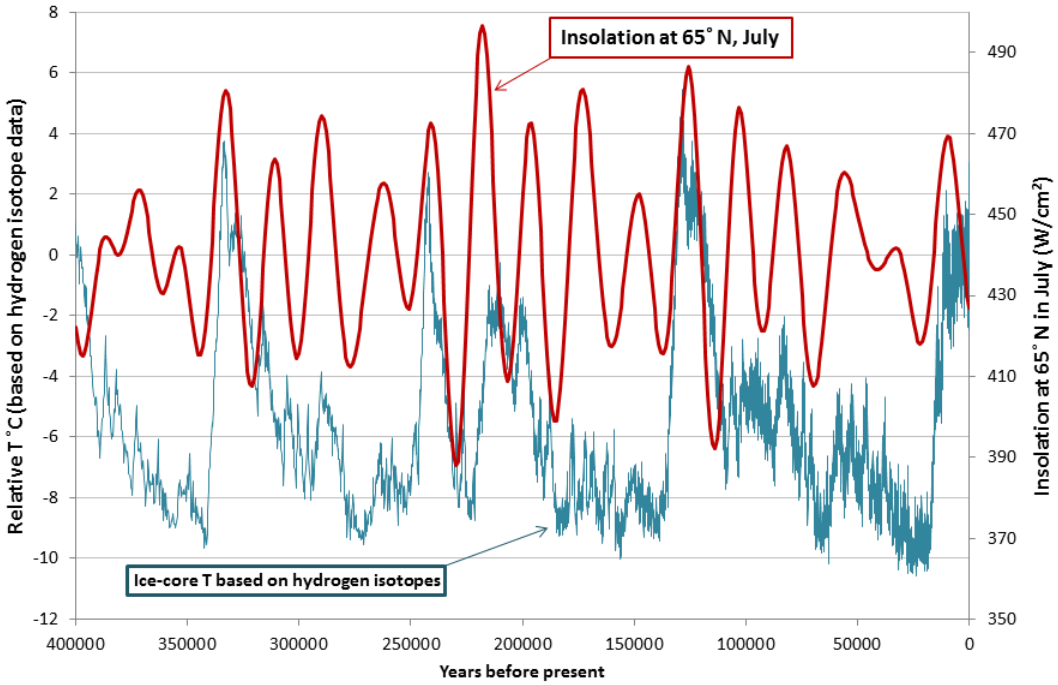



Figure **9.1.4**: Insolation at 65° N in July compared with Antarctic ice-core temperatures.
[© Steven Earle. CC BY. Based on data from Valerie Masson-Delmotte, "EPICA Dome C Ice Core 800KYr Deuterium Data and Temperature Estimates," WDCA Contribution Series Number : 2007 -091, NOAA/NCDC Paleoclimatology Program, Boulder CO, USA. Retrieved from: ftp://ftp.ncdc.noaa.gov/pub/data/pal...uttemp2007.txt and from Berger, A. and Loutre, M.F. (1991). Insolation values for the climate of the last 10 million years. Quaternary Science Reviews, 10, 297-317.]

## Plate Tectonic Processes

Plate tectonic processes contribute to climate forcing in several different ways, and on time scales ranging from tens of millions to hundreds of millions of years. One mechanism is related to continental position. For example, we know that Gondwana (South America + Africa + Antarctica + Australia) was positioned over the South Pole between about 450 and 250 Ma, during which time there were two major glaciations (Andean-Saharan and Karoo) affecting the South polar regions and cooling the rest of the planet at the same time. Another mechanism is related to continental collisions. For example, the collision between India and Asia, which started at around 50 Ma, resulted in massive tectonic uplift. The consequent accelerated weathering of this rugged terrain consumed $CO_2$ from the atmosphere and contributed to gradual cooling over the remainder of the Cenozoic. Changes in continental position can also lead to changes in ocean circulation, and therefore the distribution of energy from equator to the poles. For example, the opening of the Drake Passage — due to plate-tectonic separation of South America from Antarctica — led to the development of the Antarctic Circumpolar Current, which isolated Antarctica from the warmer water in the rest of the ocean and thus contributed to Antarctic glaciation starting at around 35 Ma.

Volcanic eruptions don't just involve lava flows and exploding rock fragments; various particulates and gases are also released, including carbon dioxide, water vapor, sulfur dioxide, hydrogen sulfide, hydrogen, and carbon monoxide. Volcanic eruptions can last a few days, but the solids and gases released during an eruption can influence the climate over a period of a few years, causing short-term climate changes. Generally, volcanic eruptions cool the climate. 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. This occurred in 1783 when volcanoes in Iceland erupted and caused the release of large volumes of sulfuric oxide. This led to haze-effect cooling, a global phenomenon that occurs when dust, ash, or other suspended particles block out sunlight and trigger lower global temperatures as a result; haze-effect cooling usually extends for one or more years. In Europe and North America, haze-effect cooling produced some of the lowest average winter temperatures on record in 1783 and 1784. 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.

## Greenhouse Gas Concentrations

Since scientists cannot go back in time to directly measure climatic variables, such as average temperature and precipitation, they must instead indirectly measure temperature. Antarctic ice cores are a key example of such evidence. These ice cores are samples of polar ice obtained by means of drills that reach thousands of meters into ice sheets or high mountain glaciers. Viewing the ice cores is like traveling backwards through time; the deeper the sample, the earlier the time period. Trapped within the ice are bubbles of air and other biological evidence that can reveal temperature and carbon dioxide data. Antarctic ice cores have been collected and analyzed to indirectly estimate the temperature of the Earth over the past 400,000 years (Figure **9.1.5**a). The 0 °C on this graph refers to the long-term average. Temperatures that are greater than 0 °C exceed Earth's long-term average temperature. Conversely, temperatures that are less than 0 °C are less than Earth's average temperature. This figure shows that there have been periodic cycles of increasing and decreasing temperature.

Before the late 1800s, the Earth has been as much as 9 °C cooler and about 3 °C warmer. Note that the graph in Figure **9.1.5**b shows that the atmospheric concentration of carbon dioxide has also risen and fallen in periodic cycles; note the relationship between carbon dioxide concentration and temperature. Figure **9.1.5**b shows that carbon dioxide levels in the atmosphere have historically cycled between 180 and 300 parts per million (ppm) by volume.

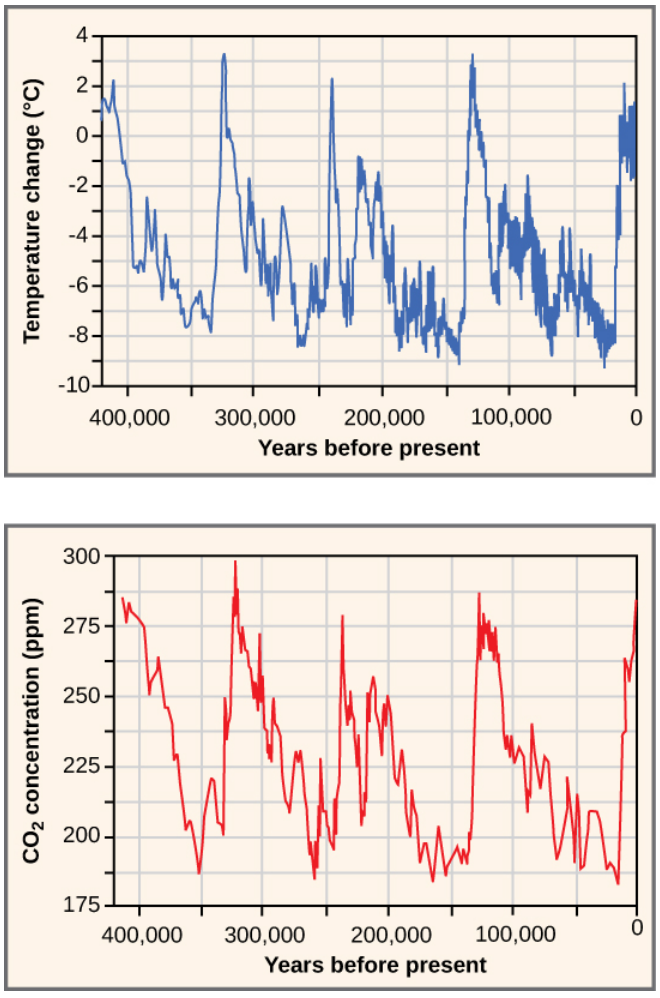



Figure **9.1.5**: Ice at the Russian Vostok station in East Antarctica was laid down over the course 420,000 years and reached a depth of over 3,000 m. By measuring the amount of $CO_2$ trapped in the ice, scientists have determined past atmospheric $CO_2$ concentrations. Temperatures relative to modern day were determined from the amount of deuterium (an isotope of hydrogen) present.

## Changes in Ocean Currents

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 (and other parts of the thermohaline circulation system) 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—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 past 65 years of ENSO index values are shown in Figure **9.1.6**. The strongest **El Niños** in recent decades were in 1983 and 1998, and those were both 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.

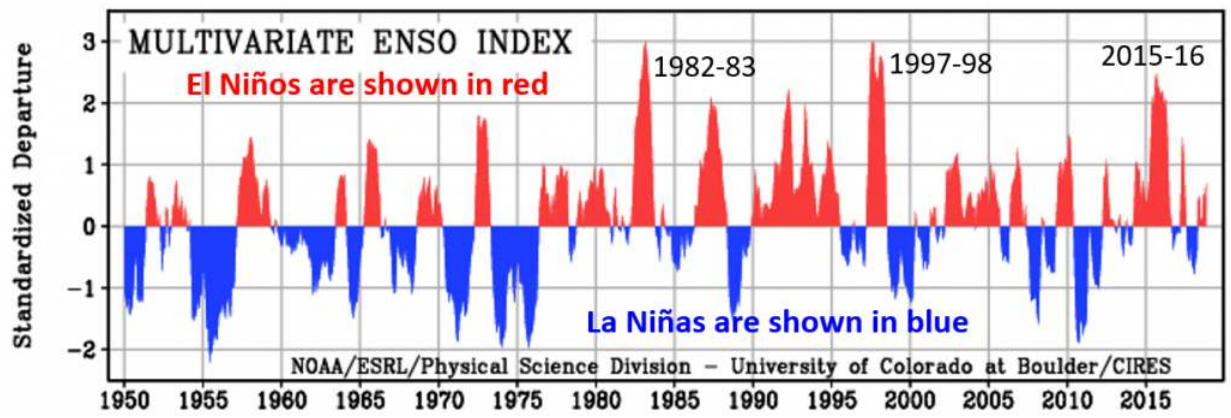



Figure **9.1.6**: Variations in the ENSO index from 1950 to early-2019.
[© Steven Earle. CC BY. after NOAA at: http://www.esrl.noaa.gov/psd/enso/mei/ "Multivariate ENSO Index (MEI)" by NOAA.]

## Contributors and Attributions

- 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
- Contributed by Paul Webb Professor (Biology) at Rodger Williams University
- Contributed by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, & Cam Mosher Faculty (Geology) at Salt Lake Community College
- Contributed by Steven Earle Faculty (Geology) at Vancover Island University
- 19.1 What Makes Climate Change from *Physical Geology* by Steven Earle (licensed under CC-BY)
- Sourced from OpenGeology from *An Introduction to Geology* Section 15.2 Earth's Temperature
- 8.1 Earth's Heat Budget from *Introduction to Oceanography* by Paul Webb (licensed under CC-BY)
- 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)

---

This page titled 9.1: What Makes the Climate Change? is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Steven Earle (BCCampus) .

- **44.5: Climate and the Effects of Global Climate Change** by OpenStax is licensed CC BY 4.0.

# 9.2: Prehistoric Climate Change and Proxys for Studying Past Climates

Over Earth history, the climate has changed a lot. For example, during the Mesozoic Era, the Age of Dinosaurs, the climate was much warmer and carbon dioxide was abundant in the atmosphere. However, throughout the Cenozoic Era (65 Million years ago to today), the climate has been gradually cooling. This section summarizes some of these major past climate changes.

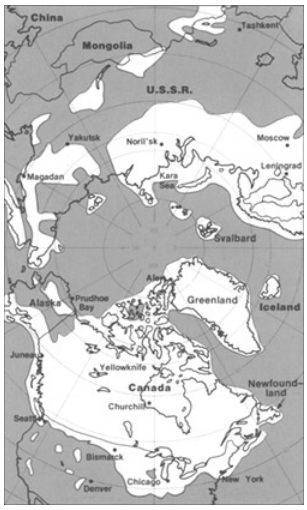



Figure **9.2.1**: Maximum extent of Laurentide Ice Sheet

## Past Glaciations

Through geologic history, the climate has changed slowly over millions of years. Before the most recent Pliocene-Quaternary glaciation, there were three other major glaciations [20]. The oldest, known as the Huronian, occurred toward the end of the Archean-early Proterozoic (~2.5 billion years ago). The major event of that time, the great oxygenation event, is most commonly associated with the cause of that glaciation. The increased oxygen is thought to have reacted with the potent greenhouse gas methane, causing cooling [21].

The end of the Proterozoic (about 700 million years ago) had another glaciation, known as the **Snowball Earth hypothesis** [22]. Glacial evidence has been interpreted in widespread rock sequences globally and even has been linked to low-latitude glaciation [23]. Limestone rock (usually formed in tropical marine environments) and glacial deposits (usually formed in cold climates) are often found together from this time in regions all around the world. In Utah, Antelope Island in the Great Salt Lake has interbedded limestone and glacial deposits (diamictites) interpreted to be formed by continental glaciation [24]. The idea of the controversial Snowball Earth hypothesis is that a runaway albedo effect (ice and snow reflecting solar radiation) might cause the complete freezing of land and ocean surfaces and a collapse of biological activity. The ice-covered earth would only melt when carbon dioxide from volcanoes reached high concentrations, due to the inability for carbon dioxide to enter the then-frozen ocean. Some studies estimated carbon dioxide was 350 times higher than today's concentrations [22]. The complete freezing [25] and the extent of the freezing [26] has come into question.

Glaciation also occurred in the Paleozoic, most notably with the Karoo Glaciation of the Pennsylvanian (323 to 300 million years ago). This also was caused by an increase of oxygen and a subsequent drop in carbon dioxide, most likely produced by the evolution and rise of land plants [27].

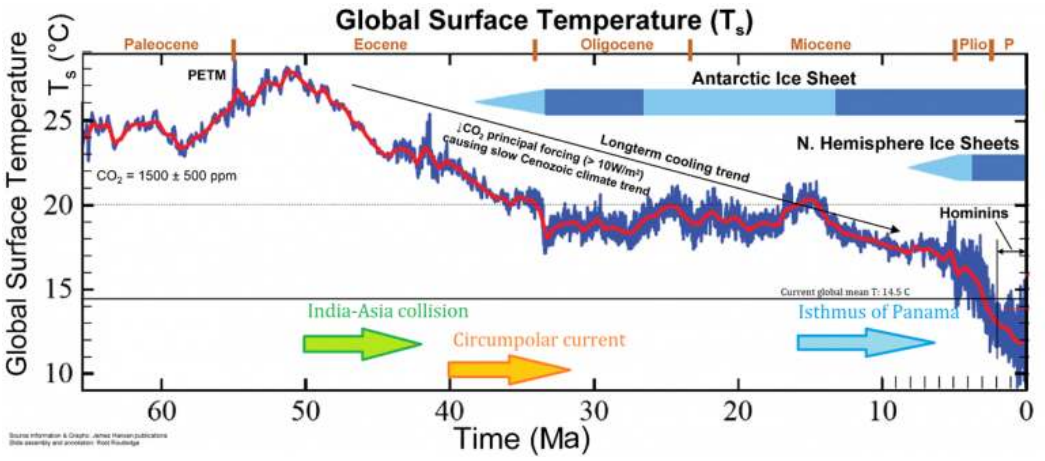



Figure **9.2.2**: Global average surface temperature over the past 70 million years.

During the Cenozoic Era (the last 65 million years), the climate started out warm and gradually cooled to today. This warm time is called the **Paleocene-Eocene Thermal Maximum** and Antarctica and Greenland were ice-free during this time (Figure **9.2.2**). Since the Eocene, tectonic events during the Cenozoic caused persistent and significant planetary cooling. For example, the collision of the Indian Plate with the Asian Plate created the Himalaya Mountains increasing weathering and erosion rates. An increased rate of weathering of silicate minerals, especially feldspar, consumes carbon dioxide from the atmosphere and therefore reduces the greenhouse effect, resulting in long-term cooling [28].

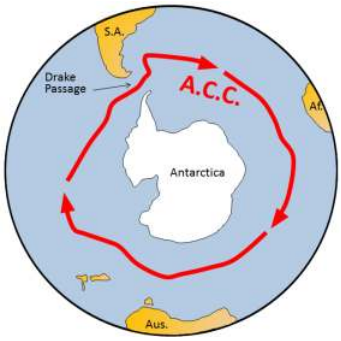
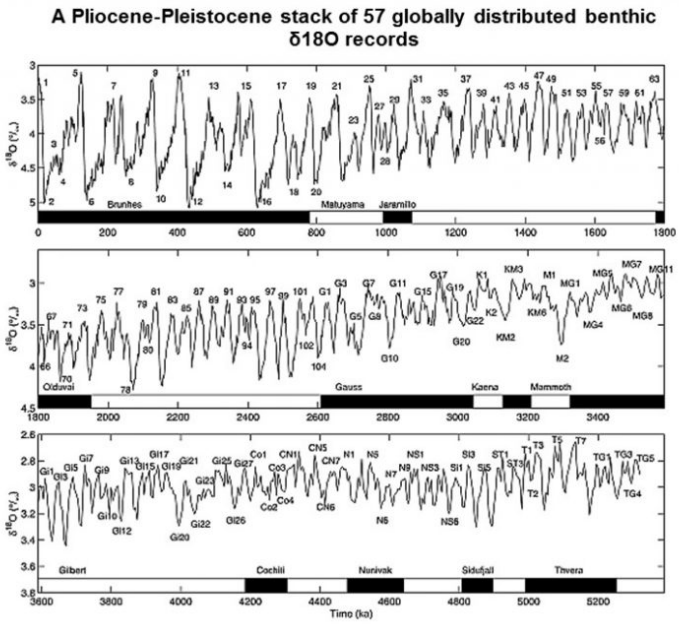



Figure **9.2.3**: The Antarctic Circumpolar Current

At about 40 Ma, the narrow gap between the South American Plate and the Antarctica Plate widened, resulting in the opening of the Drake Passage. This allowed for the unrestricted west-to-east flow of water around Antarctica, the Antarctic Circumpolar Current, which effectively isolated the southern ocean from the warmer waters of the Pacific, Atlantic, and Indian Oceans (Figure **9.2.3**). The region cooled significantly, and by 35-million-year ago (Oligocene) glaciers had started to form on Antarctica [29].

At around 15 Ma, subduction-related volcanism between Central and South America created the Isthmus of Panama that connected North and South America. This prevented water from flowing between the Pacific and Atlantic Oceans and reduced heat transfer from the tropics to the poles. This created a cooler Antarctica and larger Antarctic glaciers. The expansion of that ice sheet (on land and water) increased Earth's reflectivity (albedo), a positive feedback loop of further cooling: more reflective glacial ice, more cooling, more ice, and so on [30; 31].

By 5 million years ago (Pliocene Epoch), ice sheets had started to grow in North America and northern Europe. The most intense part of the current glaciation is the last 1 million years of the Pleistocene Epoch. The Pleistocene has significant temperature variations (through a range of almost 10°C) on time scales of 40,000 to 100,000 years, and corresponding expansion and contraction of ice sheets. These variations are attributed to subtle changes in Earth's orbital parameters called **Milankovitch cycles** [32; 33], which are explained in more detail in the section on natural climate forcings. Over the past million years, the glaciation cycles have been approximately every 100,000 years [34] with many glacial advances in the last 2 million years (Figure **9.2.4**) (Lisiecki and Raymo, 2005) [35].




**Source:** Lisiecki, L. E., & Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography*, *20*(1).

Figure **9.2.4**: A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. X-axis is time in thousands of years (ka) so 200 is actually 200,000. (Source: Lisiecki and Raymo, 2005)

Warmer portions of climate within an ice age are called **interglacials**, with brief versions called **interstadials**. These warming upticks are related to variations in Earth's climate like Milankovitch cycles. In the last 500,000 years, there have been 5 or 6 interglacials, with the most recent belonging to our current time, the Holocene (Figure **9.2.4**).

Two of the more recent climate swings demonstrate the complexity of the changes: the Younger Dryas and the Holocene Climatic Optimum. These events are more recent and yet have conflicting information. The Younger Dryas cooling is widely recognized in the Northern Hemisphere [36], though the timing of the event (about 12,000 years ago) does not appear to be equal everywhere [37]. It also is difficult to find in the Southern Hemisphere [38]. The Holocene Climatic Optimum is the warming around 6,000 years ago [39], though it was not universally warmer, and probably not as warm as current warming [40], and not at the same time everywhere [41].

Two significant temperature anomalies, or irregularities, have occurred in the last 2000 years. These are the Medieval Climate Anomaly (or the Medieval Warm Period) and the Little Ice Age. A third temperature anomaly aligns with the Industrial Era. The Medieval Climate Anomaly occurred between 900 and 1300 AD. During this time period, many climate scientists think that slightly warmer weather conditions prevailed in many parts of the world; the higher-than-average temperature changes varied between 0.10 °C and 0.20 °C above the norm. Although 0.10 °C does not seem large enough to produce any noticeable change, it did free seas of ice. Because of this warming, the Vikings were able to colonize Greenland. The Little Ice Age was a cold period that occurred between 1550 AD and 1850 AD. During this time, a slight cooling of a little less than 1 °C was observed in North America, Europe, and possibly other areas of the Earth. This 1 °C change in global temperature is a seemingly small deviation in temperature (as was observed during the Medieval Climate Anomaly); however, it also resulted in noticeable changes. Historical accounts reveal a time of exceptionally harsh winters with much snow and frost.

## Proxy Indicators of Past Climates

How do we know about past climates? Scientists use proxy indicators to understand past climate. A **proxy indicator** is a biological, chemical, or physical signature preserved in the rock, sediment, or ice record that acts like a "fingerprint" of something in the past [42]. Thus they are an indirect indicator of something like climate. For ancient glaciations from the Proterozoic and Paleozoic, there are rock formations of glacial sediments such as the diamictite (or tillite) of the Mineral Fork Formation in Utah. This dark rock has many fine-grained components plus some large out-sized clasts like a modern glacial till [43; 44].

For climate changes during the Cenozoic Era (the last 65 Ma), there is a detailed chemical record from the coring of deep-sea sediments as part of the Ocean Drilling Program (Figure **9.2.5**). Studies of deep-sea sediment use stable carbon and oxygen isotopes obtained from the shells of deep-sea benthic foraminifera that have settled on the ocean floor over millions of years. Oxygen isotopes are a proxy indicator of deep-sea temperatures and continental ice volume [45].

## Sediment Cores – Stable Oxygen Isotope



Figure **9.2.5**: Sediment core from the Greenland continental slope (Source: Hannes Grobe)

Oxygen isotopes are an indicator of past climate. The two main stable oxygen isotopes are $^{16}O$ and $^{18}O$. They both occur in water ($H_2O$) and in the calcium carbonate ($CaCO_3$) shells of foraminifera as the oxygen component of both of those molecules. The most abundant and lighter isotope is $^{16}O$. Since it is lighter, it evaporates more easily from the ocean's surface as water vapor, which later turns to clouds and precipitation on the ocean and land.

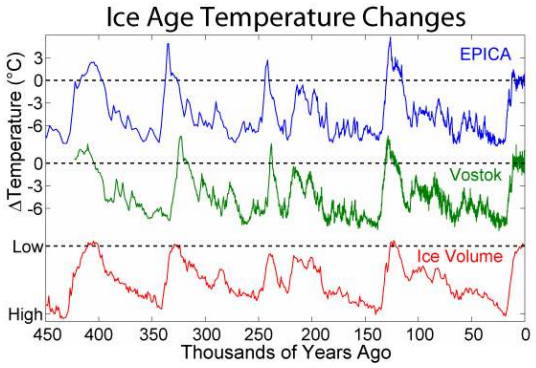



Figure **9.2.6**: Antarctic temperature changes during the last few glaciations compared to global ice volume. The first two curves are based on the deuterium (heavy hydrogen) record from ice cores (EPICA Community Members 2004, Petit et al. 1999). The bottom line is ice volume based on oxygen isotopes from a composite of deep-sea sediment cores (Lisiecki and Raymo 2005).

During geologic times when the climate is cooler, more of this precipitation is locked onto land in the form of glacial ice. Consider the giant ice sheets, more than a mile thick, that covered a large part of North America during the last ice age only 14,000 years ago. During glaciation, the glaciers effectively lock away more $^{16}O$, thus the ocean water and foraminifera shells become enriched in $^{18}O$. Therefore, a ratio of $^{18}O$ to $^{16}O$ in calcium carbonate shells of foraminifera is an indicator of past climate. The sediment cores from the Ocean Drilling Program record a continuous accumulation of sediment (Figure **9.2.6**).

## Sediment Cores – Boron Isotopes and Acidity

Boron-isotope ratios in ancient planktonic foraminifera shells in deep-sea sediment cores have been used to estimate the pH (acidity) of the ocean over the past 60 million years. Ocean acidity is a proxy for past atmospheric $CO^2$ concentrations. In the early Cenozoic, around 60 million years ago, $CO^2$ concentrations were over 2,000 ppm and started falling around 55 to 40 million years ago possibly due to reduced $CO^2$ outgassing from ocean ridges, volcanoes, and metamorphic belts and increased carbon burial due to uplift of the Himalaya Mountains. By the Miocene (about 24 million years ago), $CO^2$ levels were below 500 ppm [46] and by 800,000 years ago $CO^2$ levels didn't exceed 300 ppm [47].

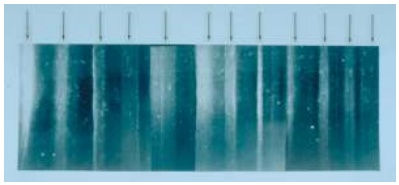


Figure 9.2.7: 19 cm long section of ice core showing 11 annual layers with summer layers (arrowed) sandwiched between darker winter layers. (Source: US Army Corps of Engineers)

For the more recent Pleistocene climate, there is a more detailed and direct chemical record from coring into the Antarctic and Greenland ice sheets (Figure 9.2.7). Snow accumulates on these ice sheets and creates yearly layers. Ice cores have been extracted from ice sheets covering the last 800,000 years. Oxygen isotopes are collected from these annual layers and the ratio of $^{18}O$ to $^{16}O$ is used to determine temperature as discussed above. In addition, the ice traps small atmospheric gas bubbles as the snow turns to ice.

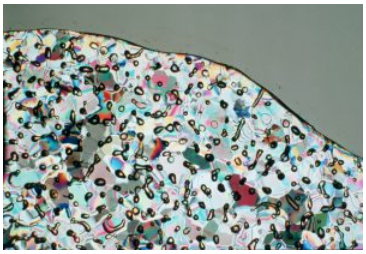


Figure 9.2.8: Antarctic ice showing hundreds of tiny trapped air bubbles from the atmosphere thousands of years ago. (Source: CSIRO)

Small pieces of this ice are crushed and the ancient air extracted into a mass spectrometer that can detect the chemistry of the ancient atmosphere. Carbon dioxide levels are recreated from these measurements. Over the last 800,000 years, the maximum carbon dioxide concentration during warm times was about 300 ppm and the minimum during cold stretches was about 170 ppm [46; 47; 48] (Figure 9.2.9). The carbon dioxide content of earth's atmosphere is currently over 400 ppm.

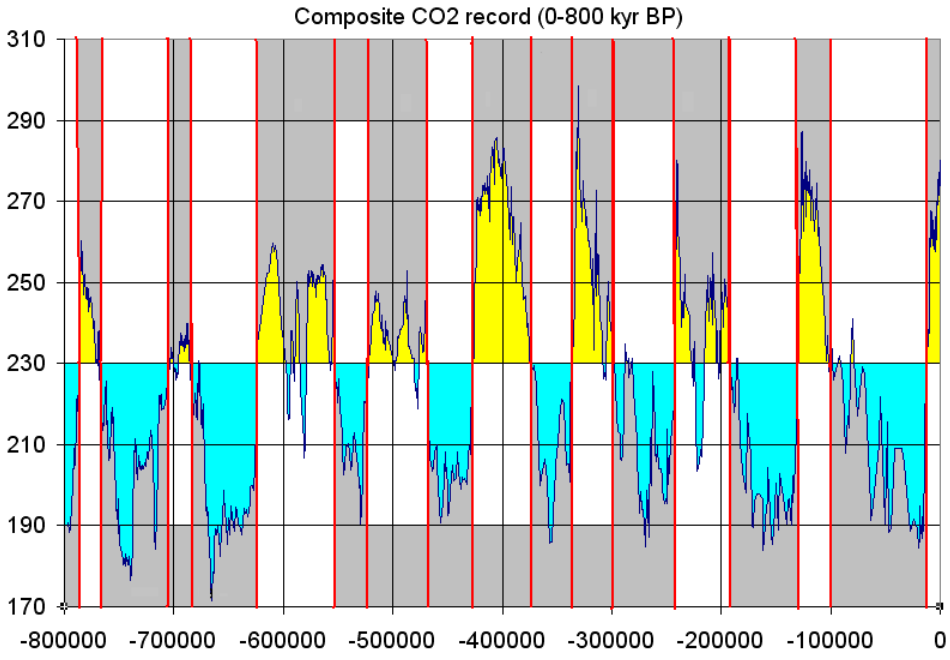



Figure 9.2.9: Composite carbon dioxide record from the last 800,000 years based on ice core data from EPICA Dome C Ice Core.


https://bio.libretexts.org/@go/page/69448

## Oceanic Microfossils

Microfossils, like foraminifera, diatoms, and radiolarians, can be used to interpret past climate records. In sediment cores, different species of microfossils are found in different layers. Groups of these microfossils are called assemblages. One assemblage consists of species that lived in cooler ocean water (in glacial times) and another assemblage found at a different level in the same sediment core is made of warmer water species [49].





## Tree Rings



Figure **9.2.10**: Tree rings form every year. Rings that are farther apart are from wetter years and rings that are closer together are from dryer years.

Every year a tree will grow one ring with a light section and dark section (Figure **9.2.10**).. The rings vary in width. Since trees need a lot of water to survive, narrower rings indicate colder and drier climates. Since some trees can be several thousands years old, we

can use their rings for regional paleoclimatic reconstructions. Further, dead trees such as those used in Puebloan ruins can be used to extend this proxy indicator, which showed long term droughts in the region and why their villages were abandoned (Figure 9.2.11)..

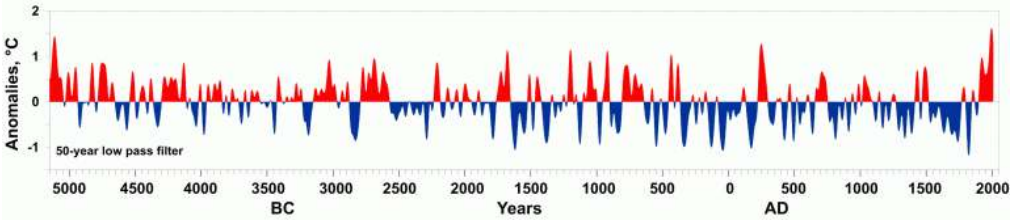



Figure 9.2.11: Summer temperature anomalies for the past 7000 years (Source: R.M.Hantemirov)

### Pollen

Flowering plants produce pollen grains. Pollen is distinctive when viewed under a microscope (Figure 9.2.5). Sometimes pollen can be preserved in lake sediments that accumulate every year. Coring of lake sediments can reveal ancient pollen. Fossil pollen assemblages are groups of pollen from multiple species such as spruce, pine, and oak. Through time (via the sediment cores and radiometric age-dating techniques), the pollen assemblage will change revealing the plants that lived in the area at the time. Thus the pollen assemblages are an indicator of past climate since different plants will prefer different climates [50]. For example, in the Pacific Northwest east of the Cascades, a region close to the border of grasslands and forest, a study tracked pollen over the last 125,000 years covering the last two glaciations. As shown in the figure (Fig. 2 from reference Whitlock and Bartlein 1997 [51]), pollen assemblages with more pine tree pollen are found during glaciations and pollen assemblages with less pine tree pollen are found during interglacial times [51].

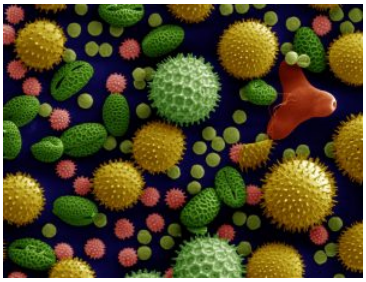


Figure 9.2.12: Scanning electron microscope image of modern pollen with false color added to distinguish plant species. (Source: Dartmouth Electron Microscope Facility, Dartmouth College)

## Other Proxy Indicators

Paleoclimatologists study many other phenomena to understand past climates such as human historical accounts, human instrument record from the recent past, lake sediments, cave deposits of calcium carbonates, and corals (both oxygen isotopes and growth rings).

## References


20. Deynoux, M., Miller, J. M. G. & Domack, E. W. *Earth's Glacial Record*. (Cambridge University Press, 2004).

21. Kopp, R. E., Kirschvink, J. L., Hilburn, I. A. & Nash, C. Z. The Paleoproterozoic snowball Earth: a climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 11131–11136 (2005).

22. Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. A neoproterozoic snowball earth. *Science* **281**, 1342–1346 (1998).

23. Schopf, J. W. & Klein, C. Late Proterozoic Low-Latitude Global Glaciation: the Snowball Earth. in *The Proterozoic biosphere : a multidisciplinary study* (eds. Schopf, J. W. & Klein, C.) 51–52 (Cambridge University Press, 1992).

24. Doelling, H. H. *et al.* Geology of Antelope Island. *Davis County, Utah: Utah Geological and Mineral Survey Open-File Release* **144**, 82 (1988).

25. Allen, P. A. & Etienne, J. L. Sedimentary challenge to Snowball Earth. *Nat. Geosci.* **1**, 817–825 (2008).

26. Eyles, N. & Januszczak, N. ‘Zipper-rift’: a tectonic model for Neoproterozoic glaciations during the breakup of Rodinia after 750 Ma. *Earth-Sci. Rev.* (2004).

27. Berner, R. A. The carbon cycle and carbon dioxide over Phanerozoic time: the role of land plants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **353**, 75–82 (1998).

28. Kump, L. R., Brantley, S. L. & Arthur, M. A. Chemical Weathering, Atmospheric CO2, and Climate. *Annu. Rev. Earth Planet. Sci.* **28**, 611–667 (2000).

29. Lagabrielle, Y., Goddéris, Y., Donnadieu, Y., Malavieille, J. & Suarez, M. The tectonic history of Drake Passage and its possible impacts on global climate. *Earth Planet. Sci. Lett.* **279**, 197–211 (2009).

30. Kuipers Munneke, P. *et al.* A new albedo parameterization for use in climate models over the Antarctic ice sheet. *J. Geophys. Res.* **116**, D05114 (2011).

31. Curry, J. A., Schramm, J. L. & Ebert, E. E. Sea Ice-Albedo Climate Feedback Mechanism. *J. Clim.* **8**, 240–247 (1995).

32. Milankovitch, M. Mathematische klimalehre und astronomische theorie der klimaschwankungen. (1930).

33. Roe, G. In defense of Milankovitch. *Geophys. Res. Lett.* **33**, L24703 (2006).

34. Abe-Ouchi, A. *et al.* Insolation-driven 100,000-year glacial cycles and hysteresis of ice-sheet volume. *Nature* **500**, 190–193 (2013).

35. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. *Paleoceanography* **20**, (2005).

36. Carlson, A. E. The younger Dryas climate event. (2013).

37. Fairbanks, R. G. The age and origin of the “Younger Dryas climate event” in Greenland ice cores. *Paleoceanography* **5**, 937–948 (1990).

38. Kaplan, M. R. *et al.* Glacier retreat in New Zealand during the Younger Dryas stadial. *Nature* **467**, 194–197 (2010).

39. Ganopolski, A., Kubatzki, C., Claussen, M., Brovkin, V., V. & Petoukhov, V., V. The influence of vegetation-atmosphere-ocean interaction on climate during the mid-holocene. *Science* **280**, 1916–1919 (1998).

40. Hewitt, C. D. & Mitchell, J. F. B. A fully coupled GCM simulation of the climate of the mid-Holocene. *Geophys. Res. Lett.* **25**, 361–364 (1998).

41. He, Y. *et al.* Asynchronous Holocene climatic change across China. *Quat. Res.* **61**, 52–63 (2004).

42. Weissert, H. Deciphering methane’s fingerprint. *Nature* **406**, 356–357 (2000).

43. Ojakangas, R. W. & Matsch, C. L. Upper Precambrian (Eocambrian) Mineral Fork Tillite of Utah: A continental glacial and glaciomarine sequence. *GSA Bulletin* **91**, 495–501 (1980).

44. Christie-Blick, N., Ojakangas, R. W. & Matsch, C. L. Upper Precambrian (Eocambrian) Mineral Fork Tillite of Utah: A Continental Glacial and Glaciomarine Sequence: Discussion and Reply. *The Geological Society of America Bulletin* **93**, 184–187 (1982).

45. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).

46. Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406**, 695–699 (2000).

47. Lüthi, D. *et al.* High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453**, 379–382 (2008).

48. Oak Ridge National Laboratory. 800,000-year Ice-Core Records of Atmospheric Carbon Dioxide (CO2). (2008). Available at: http://cdiac.ornl.gov/trends/co2/ice_core_co2.html. (Accessed: 14th September 2016)

49. Cunningham, W. L., Leventer, A., Andrews, J. T., Jennings, A. E. & Licht, K. J. Late Pleistocene–Holocene marine conditions in the Ross Sea, Antarctica: evidence from the diatom record. *The Holocene* **9**, 129–139 (1999).

50. Webb, T. & Thompson, W. Is vegetation in equilibrium with climate? How to interpret late-Quaternary pollen data. *Vegetatio* **67**, 75–91 (1986).

51. Whitlock, C. & Bartlein, P. J. Vegetation and climate change in northwest America during the past 125 kyr. *Nature* **388**, 57–61 (1997).

## Contributors and Attributions

- Contributed by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, & Cam Mosher Faculty (Geology) at Salt Lake Community College
- Sourced from OpenGeology from *An Introduction to Geology* Section 15.2 Prehistoric Climate Change

This page titled 9.2: Prehistoric Climate Change and Proxys for Studying Past Climates is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, & Cam Mosher (OpenGeology) .

# 9.3: 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) to drive machinery and trains. 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. Some climate scientists argue that anthropogenic climate change actually goes back much further than the industrial era, and that humans began to impact the climate by clearing land to grow grains in Europe and the Middle East around 8,000 years BCE and by creating wetlands to grow rice in Asia around 5,000 years BCE. Clearing forests for crops is a type of climate-forcing because the $CO_2$ storage capacity of the crops is generally lower than that of the trees they replace, and creating wetlands is a type of climate forcing because the anaerobic bacterial decay of organic matter within wetlands produces $CH_4$.

In fact, whether anthropogenic climate change started with the agricultural revolution or the industrial revolution is not important, because the really significant climate changes didn't start until the early part of the 20th century, and although our activities are a major part of the problem, our increasing numbers are a big issue as well. Figure **9.3.1** shows the growth of the world population from around 5 million, when we first started growing crops, to about 18 million when wetland rice cultivation began, to over 800 million at the start of the industrial revolution, to over 7,700 million in 2019. A big part of the incredible growth in our population is related to the availability of the cheap and abundant energy embodied in fossil fuels, which we use for transportation, heating and cooling, industry, and food production.

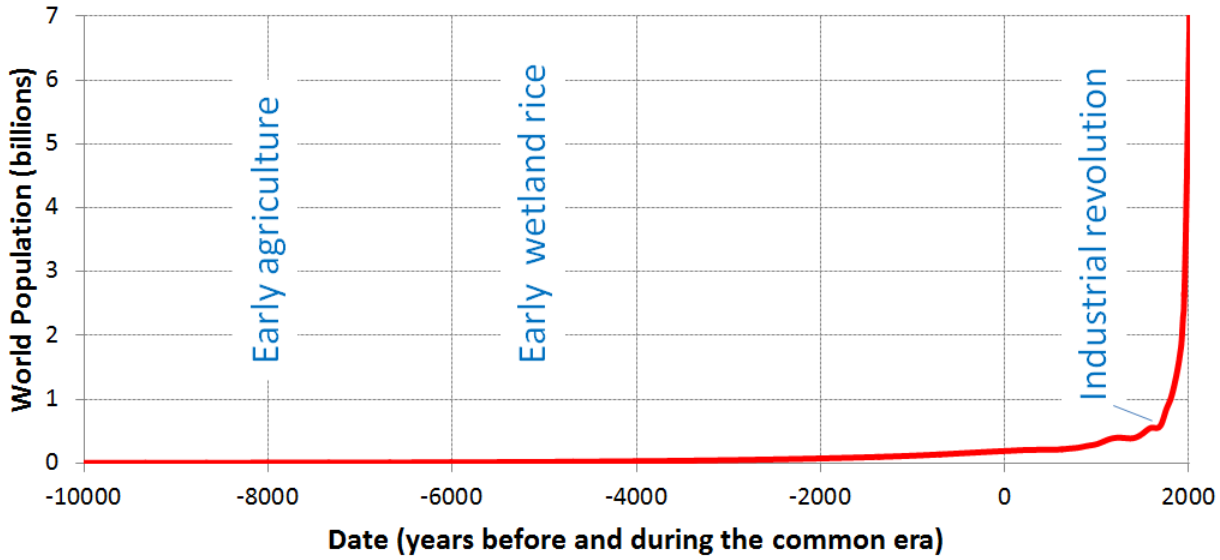



Figure **9.3.1**: World population growth over the past 12,000 years.
© Steven Earle. CC BY. Based on World Population Annual 12000 years data.

A rapidly rising population, the escalating level of industrialization and mechanization of our lives, and an increasing dependence on fossil fuels for transportation and energy generation have driven the anthropogenic climate change of the past century. The trend of mean global temperatures since 1880 is shown in Figure **9.3.2**. For approximately the past 55 years, the temperature has increased at a relatively steady and rapid rate, especially compared to past changes. The average temperature now is approximately 1.0°C higher than before industrialization, and two-thirds of this warming has occurred since 1975.

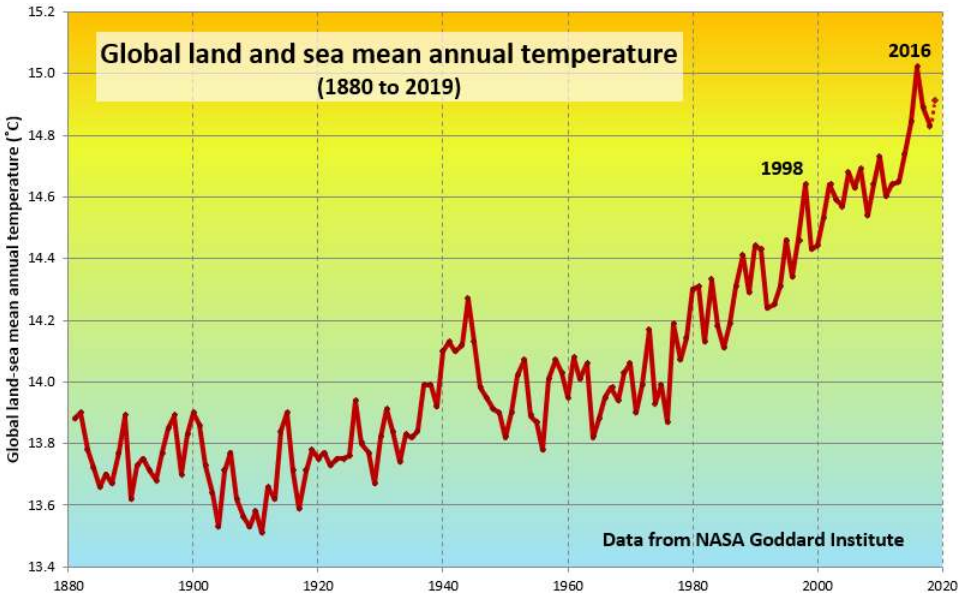



Figure **9.3.2**: Global mean annual temperatures for the period from 1880 to 2019. The value for 2019 is projected, based on only 6 months of data. 1998 and 2016 were strong El Niño years.
© Steven Earle. CC BY. Based on data from NASA.

Changes in land surface or ocean surface temperatures can be expressed as temperature **anomalies**. A temperature anomaly is the difference in average temperature measurement from a predetermined datum (or baseline). This datum is the average temperature of a particular date range, for example, 1951 to 1980. Another common datum is the last century (1900-2000). Therefore, an anomaly of 1.25 ℃ for 2015 (last century datum) means that the average temperature for 2015 was 1.25 ℃ greater than the 1900-2000 average. In 1950, the temperature anomaly was -0.28 ℃, so this is -0.28 ℃ lower than the 1900-2000 average [3]. These temperatures are annual average surface temperatures. Figure **9.3.3** shows the global mean temperature estimates for the period of 1880 to 2020 using the datum of the 1951-1980 average temperature.

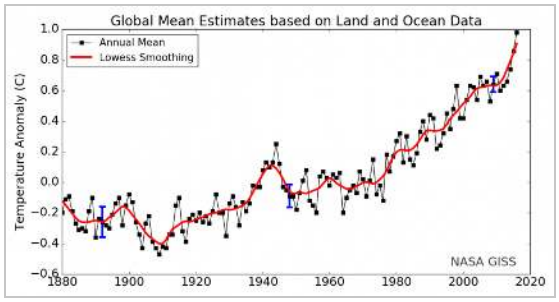



Figure **9.3.3**: Land-ocean temperature index, 1880 to present, with a base time 1951-1980. The solid black line is the global annual mean and the solid red line is the five-year lowess smooth. The blue uncertainty bars (95% confidence limit) account only for incomplete spatial sampling.
[From https://opengeology.org/textbook/15-...limate-change/ Adapted from *Source: National Climate Data Center*]

Since 1880, average global surface temperatures have trended upward and most of that warming has occurred since 1970 (see this NASA animation). Since the ocean is absorbing a lot of the additional trapped heat, surface temperatures include both land surface and ocean temperatures [13]. This video figure also shows worldwide temperature changes since 1880. The more blue, the cooler; the more yellow and red, the warmer.

## Scientific Consensus: Global Climate Change is Real

The **Intergovernmental Panel on Climate Change** (IPCC)—established by the United Nations Environment Programme and the World Meteorological Organization 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. The IPCC uses this information to evaluate current impacts and future risks, in addition to providing policymakers with assessments. These assessments are released

about once every every six years. The most recent report, the 5th Assessment, was released in 2013. Hundreds of leading scientists from around the world are chosen to author these reports. Over the history of the IPCC, these scientists have reviewed thousands of peer-reviewed, publicly available studies. The scientific consensus according to the last IPCC assessment is clear: global climate change is real and humans are very likely the cause for this change.

Additionally, the major scientific agencies of the United States, including the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA), also agree that climate change is occurring and that humans are driving it. In 2010, the US National Research Council concluded that “Climate change is occurring, is very likely caused by human activities, and poses significant risks for a broad range of human and natural systems”. Many independent scientific organizations have released similar statements, both in the United States and abroad. This doesn’t necessarily mean that every scientist sees eye to eye on each component of the climate change problem, but broad agreement exists that climate change is happening and is primarily caused by excess greenhouse gases from human activities.

The historical record shows that the climation (NASA) and the National Oceanic and Atmospheric Administration (NOAA), also agree that climate change is occurring and that humans are driving it. In 2010, tate varies naturally over a wide range of time scales. In general, climate changes prior to the Industrial Revolution in the 1700s can be explained by natural causes, such as changes in solar energy, volcanic eruptions, and natural changes in greenhouse gas (GHG) concentrations. Recent changes in climate**,** however, cannot be explained by natural causes alone. Research indicates that natural causes are very unlikely to explain most observed warming, especially warming since the mid-20th century. Rather, human activities, especially our combustion of fossil fuels, explains the current warming (figure **9.3.4**). The scientific consensus is clear: through alterations of the carbon cycle, humans are changing the global climate by increasing the effects of something known as the greenhouse effect.

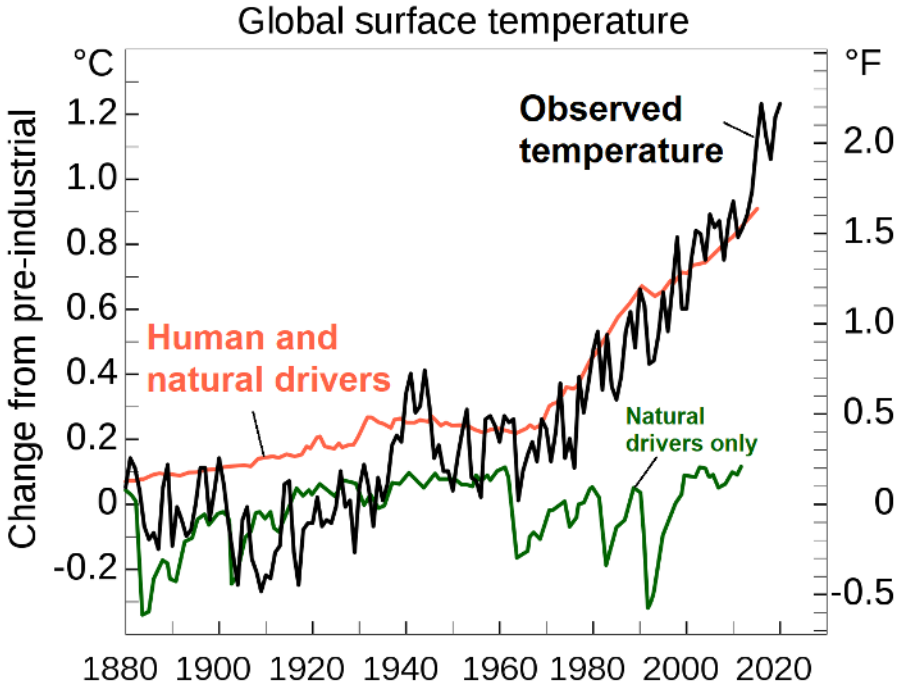



**Figure 9.3.4:** This graph shows the predicted temperatures from two climate models and observed temperatures from 1880 to 2020. The first model considered only natural factors that could influence temperature and is represented by the green line (bottom). It shows some fluctuations in temperature by no overall increase or decrease. The second model considered both human and natural factors and is represented by the orange line (top). It shows an overall increase in temperature. Actual observations (black, jagged line; middle) more closely match the second model. Overall, temperature has increased about 1.2 degrees Celsius (2.1 degrees Fahrenheit) since pre-industrial times. Image relabeled from Efbrazil (CC-BY-SA).

## Human Impact on Greenhouse Gases

Recall that greenhouse gases are probably the most significant drivers of the climate. The greenhouse gases that affect Earth include carbon dioxide, methane, water vapor, nitrous oxide, and ozone. Approximately half of the radiation from the sun passes through these gases in the atmosphere and strikes the Earth. This radiation is converted into thermal radiation on the Earth's surface, and then a portion of that energy is re-radiated back into the atmosphere as thermal energy. Greenhouse gases absorb much of the thermal energy near the Earth's surface so the more greenhouse gases there are in the atmosphere, the more thermal energy is retained by the Earth.

Water vapor is the most abundant greenhouse gas and also the most important in terms of its contribution to the natural greenhouse effect, despite having a short atmospheric lifetime. Some human activities can influence local water vapor levels. However, on a global scale, the concentration of water vapor is controlled by temperature, which influences overall rates of evaporation and precipitation. Therefore, the global concentration of water vapor is not substantially affected by direct human emissions.

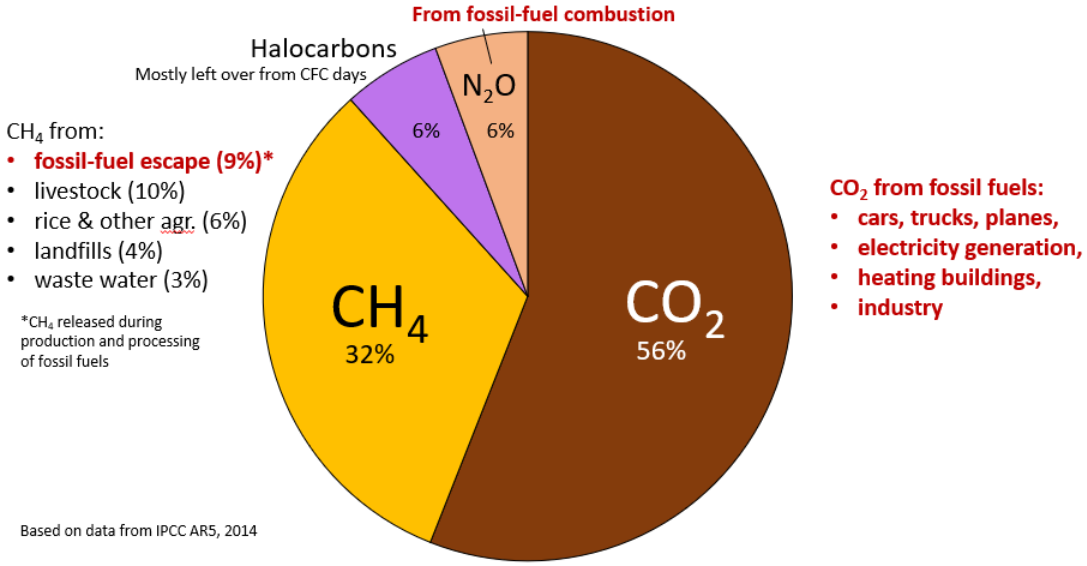



Figure **9.3.5**: The relative climate-warming effects of the various greenhouse gases. (Effects listed in red are all related to fossil fuels.)
© Steven Earle. CC BY. Based on data in IPCC AR 5, 2014.

Figure **9.3.5**, which is based on data from the fifth assessment report of the IPCC, issued in 2014, shows the relative contributions of various long-lasting anthropogenic GHGs to current climate forcing, based on the changes from levels that existed in 1750. Human activity releases carbon dioxide and methane, two of the most important greenhouse gases, into the atmosphere in several ways. The biggest anthropogenic contributor to warming is $CO_2$, which accounts for 56% of positive forcing. The primary mechanism that releases carbon dioxide is the burning of fossil fuels, such as gasoline, coal, and natural gas (Figure **9.3.6**). $CH_4$ accounts for 32%, and the halocarbon gases (mostly leaked from older air-conditioning appliances that still contain CFCs) and nitrous oxide ($N_2O$) (from burning fossils fuels) account for 6% each. $CO_2$ emissions come mostly from coal- and gas-fired power stations, motorized vehicles (cars, trucks, and aircraft), and industrial operations (e.g., smelting). $CH_4$ emissions come from production of fossil fuels (escape from coal mining and from gas and oil production and processing), livestock farming (mostly beef), landfills, waste water, and wetland rice farming. $N_2O$ is derived almost entirely from the combustion of fossil fuels. In summary, most (by far) of our current GHG emissions come from fossil fuel production and use.



Figure **9.3.6**: The burning of fossil fuels in industry and by vehicles releases carbon dioxide and other greenhouse gases into the atmosphere. (credit: “Pöllö”/Wikimedia Commons)

## Carbon Dioxide

**Carbon dioxide** ($CO_2$) is the primary greenhouse gas that is contributing to recent global climate change. $CO_2$ is a natural component of the carbon cycle, involved in such activities as photosynthesis, respiration, volcanic eruptions, and ocean-atmosphere exchange. Human activities, primarily the burning of fossil fuels and changes in land use, release very large amounts of $CO_2$ to the atmosphere, causing its concentration in the atmosphere to rise. Deforestation, cement manufacture, animal agriculture, the clearing of land, and the burning of forests are other human activities that release carbon dioxide.

Scientists look at patterns in data and try to explain differences or deviations from these patterns. The atmospheric carbon dioxide data reveal a historical pattern of carbon dioxide increasing and decreasing, cycling between a low of 180 ppm and a high of 300 ppm (Figure **9.3.4**). Scientists have concluded that it took around 50,000 years for the atmospheric carbon dioxide level to increase from its low minimum concentration to its higher maximum concentration.

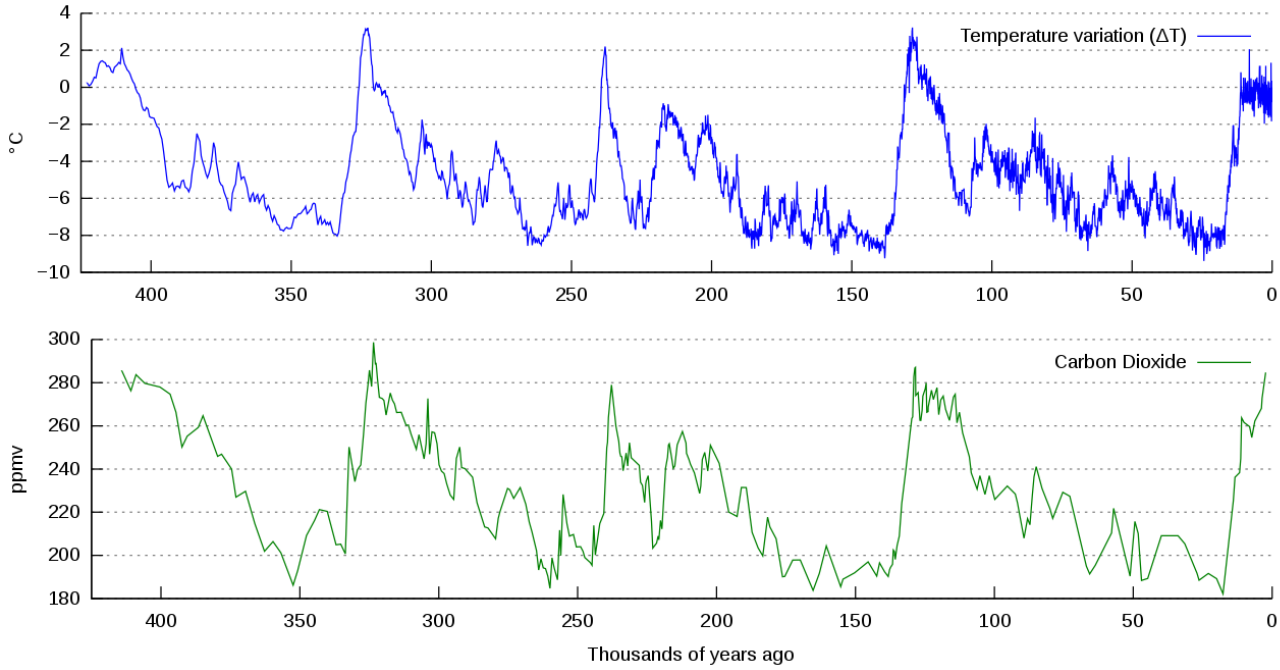



**Figure 9.3.7:** Graph of change in temperature in degrees Celsius (top, blue line) and carbon dioxide concentration in parts per million by volume (bottom, green line) measured from the Vostok, Antarctica ice core. These have been associated for over 400,000 years. As carbon dioxide concentration increased, so did temperature. As carbon dioxide concentration decreased, so did temperature. These data were collected in 1999. Since then, carbon dioxide concentrations have increased to 409.8 ppm (2019 average). Image and caption (modified) by NOAA/Autopilot (CC-BY-SA)

However, starting recently, atmospheric carbon dioxide concentrations have increased beyond the historical maximum of 300 ppm (figure **9.3.8 – 9**). Atmospheric $CO_2$ concentrations have increased by 45% since pre-industrial times, from approximately 280 parts per million (ppm) in the 18th century to 409.8 ppm in 2019 (figure **9.3.8 – 9**).

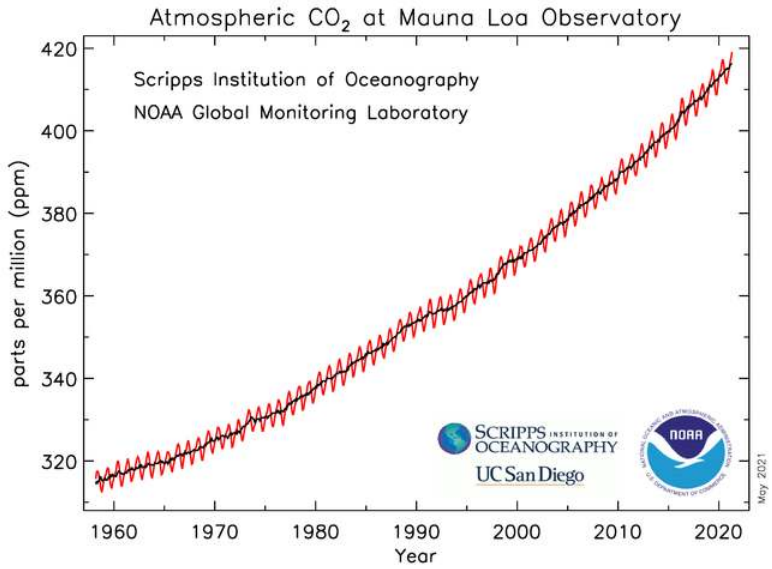



**Figure 9.3.8:** Monthly mean carbon dioxide concentration in parts per million (ppm) measured at Mauna Loa Observatory, Hawaii. The carbon dioxide data on Mauna Loa constitute the longest record of direct measurements of $CO_2$ in the atmosphere. The fluctuating red line represents the monthly mean values, centered on the middle of each month. Carbon dioxide concentrations dip

every summer due to increased photosynthesis. The smoother black line represents the same, after correction for the average seasonal cycle. Image and caption (modified) by NOAA (public domain).

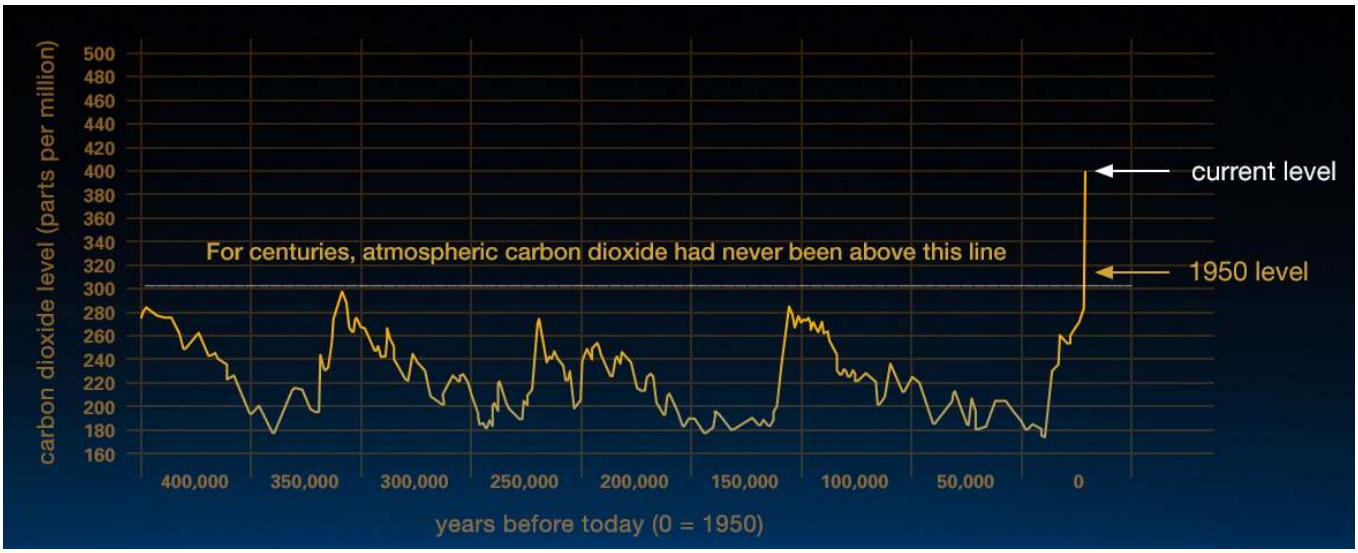
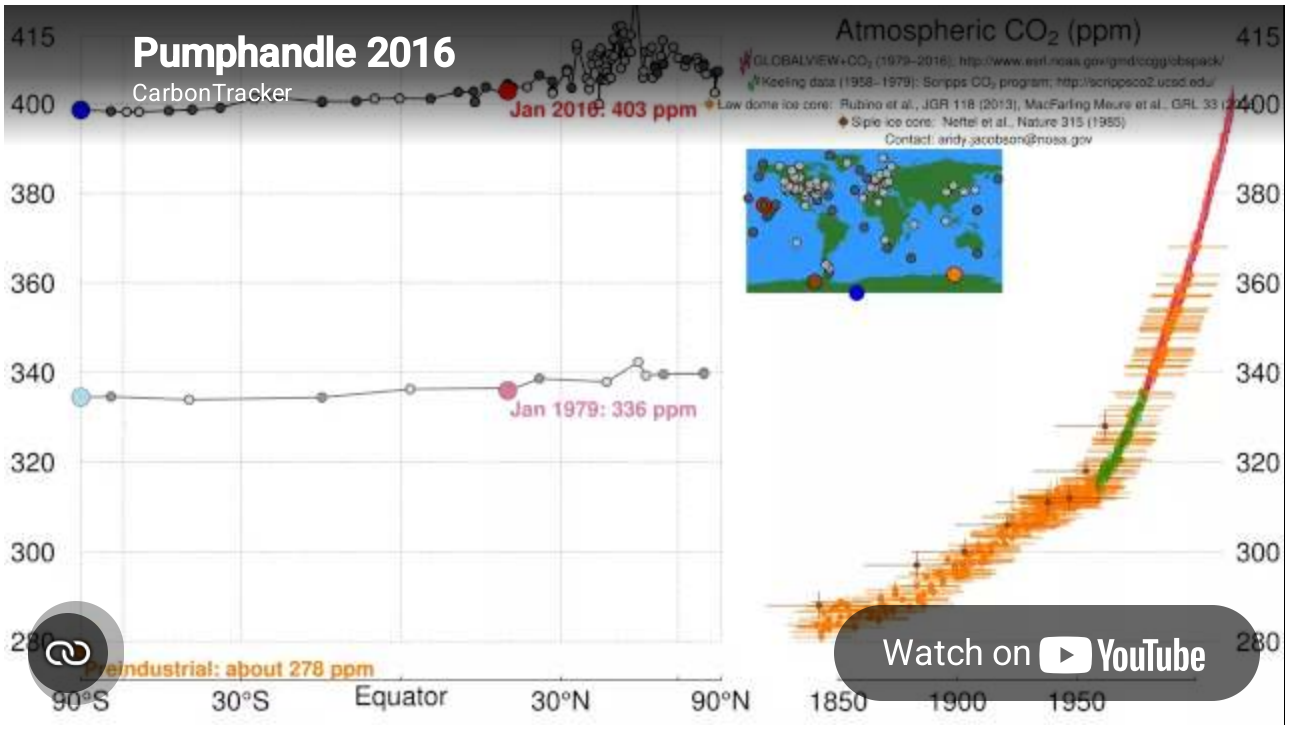
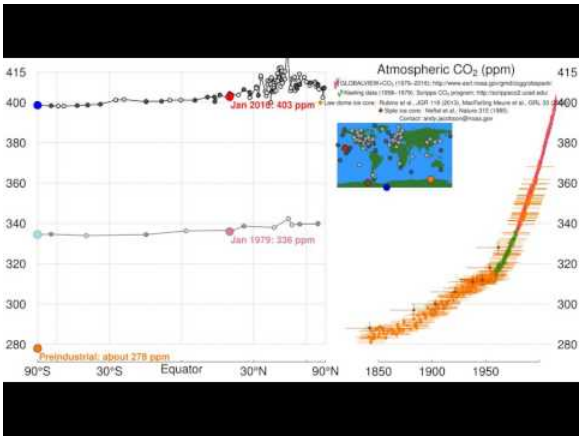



**Figure 9.3.9:** This graph, based on the comparison of atmospheric samples contained in ice cores and more recent direct measurements, provides evidence that atmospheric CO2 has increased since the Industrial Revolution. On the x-axis are the years before today (0 = 1950). It begins 400,000 years prior to 1950. On the y-axis is the carbon dioxide level in parts per million. Carbon dioxide levels have fluctuated over the years, but they never exceeded 300 parts per million until 1950. In 2018, carbon dioxide levels reached 409.8 ppm. (This graph is a few years old and shows the current $CO_2$ level at 400 ppm; Credit: Vostok ice core data/J.R. Petit et al.; NOAA Mauna Loa CO2 record.)

The current $CO_2$ level is higher than it has been in at least 800,000 years, based on evidence from ice cores that preserve ancient atmospheric gases. The current increases in atmospheric carbon dioxide have happened very quickly—in a matter of hundreds of years rather than thousands of years. What is the reason for this difference in the rate of change and the amount of increase in carbon dioxide? A key factor that must be recognized when comparing the historical data and the current data is the presence of modern human society; no other driver of climate change has yielded changes in atmospheric carbon dioxide levels at this rate or to this magnitude. Human activities currently release over 30 billion tons of $CO_2$ into the atmosphere every year. While some volcanic eruptions released large quantities of $CO_2$ in the distant past, the U.S. Geological Survey (USGS) reports that human activities now emit more than 135 times as much $CO_2$ as volcanoes each year. This human-caused build-up of $CO_2$ in the atmosphere is like a tub filling with water, where more water flows from the faucet than the drain can take away.

The following video shows how atmospheric $CO_2$ has varied recently and also over the last 800,000 years as determined by many $CO_2$ monitoring stations (shown on the insert map). It is also instructive to watch the $CO_2$ variation of the Keeling portion of the video by latitude. This shows that most of the human sources of $CO_2$ are in the Northern Hemisphere.



## Methane

Although this concentration is far less than that of $CO_2$, **methane** ($CH_4$) is 28 times as potent a greenhouse gas. Methane is produced when bacteria break down organic matter under anaerobic conditions and can be released due to natural or anthropogenic processes. Anaerobic conditions can happen when organic matter is trapped underwater (such as in rice paddies) or in the intestines of herbivores. Anthropogenic causes now account for 60% of total methane release. Examples include agriculture, fossil fuel extraction and transport, mining, landfill use, and burning of forests. Methane can also be released from natural gas fields and the decomposition that occurs in landfills. Specifically, raising cattle releases methane due to fermentation in their rumens produces methane that is expelled from their GI tract. Methane is more abundant in Earth's atmosphere now than at any time in at least the past 650,000 years, and $CH_4$ concentrations increased sharply during most of the 20th century. They are now more than two and-a-half times pre-industrial levels (1.9 ppm), but the rate of increase has slowed considerably in recent decades.

Another source of methane is the melting of clathrates. Clathrates are frozen chunks of ice and methane found at the bottom of the ocean. When water warms, these chunks of ice melt and methane is released. As the ocean's water temperature increases, the rate at which clathrates melt is increasing, releasing even more methane. This leads to increased levels of methane in the atmosphere, which further accelerates the rate of global warming. This is an example of the positive feedback loop that is leading to the rapid rate of increase of global temperatures.

### Other Greenhouse Gases

Nitrous oxide ($N_2O$) is produced through natural and human activities, mainly through agricultural activities and natural biological processes. Fuel burning and some other processes also create $N_2O$. Concentrations of $N_2O$ have risen approximately 18% since the start of the Industrial Revolution, with a relatively rapid increase towards the end of the 20th century. [3] In contrast, the atmospheric concentration of $N_2O$ varied only slightly for a period of 11,500 years before the onset of the industrial period, as shown by ice core samples.

Ground-level ozone ($O_3$), which also has a short atmospheric lifetime, is a potent greenhouse gas. Chemical reactions create ozone from emissions of nitrogen oxides and volatile organic compounds from automobiles, power plants, and other industrial and commercial sources in the presence of sunlight. In addition to trapping heat, ozone is a pollutant that can cause respiratory health problems and damage crops and ecosystems.

Chlorofluorocarbons (CFCs), hydrochlorofluorocarbons (HCFCs), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), and sulfur hexafluoride (SF6), together called F-gases, are often used in coolants, foaming agents, fire extinguishers, solvents, pesticides, and aerosol propellants. Unlike water vapor and ozone, these F-gases have a long atmospheric lifetime, and some of these emissions will affect the climate for many decades or centuries.

### Human Induced Changes in Albedo

When sunlight energy reaches Earth it can be reflected or absorbed. The amount that is reflected or absorbed depends on Earth's surface and atmosphere. Light-colored objects and surfaces, like snow and clouds, tend to reflect most sunlight, while darker objects and surfaces, like the ocean and forests, tend to absorb more sunlight. The term **albedo** refers to the amount of solar radiation reflected from an object or surface, often expressed as a percentage. Earth as a whole has an albedo of about 30%, meaning that 70% of the sunlight that reaches the planet is absorbed. Sunlight that is absorbed warms Earth's land, water, and atmosphere.

Albedo is also affected by aerosols. **Aerosols** are small particles or liquid droplets in the atmosphere that can absorb or reflect sunlight. Unlike greenhouse gases (GHGs), the climate effects of aerosols vary depending on what they are made of and where they are emitted. Those aerosols that reflect sunlight, such as particles from volcanic eruptions or sulfur emissions from burning coal, have a cooling effect. Those that absorb sunlight, such as black carbon (a part of soot), have a warming effect. In addition, human activities have generally increased the number of aerosol particles in the atmosphere. Overall, human-generated aerosols have a net cooling effect offsetting about one-third of the total warming effect associated with human greenhouse gas emissions. Reductions in overall aerosol emissions can therefore lead to more warming. However, targeted reductions in black carbon emissions can reduce warming.

**Human changes** in land use and land cover have changed Earth's albedo. Processes such as deforestation, reforestation, desertification, and urbanization often contribute to changes in climate in the places they occur. These effects may be significant regionally, but are smaller when averaged over the entire globe.

## Climate Feedbacks

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%. This is a positive feedback because 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.) 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.



Figure **9.3.10**: A degrading permafrost site on the north coast of Alaska.

Most of northern Canada has a layer of permafrost that ranges from a few centimeters to hundreds of meters in thickness; the same applies in Alaska, Russia, and Scandinavia. Permafrost is a mixture of soil and ice (Figure **9.3.10**), 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 sea floor, 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. But the solubility of $CO_2$ in water decreases as the temperature goes up. In other words, the warmer it gets, the more of that oceanic bicarbonate 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.

## Reference

3. Lindsey, R. Climate and Earth's Energy Budget: Feature Articles. (2009). Available at: http://earthobservatory.nasa.gov. (Accessed: 14th September 2016)

6. Pachauri, R. K. *et al*. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. (IPCC, 2014).

7. Oreskes, N. The scientific consensus on climate change. *Science* **306**, 1686–1686 (2004).

11. Earle, S. *Physical geology OER textbook*. (BC Campus OpenEd, 2015).

12. Callendar, G. S. The artificial production of carbon dioxide and its influence on temperature. *Q.J.R. Meteorol. Soc.* **64**, 223–240 (1938).

13. Hansen, J., Sato, M., Kharecha, P. & others. Earth's energy imbalance and implications. *Atmospheric* (2011).

14. Foster, G. & Rahmstorf, S. Global temperature evolution 1979–2010. *Environ. Res. Lett.* **6**, 044022 (2011).

15. Kosaka, Y. & Xie, S.-P. Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature* **501**, 403–407 (2013).

16. Easterling, D. R. & Wehner, M. F. Is the climate warming or cooling? *Geophys. Res. Lett.* **36**, (2009).

17. Zemp, M. *et al*. Historically unprecedented global glacier decline in the early 21st century. *J. Glaciol.* **61**, 745–762 (2015).

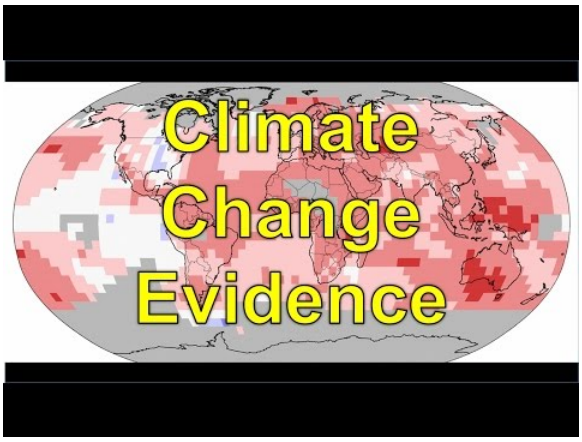
18. Karl, T. R. & Knight, R. W. Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Am. Meteorol. Soc.* **79**, 231–241 (1998).

19. Santer, B. D. *et al*. Identification of human-induced changes in atmospheric moisture content. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15248–15253 (2007).

## Further "Reading"

For more information, watch this six-minute video on climate change by two professors at a North Carolina State University.






## Contributors and Attributions

- 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)
- ”Physical Geology” by Steven Earle used under a CC-BY 4.0 international license. Download this book for free at http://open.bccampus.ca
- Contributed by Melissa Ha and Rachel Schleiger Faculty (Biological Sciences) at Yuba College & Butte College
- Contributed by Paul Webb Professor (Biology) at Rodger Williams University
- Contributed by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, & Cam Mosher Faculty (Geology) at Salt Lake Community College
- Sourced from OpenGeology from *An Introduction to Geology* Chapter Chapter 15: Global Climate Change
- 12.4 Climate Change from *Essentials of Environmental Science* by CK-12 and Kamala Doršner (licensed under CC BY 4.0.)
- 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)

---

This page titled 9.3: Anthropogenic Climate Change is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Steven Earle (BCCampus) .

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

# 9.4: Implications of Climate Change

## Past and Present-day GHG Emissions Will Affect Climate Far into the Future

Many greenhouse gases stay in the atmosphere for long periods of time. As a result, even if emissions stopped increasing, atmospheric greenhouse gas concentrations would continue to remain elevated for hundreds of years. Moreover, if we stabilized concentrations and the composition of today's atmosphere remained steady (which would require a dramatic reduction in current greenhouse gas emissions), surface air temperatures would continue to warm. This is because the oceans, which store heat, take many decades to fully respond to higher greenhouse gas concentrations. The ocean's response to higher greenhouse gas concentrations and higher temperatures will continue to impact climate over the next several decades to hundreds of years.

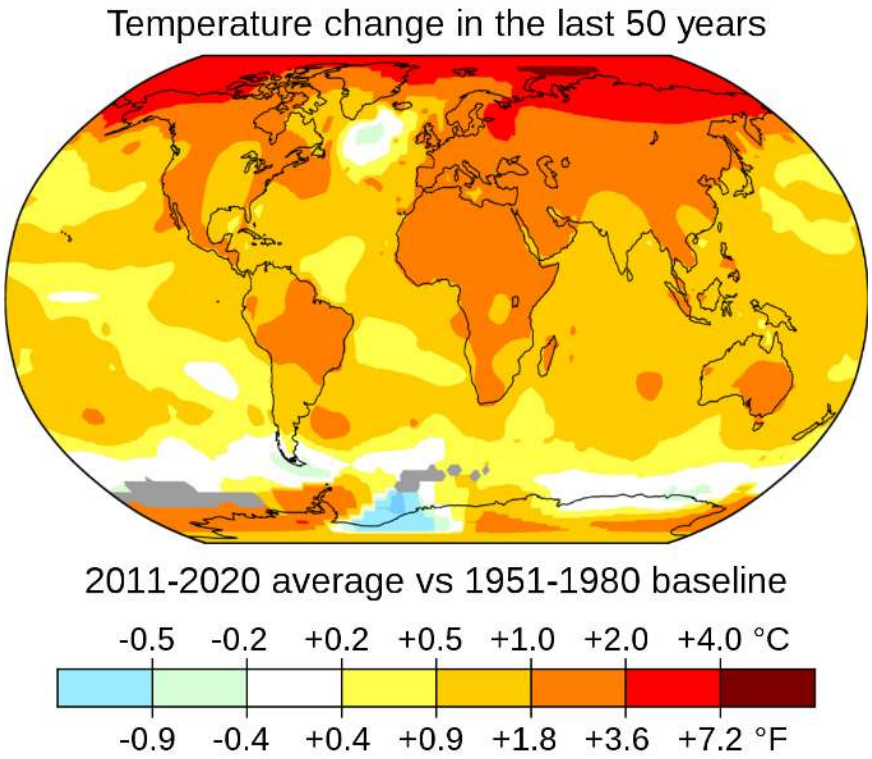



Figure **9.4.1**: Temperature increases have been most pronounced in northern latitudes and over land masses. The colors represent the temperature difference between the 2011-2020 average and the 1951-1980 baseline, with warmer colors (yellow, orange, red) representing increases, and cool colors (green, blue) representing decreases. The image uses longer term averages of at least a decade to smooth out climate variability due to factors such as El Niño. Grey areas in the image have insufficient data for rendering. Image and caption (modified) from NASA's Scientific Visualization Studio/Eric Fisk (public domain).

### Future Temperature Changes

Climate models project the following key temperature-related changes:

- Average global temperatures are expected to increase by 2°F to 11.5°F by 2100, depending on the level of future greenhouse gas emissions, and the outcomes from various climate models.
- By 2100, global average temperature is expected to warm at least twice as much as it has during the last 100 years.
- Ground-level air temperatures are expected to continue to warm more rapidly over land than oceans.
- Some parts of the world are projected to see larger temperature increases than the global average.

These changes will impact our food supply, water resources, infrastructure, ecosystems, and even our own health. The magnitude and rate of future climate change will primarily depend on the following factors:

- The rate at which levels of greenhouse gas concentrations in our atmosphere continue to increase,

- How strongly features of the climate (e.g., temperature, precipitation, and sea level) respond to the expected increase in greenhouse gas concentrations,
- Natural influences on climate (e.g., from volcanic activity and changes in the sun's intensity) and natural processes within the climate system (e.g., changes in ocean circulation patterns).

## Future Precipitation and Storm Events

Patterns of precipitation and storm events, including both rain and snowfall are likely to change. However, some of these changes are less certain than the changes associated with temperature. Because warm air is able to hold more water than cold air, the general global trend over the past century has been one of increasing precipitation (Figure **9.4.2**).

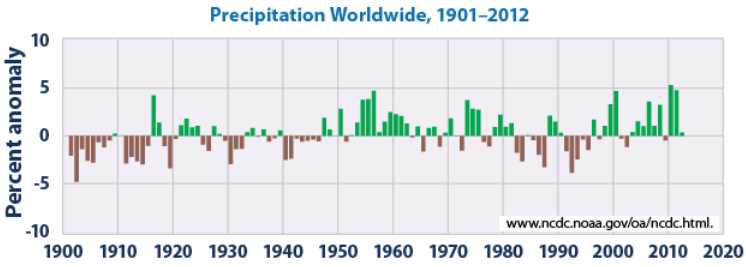



**Figure 9.4.2:** *Global precipitation anomalies compared with the average over the period from 1901 to 2012.*
["Precipitation Worldwide, 1901-2013" by NASA. Public domain.]

Projections show that future precipitation and storm changes will vary by season and region. Some regions may have less precipitation, some may have more precipitation, and some may have little or no change. The amount of rain falling in heavy precipitation events is likely to increase in most regions, while storm tracks are projected to shift towards the poles. Climate models project the following precipitation and storm changes:

- Global average annual precipitation through the end of the century is expected to increase, although changes in the amount and intensity of precipitation will vary by region.
- The intensity of precipitation events will likely increase on average. This will be particularly pronounced in tropical and high-latitude regions, which are also expected to experience overall increases in precipitation.
- The strength of the winds associated with tropical storms is likely to increase. The amount of precipitation falling in tropical storms is also likely to increase.
- Annual average precipitation is projected to increase in some areas and decrease in others.

Occurrence and intensity of extreme weather events such as hurricanes, precipitation, and heatwaves are increasing [3; 6]. Since the 1980s, hurricanes, which are generated from warm ocean water, have increased in frequency, intensity, and duration and connections to a warmer climate are likely. Since 1910, average precipitation has increased by 10% in the contiguous United States, and much of this increase is associated with heavy precipitation events like storms [18]. However, the distribution is not even and more precipitation is projected for the northern United States while less precipitation is projected for the already dry southwest [3]. Further, heatwaves have increased and rising temperatures are already affecting crop yields in northern latitudes [6]. Increased heat allows for greater moisture capacity in the atmosphere, increasing the potential for more extreme events [19].

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), 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 (Figure **9.4.3**).



Figure **9.4.3**: *Damage to the Casino Pier, Seaside Heights, New Jersey, from Hurricane Sandy, November 2012*
["Hurricane Sandy New Jersey Pier" © Master Sgt. Mark C. Olsen/U.S. Air Force/New Jersey National Guard. CC BY.]

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 **9.4.4** 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.

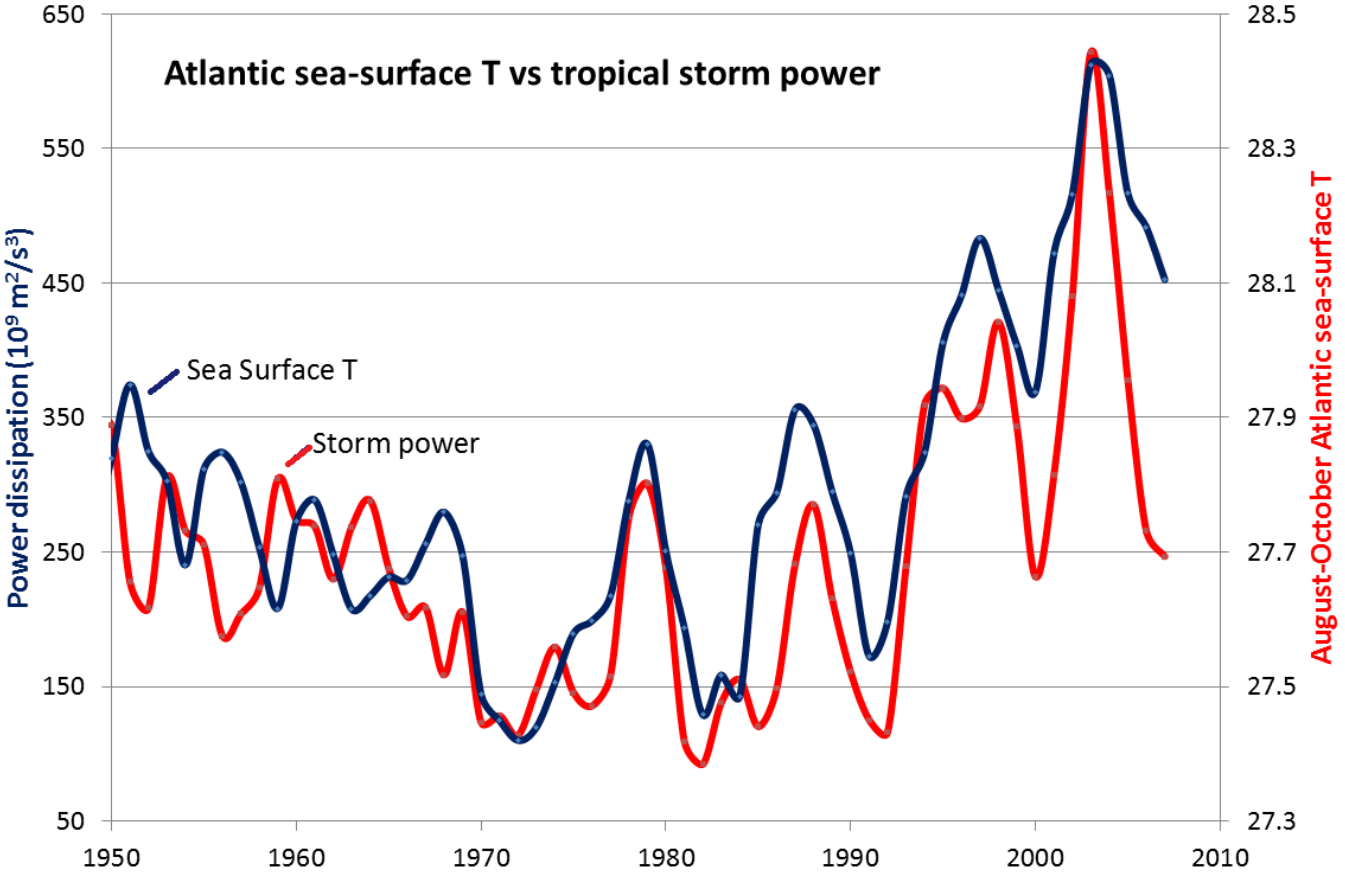



Figure **9.4.4**: Relationship between Atlantic tropical storm cumulative annual intensity and Atlantic sea-surface temperatures
["Atlantic sea-surface T vs. tropical storm power" © Steven Earle. CC BY.
Based on data from Papers, Data, and Graphics Pertaining to Tropical Cyclone Trends and Variability.]

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 **9.4.5**. 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.

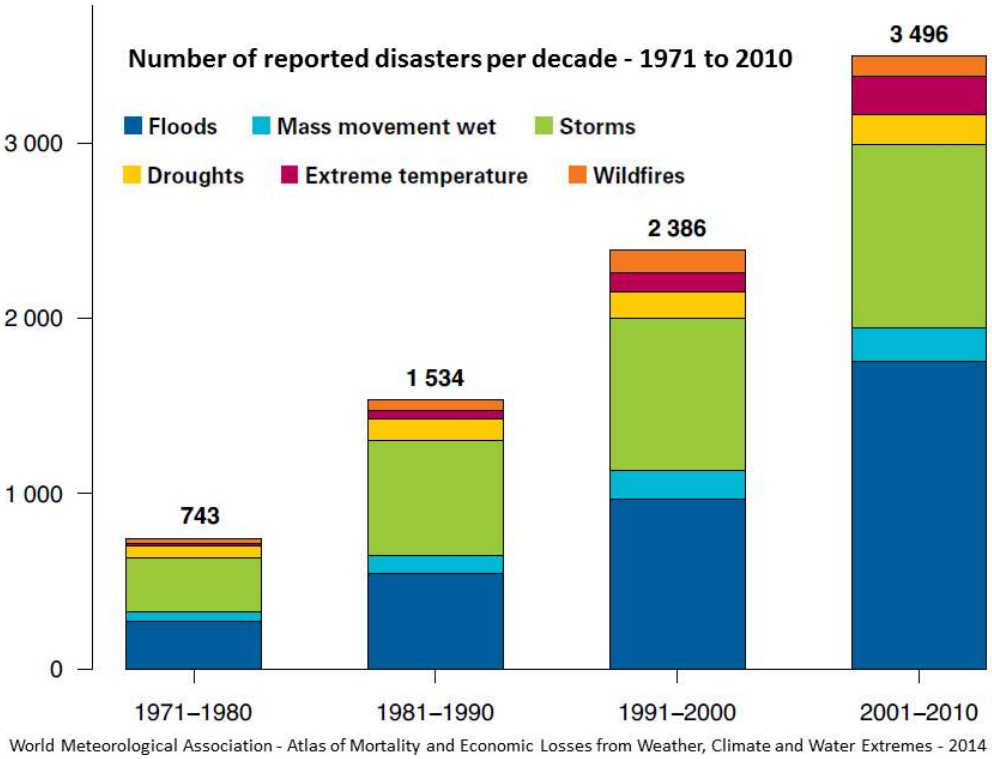



**Figure 9.4.5:** Numbers of various types of disasters between 1971 and 2010 (WMO Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes, 2014. CC BY-NC-ND).

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 **9.4.1**, or the sea level rise shown in Figure **9.4.8**.

## Future Ice, Snowpack, and Permafrost

Arctic sea ice is already declining. The area of snow cover in the Northern Hemisphere has decreased since 1970. Permafrost temperature has increased over the last century, making it more susceptible to thawing. Over the next century, it is expected that sea ice will continue to decline, glaciers will continue to shrink, snow cover will continue to decrease, and permafrost will continue to thaw.

For every 2°F of warming, models project about a 15% decrease in the extent of annually averaged sea ice and a 25% decrease in September Arctic sea ice. The coastal sections of the Greenland and Antarctic ice sheets are expected to continue to melt or slide into the ocean. If the rate of this ice melting increases in the 21st century, the ice sheets could add significantly to global sea level rise. Glaciers are expected to continue to decrease in size. The rate of melting is expected to continue to increase, which will contribute to sea level rise.

Glaciers are ice on top of the land. Alpine glaciers, ice sheets, and sea ice are all melting. Almost all major alpine glaciers are shrinking, deflating, and retreating and the rate of ice mass loss is unprecedented (never observed before) since the 1940's when quality records for most began.Before anthropogenic warming, glacial activity was variable with some retreating and some

advancing [17]. The extent of spring snow cover has decreased. Satellites have recorded that Antarctica is melting at 118 gigatons per year and Greenland is melting at 281 gigatons per year (1 gigaton is over 2 trillion pounds) (Figure **9.4.6**).

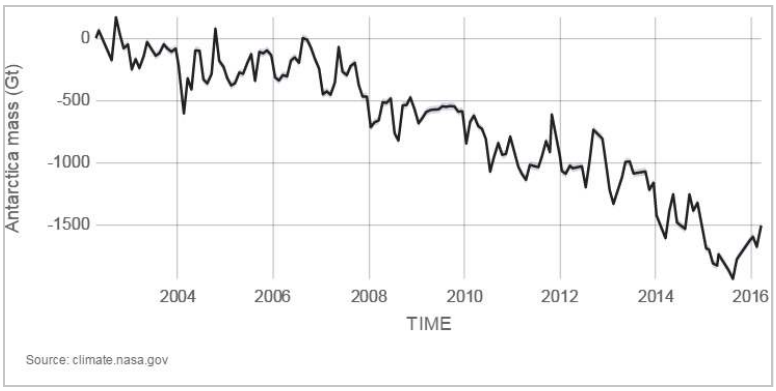
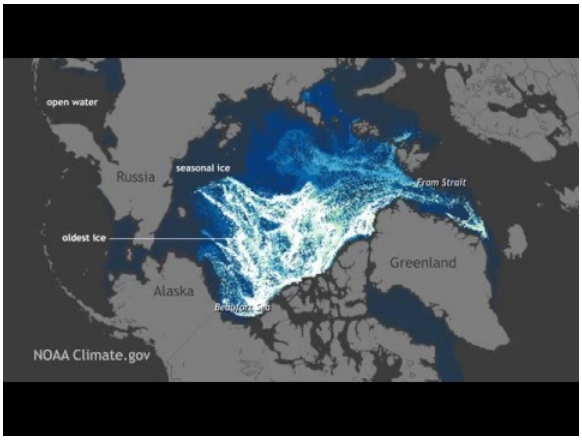



Figure **9.4.6**: Decline of Antarctic ice mass from 2002 to 2016

In addition, the extent of sea ice is shrinking. Sea ice is ice floating in the ocean (not on land like a glacier). Most sea ice is at the North Pole which is only occupied by the Arctic Ocean and sea ice [3; 6]. Below, the NOAA animation shows how perennial sea ice has declined from 1987 to 2015. The oldest ice is white and the youngest (seasonal) ice is dark blue. The amount of old ice has declined from 20% in 1985 to 3% in 2015.





This loss of ice is leading to increases in the global sea level. On average, the sea is rising at a rate of 1.8 mm per year. However, between 1993 and 2010 the rate of sea level increase ranged between 2.9 and 3.4 mm per year. A variety of factors affect the volume of water in the ocean, including the temperature of the water (the density of water is related to its temperature) and the amount of water found in rivers, lakes, glaciers, polar ice caps, and sea ice. As glaciers and polar ice caps melt, there is a significant contribution of liquid water that was previously frozen.

A number of global events have occurred that may be attributed to climate change during our lifetimes. Explore melting glaciers at NASA's interactive Global Ice Viewer. Glacier National Park in Montana is undergoing the retreat of many of its glaciers, a phenomenon known as glacier recession. In 1850, the area contained approximately 150 glaciers. By 2010, however, the park contained only about 24 glaciers greater than 25 acres in size. One of these glaciers is the Grinnell Glacier (Figure **9.4.7**) at Mount Gould. Between 1966 and 2005, the size of Grinnell Glacier shrank by 40 percent. Similarly, the mass of the ice sheets in Greenland and the Antarctic is decreasing: Greenland lost 150–250 $km^3$ of ice per year between 2002 and 2006. In addition, the size and thickness of the Arctic sea ice is decreasing.




Figure **9.4.7**: The effect of global warming can be seen in the continuing retreat of Grinnel Glacier. The mean annual temperature in the park has increased 1.33 °C since 1900. The loss of a glacier results in the loss of summer meltwaters, sharply reducing seasonal water supplies and severely affecting local ecosystems. (credit: modification of work by USGS)

## Future Sea Level Change

Warming temperatures contribute to sea level rise by expanding ocean water, melting mountain glaciers and ice caps, and causing portions of the Greenland and Antarctic ice sheets to melt or flow into the ocean. Since 1870, global sea level has risen by about 8 inches. Estimates of future sea level rise vary for different regions, but global sea level for the next century is expected to rise at a greater rate than during the past 50 years. The contribution of thermal expansion, ice caps, and small glaciers to sea level rise is relatively well-studied, but the impacts of climate change on ice sheets are less understood and represent an active area of research. Thus, it is more difficult to predict how much changes in ice sheets will contribute to sea level rise. Greenland and Antarctic ice sheets could contribute an additional 1 foot of sea level rise, depending on how the ice sheets respond.

Sea-level is rising 3.4 millimeters (0.13 inches) per year and has risen 0.19 meters (7.4 inches) from 1901 to 2010 (Figure **9.4.8**). This is thought largely to be from both the melting of glaciers and thermal expansion. Thermal expansion means that as objects such as solids, liquids, and gases heat up, they expand in volume. Since 1970, the melting of glaciers and thermal expansion account for 75% of the sea-level rise [6].

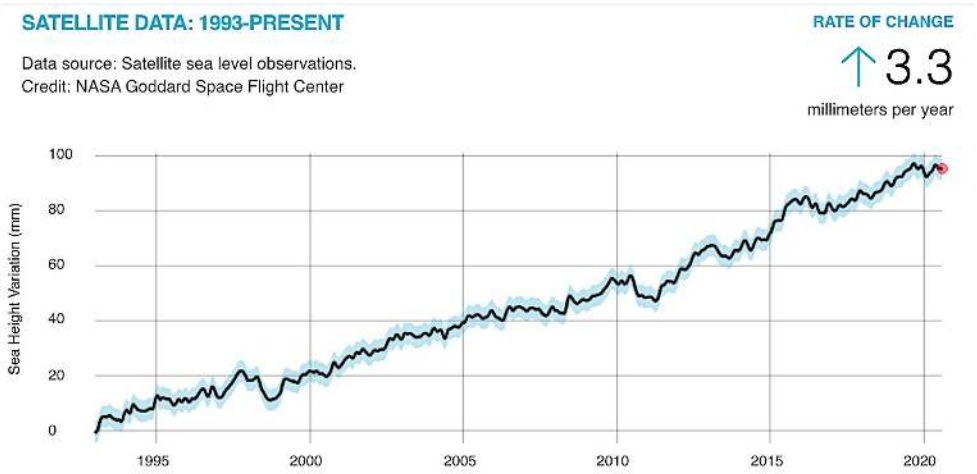



**Figure 9.4.8:** Sea height variation (mm) over time. Sea height has increased about 3.3 millimeters per year on average since 1993. Data is from satellite sea level observations by the NASA Goddard Space Flight Center. Image by NASA (public domain).

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 **9.4.9**) 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.

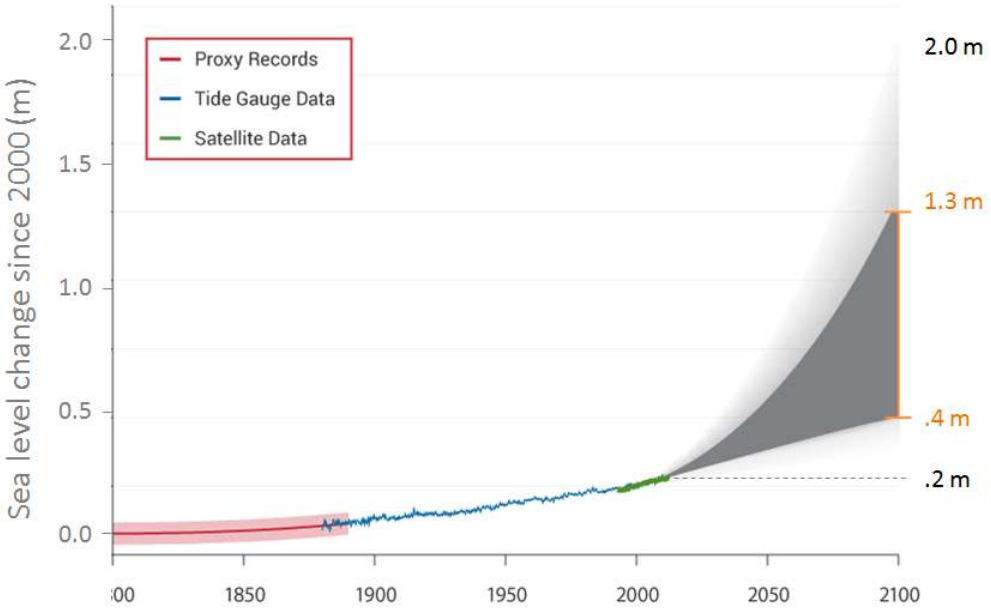



Figure **9.4.9**: Projected sea-level increases to 2100, showing likely range (grey) and possible maximum [Adapted by Steven Earle from "Past and Projected Changes in Global Sea Level Rise" based on data from Parris et al. 2012 with input from NASA Jet Propulsion Laboratory.]

Regional and local factors will influence future relative sea level rise for specific coastlines around the world (Figure **9.4.10**). For example, relative sea level rise depends on land elevation changes that occur as a result of subsidence (sinking) or uplift (rising). Relative sea level rise also depends on local changes in currents, winds, salinity, and water temperatures, as well as proximity to thinning ice sheets. Assuming that these historical geological forces continue, a 2-foot rise in global sea level by 2100 would result in the following relative sea level rise:

- 2.3 feet at New York City
- 2.9 feet at Hampton Roads, Virginia
- 3.5 feet at Galveston, Texas
- 1 foot at Neah Bay in Washington state



**Figure 9.4.10:** The United States government paid for the residents of Isle De Jean Charles, an island south of Louisiana (that is also part of Louisiana), to relocate when it became inhabitable due to sea level rise. Image by Karen Apricot (CC-BY-SA).

## Future Ocean Acidification

Since 1750, about 40% of the new anthropogenic carbon dioxide has remained in the atmosphere. The remaining 60% gets absorbed by the ocean and vegetation. Therefore, the ocean has absorbed about 30% of new anthropogenic carbon dioxide. When carbon dioxide gets absorbed in the ocean, it creates carbonic acid which makes the ocean more acidic. **Ocean acidification** is the process of ocean waters decreasing in pH. Oceans become more acidic as carbon dioxide ($CO_2$) emissions in the atmosphere dissolve in the ocean. This change is measured on the pH scale, with lower values being more acidic. The pH level of the oceans has decreased by approximately 0.1 pH units since pre-industrial times, which is equivalent to a 25% increase in acidity. The pH level of the oceans is projected to decrease even more by the end of the century as $CO_2$ concentrations are expected to increase for the foreseeable future. Ocean acidification adversely affects many marine species, including plankton, mollusks, shellfish, and corals. As ocean acidification increases, the availability of calcium carbonate will decline. Calcium carbonate is a key building block for the shells and skeletons of many marine organisms. Acidification combined with warmer temperature and lower oxygen levels is expected to have severe impacts on marine ecosystems and human-used fisheries, possibly affecting our ocean-derived food sources [6].

Corals require the right combination of temperature, light, and the presence of calcium carbonate (which they use to build their skeletons). As atmospheric carbon dioxide ($CO_2$) levels rise, some of the excess $CO_2$ dissolves into ocean water, reducing its calcium carbonate saturation. Calcium carbonate saturation has already been reduced considerably from its pre-industrial level, and model projections suggest much greater reductions in the future. If atmospheric $CO_2$ concentrations double, coral calcification rates are projected to decline by more than 30%. If $CO_2$ concentrations continue to rise at their current rate, corals could become rare on tropical and subtropical reefs by 2050. Under projections for the future, it is very unlikely that calcium carbonate saturation levels will be adequate to support coral reefs in any U.S. waters.

## Plant Productivity




Watch this NASA video to discover the mixed effects of global warming on plant growth. While scientists found that warmer temperatures in the 1980s and 1990s caused an increase in plant productivity, this advantage has since been counteracted by more frequent droughts.

### Mismatched Interactions

In addition to some abiotic conditions changing in response to climate change, many organisms are also being affected by the changes in temperature. Temperature and precipitation play key roles in determining the *geographic distribution* and *phenology* of plants and animals. **Phenology** is the study of the effects of climatic conditions on the timing of periodic life cycle events, such as flowering in plants or migration in birds.

Researchers have shown that 385 plant species in Great Britain are flowering 4.5 days sooner than was recorded earlier during the previous 40 years. In addition, insect-pollinated species were more likely to flower earlier than wind-pollinated species. The impact of changes in flowering date would be mitigated if the insect pollinators emerged earlier. This mismatched timing of plants and pollinators could result in injurious ecosystem effects because, for continued survival, insect-pollinated plants must flower when their pollinators are present. Likewise, migratory birds rely on daylength cues, which are not influenced by climate change. Their insect food sources, however, emerge earlier in the year in response to warmer temperatures. As a result, climate change decreases food availability for migratory bird species.

### Spread of Disease

This rise in global temperatures will increase the range of disease-carrying insects and the viruses and pathogenic parasites they harbor. Thus, diseases will spread to new regions of the globe. This spread has already been documented with dengue fever, a disease the affects hundreds of millions per year, according to the World Health Organization. Colder temperatures typically limit the distribution of certain species, such as the mosquitoes that transmit malaria, because freezing temperatures destroy their eggs. 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).

Not only will the range of some disease-causing insects expand, the increasing temperatures will also accelerate their lifecycles, which allows them to breed and multiply quicker, and perhaps evolve pesticide resistance faster. In addition to dengue fever, other diseases are expected to spread to new portions of the world as the global climate warms. These include malaria, yellow fever, West Nile virus, zika virus, and chikungunya.

Climate change does not only increase the spread of diseases in humans. Rising temperatures are associated with greater amphibian mortality due to chytridiomycosis (see Invasive Species). Similarly, warmer temperatures have exacerbated bark beetle infestations of coniferous trees, such as pine an spruce.

## Climate Change Affects Everyone

Our lives are connected to the climate**.** Human societies have adapted to the relatively stable climate we have enjoyed since the last ice age which ended several thousand years ago. A warming climate will bring changes that can affect our water supplies, agriculture, power and transportation systems, the natural environment, and even our own health and safety. Carbon dioxide can stay in the atmosphere for nearly a century, on average, so Earth will continue to warm in the coming decades. The warmer it gets, the greater the risk for more severe changes to the climate and Earth’s system. Although it’s difficult to predict the exact impacts of climate change, what’s clear is that the climate we are accustomed to is no longer a reliable guide for what to expect in the future

## References


3. Lindsey, R. Climate and Earth’s Energy Budget: Feature Articles. (2009). Available at: http://earthobservatory.nasa.gov. (Accessed: 14th September 2016)

6. Pachauri, R. K. *et al. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. (IPCC, 2014).

7. Oreskes, N. The scientific consensus on climate change. *Science* **306**, 1686–1686 (2004).

11. Earle, S. *Physical geology OER textbook*. (BC Campus OpenEd, 2015).

12. Callendar, G. S. The artificial production of carbon dioxide and its influence on temperature. *Q.J.R. Meteorol. Soc.* **64**, 223–240 (1938).

13. Hansen, J., Sato, M., Kharecha, P. & others. Earth’s energy imbalance and implications. *Atmospheric* (2011).

14. Foster, G. & Rahmstorf, S. Global temperature evolution 1979–2010. *Environ. Res. Lett.* **6**, 044022 (2011).

15. Kosaka, Y. & Xie, S.-P. Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature* **501**, 403–407 (2013).

16. Easterling, D. R. & Wehner, M. F. Is the climate warming or cooling? *Geophys. Res. Lett.* **36**, (2009).

17. Zemp, M. *et al*. Historically unprecedented global glacier decline in the early 21st century. *J. Glaciol*. **61**, 745–762 (2015).

18. Karl, T. R. & Knight, R. W. Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Am. Meteorol. Soc*. **79**, 231–241 (1998).

19. Santer, B. D. *et al*. Identification of human-induced changes in atmospheric moisture content. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15248–15253 (2007).

## Suggested Supplementary Reading

Intergovernmental Panel on Climate Change. 2013. 5th Assessment: Summary for Policymakers.

NASA. 2018. Global Climate Change: Vital Signs of the Planet. This website by NASA provides a multi-media smorgasbord of engaging content. Learn about climate change using data collected by NASA satellites and more.

## Attributions

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

- Contributed by Chris Johnson, Matthew D. Affolter, Paul Inkenbrandt, & Cam Mosher Faculty (Geology) at Salt Lake Community College
- Sourced from OpenGeology from *An Introduction to Geology* Chapter Chapter 15: Global Climate Change
- ”Physical Geology” by Steven Earle used under a CC-BY 4.0 international license. Download this book for free at http://open.bccampus.ca
- 12.4 Climate Change from *Essentials of Environmental Science* by CK-12 and Kamala Dorsner (Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.)
- 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)

---

This page titled 9.4: Implications of Climate Change is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

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

# CHAPTER OVERVIEW

## 10: Urbanization


10.1: Urbanization

10.2: Environmental Challenges of Urbanization

10.3: Sustainable Cities

10.4: Case Study- Sustainable Stormwater Management


---


10: Urbanization is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 10.1: Urbanization

**Urbanization** is the study of the social, political, and economic relationships in cities, and someone specializing in **urban sociology** would study those relationships. Urban areas are cities, towns, and other places where people live in a compact population. Most urban people are engaged in economic activities that function efficiently in dense populations, such as commerce, education, manufacturing, and services. In contrast, resource-based economic activities, such as cultivating food and harvesting minerals, fossil fuels, timber, or wildlife, occur primarily in rural areas. There is no strict dividing line between rural and urban; rather, there is a continuum where one bleeds into the other. However, once a geographically concentrated population has reached approximately 100,000 people, it typically behaves like a city regardless of what its designation might be.

Urban areas import energy, food, and materials from the local countryside, as well as from more distant regions of their country, in addition to other nations around the world in an increasingly global commerce. The resource-related connections between cities and rural areas exist at all spatial scales (local, regional, and global) and are an integral but often insufficiently appreciated aspect of urban ecology.

The development of dense habitations (initially small villages) during human cultural evolution began about 10-thousand years ago, when early agricultural practices allowed the production of local surpluses of food (harvests that exceeded the subsistence needs of the farmers themselves). The excess food encouraged the development of specialized occupations that were most efficiently performed in a central place such as a village or town. New societal activities included organized religion and political systems, artisanal manufacturing, and the means of administering these over wide areas through hierarchical economic, political, religious, and military structures.

The first villages were probably supported by both settled agriculture and local hunting-and-gathering activities. By about 9,000 years ago, small farming villages were relatively widespread in some regions, particularly in the Fertile Crescent of the Middle East (see Chapter 24). By 8,500 years ago, that region boasted sizeable walled towns. Similar developments also occurred in China and probably elsewhere in southern Asia.

Most scholars agree that the first cities were developed somewhere in ancient Mesopotamia, though there are disagreements about exactly where. Most early cities were small by today's standards, and the largest city around 100 CE was most likely Rome, with about 650,000 inhabitants. The factors limiting the size of ancient cities included lack of adequate sewage control, limited food supply, and immigration restrictions. For example, serfs were tied to the land, and transportation was limited and inefficient. Today, the primary influence on cities' growth is economic forces.

## Growth of Urban Populations

In 1750, few cities supported more than 50-thousand people – London was the only one in England, and there were none in North America. In 1830, New York and Philadelphia were the only American cities with a population greater than 100-thousand. By 1910, however, 52 cities in North America supported more than 100-thousand people (New York was the largest, with more than 1 million). In 1950, New York and Tokyo were the only two megacities in the world (that is, having a population greater than 10 million). In 2014, there were 537 cities supporting more than 1 million people, including 35 megacities, of which 29 were in the developing world (Brinkhoff, 2015).

Urbanization levels are affected by two things – migration and natural increase. **Migration** is the movement of population from one area to another. Some migrations are forced, voluntary, permanent and temporary, international and regional. Rural to urban migration, is the movement of people from countryside to city areas. This type of migration happened in developed countries from the *18th century onwards* on a large scale, and has gradually slowed down. However, many developing countries are experiencing massive rural to urban migration, mainly of young males, into the major cities.

As was the case in North America, other urban centers experienced a growth spurt during the Industrial Era. In 1800, the only city in the world with a population over 1 million was Beijing, but by 1900, there were 16 cities with a population over 1 million. The development of factories brought people from rural to urban areas, and new technology increased the efficiency of transportation, food production, and food preservation. For example, from the mid-1670s to the early 1900s, London increased its population from 550,000 to 7 million. The growth in global urbanization in the 20th and 21st centuries is following the blueprint of North American cities, but is occurring much more quickly and at larger scales, especially in peripheral and semi-peripheral countries. Shanghai almost tripled its population from 7.8 million to 20.2 million between 1990 and 2011, adding the equivalent of the population of

New York City in 20 years. It is projected to reach 28.4 million by 2025, third in size behind Tokyo (38.7 million) and New Delhi (32.9 million).

The major reasons for migration can be classified into push and pull factors. A **push factor** is something that can force or encourage people to move away from a country. Push factors can include famine (as in Ethiopia in the 1980s), drought, flooding, a lack of employment opportunities, population growth, over population, and civil war. A **pull factor** is one in which encourages people to move to a city. Pull factors include the chance of a better job, better access to education and services, a higher standard of living. These factors have contributed to millions of people in developing countries moving to cities, creating mass urbanization. **Natural increase** (a population increase due to more births and fewer deaths) also has a major effect on rates of urbanization. Natural increase is stimulated by better access to medical care, improved water supplies, sanitary conditions and wealth.

### Suburbs and Exurbs

As cities grew and became more crowded (and often more impoverished and costly) more and more people began to migrate back out of them. But instead of returning to rural small towns (like they had resided in before moving to the city), these people needed close access to the cities for their jobs. In the 1850s, as the urban population greatly expanded and transportation options improved, suburbs developed. **Suburbs** are the communities surrounding cities, typically close enough for a daily commute in, but far enough away to allow for more space than city living affords. The bucolic suburban landscape of the early 20th century has largely disappeared due to sprawl.

**Urban sprawl** contributes to *traffic congestion*, which in turn contributes to commuting time. Commuting times and distances have continued to increase as new suburbs developed farther and farther from city centers. Simultaneously, this dynamic contributed to an exponential increase in natural resource use, like petroleum, which sequentially increased *pollution* in the form of carbon emissions (negative aspects of urban sprawl will be explored further in the following section).

As the suburbs became more crowded and lost their charm, those who could afford it turned to the **exurbs**, communities that exist outside the ring of suburbs and are typically populated by even wealthier families who want more space and have the resources to lengthen their commute. It is interesting to note that unlike U.S. cities, Canadian cities have always retained a fairly large elite residential presence in enclaves around the city centres, a pattern that has been augmented in recent decades by patterns of inner-city resettlement by elites (Caulfield 1994; Keil and Kipfer 2003). As cities evolve from industrial to postindustrial, this practice of **gentrification** becomes more common. Gentrification refers to members of the middle and upper classes entering city areas that have been historically less affluent and renovating properties while the poor urban underclass are forced by resulting price pressures to leave those neighbourhoods. This practice is widespread and the lower class is pushed into increasingly decaying portions of the city.

## Urban Ecosystems

Any urbanized area can be viewed as being an ecosystem, because it has the following ecological attributes:

1. a need for enormous inputs of energy and materials to sustain its human population and its diverse economic activities, and to maintain its structure and grow
2. a complex metabolism, including well-developed webs of transfer, processing, and storage of materials, energy, and information among interacting organisms and economic sectors
3. and immense outputs of heat and other waste materials, which are disposed of in surrounding ecosystems, causing pollution and other environmental problems

Of course, the habitats of cities and towns are very strongly influenced by human activities. Collectively they comprise an urban–industrial techno-ecosystem. Although humans are the dominant species in the urban ecosystem, many other species are also supported, most of which are not native to the region. The structure of this anthropogenic ecosystem is dominated by the businesses, dwellings, factories, roads, and other infrastructure of the human economy, while also supporting manicured green space as well as remnants of natural habitats in parks and other less-developed spaces. Even parking lots, sidewalks, and industrial areas, which from an ecological perspective are extremely degraded habitats, do manage to support some biodiversity. Important ecological functions also occur within the urban ecosystem, such as biological productivity and water and nutrient cycling, but these processes are greatly influenced by humans. Still, some urban places are relatively natural in character and are maintained in this condition as parks. Some prominent examples of “natural-area parks” include Stanley Park in Vancouver; a series of parks

along the Bow River in Calgary and on the North Saskatchewan River in Edmonton; Assiniboine Park in Winnipeg; High Park and Hanlan's Point in Toronto; Bois de Liesse in Montreal; Point Pleasant Park in Halifax; and Signal Hill in St. John's. These greenspaces are remnants of natural habitat that have survived the urbanization process, and they contain ecological communities that are mostly dominated by native species.

All urban areas are intrinsically dependent on surrounding ecosystems to provide them with necessary resources and to assimilate wastes that are generated. The ecological footprint (or eco-footprint) of an urban population is the area of ecoscape (landscape and seascape) that is needed to supply the necessary food, energy, materials, waste disposal, and other crucial goods and services. As a global average, the average human has an eco-footprint of about 2.7 hectares, but there are only 1.8 ha of bio-productive land and water on Earth (Ewing et al., 2010). This means that the human enterprise has already overshot global bio-capacity by 30%, and is now operating on an unsustainable basis by depleting the remaining stocks of "natural capital". Note that, in the context of ecological footprints, these data are measures in "global hectares" (gha), which represent the average productivity of all bio-productive habitats on Earth or in a country, including agricultural land, forests, and fishing grounds, but not including desert, glaciers, or the open ocean.

An average Canadian has an ecological footprint of about 7.3 global hectares (2007 data; Ewing et al., 2010). This is the seventh-most intensive national per-capita footprint in the world, after the United Arab Emirates (10.7 gha), Qatar (10.5 gha), Denmark (8.3 gha), Belgium (8.0 gha), the United States (8.0 gha), and Estonia (7.9 gha). Here are some additional comparisons among wealthier countries: Australia (6.8 gha), Kuwait (6.3 gha), Ireland (6.3 gha), Norway (5.6 gha), France (5.0 gha), Germany (5.0 gha), United Kingdom (4.9 gha), and Japan(4.7 gha). The rapidly-growing economies include Russia (4.4 gha), Brazil (2.2 gha), China (2.2 gha), and India (0.9 gha). Of course, people living in poorer countries have much smaller ecological footprints: Afghanistan (0.6 gha), Bangladesh (0.6 gha), Haiti (0.7 gha), Burundi (0.9 gha), Ethiopia (1.1 gha), Vietnam (1.4 gha), and Peru, (1.5 gha).

## Urban Planning

Urban ecosystems have extremely complex structures and functions (although not more so than natural ecosystems). To some degree, their development has occurred in an orderly fashion, with certain areas being designated for particular kinds of structures and activities. Urban planning is the active process of designing and organizing the structure and function of cities. As such, urban planning contributes to the information needed by legislators and other decision makers as they develop sensible and efficient siting of the following:

- buildings, including homes, commercial properties, institutions (such as hospitals and schools), and industrial facilities
- infrastructure for transportation, utilities, and waste management, such as roads, railways, public-transit routes, electrical transmission and pipeline corridors, sewers and sewage-treatment facilities, and solid-waste disposal areas
- greenspaces, including playing fields and horticultural and natural-area parks

Well-planned urban areas have relatively pleasant neighborhoods where people live and work. In contrast, poorly planned cities are chaotic, dirty, and unpleasant. In general, urban planning is most effective in wealthy developed countries such as Canada, but much less so in poorer developing countries where the population is growing and urbanization is proceeding most rapidly.

The dominant planning paradigm of the past 60 years has involved the segregation of major land-uses and economic activities into different areas. This kind of strategy has greatly influenced the design of modern cities, including all those in Canada. It has resulted in many urban people living in discrete inner-city neighborhoods or more distant suburbs, while shopping in large malls, working in factory or office complexes and industrial parks, and commuting long distances among these land-use types. However, this type of planning has contributed to some important urban problems, including the following:

- the rapid growth of huge, multi-city, urbanized regions (sometimes known as conurbations)
- the inefficient segregation of residences from places of work and commerce
- long commuting times for workers
- congested transportation systems
- a decay of neighborhood life
- environmental problems such as air and water pollution, wasteful use of energy and materials, paving of valuable farmland (many urban areas are located on excellent agricultural land), and losses of natural habitat

The paradigm of widely segregated land-use is now being challenged by the concept of a more integrated “neighborhood” design. This involves the development of relatively compact, self-sufficient communities that contain a mixture of residential and commercial land-uses. In some respects, this harkens back to more traditional elements of community design, in which housing, employment, local commerce, small-scale manufacturing, and recreation were all within easy walking distance. The re-emergence of this concept in urban planning has been substantially influenced by the ideas of Jane Jacobs, a geographer who taught at the University of Toronto.

## References Cited and Further Reading


American Public Transportation Association (APTA). 2014. Public Transportation Fact Book, Appendix A: Historical Tables. APTA, Washington, DC.

Berkowitz, A.R., C.H. Nilon, and K.S. Hollweg. 2002. Understanding Urban Ecosystems. Springer Verlag, Berlin, Germany.

Breuste, J., H. Feldmann, and O. Ohlmann (eds.). 1998. Urban Ecology. Springer Verlag, Berlin, Germany.

Brinkhoff, T. 2015. Major Agglomerations of the World. http://www.citypopulation.de/world/Agglomerations.html

Environment Canada. 1996. The State of Canada’s Environment. Government of Canada, Ottawa, ON.

Environment Canada. 1999. Air Quality Trends in Canadian Cities, 1979–1992. Ottawa, ON: State of the Environment Reporting, Environment Canada. https://web.archive.org/web/20010214040646/http://www.ec.gc.ca/pdb/uaqt/aqfact_e.html

Environment Canada. 2007. Municipal population served by wastewater treatment. State of Environment Infobase Ottawa, ON. https://web.archive.org/web/20040621114650/http://www.ec.gc.ca:80/soer-ree/English/Indicators/Issues/Urb_H2O/Bulletin/uwind3_e.cfm

Environment Canada. 2014. Air Quality Indicators. Environment Canada, Ottawa, ON. http://www.ec.gc.ca/indicateurs-indicators/default.asp?lang=en&n=7DCC2250-1

Ewing, B., D. Moore, A. Ourser, A. Reed, and M. Wackernagel. 2010. Ecological Footprint Atlas 2010. Global Footprint Network, Oakland, CA. https://www.footprintnetwork.org/content/images/uploads/Ecological_Footprint_Atlas_2010.pdf

Federation of Canadian Municipalities (FCM). 2009. Waste Diversion Success Stories from Canadian Municipalities. FCM, Ottawa, ON. http://www.fcm.ca/Documents/tools/GMF/Getting_to_50_percent_en.pdf

Fincher, R. and J.M. Jacobs (eds.). 1998. Cities of Difference. Guilford Press, New York, NY.

Freedman, B., S. Love, and B. O’Neil. 1996. Tree species, biomass, and carbon storage in stands of urban forest of varying character in Halifax, Nova Scotia, Canada. Canadian Field-Naturalist, 110: 675–682.

Gill, D. and P. Bonnett. 1973. Nature in the Urban Landscape: A Study of City Ecosystems. York Press, Baltimore, MD.

Global Footprint Network. 2007. Canadian Living Planet Report. World Wildlife Fund Canada, Toronto, ON, and the Global Footprint Network, Oakland, CA. https://web.archive.org/web/20180618150932/https://www.footprintnetwork.org/content/documents/2007_Canadian_Living_Planet_Report.pdf

Nancy B. Grimm, Stanley H. Faeth, Nancy E. Golubiewski, Charles L. Redman, Jianguo Wu, Xuemei Bai, and John M. Briggs (2008). “Global Change and the Ecology of Cities”, Science 8 February 2008: Vol. 319 no. 5864 pp. 756-760 DOI: 10.1126/science.1150195.

Holcomb, M.C. 1987. Transportation Energy Data Book, 9th ed. Oak Ridge National Laboratory, Oak Ridge, TN.

Jacobs, J. 1961. The Death and Life of Great American Cities. Random House, New York, NY.

Jacobs, J. 1969. The Economy of Cities. Random House, New York, NY.

Kreith, F. (ed.). 1994. Handbook of Solid Waste Management. McGraw-Hill, New York, NY.

Kryter, K.D. 1985. The Effects of Noise. 2nd ed. Academic Press, Orlando, FL.

Landsberg, H.E. 1981. The Urban Climate. Academic Press, New York, NY.

Langford, T.E.L. 1990. Ecological Effects of Thermal Discharges. Elsevier Science, Amsterdam, The Netherlands.

Lowe, M.D. 1989. The Bicycle: Vehicle for a Small Planet. Worldwatch Institute, Washington, DC. Nathanson, J.A. 2002. Basic Environmental Technology: Water Supply, Waste Management, and Pollution Control. 3rd ed. Prentice Hall, Upper Saddle River, NJ.

National Round Table on the Environment and the Economy (NRTEE). 1998. Greenhouse Gas Emissions from Urban Transportation: Backgrounder. NRTEE, Ottawa, ON.

Platt, R.H., R.A. Rowntree, and P.C. Muick. 1994. The Ecological City: Preserving and Restoring Urban Biodiversity. University of Massachusetts Press Amherst, MA.

Ponting, C. 1991. A Green History of the World. Sinclair-Stevenson London, UK.

Rees, W.E. 1992. Ecological footprints and appropriated carrying capacity: What urban economics leaves out. Environment and Urbanization, 4: 121–130.

Rees, W.E. 1997. Is “sustainable city” an oxymoron? Local Environment, 2: 303–310.

Rees, W.E. 2010. The Human Nature of Unsustainability. The Post Carbon Institute, Santa Rosa, CA.

Savard, J.-P. 1978. Birds in Metropolitan Toronto: Distribution, Relationships with Habitat Features, and Nesting Sites. Unpublished M.Sc. Thesis, University of Toronto.

Statistics Canada. 2006. Human Activity and the Environment. Annual Statistics 2006. www.statcan.ca/english/freepub/16-201-XIE/16-201-XIE2006000.pdf

Statistics Canada. 2006. Road motor vehicle, trailer and snowmobile registration, annual (registrations). Table 405-0001, Statistics Canada, Ottawa, ON. http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=4050001&pattern=automobiles&tabMode=dataTable&srchLan=-1&p1=1&p2=50

Statistics Canada. 2012. Human Activity and the Environment, Waste Management in Canada. Statistics Canada, Ottawa, ON. http://www5.statcan.gc.ca/olc-cel/olc.action?ObjId=16-201-X&ObjType=2&lang=en&limit=0

Statistics Canada. 2017. Road motor vehicle registrations, by type of vehicle. Table 23-10-0067-01 (formerly CANSIM 405-0004), Statistics Canada, Ottawa, ON. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2310006701

Statistics Canada. 2014b. Leading causes of death, by sex. CANSIM Table 102-0561, Statistics Canada, Ottawa, ON. http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/hlth36a-eng.htm Accessed November, 2014.

Timerson, B.J. 1999. A Guide to Noise Control in Minnesota. Minnesota Pollution Control Agency. Minneapolis, MN.

Transport Canada. 2012. Canadian Motor Vehicle Traffic Collision Statistics, 2004. Transport Canada, Ottawa, ON. http://www.tc.gc.ca/eng/roadsafety/tp-tp3322-2004-menu-691.htm

Transport Canada. 2017. Annual Reports: Transportation in Canada. Transport Canada Ottawa, ON. https://www.tc.gc.ca/eng/policy/anre-menu.htm

Turner, K., L. Lefler, and B. Freedman. 2004. Plant communities of selected urbanized areas of Halifax, Nova Scotia, Canada. Landscape and Urban Planning, 71: 191–206.

Wackernagel, M. and W.E. Rees. 1996. Our Ecological Footprint: Reducing Human Impact on the Earth. New Society, Gabriola Island, BC.

## Attribution

Modified by Kyle Whittinghill from the following sources:

- Urbanization and Cities from *Essentials of Environmental Science* by Kamala Dorsner and CK-12 (CC-BY)
- Urbanization around the World from *Essentials of Environmental Science* by Kamala Dorsner and CK-12 (CC-BY)
- Chapter 12: Urban Ecology from *Environmental Science: A Canadian Perspective* by Bill Freedman (Creative Commons Attribution NonCommercial)

10.1: Urbanization is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 10.2: Environmental Challenges of Urbanization

## Impacts of Urban Sprawl

Urban sprawl is the extension of low-density residential, commercial, and industrial development into areas beyond a city's boundaries that occurs in an unplanned or uncoordinated manner. It is generally characterized by:

- low-density development that is dispersed and situated on large lots (greater than one acre)
- geographic separation of essential places such as work, home, school, and shopping
- high dependence on automobiles for travel
- increased impervious surface area in watersheds
- habitat fragmentation and degradation

Urban sprawl combines low density (Figure **10.2.1**) and fragmentation of the urban area (Figure **10.2.2**), increases the average travel distances for daily trips, and hinders a shift toward less energy-intensive transportation modes. The sprawling nature of cities is critically important because of the major impacts that are evident in increased energy, land and soil consumption. These impacts threaten both the natural and rural environments, raising greenhouse gas emissions that cause climate change, and elevated air and noise pollution levels which often exceed the agreed human safety limits. Thus, urban sprawl produces many adverse impacts that have direct effects on the quality of life.

| | Sprawling City | Compact City |
|---|---|---|
| |  |  |
| Location | Atlanta | Barcelona |
| Urban Population (million inhabitants) | 2.5 | 2.8 |
| Urban Area ($km^2$) | 4,280 | 162 |
| Urban Density (pph) | 6 | 173 |
| Energy Consumption Per Capita for Private Transportation (MJoules) | 80 | 9 |

[Figure 1]

Figure **10.2.1**: Sprawl vs. Compact Cities. An often cited example of urban sprawl is Atlanta, GA (US), which has a similar population as Barcelona but occupies an urban area that is 26 times as large



Figure 10.2.2: Sprawling area. An aerial photo of Bellefonte, Pennsylvania, USA. Looking southeast taken from a glider on July 1, 2007 by Dhaluza at English Wikipedia

### Health

If communities are not walkable or bikeable, we need to drive to schools, shops, parks, entertainment, play dates, etc. Thus we become more sedentary. Residents of sprawling counties were likely to walk less during leisure time and weigh more than residents of compact counties. A sedentary lifestyle increases the risk of overall mortality, cardiovascular disease, and some types of cancer. The effect of low physical fitness is comparable to that of hypertension, high cholesterol and diabetes.

### Transportation and Consumption of Energy

A consequence of the increasing consumption of land and reductions in population densities as cities sprawl is the growing consumption of energy. Generally, compact urban developments with higher population densities are more energy efficient. Evidence from 17 cities around the world shows a consistent link between population density and energy consumption (Figure 10.2.3), and in particular high energy consumption rates that are associated with lower population densities, characteristic of sprawling environments, dependent on lengthy distribution systems that undermine efficient energy use.

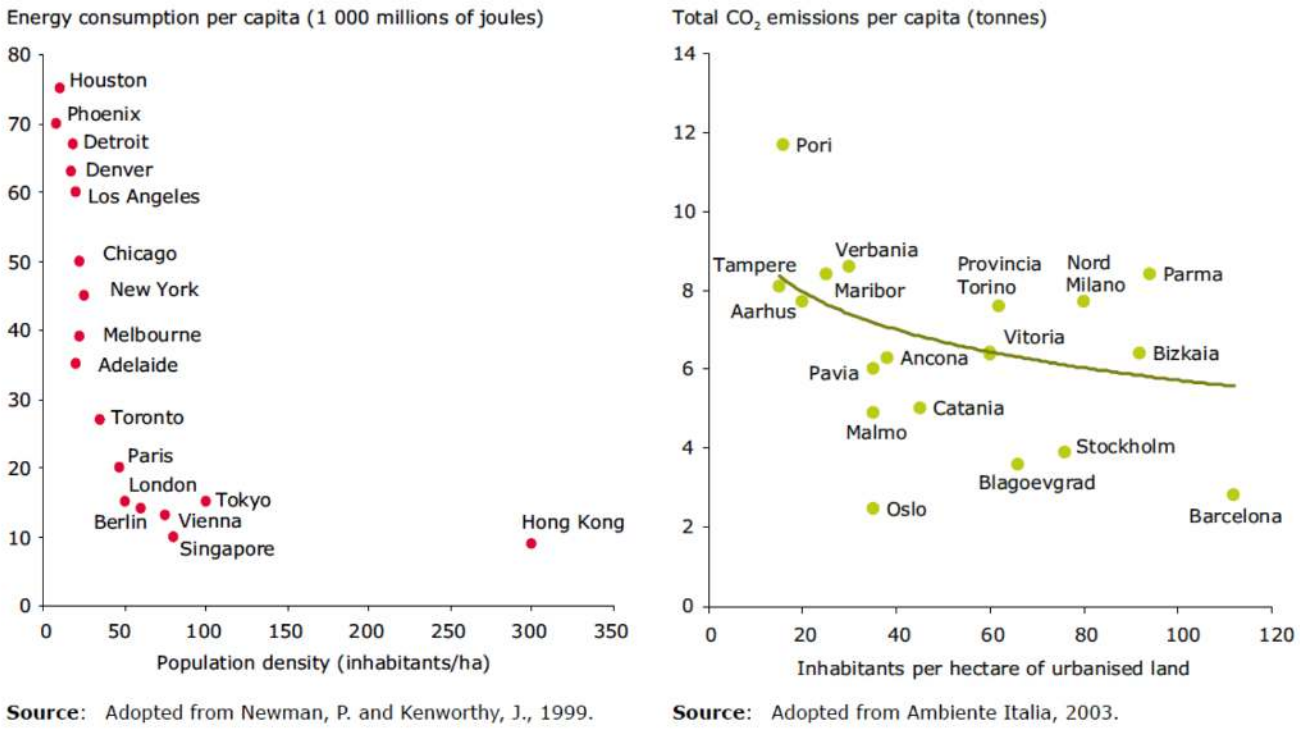



Figure **10.2.3**: Left: Population density and energy consumption, selected World cities. Right: Population density and CO2 emissions, selected European cities.

Transport related energy consumption in cities depends on a variety of factors including the nature of the rail and road networks, the extent of the development of mass transportation systems, and the modal split between public and private transport. From the turn of the twentieth century to the 1950s, public transit systems were the most commonly used means of longer-distance personal transportation in North American cities. Public transit is still an important means of transportation in cities, although its growth is moderate and not nearly as impressive as that of automobile use. Evidence shows that there is a significant increase in travel related energy consumption in cities as densities fall. Essentially, the sprawling city is dominated by relatively energy inefficient car use, as the car is frequently the only practical alternative to more energy efficient, but typically inadequate, relatively and increasingly expensive public transportation systems. Increased transport related energy consumption is in turn leading to an increase in the emission of $CO_2$ to the atmosphere. Urban sprawl therefore poses significant threats to the commitments to reduce GHG gas emissions.

Although owning a car or truck is widely viewed as a desirable aspect of our lifestyle, the reliance on personal motor vehicles seriously affects environmental quality in cities and towns. Compared with other means of urban transportation, personal motor vehicles emit more air pollution, require more physical infrastructure (such as roads and parking space), use more material and energy resources, and are more costly to own and operate.

## Air pollution

Using fossil fuels also results in the emission of other gases and particulates that degrade air quality (note that commuters generate emissions of air pollution, which lowers the ambient quality of the air in areas they pass through and causes health problems for other people). Longer transportation distances intensify traffic congestion, resulting in lost productivity, and increase the need for more extensive infrastructure (such as more highways) that negatively impact the environment by increasing the amount of impervious cover and by requiring more natural resources. Finally, traffic congestion and air pollution from driving contribute to an estimated 900,000 fatalities per year worldwide.

Urban air pollution can cause life-threatening health problems for many people, particularly the old or very young and those with respiratory diseases. Severe pollution by sulphur dioxide and particulates was once common in North American and European cities, when coal was burned as an industrial and residential fuel. This problem has, however, been greatly alleviated since the early

1960s because of clean-air regulations. However, this kind of reducing smog is still severe and even worsening in many rapidly developing countries, where a lower priority is given to enforcing clean-air laws. Cities in developed countries continue to be affected by significant levels of air pollution. The following are particularly important.

- **$SO_2$ and NOx gases**: Although so-called reducing smog is much less of a problem than before, urban atmospheres still have higher concentrations of $SO_2$ and NOx than rural areas. The dry deposition of these gases damages marble, limestone, sandstone, and other acid-sensitive building materials and statues, and it contributes to the acidification of aquatic habitats (Chapter 19). In addition, these gases may be directly toxic to sensitive urban plants and lichens.
- **Suspended particulates**: Largely originating from emissions from vehicles, furnaces, and other combustion sources, suspended particulates are mostly fine aerosols of ammonium sulphate, ammonium nitrate, and sooty organics. At high concentrations they form a haze that interferes with visibility and is a hazard to people with respiratory problems. In general, air pollution by particulates has decreased markedly during the past several decades, mainly because of better emission controls on power plants, other industrial sources, and vehicles, and because of widespread switching to natural gas as a fuel.
- **Oxidants**: Cities with a sunny climate and a large number of motor vehicles often develop a photochemical smog during the day, characterized by high concentrations of ozone, peroxy acetyl nitrate (PAN), NOx, and volatile organic compounds (Chapter 16). The ozone component is responsible for most of the oxidant damage because it is toxic to many plants, irritates people's eyes and respiratory tracts, and degrades many materials and pigments. The worst photochemical smog occurs in southern Ontario and Quebec, extending from Windsor to Quebec City, and to a lesser degree in the greater Vancouver area and around Saint John. The air-quality objective for ozone (63 ppb) is exceeded on many summer days in the Quebec-Windsor corridor, where maximum hourly concentrations can exceed 160 ppb. Niagara Falls, Sarnia, Windsor, and Guelph are among the smoggiest cities in Canada, exceeding ozone guidelines for more than 50 days per year. The photochemical smog in this region occurs because of local emissions of NOx and volatile organics (these are precursors in the photochemical production of ozone; see Chapter 16), as well as those blowing in from urban areas in the nearby United States. In contrast, local emissions are the major cause of ozone smog in the Vancouver area.

### Natural and Protected areas

The impacts of sprawl on natural areas are significant. The considerable impact of urban sprawl on natural and protected areas is exacerbated by the increased proximity and accessibility of urban activities to natural areas, imposing stress on ecosystems and species through noise and air pollution. Immediate impacts such as the loss of natural land or the fragmentation of forests, wetlands and other habitats are well known direct and irreversible impacts. Urban land fragmentation, with the disruption of migration corridors for wildlife species, isolates these populations and can reduce natural habitats to such an extent that the minimum area required for the viability of species populations is no longer maintained.

The environmental impacts of sprawl are evident in a number of ecologically sensitive areas located in coastal zones and mountain areas. The Mediterranean coast, one of the world's 34 biodiversity hotspots, is particularly affected, and the increased demand for water for urban use, competes with irrigation water for agricultural land. This problem has been exacerbated by the increased development of golf courses in Spain, where the over-extraction of groundwater has led to salt water intrusion into the groundwater. Increased transit and tourist traffic, particularly day tourism from the big cities, also adds to the exploitation of the mountain areas as a natural resource for 'urban consumption' by the lowland populations.

In another example, the area around Victoria is the only place in Canada where dry forest dominated by Garry oak (Quercus garryana) occurs. This is habitat to many rare species and is one of our most endangered natural ecosystems. The expansion of residential areas is the greatest risk to this coastal oak forest. Conservation agencies in government and the private sector, including the Nature Conservancy of Canada, are trying to protect the surviving patches of this rare forest.

### Agricultural Land Area

The growth of European cities in recent years has primarily occurred on former agricultural land. Typically, urban development and agriculture are competing for the same land, as agricultural lands adjacent to existing urban areas are also ideal for urban expansion. The loss of agricultural land has major impacts on biodiversity with the loss of valuable habitat for many animals, and particularly birds. Sprawling cities also threaten to consume the best agricultural lands, displacing agricultural activity to both less productive areas (requiring higher inputs of water and fertilisers) and more remote upland locations (with increased risk of soil erosion).

## Soil

Urban sprawl and the development of urban land dramatically transform the properties of soil, reducing its capacity to perform its essential functions. These impacts are evident in the extent of compaction of soil leading to impairment of soil functions; loss of water permeability (soil sealing) which dramatically decreases; loss of soil biodiversity, and reductions of the capacity for the soil to act as a carbon sink. In Germany, for example, it is estimated that 52 % of the soil in built-up areas is sealed (or the equivalent of 15 m2 per second over a decade). In addition, rainwater which falls on sealed areas is heavily polluted by tire abrasion, dust and high concentrations of heavy metals, which when washed into rivers degrade the hydrological system.

Urban soil can become polluted by metals in many ways. For example, people who live near metal-recycling factories may be exposed to pollution through dustfall from the atmosphere. In one case, severe pollution was found in a downtown Toronto residential neighborhood near a car-battery recycling factory. Lead in surface soil near the factory exceeded several thousand ppm (one sample had >5% lead), and some residents had elevated lead in their blood and hair. This astonishing land-use conflict occurred because of earlier, inappropriate land-use zoning, when less attention was placed on environmental concerns. However, the battery-recycling factory still remains in the residential neighborhood, although it operates much more cleanly than in the past. Lead pollution of urban soil has also been caused by residues of old paint, which often had very high concentrations of lead-containing pigments (commonly reaching 38% in the dried residue), especially in white and red colours. (Lead is now mostly replaced in household paints by titanium and other less-toxic metals). Soil close to busy highways is also affected by leaded gasoline (this problem has dissipated some since the ban on leaded gasoline).

## Water Quality

Increasing numbers of roads and parking lots are needed to support an automobile transportation system, which lead to increased non-point source water pollution and contamination of water supplies (road runoff of oil/gas, metals, nutrients, organic waste, to name a few) with possible impacts on human health. Increased erosion and stream siltation causes environmental damage and may affect water treatment plants and thus affect water quality.

Eutrophication is another kind of urban pollution. Lakes and rivers in urban areas often have an enriched nutrient supply, which results in increased productivity and eutrophication. The nutrient inputs originate from fertilizer use in horticulture, leaching from septic fields, and sometimes the dumping of sewage. Urban waterbodies are also commonly polluted with coliform bacteria and other intestinal pathogens and parasites that originate with sewage and surface runoff contaminated by pet and bird feces. Lakes and rivers may also be affected by spilled fuel, metals, garbage dumping, and sedimentation by soil eroded during construction activities. Any of these influences on urban lakes may render them less suitable or even unfit as a source of drinking water or for recreational purposes. Eutrophication and sedimentation also cause significant ecological damage in many urban waterbodies.

Urban sprawl can also be a threat to the ability of landscapes to provide key environmental “services,” such as clean water. One such case involves an area known as the Oak Ridges Moraine, which has become a rallying point for habitat protection against further urbanization in the Greater Toronto Area. The moraine is a 160-km-long, 1950-$km^2$ ridge located north of the city, and it is composed of hilly terrain underlain by glacial sand and gravel. Because of its rough topography and poor fertility, much of its area has remained forested or is used for pasture or other low-intensity agricultural purposes. Groundwater originating in the moraine is a source of well-water recharge for about 250-thousand people and is a source of 65 rivers and streams and many wetlands. However, with the rapid growth of the greater Toronto region, the moraine has been subjected to increasing pressure from residential and commercial development. Extensive clearing of its forested and pasture areas would degrade the ability of the moraine to provide clean groundwater for use by people and to support streams and wetlands. Loss of the remaining forest would also destroy habitat for native species that are rare in the region. In response to intense lobbying from groups seeking to limit new developments on the moraine, the Government of Ontario formed an advisory panel to provide advice about the regulation of land-use. The panel provided many recommendations, including the need to protect 92% of the moraine from intensive development, and these were key in the preparation of The Oak Ridges Moraine Conservation Act, 2001. The Act was used to prepare a land-use plan, including provisions for core natural areas and linkages among them, comprising about 62% of the moraine. In 2002, the premier of Ontario announced the formation of a non-profit Oak Ridges Moraine Foundation and provided it with $15 million to fund public education, monitor the moraine, develop trails, and secure natural habitats. The federal government also announced that land it owns on the moraine, about 30 $km^2$, would be kept as greenspace. In addition, the Nature Conservancy of Canada has been acquiring properties of high conservation value and is setting them aside as protected areas.

## Noise Pollution

Urban areas are typically much noisier than rural ones. Most of the noise comes from the operation of various kinds of machines that are abundant in urban places, such as air conditioners, lawnmowers, automobiles, heavy trucks, and air compressors. Loud music may also be important. Noise pollution begins when the level of ambient sound becomes distracting to the normal activities of people, for example, by making it difficult to understand a conversation or to have a restful sleep.

Noise intensity (also called sound pressure or loudness) is measured in units of decibels (dB). An increase of 10 dB is approximately equivalent to a doubling of the loudness. When the distance from a point source is doubled, the noise level decreases by about 6 dB. However, when the distance from a linear source of noise (such as a busy highway) is doubled, the sound level decreases by only about 3 dB.

At a high intensity, noise pollution can cause a progressive and enduring hearing impairment, which begins with a decreased ability to perceive higher-pitched sound. Under prolonged exposure, the hearing loss may progress through much of the auditory range. Hearing loss is caused by both the duration and intensity of exposure. For example, an occupational exposure to 115 dB may be permitted for no longer than 15 minutes, while exposure to 100 dB may be allowed for two hours and 90 dB for up to eight hours. Noise becomes painful at about 140 dB, although prolonged exposure to levels above 80-90 dB can result in long-term hearing loss. The noise levels typically associated with various sources are shown in Figure 25.7.

People may be exposed to excessive noise in various ways. Non-voluntary exposures occur in the ambient environment, such as near a road with heavy traffic. Occupational exposures occur in noisy factories or in hangars where jet engines are serviced. Voluntary exposures to excessive noise are also common, such as when people attend thunderous rock concerts or use headphones to listen to loud music. It is well known that permanent hearing loss is common among rock musicians as well as people who regularly listen to loud music at concerts, clubs, or using headphones. Frequent target-shooting with firearms can also result in long-term hearing loss unless effective ear protection is used.

## Urban Biodiversity

Some non-native animals and plants have found good habitat in urban areas, where they have wild, self-maintaining populations and can out-compete native species. Urban habitats tend to be dominated by alien species of plants, which were introduced in various ways including:

- as seeds contained in soil carried by ships as ballast, which was dumped in a port when the cargo was discharged (this was particularly important before the twentieth century)
- as seeds that contaminated the seedstock of crop plants (this is now less of a problem because weed seeds are “cleaned” from commercial seedstock)
- or as plants used in agriculture, forestry, or horticulture.

Some animals, such as the house mouse (*Mus musculus*) and Norway rat (*Rattus norvegicus*), were introduced accidentally when they escaped from infested ships and cargoes. Others were deliberately introduced. For example, the rock dove (pigeon; *Columba livea*), starling (*Sturnus vulgaris*), and house sparrow (*Passer domesticus*) were introduced by a nineteenth-century society of gentlemen who were dedicated to bringing all of the birds mentioned in the plays of William Shakespeare to North America. Of course, the domestic dog (*Canis familiaris*), domestic cat (*Felis catus*), and other pets are also common non-natives in our cities and towns.

Some other examples of these invasive plants and animals and the ecological problems that are associated with them include the following:

- The dandelion (Taraxacum officinale) is a perennial, herbaceous plant that was originally native to alpine habitat in Europe, but now occurs in temperate regions throughout the world, probably having been introduced in marine ballast. The dandelion is considered an important weed of lawns and pastures.
- Japanese knotweed (Polygonum cuspidatum) is a perennial, herbaceous plant that grows up to 2-m tall and is native to Japan. It has attractive foliage and was widely introduced for horticultural use. It can be invasive in disturbed areas.
- The sticky touch-me-not (Impatiens glandulifera) is an annual wildflower native to the Himalayas. It was introduced through horticulture and can be invasive in gardens and wetlands.

- Purple loosestrife (Lythrum salicaria) is a perennial, herbaceous plant of Eurasia that was introduced through ships' ballast and horticulture and is a serious invader of wetlands.
- St. John's wort (Hypericum perforatum) is a perennial, herbaceous plant of Europe that was introduced probably through marine ballast and is now a serious weed of pastures that causes a photosensitivity disease in cattle.
- Norway maple (Acer platanoides) is a native tree of Europe that was introduced for horticultural use. It invades natural hardwood forest.
- The oriental cockroach (Blatta orientalis) is an insect native to eastern Asia. It was probably introduced accidentally with ship cargo and is a serious pest in homes and other places where food is stored.

Many urban habitats are highly disturbed and are managed to keep them in an early stage of succession (Figure **10.2.4**). Urban lawns, for example, are mown frequently to prevent their vegetation from developing beyond a stage that is dominated by low-growing, herbaceous plants. Moreover, because most lawn-growers want a monoculture of only one or two species of grasses, they may apply herbicide to kill unwanted dicotyledonous plants, such as clovers, dandelion, and other "weeds." The most commonly grown grass in lawns in temperate-zone countries is the Kentucky blue grass (*Poa pratensis*), which, despite its common name, is actually a Eurasian species. Other aspects of the intensive management of lawns include the use of fertilizer, insecticide, and irrigation. Some ecologists recommend less-intensive systems of lawn management, including tolerance of such non-grass plants as red clover (*Trifolium pratense*), black medic (*Medicago lupulina*), and other legumes, which fix atmospheric nitrogen and so reduce the need for fertilizer.



Figure **10.2.4**: Urban biodiversity can be pleasant, but it is often managed in a highly contrived manner and is dominated by alien plants and animals. Almost none of the species in this cemetery in Toronto are native to Canada. Source: B. Freedman.

## Urban Climate

The climate of urban regions is different from that of nearby rural areas. Compared with rural areas, cities typically exhibit the following characteristics:

- they are warmer by 3-6°C
- have 5-10% more cloud cover
- receive about 20% less solar radiation
- have 20-30% lower wind speeds (although wind-tunnelling may increase windspeeds near large buildings)
- have 5-15% lower relative humidity
- receive 5-15% more precipitation The best known of the climatic differences, called the "heat-island" effect, refers to the warmer temperature of cities, which occurs because of the following influences on the urban energy budget:
- an emission of large amounts of heat (thermal energy) from buildings and machines (including motor vehicles)
- interference by buildings with the dispersal of warmed air by wind
- the absorption of solar radiation by dark surfaces (especially asphaltic roads and parking lots ), followed by the re-radiation of long-wave infrared

- a relative lack of plant foliage in urban areas (the atmosphere is cooled by the evaporation of water from foliage, also known as transpiration)

### Socio-economic Impacts

From a social perspective urban sprawl generates greater segregation of residential development according to income. Consequently, it can exacerbate urban social and economic divisions. The socio-economic character of suburban and peripheral areas is typified by middle and upper income families with children, who have the necessary mobility and lifestyle to enable them to function effectively in these localities. However, the suburban experience for other groups, including the young and old, who lack mobility and resources can be very different and can reduce social interaction. Furthermore, large segments of urban society are excluded from living in such areas.

From an economic perspective urban sprawl is at the very least a more costly form of urban development due to:

- increased household spending on commuting from home to work over longer and longer distances;
- the cost to business of the congestion in sprawled urban areas with inefficient transportation systems;
- the additional costs of the extension of urban infrastructures including utilities and related services, across the urban region.

Urban sprawl inhibits the development of public transport and solutions based on the development of mass transportation systems, and the provision of alternative choices in transportation that are essential to ensure the efficient working of urban environments. These conclusions are reinforced by experience from both Munich and Stockholm where the efficient control of urban sprawl and resulting increase in population densities fosters the use of public transport and reduces the growth of car use.

### Social Capital

On the social sustainability side, we can look at social capital otherwise defined as the “connectedness” of a group built through behaviors such as *social networking* and *civic engagement*, along with attitudes such as trust and reciprocity. Greater social capital has been associated with healthier behaviors, better self-rated health, and less negative results such as heart disease. However, social capital has been diminishing over time. Proposed causes include long commute times, observed in sprawling metropolitan areas. As of 2011, according to an article in the Chicago Tribune, Chicago commuting times are some of the worst – with Chicagoans spending 70 hours per year more on the road than they would if there was no congestion – up from 18 hours in 1982. They have an average commute time of 34 minutes each way. These drivers also use 52 more gallons per year per commuter.

## Urban Environmental Problems of the Developing World

Global urbanization reached the 50 percent mark in 2008, meaning that more than half of the global population was living in cities compared to only 30 percent 50 years ago. The access to basic services— clean water, sanitation, electricity, and roads—are some of the main urbanization challenges facing the developing world.

Municipal **waste** management is a crucial service provided by cities around the world, but is often inefficient and underperforming in developing countries. Low income countries face the most acute challenges with solid waste management. In low income countries, cities collect less than half the waste stream. Of this, only about half is processed to minimum acceptable standards. Improper waste management, especially open dumping and open burning, has significant adverse effects on water bodies, air and land resources. People who live near or work with solid waste have increased disease burdens. Unmanaged waste also frequently blocks drainage systems and worsens flooding. Even when collected and transported, waste in dumpsites and landfills contributes to greenhouse gas emissions.

Roughly 2.6 billion people in the developing world are without adequate **sanitation**, and facilities are often overloaded, in disrepair, or unused. Even though the sanitation gap is twice as large as that of water supply, investments in sanitation and hygiene have lagged far behind those in water and other “social” sectors, such as health and education. The main costs of urban sanitation services are those of sewers and sewage treatment. Whereas sewers contribute to public health through reducing everyday contact with sewage (especially by children), wastewater treatment is designed largely to meet ecological objectives and not those of public health. Urban utilities, by and large, are not well designed or staffed to address off-network solutions for water supply or sanitation, yet those solutions are likely to be the most important first steps of progress in environmental health for many of the urban poor.

Experience suggests that demand for car ownership increases dramatically at annual household incomes of $6,000–$8,000. If history repeats itself, an additional 2.3 billion cars will be added by 2050, mostly in developing countries, given expected economic

growth and past patterns of motorization. For instance, in the six largest cities in India, the population doubled between 1981 and 2001, but the number of motor vehicles increased eight times over the same period. Between 2000 and 2013, car ownership in China increased more than six times. Similar trends are seen in other fast growing economies. Increased income levels and the availability of cheaper personal vehicles, coupled with increased travel distances and inadequate public transport systems, have made the personal motorcar an increasingly attractive travel option. The associated health costs are high—in Beijing, the health costs from local **air pollution** are estimated at $3.5 billion annually. In Pakistan, more than 22,600 adult deaths were attributable to urban ambient air pollution in 2005 where air pollution alone causes more than 80,000 hospital admissions per year, nearly 8,000 cases of chronic bronchitis, and almost 5 million cases of lower respiratory cases among children under five.

## References Cited and Further Reading


American Public Transportation Association (APTA). 2014. Public Transportation Fact Book, Appendix A: Historical Tables. APTA, Washington, DC.

Berkowitz, A.R., C.H. Nilon, and K.S. Hollweg. 2002. Understanding Urban Ecosystems. Springer Verlag, Berlin, Germany.

Breuste, J., H. Feldmann, and O. Ohlmann (eds.). 1998. Urban Ecology. Springer Verlag, Berlin, Germany.

Brinkhoff, T. 2015. Major Agglomerations of the World. http://www.citypopulation.de/world/Agglomerations.html

Environment Canada. 1996. The State of Canada's Environment. Government of Canada, Ottawa, ON.

Environment Canada. 1999. Air Quality Trends in Canadian Cities, 1979–1992. Ottawa, ON: State of the Environment Reporting, Environment Canada. https://web.archive.org/web/20010214040646/http://www.ec.gc.ca/pdb/uaqt/aqfact_e.html

Environment Canada. 2007. Municipal population served by wastewater treatment. State of Environment Infobase Ottawa, ON. https://web.archive.org/web/20040621114650/http://www.ec.gc.ca:80/soer-ree/English/Indicators/Issues/Urb_H2O/Bulletin/uwind3_e.cfm

Environment Canada. 2014. Air Quality Indicators. Environment Canada, Ottawa, ON. http://www.ec.gc.ca/indicateurs-indicators/default.asp?lang=en&n=7DCC2250-1

Ewing, B., D. Moore, A. Ourser, A. Reed, and M. Wackernagel. 2010. Ecological Footprint Atlas 2010. Global Footprint Network, Oakland, CA. https://www.footprintnetwork.org/content/images/uploads/Ecological_Footprint_Atlas_2010.pdf

Federation of Canadian Municipalities (FCM). 2009. Waste Diversion Success Stories from Canadian Municipalities. FCM, Ottawa, ON. http://www.fcm.ca/Documents/tools/GMF/Getting_to_50_percent_en.pdf

Fincher, R. and J.M. Jacobs (eds.). 1998. Cities of Difference. Guilford Press, New York, NY.

Freedman, B., S. Love, and B. O'Neil. 1996. Tree species, biomass, and carbon storage in stands of urban forest of varying character in Halifax, Nova Scotia, Canada. Canadian Field-Naturalist, 110: 675–682.

Gill, D. and P. Bonnett. 1973. Nature in the Urban Landscape: A Study of City Ecosystems. York Press, Baltimore, MD.

Global Footprint Network. 2007. Canadian Living Planet Report. World Wildlife Fund Canada, Toronto, ON, and the Global Footprint Network, Oakland, CA. https://web.archive.org/web/20180618150932/https://www.footprintnetwork.org/content/documents/2007_Canadian_Living_Planet_Report.pdf

Holcomb, M.C. 1987. Transportation Energy Data Book, 9th ed. Oak Ridge National Laboratory, Oak Ridge, TN.

Jacobs, J. 1961. The Death and Life of Great American Cities. Random House, New York, NY.

Jacobs, J. 1969. The Economy of Cities. Random House, New York, NY.

Kreith, F. (ed.). 1994. Handbook of Solid Waste Management. McGraw-Hill, New York, NY.

Kryter, K.D. 1985. The Effects of Noise. 2nd ed. Academic Press, Orlando, FL.

Landsberg, H.E. 1981. The Urban Climate. Academic Press, New York, NY.

Langford, T.E.L. 1990. Ecological Effects of Thermal Discharges. Elsevier Science, Amsterdam, The Netherlands.

Lowe, M.D. 1989. The Bicycle: Vehicle for a Small Planet. Worldwatch Institute, Washington, DC. Nathanson, J.A. 2002. Basic Environmental Technology: Water Supply, Waste Management, and Pollution Control. 3rd ed. Prentice Hall, Upper Saddle River, NJ.

National Round Table on the Environment and the Economy (NRTEE). 1998. Greenhouse Gas Emissions from Urban Transportation: Backgrounder. NRTEE, Ottawa, ON.

Platt, R.H., R.A. Rowntree, and P.C. Muick. 1994. The Ecological City: Preserving and Restoring Urban Biodiversity. University of Massachusetts Press Amherst, MA.

Ponting, C. 1991. A Green History of the World. Sinclair-Stevenson London, UK.

Rees, W.E. 1992. Ecological footprints and appropriated carrying capacity: What urban economics leaves out. Environment and Urbanization, 4: 121–130.

Rees, W.E. 1997. Is “sustainable city” an oxymoron? Local Environment, 2: 303–310.

Rees, W.E. 2010. The Human Nature of Unsustainability. The Post Carbon Institute, Santa Rosa, CA.

Savard, J.-P. 1978. Birds in Metropolitan Toronto: Distribution, Relationships with Habitat Features, and Nesting Sites. Unpublished M.Sc. Thesis, University of Toronto.

Statistics Canada. 2006. Human Activity and the Environment. Annual Statistics 2006. www.statcan.ca/english/freepub/16-201-XIE/16-201-XIE2006000.pdf

Statistics Canada. 2006. Road motor vehicle, trailer and snowmobile registration, annual (registrations). Table 405-0001, Statistics Canada, Ottawa, ON. http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=4050001&pattern=automobiles&tabMode=dataTable&srchLan=-1&p1=1&p2=50

Statistics Canada. 2012. Human Activity and the Environment, Waste Management in Canada. Statistics Canada, Ottawa, ON. http://www5.statcan.gc.ca/olc-cel/olc.action?ObjId=16-201-X&ObjType=2&lang=en&limit=0

Statistics Canada. 2017. Road motor vehicle registrations, by type of vehicle. Table 23-10-0067-01 (formerly CANSIM 405-0004), Statistics Canada, Ottawa, ON. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2310006701

Statistics Canada. 2014b. Leading causes of death, by sex. CANSIM Table 102-0561, Statistics Canada, Ottawa, ON. http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/hlth36a-eng.htm Accessed November, 2014.

Timerson, B.J. 1999. A Guide to Noise Control in Minnesota. Minnesota Pollution Control Agency. Minneapolis, MN.

Transport Canada. 2012. Canadian Motor Vehicle Traffic Collision Statistics, 2004. Transport Canada, Ottawa, ON. http://www.tc.gc.ca/eng/roadsafety/tp-tp3322-2004-menu-691.htm

Transport Canada. 2017. Annual Reports: Transportation in Canada. Transport Canada Ottawa, ON. https://www.tc.gc.ca/eng/policy/anre-menu.htm

Turner, K., L. Lefler, and B. Freedman. 2004. Plant communities of selected urbanized areas of Halifax, Nova Scotia, Canada. Landscape and Urban Planning, 71: 191–206.

Wackernagel, M. and W.E. Rees. 1996. Our Ecological Footprint: Reducing Human Impact on the Earth. New Society, Gabriola Island, BC.

## Attribution

Modified by Kyle Whittinghill from the following sources:

- Impacts of Urban Sprawl from *Essentials of Environmental Science* by Kamala Dorsner and CK-12 (CC-BY)
- Urbanization around the World from *Essentials of Environmental Science* by Kamala Dorsner and CK-12 (CC-BY)
- Chapter 12: Urban Ecology from *Environmental Science: A Canadian Perspective* by Bill Freedman (Creative Commons Attribution NonCommercial)

10.2: Environmental Challenges of Urbanization is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

https://bio.libretexts.org/@go/page/76237

# 10.3: Sustainable Cities

## Introduction

Sustainability, from science to philosophy to lifestyle, finds expression in the way we shape our cities. Cities are not just a collection of structures, but rather groups of people living different lifestyles together. When we ask if a lifestyle is sustainable, we're asking if it can endure. Some archeologists posit that environmental imbalance doomed many failed ancient civilizations. What could the sustainable city look like, how would it function, and how can we avoid an imbalance that will lead to the collapse of our material civilization? While attempting to envision the "sustainable city" we must discern what factors will influence its shape and form in the future.

Throughout history settlement patterns have been set by technology and commerce. Civilizations have produced food, clothing and shelter, and accessed foreign markets to purchase and sell goods. Workers traditionally had to live near their place of occupation, although in modern industrial times advanced transportation systems have enabled us to live quite a distance from where we work. In hindsight we can see how reliance on water and horse-drawn transportation shaped historical civilizations and how this equation was radically altered with the rise of the automobile following World War II.

The urban environment is an example of a complex human-natural system. The resiliency of such systems lies in their capacity to maintain essential organization and function in response to disturbances (of both long and short duration). A complimentary view, inspired by traditional ecological and economic thought focuses on the degree of damage a system can withstand without exhibiting a "regime" shift, defined as a transition that changes the structure and functioning of the system from one state to another as a result of one or more independent factors. Upon exceeding a given threshold, the system shifts to a new alternative state which may not be readily reversed through manipulation of causative factors. In the context of human-natural systems, regime shifts can have significant consequences, and not all shifts are preferred by the human component of the system. To the extent that change of some order is a given property of essentially all dynamic systems, "preferred" resiliency might be viewed as the extent to which human societies can adapt to such shifts with acceptable levels of impacts. Resilient infrastructures, then, are those which most readily facilitate such adaptation. Much of the foregoing discussion also applies to sustainability, with the added constraints of the sustainability paradigm: the equitable and responsible distribution of resources among humans, present and future, in ways that do not harm, and ideally reinforce, the social and biological systems upon which human society is based. Although there are important differences between those two concepts, there remains a close interrelationship that stems from the same need: to understand and design urban infrastructural systems that enhance human interactions with the environment.

## So what is the major hindrance for designing/updating for a "sustainable city"?

**Sprawl:** The outward expansion of human populations (typically in low-density residential and commercial development) into the outer edges of cities/towns.

As sprawl expands, it eats at natural and agricultural lands and isolates people. People have to travel further and further to go to work, the grocery store, a hospital, etcetera. What this means is that cities/towns have to find solutions to moving water, waste, power, services, road development/upkeep, etcetera to these communities. As such, there are chronic issues with sprawl on both sides that make opportunity and progress difficult due to economic constraints.

It's important to note that sprawl can happen at different scales:

- **Low-density sprawl:** Small 1-2 story buildings in high densities. Low-density sprawl can be further distinguished by income as neighborhoods tend to be segregated by this. (Ex: North America, (Figure **10.3.1**)
- **High-density sprawl:** High rise buildings in high densities (Ex: China, Figure **10.3.1**)

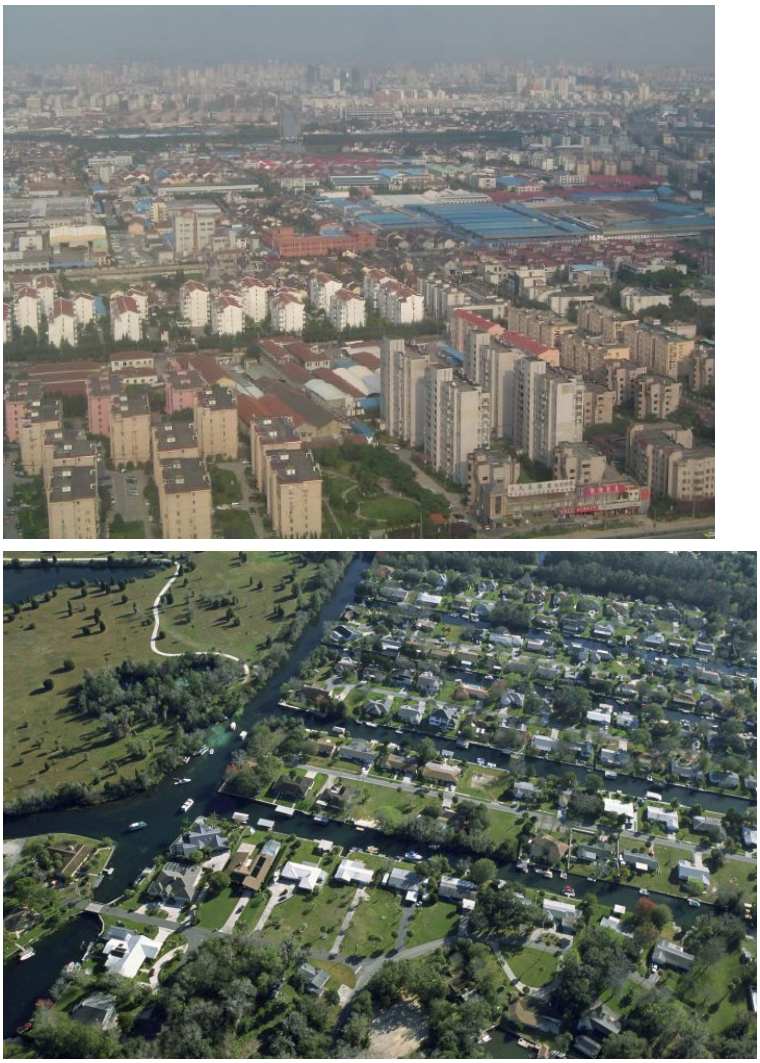


Figure 10.3.1: Left - High density sprawl in Shanghai China by John Patrick Robichaud in Wikimedia Commons (CC-BY), Right - Low density sprawl in North America by Pixnio (Public Domain).

## So what can be done to design/update and move towards “sustainable cities”?

- **Mixed use development:** Lends residential, commercial, cultural, institutional, or industrial uses, where those functions are physically and functionally integrated, and that provides pedestrian connections. Making cities walkable!
- **Geographic equity:** Circulation system defined by ratios of how people move. This means more bikes, walking, public transportation (in many forms) and not cars. Equitability of the streets!
- **Preserve:** Preserve land, culture, and biodiversity as much as possible.
- **Mix**: Mix incomes, cultures, green spaces, public transport options, etcetera as much as possible. Allow people to connect and become neighbors not strangers.
- **Focus:** Match density of a community to meet transit/service needs.

Public space upgrading through **urban renewal** programs will bring people back to the city centre. This will need some strategic thinking about how to use brownfield and greyfield developments and also the adaptive reuse of existing buildings. Remodeling and re-energizing existing city centers to bring about diverse and vibrant communities requires people to move back into downtown areas. This can be achieved through mixed-use urban infill projects, building the "city above the city" by converting low density districts into higher density communities; and by revitalizing underutilized land for community benefit and affordable housing.

**Video 10.3.*a*:** Perspectives: Peter Calthrope describes 7 principles for building better cities.



## Renewable Energy for Zero $CO_2$ Emissions

The various aspects of the principle of renewable energy for zero $CO_2$ emissions include: Energy supply systems and services, as well as energy efficient use and operation, promoting increased use of renewable power, and perhaps natural gas as a transition fuel in the energy mix, but always moving quickly away from heavy fossil-fuels such as coal and oil; and the transformation of the city district from an energy consumer to an energy producer, with local solutions for renewables and the increasing de-carbonizing of the energy supply. The local availability of a renewable source of energy is the first selection criteria for deciding on energy generation. In general, a well-balanced combination of energy sources can sensibly secure future supply. A necessary aim is also to have a distributed energy supply through a decentralized system, utilizing local renewable energy sources. This will transform city districts into local power stations of renewable energy sources, which will include solar photovoltaic, solar thermal, wind (on- and off-shore), biomass, geothermal power, mini-hydro energy and other new technologies. That is to say, there should be on-site electrical generation and energy storage in combination with a smart grid, which integrates local solar and wind generation, utilizing energy-efficiency in all its forms. High building insulation, high energy-efficiency standards and the use of smart metering technology is essential, so that if a part of an office building is not in use, the intelligent building management system will shut down lights and ventilation.

Some of the most promising technologies are in building-integrated photovoltaics, urban wind turbines, micro CHP (combined heat and power) systems and solar cooling. Co-generation technology utilizes waste heat through CHP combined-heat-and-power plants. Energy-efficiency programs are not enough. Too often we find that savings from energy-efficiency programs are absorbed by a rise in energy use. Eco-districts will need to operate on renewable energy sources as close to 100 per cent as possible. As a minimum, at least 50 per cent of on-site renewable energy generation should be the aim of all urban planning, where the energy mix comes from decentralized energy generation and takes into account the resources that are locally available, as well as the cost and the availability of the technology. It is, therefore, essential that the fossil-fuel powered energy and transportation systems currently supporting our cities are rapidly turned into systems that are supplied by renewable energy sources.

Homes use about 23% of all energy in the United States.[27] In the future many of our homes will generate their own power (See **Case Study: A Net-Zero Energy Home in Urbana, Illinois**). Today ultra-efficient homes combine state-of-the-art, energy-efficient construction and appliances with commercially available renewable energy systems, such as solar water heating and solar electricity, so that the net energy use is zero or even less than zero (positive energy production).[28] There have been efforts since the 70's oil crisis to promote (mandatory) energy codes, but voluntary efforts such as Energy Star and LEED are the ones that have made substantial headway.

## Zero-Waste

Sustainable waste management means to turn waste into a resource. All cities should adopt nature's zero-waste management system (See Box below). Zero-waste urban planning includes reducing, recycling, reusing and composting waste to produce energy. All material flows need to be examined and fully understood, and special attention needs to be given to industrial waste and e-waste treatment. We need to plan for recycling centers, for zero landfill and 'eliminating the concept of waste' and better understanding nutrient flows. Eco-districts are neighborhoods where we reuse and recycle materials and significantly reduce the volume of solid

waste and toxic chemical releases. All construction materials as well as the production of goods (and building components) need to be healthy and fully-recyclable.

Waste prevention is always better than the treatment or cleaning-up after waste is formed. Some other systems that need to be put in place are: the remanufacturing of metals, glass, plastics, paper into new products needs to be a routine (without down-grading the product); waste-to-energy strategies are needed for residual waste; and an `extended producer responsibility' clause is needed for all products. In this context of waste, better management of the nitrogen cycle has emerged as an important topic: to restore the balance to the nitrogen cycle by developing improved fertilization technologies, and technologies in capturing and recycling waste. Controlling the impact of agriculture on the global cycle of nitrogen is a growing challenge for sustainable development. Essentially, we need to become (again) a *recycling society*, where it is common that around 60 to 90 per cent of all waste is recycled and composted.

## Water

The various aspects of this principle include, in general, reducing water consumption, finding more efficient uses for water resources, ensuring good water quality and the protection of aquatic habitats. The city can be used as a water catchment area by educating the population in water efficiency, promoting rainwater collection and using wastewater recycling and storm water harvesting techniques. Storm water and flood management concepts need to be adopted as part of the urban design, and this includes storm water run-offs and improved drainage systems and the treatment of wastewater. As part of the eco-district's adequate and affordable health care provisions, it needs to ensure the supply of safe water and sanitation. This includes such things as algae and bio-filtration systems for grey water and improving the quality of our rivers and lakes so that they are fishable and swimmable again. An integrated urban water cycle planning and management system that includes a high-performance infrastructure for sewage recycling (grey and black water recycling), storm water retention and harvesting the substantial run-off through storage, must be a routine in all design projects. On a household level we need to collect rain water and use it sparingly for washing and install dual-water systems and low-flush toilets. On a food production level we need to investigate the development of crops that need less water and are more drought resistant.

Traditional approaches have sought to rapidly move stormwater away from what we've built via gutters, sewers and artificial channels. While this approach on the micro scale is intended to prevent local flooding and undesired ponding, on the macro scale it may actually cause area wide flooding. It also short-circuits the opportunity for water to naturally soak into the ground – to water plants and recharge groundwater resources[18], and, with traditional planting of lawns and other exotics, necessitates bringing more water in for irrigation. Best water management practices for sustainable cities would include:

- Green Roofs (Figure **10.3.2**)
- Downspouts, Rain Barrels and Cisterns
- Permeable Paving
- Natural Landscaping
- Filter Strips
- Bioinfiltration using: Rain Gardens
- Drainage Swales
- Naturalized Detention Basins



Figure 10.3.2: **Green Roof** The green roof of City Hall in Chicago, Illinois *Source: TonyTheTiger [CC-BY-SA-3.0] via Wikimedia Commons*

The sustainable city would recharge its local aquifer and surface water would flow primarily from groundwater and not storm water discharge. Instead of charging a flat tax for storm and sanitary sewer services, technology allows districts to charge usage fees based upon volume, thus providing a financial incentive for sustainable design. Governmental bodies could use these tools to encourage the reorientation and designers could use the techniques outlined above as divert stormwater into the ground rather than directly into surface water.[19]

Some visionaries see even more dramatic transformations in the way we deal with water. Sarah Dunn and Martin Felsen of Urbanlab envision Chicago's evolution into a model city for "growing water" by creating a series of Eco-Boulevards that function as a giant Living Machine – treating the city's waste and storm water naturally, using micro-organisms, small invertebrates such as snails, fish, and plants. Under their plan treated water would be returned to the Great Lakes Basin and create a closed water loop within Chicago, instead of being exported to the Mississippi and Gulf Coast.

## Sustainable Transport: Compact and Poly-Centric Cities

Good access to basic transport services is crucial, as it helps to reduce automobile dependency, as does reducing the need to travel. We need to see integrated non-motorized transport, such as cycling or walking, and, consequently, bicycle/pedestrian-friendly environments, with safe bicycle ways, free rental bike schemes and pleasant public spaces. It is important to identify the optimal transport mix that offers inter-connections for public transport and the integration of private and public transport systems. Some ideas here include: eco-mobility concepts and smart infrastructure (electric vehicles); integrated transport systems (bus transit, light railway, bike stations); improved public space networks and connectivity, and a focus on transport-oriented development. It is a fact that more and wider roads result in more car and truck traffic, and $CO_2$ emissions, and also allows for sprawling development and suburbs that increases electricity-demand and provides less green space. The transport sector is responsible for causing significant greenhouse-gas emissions (over 20 per cent). To combat this effect we need to change our lifestyles by, for example, taking public

transport, driving the car less, or car-pooling. Alternatively, we can ride a bike or walk, if the city district has been designed for it. Personal arrangements have the potential to reduce commuting and to boost community spirit. We want a city district which is well-connected for pedestrians, a city with streetscapes that encourage a healthy, active lifestyle and where residents travel less and less by car.

## Charge Full Costs

The economist Anthony Downs has written that traffic jams during rush hour are a good thing, because they indicate that infrastructure is useful and a lot of people are using it (Downs, 1992). He also notes that building more lanes on a highway is not a solution to congestion, because people who were staying away from the road during rush hour (by traveling at different times, along different routes, or by a different mode) will now start to use the wider road, and it will become just as congested as it was before it was widened. His point is that the road itself is a resource, and when people are using it for free, they will overuse it. If instead, variable tolls were charged depending on how crowded the road was—in other words, how much empty pavement is available—people would choose to either pay the toll (which could then be invested in alternative routes or modes) or stay off the road during congested times. The point is that every car on the road is taking up space that they aren't paying for and therefore slowing down the other people around them; charging a small amount for that space is one way of recovering costs.



Figure **10.3.3**: **Freeway Traffic** Typical congested traffic on an urban freeway – I-80 in Berkeley, California. Residents of U.S. cities typically require automobiles to experience mobility. Note the externalities that the drivers are imposing on others such as air pollution and congestion. The left lane is for car-pooling – as marked by the white diamond – an attempt to address the congestion externality. *Source: By User Minesweeper on en.Wikipedia (Minesweeper)* *CC-BY-SA-3.0*, *via Wikimedia Commons*

Traffic congestion is an example of what economists call **externalities**, the costs of an activity that aren't paid by the person doing the activity. Suburbanites who drive into the city every day don't breathe the polluted air produced by their cars; urban residents suffer that externality. People around the country who use gasoline derived from oil wells in the Gulf of Mexico didn't experience oil washing up on their beaches after the BP disaster in 2010. By charging the full cost of travel via taxes on gas or insurance, we could, for example, pay for children's hospitalization for asthma caused by the cars speeding past their neighborhoods. Or we could purchase and preserve wetland areas that can absorb the floodwaters that run off of paved streets and parking lots, keeping people's basements and yards drier. Not only would this help to deal with some of the externalities that currently exist, but the higher cost of gas would probably lead us to focus on accessibility rather than mobility, reducing overall demand.

## Walkability

Most agree that **walkability** is a key component of any sustainable neighborhood. Walkability not only reduces energy use, but also increases public health. How can we measure walkability? Walkscore.com identifies and measures nearby amenities and provides a rating for specific locations and neighborhoods. Try it for where you live.

Sustainable cities could consist of walkable neighborhoods that separate pedestrian, bike and vehicular traffic and are connected to each other through multiple transportation modes, with biking and mass transit choices in addition to the automobile. Instead of

averaging 10 (auto) trips per unit per day[15] most residents would not use an automobile daily and instead walk or bike or order online. Trip generation would be reduced through telework and ecommerce. Goods would be brought to residents in bulk delivery trucks using services like UPS, Fedex and Peapod. Under telework many would only visit an office to attend meetings once or twice a week (or less). Streets and intersections would be made less daunting to pedestrians through traffic calming techniques.

Most trips to other parts of the city would be made via mass transit. When an individual car is need it would be provided through car sharing, or taxis. Traffic congestion would be relieved not through artificial subsidies (overbuilding roads and providing free parking) but through congestion pricing, removal of free street parking, and providing viable bicycle and mass transit alternatives.

### Bicycles

Unlike motorized transportation, riding a bicycle does not emit air pollutants such as $NO_x$, $SO_2$, hydrocarbons, or particulates. Moreover, bicycle manufacturing requires far less material and energy resources, and less infrastructure is needed to support these vehicles in terms of space for roads and parking. Therefore, bicycles represent an environmentally soft (or "green") alternative to motorized vehicles for personal transportation. There are more than 1-billion million bicycles in the world, about half of them in China. Canadians own about 13 million bicycles, or 0.42 per person. In spite of this number, only 1% of all trips are made using a bicycle, which suggests that many bicycles are not well used. Great environmental benefits would result if more Canadians used bicycles instead of motor vehicles as a means of personal transportation. These benefits include less air pollution, smaller costs of transportation infrastructure, and less use of energy and material resources. Personal benefits include the convenience of bicycle travel and improved health from frequent exercise.

However, bicycle riding can be dangerous, mostly because of the risk of collisions with motor vehicles. About 2% of vehicle-related deaths in North America involve bicyclists (motor-vehicle accidents are the other 98%, but the per-trip and per-kilometre risk of death associated with bicycle use is more than twice as great; Insurance Institute for Highway Safety, 2014). In the United States in 2012, there were 772 bicycle-related fatalities, and Canadian data are about 10% of that. However, bicycle fatalities have decreased by 24% since 1975 because most cyclists now wear a helmet (98% of fatalities involve people without a helmet). Safer bicycle use requires separate lanes along busy roadways, improved awareness by automobile and truck drivers of the need to share the road with bicycle riders (and vice versa), and increased use of safety equipment by bicyclists, especially helmets and lights and reflective devices at night. The use of bicycles in cities can be encouraged in the following ways:

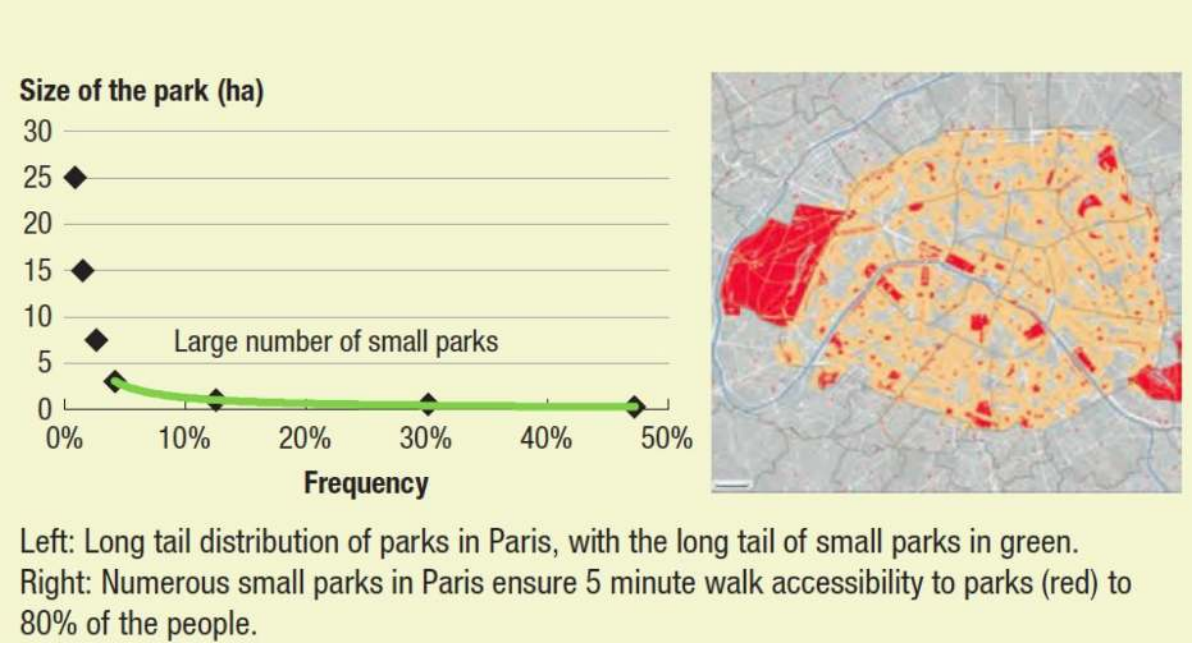
- by developing a network of cycleways (separate paths for bicycles), bike lanes, and paved shoulders on major roads
- by providing secure parking facilities at workplaces and public-transit stations (the latter is known as "bike and ride")
- by having a fleet of bicycles for free use or inexpensive rental from a network of pick-up centres (this system, first used in several European cities, is being tried on a smaller scale in North America, including Montreal)
- by educating people about the advantages of bicycle transport and safety issues.

Other ways to increase bicycle use would be much more difficult to implement. For example, urban-planning options could encourage people to live closer to their workplaces, giving them shorter commutes. In addition, legislation could ensure that people using motorized transport pay the full costs of their choice of transportation by withdrawing subsidies for the construction of roads, bridges, parking lots, and other infrastructure and by applying realistic taxes (such as user fees) to offset the costs of associated air pollution and resource depletion.

## Landscape, Gardens and Urban Biodiversity

A sustainable city takes pride in its many beautiful parks and public gardens. This pride is best formed through a strong focus on local biodiversity, habitat and ecology, wildlife rehabilitation, forest conservation and the protecting of regional characteristics. Ready access to these green spaces: public parks, gardens, with opportunities for leisure and recreation, are essential components of a healthy city. As is arresting the loss of biodiversity by enhancing the natural environment and landscape, and planning the city using ecological principles based on natural cycles (not on energy-intensive technology) as a guide, and increasing urban vegetation. A city that preserves and maximizes its open spaces, natural landscapes and recreational opportunities is a more healthy and resilient city. It needs to maximize the resilience of the eco-system through urban landscapes that mitigate the `urban heat island' (UHI) effect, using plants for air-purification and urban cooling. Further, the narrowing of roads, which calms traffic and lowers the UHI effect, allows for more (all-important) tree planting. Preserving green space, gardens and farmland, maintaining a green belt around the city, and planting trees everywhere (including golf courses), as trees absorb $CO_2$, is an important mission.

Green spaces are an essential feature of energy efficient and livable cities. However, many urban policies concerning green spaces in emerging countries' cities boil down to a predetermined percentage of green space. What really matters is access and proximity of green spaces with diversified social activities instead of proportion only. The spatial distribution of green spaces in Paris results from policies and regulations that have ensured that every citizen lives less than 400 m from a public park, square, or garden (Figure **10.3.4**). This target has been reached whereas green spaces only represent 5 percent of the urban area. There are a few large parks that allow a wide range of activities for residents, and a large number of very successful pocket parks for daily family activities and intergenerational mix (300 green spaces less than ó ha). As a comparison, green spaces in Beijing represent 30 percent of the urban area. Most of the green spaces are very big parks, and the long tail of small parks is lacking. As a result, residents live more than 3 km away from public parks on average.




Left: Long tail distribution of parks in Paris, with the long tail of small parks in green. Right: Numerous small parks in Paris ensure 5 minute walk accessibility to parks (red) to 80% of the people.

Figure **10.3.4**: Greenspace in Paris.

## Local Food and Green Roofs

The sustainable city also needs to introduce inner-city gardens, **urban farming** (Figure **10.3.5**) and **green roofs** (Figure **10.3.2**) in all its urban design projects (using the city for food supply). Other aspects of the local food principle include: local food production; regional supply; and the `eat local' and `slow food' initiatives. The sustainable city makes provision for adequate land for food production in the city, a return to the community and to the allotment gardens of past days, where roof gardens become an urban market garden. It is essential that we bridge the urban-rural disconnect and move cities towards models that deal in natural eco-systems and healthy food systems. The people of the eco-city would garden and farm locally, sharing food, creating compost with kitchen scraps and garden clippings and growing `community' vegetables. Buying and consuming locally will be necessary to cut down on petrol-based transport. Such things as re-using bags and glass containers, paper recycling and the cost of food processing will need reconsideration.



Figure 10.3.5: Some new crops being started, protected by shade cloth barriers to the west. Note the new construction in the background. This area used to be all public housing and are being replaced with mix of market-rate and low-income housing (also called mixed income housing.) The 1.5 acre parcel that City Farm sits on is owned by the City of Chicago and provided, rent-free, to this non-profit initiative. Taken on 1 August 2008 by Linda from Chicago, USA.

USEPA is encouraging urban agriculture as one of the solutions to the shrinking city.[21] It sees urban agriculture as not only providing a use for vacant land, (thus addressing blight and the deleterious affect of neglect on property values) but also as a potential cleanup strategy for contamination. It addresses the problem of food deserts (lack of healthy, affordable, fresh produce) in blighted inner city neighborhoods while educating children and adults about farming and local enterprise. Practitioners have found that urban farming enhances social capital and community connections. Victory gardens produced about 40% of American vegetables consumed during World War II[22] and urban gardens could be a prime user of the compost that we could generate either through individual compost bins or through collective efforts performed on a large scale by our waste haulers.

## Urban Forests

The urban forest is another prominent habitat in cities and towns, although it does not always support native species. Neighborhoods with abundant trees can be studied using the same methods that are used to examine natural forest. One study looked at the urban forest in residential, institutional, and parkland areas of Halifax (Figure 10.3.6 – 7). Older residential neighborhoods had a tree biomass similar to that of natural forest in the region, although the stem density was considerably less (many trees in the mature urban forest are relatively large). Younger neighborhoods had a smaller tree biomass but a higher density, suggesting a pattern of succession in the urban forest. Even an institutional neighborhood, consisting mostly of several hospitals and a university, had a substantial population of trees.



Figure **10.3.6**: Streets with large numbers of mature trees are relatively pleasant, compared with places in which the urban forest is sparse or missing. This residential street in Halifax is mostly forested with alien trees, such as Norway maple (*Acer platanoides*) and linden (*Tilia cordata*). Source: B. Freedman.

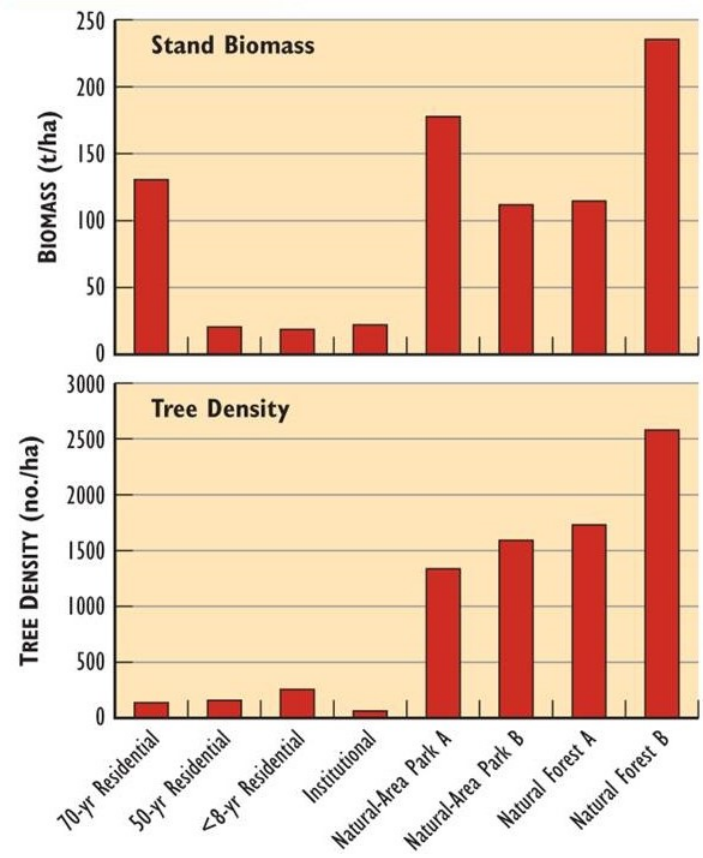



Figure **10.3.7**: Characteristics of the Urban Forest in Halifax. Tree density and biomass are shown for stands of urban forest in different kinds of neighborhoods. The data for natural forest are for rural mature hardwood stands. Source: Freedman et al. (1996).

The urban forest provides many useful ecological services. For example, large amounts of carbon are stored in urban trees, which helps offset some emissions of $CO_2$ from the combustion of fossil fuels. Urban parks and well-treed residential areas can store as much carbon in tree biomass as can a natural forest (Table 25.8). In addition, well-placed trees reduce the wind speed near

buildings, which decreases the air-infiltration rate and thereby helps to conserve heat during cold weather. Trees can also shade buildings, reducing the energy needed for air conditioning during warm weather. Transpiration from tree foliage also helps to cool the ambient urban atmosphere, as does the shading of streets, lawns, and other areas. Urban trees also absorb some air pollutants (such as $SO_2$ and particulates), help to reduce noise levels, and greatly improve outdoor aesthetics. A well-developed urban forest also provides habitat for lower-growing plants and for animals in built-up areas.

Waterfowl are also abundant in urban areas with aquatic habitat, such as beside an ocean or near a lake or river. For example, many cities have wild breeding populations of “giant” Canada goose (*Branta canadensis maxima*), which are descended from birds that were released in those places since the 1950s (Figure **10.3.8**). The geese find that grassy lawns near water provide suitable feeding habitat, and their population in some cities has increased enormously. However, many of these geese have lost the habit of migrating, and some people consider their abundance and year-round presence to be a nuisance. Some cities have attempted to alleviate their “over-population” of geese by capturing animals and shipping them to willing host cities elsewhere. Some cities are culling the geese by killing part of the population. During the spring and fall migrations, some cities with aquatic habitat support large numbers of native waterfowl and other birds. Some waterfowl may remain during the winter if open water is present. Urban birders from Victoria to St. John’s often spend time viewing the numerous native ducks, geese, swans, gulls, shorebirds, and other migrating and wintering birds in local aquatic habitats. Each autumn, for example, the Toronto waterfront provides habitat for thousands of long-tailed ducks (*Clangula hyemalis*) that are migrating from their arctic breeding grounds. St. John’s is the only city in the world where ivory gulls (*Pagophila eburnea*) can be regularly seen.



Figure **10.3.8**: Stanley Park in Vancouver is a greenspace with a mixture of natural and cultural values. During the autumn and winter, geese and other migrating waterfowl use the local ponds and other habitats for feeding and other purposes. Source: B. Freedman.

Some native mammals also occur in urban habitats. Grey (black) squirrels (*Sciurus carolinensis*) and raccoons (*Procyon lotor*) may even inhabit downtown neighbourhoods as long as there is some open forest available. Urban areas with extensive shrubby and forested habitat may sustain white-tailed deer (*Odocoileus virginianus*), red fox (*Vulpes vulpes*), striped skunk (*Mephites mephites*), and coyote (*Canis latrans*). The key to sustaining populations of native animals in urban areas is to maintain as much relatively natural habitat as possible, dominated by indigenous plants. If this is done, even inner-city backyards can provide habitat for some native animals. This function is enhanced by the deliberate naturalization of urban habitats by growing native plants in horticulture.

## Social Equity

Today we recognize that safe, sound and sanitary housing is not just indoor plumbing and more bedrooms, and that affordable housing is not just rent but includes utility and transportation costs and the right to live in a safe, mixed income, stably integrated neighborhood. Our sustainable city should stand upon the leg of social equity and include ethnic and income diversity.

Perhaps the most significant development that separates sustainability from its conservation antecedents is the element of social equity. The environmental and conservation movements have been criticized in the past for being too "white collar" and promoting the interests of the “haves”; these movements have traditionally not dealt with the needs of the underclass in the U.S. and, especially, developing countries. In Agenda 21, the manifesto of the Earth Summit conference on the environment held in Rio de Janeiro in 1992, sustainable development was viewed as the strategy that would be needed to increase the basic standard of living of the world's expanding population without unnecessarily depleting our finite natural resources and further degrading the environment upon which we all depend.[9] The challenge, as viewed at the Earth Summit, was posed in terms of asking humanity to collectively step back from the brink of environmental collapse and, at the same time, lift its poorest members up to the level of basic human health and dignity.

The concept of **ecological footprint** asks each of us to limit resource use to our equitable share. Sitting on the apex of civilization the Western world is being asked to share the earth’s bounty with the masses of Asia, South America and Africa. If technology continues to advance we can do this without significant long-term degradation of our standard of living. Short-term economic dislocations are inevitable, however, as increasing demand from China and India bring us to peak oil and rising transportation costs will highlight the nexus between location efficiency and affordable housing. The Center for Neighborhood Technology, for instance, has mapped 337 metropolitan areas covering 80% of the United States population showing how efficient (near mass transit) locations reduce the cost of living (housing + utilities transportation) and vice versa. The reality of rising transportation costs could have a significant impact on the shape of the city. Looking backward we also realize that racial politics were one of the dynamics that fueled suburban expansion in the 50’s and 60’s decimating many of our urban centers. The Sustainable City of the future, if it works, will have stably integrated mixed income neighborhoods.

## References Cited and Further Reading


American Public Transportation Association (APTA). 2014. Public Transportation Fact Book, Appendix A: Historical Tables. APTA, Washington, DC.

Berkowitz, A.R., C.H. Nilon, and K.S. Hollweg. 2002. Understanding Urban Ecosystems. Springer Verlag, Berlin, Germany.

Breuste, J., H. Feldmann, and O. Ohlmann (eds.). 1998. Urban Ecology. Springer Verlag, Berlin, Germany.

Brinkhoff, T. 2015. Major Agglomerations of the World. http://www.citypopulation.de/world/Agglomerations.html

Downs, A. 1992. Stuck in Traffic: Coping With Peak-Hour Traffic Congestion. Washington, DC: Brookings Institution Press.

Environment Canada. 1996. The State of Canada’s Environment. Government of Canada, Ottawa, ON.

Environment Canada. 1999. Air Quality Trends in Canadian Cities, 1979–1992. Ottawa, ON: State of the Environment Reporting, Environment Canada. https://web.archive.org/web/20010214040646/http://www.ec.gc.ca/pdb/uaqt/aqfact_e.html

Environment Canada. 2007. Municipal population served by wastewater treatment. State of Environment Infobase Ottawa, ON. https://web.archive.org/web/20040621114650/http://www.ec.gc.ca:80/soer-ree/English/Indicators/Issues/Urb_H2O/Bulletin/uwind3_e.cfm

Environment Canada. 2014. Air Quality Indicators. Environment Canada, Ottawa, ON. http://www.ec.gc.ca/indicateurs-indicators/default.asp?lang=en&n=7DCC2250-1

Ewing, B., D. Moore, A. Ourser, A. Reed, and M. Wackernagel. 2010. Ecological Footprint Atlas 2010. Global Footprint Network, Oakland, CA. https://www.footprintnetwork.org/content/images/uploads/Ecological_Footprint_Atlas_2010.pdf

Federation of Canadian Municipalities (FCM). 2009. Waste Diversion Success Stories from Canadian Municipalities. FCM, Ottawa, ON. http://www.fcm.ca/Documents/tools/GMF/Getting_to_50_percent_en.pdf

Fincher, R. and J.M. Jacobs (eds.). 1998. Cities of Difference. Guilford Press, New York, NY.

Freedman, B., S. Love, and B. O'Neil. 1996. Tree species, biomass, and carbon storage in stands of urban forest of varying character in Halifax, Nova Scotia, Canada. Canadian Field-Naturalist, 110: 675–682.

Gill, D. and P. Bonnett. 1973. Nature in the Urban Landscape: A Study of City Ecosystems. York Press, Baltimore, MD.

Global Footprint Network. 2007. Canadian Living Planet Report. World Wildlife Fund Canada, Toronto, ON, and the Global Footprint Network, Oakland, CA. https://web.archive.org/web/20180618150932/https://www.footprintnetwork.org/content/documents/2007_Canadian_Living_Planet_Report.pdf

Holcomb, M.C. 1987. Transportation Energy Data Book, 9th ed. Oak Ridge National Laboratory, Oak Ridge, TN.

Jacobs, J. 1961. The Death and Life of Great American Cities. Random House, New York, NY.

Jacobs, J. 1969. The Economy of Cities. Random House, New York, NY.

Kreith, F. (ed.). 1994. Handbook of Solid Waste Management. McGraw-Hill, New York, NY.

Kryter, K.D. 1985. The Effects of Noise. 2nd ed. Academic Press, Orlando, FL.

Landsberg, H.E. 1981. The Urban Climate. Academic Press, New York, NY.

Langford, T.E.L. 1990. Ecological Effects of Thermal Discharges. Elsevier Science, Amsterdam, The Netherlands.

Lowe, M.D. 1989. The Bicycle: Vehicle for a Small Planet. Worldwatch Institute, Washington, DC. Nathanson, J.A. 2002. Basic Environmental Technology: Water Supply, Waste Management, and Pollution Control. 3rd ed. Prentice Hall, Upper Saddle River, NJ.

National Round Table on the Environment and the Economy (NRTEE). 1998. Greenhouse Gas Emissions from Urban Transportation: Backgrounder. NRTEE, Ottawa, ON.

Platt, R.H., R.A. Rowntree, and P.C. Muick. 1994. The Ecological City: Preserving and Restoring Urban Biodiversity. University of Massachusetts Press Amherst, MA.

Ponting, C. 1991. A Green History of the World. Sinclair-Stevenson London, UK.

Rees, W.E. 1992. Ecological footprints and appropriated carrying capacity: What urban economics leaves out. Environment and Urbanization, 4: 121–130.

Rees, W.E. 1997. Is "sustainable city" an oxymoron? Local Environment, 2: 303–310.

Rees, W.E. 2010. The Human Nature of Unsustainability. The Post Carbon Institute, Santa Rosa, CA.

Savard, J.-P. 1978. Birds in Metropolitan Toronto: Distribution, Relationships with Habitat Features, and Nesting Sites. Unpublished M.Sc. Thesis, University of Toronto.

Statistics Canada. 2006. Human Activity and the Environment. Annual Statistics 2006. www.statcan.ca/english/freepub/16-201-XIE/16-201-XIE2006000.pdf

Statistics Canada. 2006. Road motor vehicle, trailer and snowmobile registration, annual (registrations). Table 405-0001, Statistics Canada, Ottawa, ON. http://www5.statcan.gc.ca/cansim/a26?lang=eng&retrLang=eng&id=4050001&pattern=automobiles&tabMode=dataTable&srchLan=-1&p1=1&p2=50

Statistics Canada. 2012. Human Activity and the Environment, Waste Management in Canada. Statistics Canada, Ottawa, ON. http://www5.statcan.gc.ca/olc-cel/olc.action?ObjId=16-201-X&ObjType=2&lang=en&limit=0

Statistics Canada. 2017. Road motor vehicle registrations, by type of vehicle. Table 23-10-0067-01 (formerly CANSIM 405-0004), Statistics Canada, Ottawa, ON. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2310006701

Statistics Canada. 2014b. Leading causes of death, by sex. CANSIM Table 102-0561, Statistics Canada, Ottawa, ON. http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/hlth36a-eng.htm Accessed November, 2014.

Timerson, B.J. 1999. A Guide to Noise Control in Minnesota. Minnesota Pollution Control Agency. Minneapolis, MN.

Transport Canada. 2012. Canadian Motor Vehicle Traffic Collision Statistics, 2004. Transport Canada, Ottawa, ON. http://www.tc.gc.ca/eng/roadsafety/tp-tp3322-2004-menu-691.htm

Transport Canada. 2017. Annual Reports: Transportation in Canada. Transport Canada Ottawa, ON. https://www.tc.gc.ca/eng/policy/anre-menu.htm

Turner, K., L. Lefler, and B. Freedman. 2004. Plant communities of selected urbanized areas of Halifax, Nova Scotia, Canada. Landscape and Urban Planning, 71: 191–206.

Wackernagel, M. and W.E. Rees. 1996. Our Ecological Footprint: Reducing Human Impact on the Earth. New Society, Gabriola Island, BC.

15 based upon Institute of Traffic Engineers Trip Generation Manual, 1997 from http://www.maine.gov/mdot/planning-process-programs/trips.php accessed 4/25/11

18 Conservation Design Forum from A Guide to Stormwater Best Management Practices, http://www.cdfinc.com/xm_client/client_documents/Chicago_GuideTo_Stormwater_BMPs.pdf, accessed 4/25/11

19 Newport, Robert, USEPA lecture to UPP 594 class 9/7/10

21 http://www.epa.gov/brownfields/urbanag/index.html accessed 4/25/11

22 http://healthfreedoms.org/2011/03/10/food-crisis-triumph-part-2-victory-gardens/ accessed 4/25/11

27 http://energyfuture.wikidot.com/us-energy-use, accessed 4/25/11

28 http://www.energysavers.gov/your_home/designing_remodeling/index.cfm/mytopic=10360, accessed 4/25/11

# Attribution

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

- Environment and Society in a Changing World by PennState OER Initiative (CC-BY-NC-SA)
- The Sustainable City from *Essentials of Environmental Science* by Kamala Dorsner and CK-12 (CC-BY)
- Chapter 12: Urban Ecology from *Environmental Science: A Canadian Perspective* by Bill Freedman (Creative Commons Attribution NonCommercial)
- Sustainable Infrastructure from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (CC-BY). Download for free at CNX.
- The Sustainable City from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (CC-BY). Download for free at CNX.
- Sustainable Transportation: Accessibility, Mobility, and Derived Demand from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (CC-BY). Download for free at CNX.

This page titled 10.3: Sustainable Cities is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 10.4: Case Study- Sustainable Stormwater Management

## Learning Objectives

After reading this module, students should be able to

- describe how stormwater runoff affects water quality in urban watersheds
- explain how stormwater is currently managed in the United States
- analyze some of the conventional and innovative techniques that have been developed to address the water pollution and flood risks associated with urban stormwater runoff

## Introduction

This module reviews some of the complex issues of urban stormwater management. It first examines the hydrological issues affecting the discharge of stormwater runoff to our urban rivers and streams, and then provides an overview of how urban stormwater is managed under the Clean Water Act. After describing the conventional approaches to urban stormwater management, the final section provides an overview of various "sustainable" strategies, especially the use of "green infrastructure," that can be considered to reduce the water pollution and flooding risks generated by urban stormwater runoff.

## The Hydrological Context of Urban Stormwater

**Stormwater runoff** (or overland flow) is the portion of precipitation reaching the ground that does not infiltrate into soils, is not taken up and transpirated by plants, nor is it evaporated into the atmosphere. It is an especially important component of the hydrological cycle in urban areas, since it can cause both pollution and flooding risks to nearby waterways and their adjacent communities. It should also be noted that many of the current models of global climate change predict changes in the hydrological cycle in the future. They predict many more severe storms likely in parts of the Midwest as a result of the moisture and energy in the atmosphere increasing over the next century because of increasingly higher concentrations of greenhouse gases. Higher frequencies of more severe storms are likely to further increase the pollution and flooding risks posed by stormwater runoff, especially in urban areas (USGCRP, 2009).

Current strategies to manage these risks employ the concept of a **watershed** – the variations in natural topography that cause both surface water and surficial ground water to flow downhill towards lower-lying areas or points of discharge, usually to a stream or river. Watershed boundaries are defined topographically by mapping variations in land elevations around waterways that create hydrologic divides between adjacent watersheds and between sub-watersheds. The amount of stormwater that ends up as runoff within a watershed not only depends on the intensity and amount of precipitation reaching the ground in the form of rain or snow, but also on the characteristics of the watershed itself. State and federal environmental protection agencies have developed a number of sophisticated hydrological simulation models that enable the amount and characteristics of stormwater runoff (in terms of its volume and the pollutant load that would be carried by the stormwater to rivers and streams within the watershed) to be forecasted. They forecast this based on historical estimates of the amount of precipitation entering the watershed, the characteristics of a watershed's terrain and soils, the amount and location of impermeable surfaces associated with the development of the watershed, and the extent and types of ground cover within the watershed's drainage area (NRC 2008, Appendix D). A change in any of these factors will affect the amount and extent of flooding and water pollution attributable to the discharge of stormwater runoff into a river or stream.

Since the pattern of precipitation varies seasonally the water pollution and flooding risks posed by stormwater runoff also tend to vary seasonally. Generally, larger flood and pollution risks will occur in the spring, when rapid snowmelt can generate a lot of runoff volume (especially if the ground is still frozen), which can carry pollutants that have accumulated within the snow cover over the winter months to nearby streams and rivers. There can also be storm-related flood and pollution "spikes" when heavy rain strikes the ground at a faster rate than it can be infiltrated into the soils, or when it is prevented from infiltrating into the soils by roofs, paving, or other impermeable surfaces. This initially high volume of stormwater runoff can carry greater amounts of contaminants – a process often described as the **"first flush" phenomenon**. Usually, the first half-inch of stormwater will be carrying the highest pollution load, so its capture and management becomes a priority for water quality protection.

How some of these features, especially the amount of impervious surface associated with different densities of development, affect the generation of urban runoff are illustrated in Figure **10.4.1**. Research by the Center for Watershed Protection has found that stream quality becomes impaired when 10% of the stream's watershed is impervious and that an urban stream's ecology is severely impacted when more than 25% of its watershed is impervious.

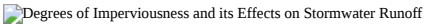



Figure **10.4.1**: **Degrees of Imperviousness and its Effects on Stormwater Runoff** These four images show increasing amount of stormwater runoff as the area becomes developed with more impervious surfaces. *Source: In Stream Corridor Restoration: Principles, Processes, and Practices (10/98) By the Federal Interagency Stream Restoration Working Group (FISRWG) (15 Federal agencies of the U.S.)*

When flowing downhill within a watershed, stormwater runoff can pick up pollutants from various anthropogenic sources and activities. It can also collect pollutants from the atmospheric deposition of particulates and air pollutants carried to the earth's surface by precipitation, by windblown dust, or by simply settling out of the atmosphere. Urban runoff can also dissolve or transport chemicals that may be found naturally in soil or nutrients which may have been deliberately added to lawns. Common urban pollutants can include such things as pesticides and fertilizers applied to residential lawns, parks and golf courses, enteric microbes from animal waste, industrial chemicals that may have been accidentally spilled on the ground or improperly stored, or oils and greases leaking from cars parked in lots or on driveways.

As stormwater runoff flows towards lower-lying areas of the watershed, it carries these contaminants with it and therefore contributes to the pollution of the stream, river or lake into which it is discharging. Once it reaches a river or stream, the concentrations of pollutants in the receiving waters are naturally reduced as the contaminants are carried downstream from their sources, largely through dilution but also by settlement, by uptake by posure to sunlight and oxygen, and by interactions with various chemical and physical proplants and animals (including bacteria and other microorganisms), through degradation by excesses occurring within the waterway and its streambed.

## Regulating Urban Runoff

Water pollution risks within watersheds are managed under the federal Clean Water Act, which requires state environmental protection agencies to regulate the discharge of pollutants into navigable waterways and waterbodies pursuant to federal guidelines (NRC, 2008). The Clean Water Act employs maximum concentration standards for common pollutants that can impair the recreational or ecological functions of a river or stream. One class of polluters regulated under the Clean Water Act consists of those that are directly discharging pollutants into a waterway from an industry or sewage treatment plant through a pipe, ditch, outfall or culvert – these are called **point sources**.

Point sourcesare managed under the Clean Water Act by the requirement that each direct source have a renewable discharge permit, called a National Pollution Discharge Elimination System (NPDES) permit. NPDES permits set limits for the various pollutants being discharged by that source based on the **ambient water quality** of the waterway and its proposed use (e.g. its use as a public water supply source, or for fishing, or recreational use). The other regulated class of polluters managed under the Clean Water Act consists of those sources that introduce contaminants into a waterway through overland or subsurface flow – these are called **non-point sources**, and include most of the water pollution loads carried by urban stormwater runoff.

Since the 1970s, the principal approach used by state and federal environmental protection agencies to control water pollution is to try to simply reduce the quantity of pollutants being released into our rivers and streams (NRC, 2008). NPDES permits control the direct discharge of contaminants into our waterways, while non-point sources are managed through Best Management Practices (BMPs) that are designed to limit the amount of pollutants released into a watershed, where they could later be carried by stormwater runoff or by groundwater flow to a receiving stream or river. Depending on the pollutant of concern, BMPs could be as simple as requiring pet owners to clean up after their pets or as complex as requiring that industries using toxic materials design, construct and manage loading and storage areas in order to keep spilled materials from being transported off-site by stormwater or groundwater flow. BMPs can even include encouraging some industries to change their production processes in order to reduce the total amount of toxic materials they use, a pollutant reduction strategy known as **pollution prevention** (since the fewer toxics used, the lower the risk that they will inadvertently be released into the environment).

The strategy of simply reducing the amount of pollutants entering the environment is complicated by the fact that many of the non-point pollutants are not amenable to management through local BMPs. For example, agricultural activities are expressly exempted from the Clean Water Act, even though stormwater runoff from farms and animal feedlots can carry agricultural chemicals, fertilizers and manure into adjacent waterways, along with topsoil from freshly-plowed fields. Pollutants could also be introduced into an urban watershed by the deposition of air pollutants. Airborn particulate matter, for example, can be transported very long distances by the wind, making most locally administered BMPs (except possibly instituting regular street-sweeping programs) ineffective in reducing the distribution and quantities of these types of urban stormwater pollutants.

In response to these challenges, the Clean Water Act was amended to require state environmental protection agencies to calculate pollution budgets for the impaired segments of their streams and rivers. The "impaired segments" were those reaches of a stream or river that did not meet the water quality standards for their intended uses. Models were used to calculate the "total maximum daily load" (TMDL) of pollutants entering the waterway through both point and non-point sources that would enable the stream segments to achieve their highest proposed use. The Clean Water Act's new TMDL program provides a more sophisticated framework for evaluating the impacts of non-point pollution on water quality. However, given the limitations of trying to put more and better BMPs into place, environmental protection agencies have begun to refocus some of their attention from reducing the total amount of pollutants being released within a watershed to also reducing the amount of stormwater runoff.

Environmental protection agencies have developed strategies for urban stormwater management that involve modifying a development site so that more precipitation would be retained on-site rather than flowing off of it into nearby waterways or waterbodies. These stormwater retention strategies initially stressed traditional engineering solutions, such as installing a stormwater collection system that temporarily stores the stormwater on-site in order to reduce the rate and amount of stormwater being released to a waterway. The strategies were later expanded to include various site modifications, such as constructing vegetated buffer strips or **swales** (ditches),in order to encourage more stormwater to infiltrate into the ground.

Reducing the volume of urban stormwater leaving a site as runoff also offers an additional hydrologic benefit in urban watersheds – reducing flood risks (NRC 2008). Besides having the potential to carry pollutants, stormwater runoff discharge increases the amount of water entering into a lake, stream or river, increasing both the water volume and flow velocity of the waterway. A relatively large amount of stormwater runoff entering a waterway over a relatively short time can quickly raise a stream's water levels beyond its banks, causing flooding that could threaten adjacent development. Stormwater contribution to a river or stream can also increase the velocity of the stream's flow, causing increased channel and bank erosion, undercutting or damaging dikes, levees and other water control structures, and scouring the stream or river bed. Stream edge or streambed erosion can impair water quality by increasing the cloudiness (or turbidity) of the waterway, which can also damage aquatic and riparian habitats.

Stormwater-induced flood risks are managed by the National Flood Insurance Act, where hydrologic models (adjusted by historical flood events) are used to forecast the potential flooding caused by a 100-year storm (a storm that has a one percent chance of occurring in any given year). The Act forces financial institutions to require homeowners within the designated 100-year floodplains to purchase flood insurance in order to get a mortgage, with the federal government subsidizing the insurance premiums if the community adopts a flood management program restricting development from extremely hazardous areas and instituting building code changes to lessen flood damage.

In assessing flood risks, it is important to realize that managing the volume and rate of urban stormwater being discharged from developed areas does not affect the total amount of stormwater that is being discharged to a river or stream within a watershed – they only affect the timing of when a storm's precipitation will be discharged to the waterway (NRC, 2008). Both the conventional

and the newer, more sustainable, ways of managing stormwater discussed below seek to delay the time it takes for stormwater runoff to reach a waterway in order to reduce the water levels and flow velocities of the receiving streams after a storm. Slowing the rate by which stormwater is being contributed to a stream spreads out the peak of the resultant flood levels over a longer time period, allowing many flood risks to be substantially reduced.

## Conventional Stormwater Management

Urban stormwater is traditionally managed by the construction of engineered stormwater facilities, such as storm sewers and detention basins, as part of the land development process. These engineering processes are specifically designed to modify the natural **hydrology** of a site. For example, when land is being developed, the parcel is usually graded for development and stormwater infrastructure is installed to channel the stormwater from individual lots into a separate stormwater sewer system connected to a detention basin where it is retained until it can be discharged off-site. Site preparation also includes elevating building sites so that they are constructed on slightly elevated "pads" to encourage stormwater to flow away from building foundations and toward the streets. After reaching the street, stormwater is then directed to the stormwater sewers by curbs and gutters.

Conventional stormwater detention facilities were historically built to reduce off-site flood risks, and were not expressly designed to reduce off-site water pollution risks. Any stormwater detention that was provided was only temporary, often providing an insufficient retention time to allow the natural attenuation of any pollutants that were carried by the runoff into the detention basin – unlike the natural attenuation processes occurring in a river or riparian wetland (where ambient pollution levels are gradually reduced through dilution, oxidation, chemically binding to rocks and soils, being gobbled up by microorganisms, etc.). Stormwater is usually detained on-site after a storm only for a period of hours or, at most, days and then released to a waterway. Some of the particulate contaminants in the stored runoff might settle out if they are large or heavy enough to do so during that short time, some might infiltrate into the soils in the bottom of the detention basin, and some pollutants might be taken up by grass lining the basin, but many pollutants still end up being carried into the waterway along with the released stormwater.

Since the 1990s, environmental protection agencies have begun to consider the water pollution impacts of releases from stormwater detention facilities, after the Clean Water Act was amended to require states to treat stormwater discharges from detention basins as a type of direct source and to require that NPDES permits be phased in for discharges from Municipal Separate Stormwater Sewer Systems ("MS4") in cities and urban areas above certain population thresholds (NRC, 2008). The NPDES permits issued under the U.S. Environmental Protection Agency's (U.S. EPA) MS4 program now require the water pollution loads from stormwater detention basin discharges to be assessed through the creation and adoption of local stormwater management plans and that the contaminants carried by the stormwater runoff to the basins for later re-release to a waterway be better managed and reduced through the adoption of local BMPs. MS4 permit regulations issued by state environmental protection agencies usually involve the issuance of a "general permit" by the agency, applying to all applicable Municipal Separate Stormwater Sewer Systems located within the state’s designated urban areas.

### Stormwater Sewer Systems Located within the State's Designated Urban Areas

A different set of stormwater management issues arise in older urban areas that are already developed. Most of the United States' older cities and suburbs, especially those established in the late-19$^{th}$ and early 20$^{th}$ centuries, do not have Municipal Separate Stormwater Sewer Systems. Instead, they have what are known as **combined sewer systems** – sewers that carry both the stormwater runoff from paved streets and the wastewater (sewage) from homes, stores and factories. These combined sewers transport the mixed wastewater and stormwater to municipal sewage treatment plants where the diluted sewage is treated and then discharged to a waterway under an NPDES permit (NRC, 2008).

Water quality problems arise when rainstorms deposit more precipitation in the city than can be handled by the sewage treatment plant. As the diluted wastewater begins to fill up the combined sewer system at a faster rate than it can be treated, the sewage treatment plant operators are faced with a difficult choice – they can either allow the diluted sewage to continue to back up in the sewers, eventually flooding residents' basements (a politically unpopular as well as unhealthy option), or they can allow the diluted wastewater to bypass the sewage treatment plant and be discharged directly into the waterway, with the untreated wastewater's pollutant levels usually exceeding the limits set forth in the plant's NPDES permit. Most treatment plant operators choose the more politically acceptable option of releasing the wastewater in violation of their NPDES permit, creating water pollution incidents called **combined sewer overflows (CSOs)**.

## Strategies to Manage CSOs

CSO problems are very difficult and expensive to resolve in older cities. One approach to managing stormwater off-site is to tear up the city's streets, digging up the old combined sewers and replacing them with separate stormwater and wastewater sewer systems. The high costs of retrofitting new separate sewer systems are often prohibitively expensive, especially in these times of stressed state and local budgets. Moreover, the extensive traffic disruptions involved in replacing most streets would not make this a politically popular choice.

A second approach to managing CSO issues off-site in developed areas is to keep the combined sewer system, but to construct a reservoir system large enough to store the diluted wastewater until it can be treated by the sewage treatment plant. This is the approach used by both the City of Milwaukee, Wisconsin and by the Metropolitan Water Reclamation District of Greater Chicago in its Tunnel and Reservoir Plan, or TARP. Although most of TARP has been built, all of the reservoirs have not yet been completed because of federal budgetary cutbacks. The tunnels themselves and one reservoir are currently able to temporarily store the combined sewage and the runoff from only the first 3/8-inch (.95 cm) of rain falling in the Metropolitan Water Reclamation District's service area. The extremely high expense of installing such a supplementary sewage and stormwater storage system would make it unaffordable to most cities unless very substantial federal and state grants are provided.

A third way to address CSO issues off-site is to use the streets themselves to temporarily store stormwater by installing low speed bump-like structures at intersections and by restricting the streets' sewer intakes to the combined sewer system (US EPA, 2000). This urban retrofit strategy would allow stormwater to flow from lots into the streets, which would flood up to their gutter tops during heavy storms, functioning as stormwater reservoirs. The stored stormwater would then slowly be discharged to the combined sewers through the restricted grates over a period of hours after the storm, reducing the amount of diluted sewage flow to a quantity that could be adequately treated by sewage treatment plants. The flooding of streets, impairing automobile access, and the possibility of stormwater overflowing the curbs and damaging parked cars and adjacent property during very heavy rainstorms may not make this a politically popular option, though.

## Managing Urban Stormwater More Sustainably

There is a fourth approach to dealing with CSO problems, which involves intercepting and delaying the discharge of precipitation from a parcel of land before it flows off-site to a separate or combined sewer system, or to an adjacent waterway. Encouraging on-site storage or infiltration reduces the stormwater contribution to a combined sewer's flow in developed areas, thereby reducing the amount of diluted wastewater being generated and enabling combined sewer systems to better handle their wastewater loads during rainstorms. These decentralized on-site approaches to managing stormwater could also be used to reduce the amount of conventional stormwater infrastructure needed in new developments using separate stormwater sewer systems. Because these on-site approaches are less resource-intensive and more cost-effective than conventional stormwater management approaches, they are also more sustainable investments.

On-site stormwater management techniques are also often known as "green infrastructure" (Jaffe et al., 2010). Development projects using "green infrastructure" for urban stormwater management are commonly known as "**Low Impact Developments**." Low Impact Development projects using green infrastructure usually allow stormwater to be managed at lower costs than by using conventional detention practices (US EPA, 2007).

There are essentially three strategies for on-site stormwater management: (1) techniques that encourage the infiltration of stormwater into soils to reduce its volume before it reaches a sewer system, or which employ more selective grading and the planting of vegetation to reduce its rate of flow from the site; (2) techniques that encourage the temporary storage of stormwater on-site, instead of transporting it off-site for centralized detention within a development project or a municipality; and (3) techniques, such as the construction of artificial wetlands, which also allow some degree of longer-term retention and treatment of the stormwater by natural processes before it is discharged. Infiltration techniques might also provide some water treatment capabilities due to the longer retention times of groundwater before discharge, but the degree of such treatment would largely depend on soil characteristics, the amount of overlying vegetation and the depth of the soil's unsaturated zone.

## Increasing Stormwater Infiltration

Techniques to decrease the volume of stormwater runoff and to reduce the rates at which it is discharged include the use of permeable paving and the construction of "rain gardens" and vegetated swales (see Figure 10.4.2). Permeable paving uses materials which are specially formulated to have air voids in their matrix, allowing water to flow into and through the paving materials after they are installed. It also includes the more common installation of precast porous pavers that are designed with holes through their surfaces, allowing stormwater to flow through their holes into the soils beneath them. Permeable paving needs to be periodically maintained because its pores can be clogged by fine grains of topsoil or with other small particles (such as soot from atmospheric deposition) carried along by the runoff. Maintenance includes periodically sweeping or vacuuming the paving to control the build-up of clogging particles.

.

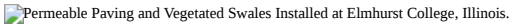



Figure 10.4.2: **Permeable Paving & Vegetated Swales** Permeable paving drains into a vegetated swale as part of Elmhurst College's (in Illinois) parking lot's "green" stormwater management system. *Source: Jaffe, M., et al. (2010), Fig. 14, p. 117.*

"Rain gardens" can also be used to encourage stormwater to infiltrate into the soils, where it can be taken up by plants and transpired to the atmosphere, evaporated from the soils, or allowed to infiltrate deeper into the soils to become groundwater. Rain gardens are created in areas of low-lying terrain that are expressly designed for, or engineered with, well-drained soils and are usually planted with deep-rooted **native vegetation** that often can survive the drier soil conditions between rains. Rain gardens can be quite effective in intercepting and infiltrating stormwater being discharged from roofs, with roof downspouts directing the discharge of stormwater into a rain garden instead of allowing it to flow across the lot and into the street sewer system. Some native vegetation, however, may have special maintenance requirements, such as the periodic burning needed to manage some prairie plants.

Vegetated ditches or swales can also be used to transport stormwater runoff to a conventional stormwater management system, with the vegetation planted in the ditch slowing the rate of stormwater flow while also allowing a portion of the runoff to be infiltrated into the soils or taken up by plants. In many cases, vegetated swales and rain gardens can provide less-expensive alternatives to the installation of separate stormwater sewer system, since it reduces the need for the construction of street gutters, grates, street catchment basins and sewer pipes (US EPA, 2007). Interception of the stormwater by infiltration and plant uptake in a rain garden or vegetated swale may also reduce the amount, capacity and size of the sewers that would have to be built to manage a predicted volume of stormwater, if these green infrastructure techniques are used to supplement a conventional stormwater collection system.

## Increasing Interim On-site Storage

Sustainable management techniques that can temporarily store stormwater on-site until it can be released off-site to a sewer system or to conventional stormwater detention facilities include the use of "green roofs" and **rain barrels** connected to roof downspouts. Rain barrels allow precipitation to be collected and stored, and then used for non-potable purposes (lawn irrigation, for instance) allowing the captured stormwater to substitute for more expensive, treated water (Figure 10.4.3).

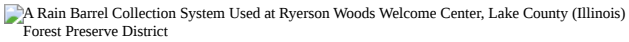

A Rain Barrel Collection System Used at Ryerson Woods Welcome Center, Lake County (Illinois) Forest Preserve District

Figure **10.4.3**: **A Rain Barrel Collection System** This "green" building (Ryerson Woods Welcome Center, Lake County (Illinois) Forest Preserve District )uses both a rain barrel to collect stormwater draining from the roof, and a rain garden to help infiltrate precipitation. *Source: Jaffe, M., et al. (2010), Fig. 12, p. 116.*

A **green roof** is a flat roof surface that uses amended soil materials installed above a layer of waterproof roofing materials to allow shallow-rooted plants to be planted. While still being an impermeable feature of a development site (because of its waterproof layer), a green roof can temporarily store rainwater before it is discharged to the ground by the roof gutters and downspouts (Figure **10.4.4**). Just as a rain barrel can store (and re-use) a portion of the stormwater precipitation being discharged from impervious roofs, the soils of a green roof can capture and temporarily store stormwater precipitation as the pores between the soil particles fill up with rainwater. Green roofs can even partially reduce the runoff's pollution load through plant uptake and by other biological and physical processes within the roofs' soil materials while they are saturated. Because of the need to both water-proof the roof while installing a biological system on top of it, green roofs tend to cost more than conventional roofs, even ignoring the additional structural engineering that might be necessary to accommodate the weight of the green roof's soil and plantings.

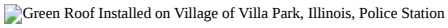

Green Roof Installed on Village of Villa Park, Illinois, Police Station

Figure **10.4.4**: **A Green Roof** The green roof on this police station in Village of Villa Park, Illinois has shallow-rooted plants placed in a thin layer of growing medium installed on top of a waterproof roof membrane. *Source: Jaffe, M. et al. (2010), Fig. 13, p. 116.*

The stormwater management benefits of rain barrels and green roofs depend on their storage capacity relative to the amount of impervious surface area of the roof with which they are associated. Rain barrels might be able to capture only a fraction of an inch of the stormwater falling on a roof and being discharged from a downspout, while several inches of amended soils on a rooftop might be able to store substantially more precipitation before it evaporates, is taken up by the roof's plants, or is discharged from the green roof via its gutters and downspouts. In both cases, however, the interception and temporary retention of stormwater by these green technologies may allow conventional stormwater management systems to function more efficiently by reducing the amount of stormwater being discharged into the systems. They would also certainly reduce some of the "**peakiness**" of stream flooding by being able to temporarily store and then release stormwater from impermeable roof surfaces later after a storm event.

## Treating Urban Stormwater

Some sustainable stormwater management approaches have the potential to actually treat the water to remove pollutants as well as control its volume and rate of discharge. These strategies include constructing wetlands and planting trees. Wetlands have proven to be very effective in both temporarily storing stormwater runoff and reducing flooding risks, while also reducing the pollutant load carried to the wetland (because of its high biological activity that can capture and degrade the contaminants). As a result, the

federal government has adopted a "no net loss" policy with respect to protecting existing wetlands. Section 404 of the federal Clean Water Act requires that the U.S. Army Corps of Engineers (under U.S. EPA oversight) review any proposals to fill or damage any wetlands that are directly hydrologically associated with navigable waterways. Any actions affecting existing wetlands will need a Corps 404 permit in addition to any local or state approvals.

Besides preserving existing wetlands, new wetlands can also be designed, created and maintained as part of a "green" stormwater management strategy (NRC, 2008). The constructed wetland can be designed and used to intercept, temporarily store and treat stormwater runoff before it is released to a stream or river. Water control structures are also usually installed to ensure that the constructed wetlands remain flooded for long enough periods of time to support wetland vegetation. If appropriate plants are selected, they can also provide important habitats. Wetland maintenance involves the control of invasive plant species (e.g. Purple Loosestrife) and the management of any sediment that can be carried by stormwater runoff into the wetland, since the sedimentation of wetlands can fill them in, impairing their ecological and treatment functions.

The planting of trees is an especially valuable strategy to manage urban stormwater, especially when the trees become mature. Tree canopies break rain velocity, reducing runoff flow rates, while tree roots can stabilize soils against being eroded by urban runoff. Tree canopies reduce temperatures, mitigating urban heat island effects, by providing shade and through their transpiration processes. Their leaves and roots can also capture some stormwater contaminants and provide carbon sequestration to reduce climate change impacts. Moreover, trees provide a valuable soil amendment as their fallen leaves decay into mulch, improving the infiltration rate and biological activity of surrounding soils, while larger broken branches falling into urban streams can slow stream velocities and provide improved riparian and aquatic habitat. The shading of streams by riparian trees is particularly important in ensuring that a stream's ecological functions remain resilient in the face of rising temperatures caused by global climate change.

## Conclusions

All of the green infrastructure and Low Impact Development techniques that provide interim on-site stormwater storage to reduce flood risks can also provide some pollution removal capabilities, as well. The American Society of Civil Engineers and U.S. EPA maintain an International Stormwater BMP Database of development projects using green infrastructure. This on-line resource reviews the effectiveness of various stormwater management practices and makes these sustainable techniques more accessible to local officials and municipal public works departments charged with managing stormwater runoff in their communities.

There is increasing public interest in using sustainable stormwater management techniques to replace or supplement conventional stormwater facilities. The U.S. federal government, for example, is now requiring that green infrastructure be used in all federal projects above a certain size to manage urban stormwater runoff. Local officials are also showing a greater interest in these sustainable approaches, since they are often less expensive to install and maintain over their life-spans than conventional stormwater sewer systems and detention facilities. Finally, state governments are beginning to set aside money in their revolving loan funds for public infrastructure that is earmarked for green infrastructure projects. It is likely that this interest in sustainable urban stormwater management will continue to grow.

## Resources

For more information about the:

- Clean Water Act, visit http://www.epa.gov/agriculture/lcwa.html.
- International Stormwater BMP Database, visit http://www.bmpdatabase.org.
- Metropolitan Water Reclamation District of Greater Chicago, visit http://www.mwrd.org/irj/portal/anonymous/tarp.
- Milwaukee Metropolitan Sewerage District, visit http://v3.mmsd.com/DeepTunnel.aspx.
- National Flood Insurance Act, visit http://www.fema.gov/library/viewRecord.do?id=2216.

## References

Gulliver, G.S. & Anderson, J.L. (eds.). (2008). Assessment of Stormwater Best Management Practices. *Stormwater Management Practice Assessment Study*. Minneapolis: University of Minnesota.

Jaffe, M., Zellner, M., Minor, E., Gonzalez-Meler, M., Cotner, L., Massey, D., Ahmed, H., Elbert M., Wise, S., Sprague, H., & Miller, B. (2010). *Using Green Infrastructure to Manage Urban Stormwater Quality: A Review of Selected Practices and State*

*Programs*. Springfield, IL: Illinois Environmental Protection Agency. Retrieved June 23, 2011 from http://www.epa.state.il.us/green-infrastructure/docs/public-act-recommendations.pdf

National Research Council. (2008). *Urban Stormwater Management in the United States*. Washington, DC: National Academies Press. Retrieved June 23, 2011 from http://www.epa.gov/npdes/pubs/nrc_stormwaterreport.pdf

U.S. Environmental Protection Agency. (2000, October). *Street Storage for Combined Sewer Surcharge Control: Skokie and Wilmette, Illinois (Factsheet).* (EPA Publication No. EPA-841-B-00-005C). Washington, D.C. Retrieved May 17, 2011 from http://www.lowimpactdevelopment.org/pubs/Street_Storage_Factsheet.pdf

U.S. Environmental Protection Agency. (2007, December). *Reducing Stormwater Costs through Low Impact Development (LID) Strategies and Practices*. (EPA Publication No. EPA 841-F-07-006). Washington. D.C. Retrieved June 23, 2011 from http://www.epa.gov/owow/NPS/lid/costs07/documents/reducingstormwatercosts.pdf

U.S. Global Climate Change Research Program (USGCCRP). 2009. *Global Climate Change Impacts in the United States*. Cambridge: Cambridge University Press. Retrieved May 18, 2011 from http://downloads.globalchange.gov/usimpacts/pdfs/climate-impacts-report.pdf

# Attribution

Modified by Kyle Whittinghill from the following sources:

- Sustainable Stormwater Management from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (CC-BY). Download for free at CNX.

10.4: Case Study- Sustainable Stormwater Management is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 11: Solid Waste Management

This page titled 11: Solid Waste Management is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 11.1: Waste Generation

In natural systems, there is no such thing as waste. Everything flows in a natural cycle of use and reuse. Living organisms consume materials and eventually return them to the environment, usually in a different form, for reuse. **Solid waste** (or trash) is a human concept. It refers to a variety of discarded materials, not liquid or gas, that are deemed useless or worthless (figure **11.1.*a***). Humans modify natural substances, making them more difficult to breakdown, or store matter under conditions that slow its degradation. However, what is worthless to one person may be of value to someone else, and solid wastes can be considered misplaced resources. Learning effective ways to reduce the amount of wastes produced and to recycle valuable resources contained in the wastes is important if humans wish to maintain a livable and sustainable environment.



Figure **11.1.*a***: Trash is a uniquely human concept. Image by Piqsels (public domain).

## Types of Waste

There are many subcategories of waste (figure **11.1.*b***). **Biodegradable** waste can be broken down by microbes whereas **nondegradable** waste does not readily break down. **Hazardous waste** is trash that presents a health risk. Specifically, hazardous wastes are defined as materials which are toxic, **carcinogenic** (cause cancer), **mutagenic** (cause DNA mutations), **teratogenic** (cause birth defects), highly flammable, corrosive, or explosive. Examples include batteries, fluorescent lights, various cleaners, and **e-waste**, which arises from discarded electronics. Precious metals can be extracted and recycled from hazardous waste, but it must be done safely.



Figure **11.1.*b***: Plastic cutlery is nondegradable, but these spoons and forks and made of wood, a biodegradable material. Image by BRRT/Pixabay (Pixabay license).

## Solid Waste Sources and Composition

Agriculture and industry generate most of the waste globally followed by mining (figure **11.1.*c***). Together, these generate **non-municipal solid waste**. Some common items that are classified as non-municipal waste are: construction materials (roofing shingles, electrical fixtures, bricks); waste-water sludge; incinerator residues; ash; scrubber sludge; oil/gas/mining waste; railroad ties, and pesticide containers. The remaining waste, **municipal solid waste** (MSW), is made up of discarded solid materials from

residences, businesses, and city buildings. Globally, waste generated from industry generated was nearly 18 times that of MSW, and waste generated from agriculture was more than 4.5 times that of MSW.

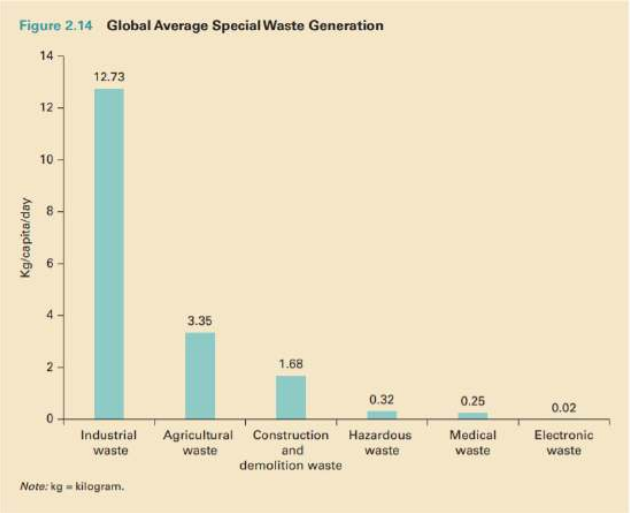



Figure **11.1.*c***: Global average special waste generation showing selected categories of waste in kilograms (kg) per person (capita) per day. 1 kg = 2.2 lbs. In order of greatest to least waste generation is industrial waste (12.73 kg/capita/day), agricultural waste (3.35), construction and demolition waste (1.68), hazardous waste (0.32), medical waste (0.25), and electronic waste (0.02). Image from What a Waste 2.0 : A Global Snapshot of Solid Waste Management to 2050 (CC-BY)

The world generates 2.21 billion tons of municipal solid waste annually. Municipal solid waste makes up 3-9% of all waste in the U.S., and 292.4 million tons were generated in U.S. in 2018 (4.9 lbs per person per day). This contrasts with average values in Sub-Saharan Africa (1.01 lbs per person per day), South Asia (1.15 lbs per person per day), and East Asia and the Pacific (1.23 lbs per person per day). Daily waste generation per person in high income countries is projected to increase by 19 percent by 2050, compared to low- and middle-income countries where it is anticipated to increase by approximately 40 percent or more.

Municipal solid waste consists of materials from plastics to food scraps. The most common waste product is paper (about 23% of the total; figure **11.1.*d***). Other common components are yard waste (green waste), plastics, metals, wood, glass and food waste. The composition of the municipal wastes can vary from region to region and from season to season.

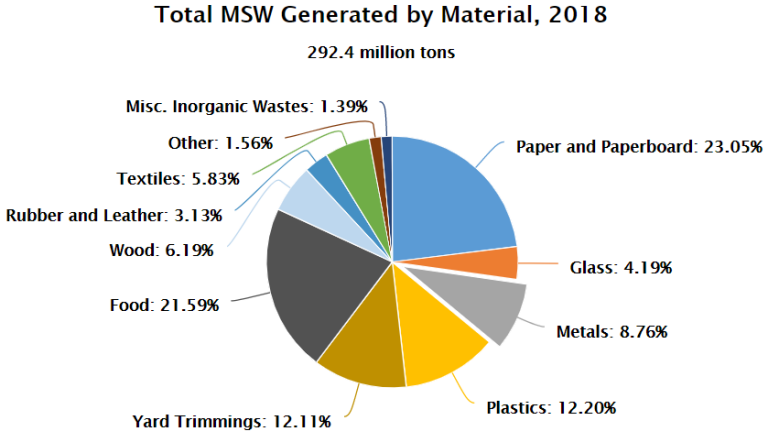



Figure **11.1.*d***: Composition of municipal solid waste (MSW) in the U.S. in 2018 before recycling. Paper and paperboard comprise 23.05% followed by food (21.59%), plastics (12.20%), yard trimmings (12.11%), metals (8.76%), wood (6.19%), textiles (5.83%), glass (4.19%), rubber and leather (3.13%), other (1.56%), and miscellaneous inorganic waste (1.39%). Image from EPA (public domain).

## Attribution

Modified by Melissa Ha from the following sources:

- Systems of Waste Management from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (CC-BY). Download for free at CNX.
- Solid Waste form *AP Environmental Science* by University of California College Prep (CC-BY). Download for free at CNX.
- Kaza, Silpa; Yao, Lisa C.; Bhada-Tata, Perinaz; Van Woerden, Frank. 2018. What a Waste 2.0 : A Global Snapshot of Solid Waste Management to 2050. Urban Development;. Washington, DC: World Bank. © World Bank. (CC-BY)


---


This page titled 11.1: Waste Generation is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **19.1: Waste Generation** by Melissa Ha and Rachel Schleiger is licensed CC BY-NC 4.0.

# 11.2: Waste Reduction

The **waste management hierarchy** lists processes to address waste in order of preference with the goals of limiting waste, minimizing environmental impact, and maximizing additional benefits (such as electricity generation or production of recycled goods). It is symbolized as an upside triangle, meaning that the largest volume of waste should be handled using the first process (most preferred), and the smallest volume should be handled using the last process (least preferred). These processes in order of preference are reduction, recycling, energy recovery (such as incineration), treatment, and disposal (figure **11.2.*a***). Individuals can exercise the first two processes by applying the four R's (see below). Unfortunately, the U.S. and the world overall, handles a large volume of waste using the least preferred method (disposal). As previously mentioned, about half of municipal solid waste (MSW) goes to a sanitary landfill for disposal. In contrast, only 23.5% of municipal solid waste (MSW) was recovered and recycled and another 8.5% was composted in 2018. Waste stream percentages also vary widely by region. As an example, San Francisco, California recycles or composts 80% of its waste material, whereas Houston, Texas recycles or composts only 19%.

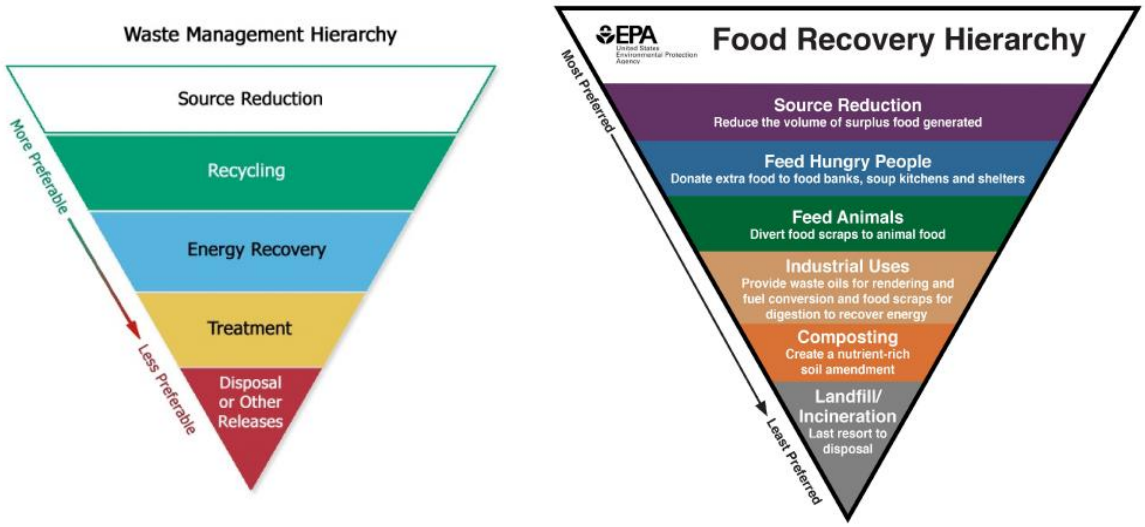



Figure **11.2.*a***: Left: The waste management hierarchy lists processes for handling waste from most to least preferable. They are source reduction, recycling, energy recovery, treatment, and disposal or other releases. Right: The food recovery hierarchy is a waste management hierarchy specific to food waste. It begins with source reduction (reduce the volume of surplus food generated). Next is feed hungry people (donate extra food to food banks, soup kitchens, and shelters. Third is feed animals (divert food scraps to animal food. Fourth is industrial uses (provide waste oils for rendering and fuel conversion and food scraps for digestion to recover energy. Fifth is composting (create a nutrient-rich soil amendment). The remaining waste goes to landfill or incineration as a last resort to disposal. Images by EPA (public domain).

The ideal waste management alternative is to prevent waste generation in the first place. Hence, **waste prevention** is a basic goal of all the waste management strategies. Numerous technologies can be employed throughout the manufacturing, use, or post-use portions of product life cycles to eliminate waste and, in turn, reduce or **prevent pollution**. Some representative strategies include environmentally conscious manufacturing methods that incorporate less hazardous or harmful materials, the use of modern leakage detection systems for material storage, innovative chemical neutralization techniques to reduce reactivity, or water saving technologies that reduce the need for fresh water inputs.

**Source reduction** (waste minimization) refers to the strategies that minimize the amount of generated waste and/or reduce the toxicity of the resultant waste when designing or manufacturing products or services. Waste can be reduced by reusing materials, using less hazardous substitute materials, or by modifying components of design and processing. Source reduction in manufacturing not only saves resources, but it also reduces costs for the manufacturer and indirectly for the consumer. For example, minimal packaging reduces material use, increases distribution efficiency, and reduces fuel consumption and resulting air emissions. Similarly, building materials can be designed to reduce the overall mass of material needed for a given structure.

The dry cleaning industry provides an excellent example of product substitution to reduce toxic waste generation. For decades, dry cleaners used tetrachloroethylene, or "perc" as a dry cleaning solvent. Although effective, tetrachloroethylene is a relatively toxic compound. Additionally, it is easily introduced into the environment, where it is highly recalcitrant due to its physical properties. Further, when its degradation occurs, the intermediate daughter products generated are more toxic to human health and the environment. Because of its toxicity and impact on the environment, the dry cleaning industry has adopted new practices and increasingly utilizes less toxic replacement products, including petroleum-based compounds. Further, new emerging technologies are incorporating carbon dioxide and other relatively harmless compounds. While these substitute products have in many cases been mandated by government regulation, they have also been adopted in response to consumer demands and other market-based forces.

**Recycling** refers to recovery and reprocessing of useful materials such as glass, paper, plastics, wood, and metals from the waste stream so they may be incorporated into the fabrication of new products. Recycling reduces the need of natural resource exploitation for raw materials, but it also allows waste materials to be recovered and utilized as valuable resource materials. Recycling of wastes reduces energy consumption and emissions generated by extraction of virgin materials and their subsequent manufacture into finished products, reduces overall energy consumption and greenhouse gas emissions that contribute to the global climate change, and reduces the incineration or landfilling of the materials that have been recycled. Moreover, recycling creates several economic benefits, including the potential to create job markets and drive growth. The success of recycling depends on effective collection and processing of recyclables, markets for reuse (e.g. manufacturing and/or applications that utilize recycled materials), and public acceptance and promotion of recycled products and applications utilizing recycled materials.

Numerous examples of successful recycling and reuse efforts are encountered every day. In some cases, the recycled materials are used as input materials and are heavily processed into end products. Common examples include the use of scrap paper for new paper manufacturing, or the processing of old aluminum cans into new aluminum products. In other cases, reclaimed materials undergo little or no processing prior to their re-use. Many construction materials can be reused, including concrete, asphalt materials, masonry, and reinforcing steel. "Green" plant-based wastes are often recovered and immediately reused for mulch or fertilizer applications. Many industries also recover various byproducts for reuse. In some cases, the recycled materials are used as input materials and are heavily processed into end products. Common examples include the use of scrap paper for new paper manufacturing, or the processing of old aluminum cans into new aluminum products. In other cases, reclaimed materials undergo little or no processing prior to their reuse. Some common examples include the use of tree waste as wood chips, or the use of brick and other fixtures into new structural construction. In any case, the success of recycling depends on effective collection and processing of recyclables, markets for reuse, and public acceptance and promotion of recycled products and applications utilizing recycled materials.

## The Four R's

The **four R's** (refuse, reduce, reuse, and recycle) are strategies that allow individuals to limit the volume and negative impacts of the waste they generate. They resemble the waste management hierarchy, but focus at the individual scale rather than the scale of a company or waste management system. They are listed in order of most to least environmentally beneficial, and they are all better alternatives than discarding trash into the landfill. Recycling is last because it requires energy to reprocess waste.

To **refuse** is to decline products or processes that harm the environment if you do not need them. Choosing products without packaging is an example of refusing. To **reduce** is to choose products or processes with lower **ecological footprint** (the area of land required to produce them). Examples include products with limited packaging or durable items rather than flimsy or disposable ones. **Reuse** refers to using a product multiple times or finding an alternative use for it. For example, one could share, borrow, or rent items. A plastic yogurt or pudding container could be repurposed for storage or gardening. As noted above, **recycling** means to return used items to be reprocessed (figure **11.2.*b***). Commonly recycled materials include cardboard, glass, cans, and certain plastics.




Figure 11.2.*b*: Recycling, or returning materials to be reprocessed, is the last of the four R's. If items cannot be refused, reduced, or reused, recycling is a better alternative than discarding into the landfill. Image by Intel Free Press (CC-BY).

**Resin identification codes** are the triangular symbols on recyclable plastics that indicate their composition (figure 11.2.*c*). A variety of materials make up plastics, and they are not all recycled the same way. Furthermore, your local recycling service may recycle some types of plastics but not others. Plastics with codes 1 and 2 are commonly recycled. For the codes, check with your local recycling service. Local businesses may house collection bins for plastics that are not commonly recycled. For example, grocery stores commonly have receptacles for recycling plastic bags, which fall under code 4.

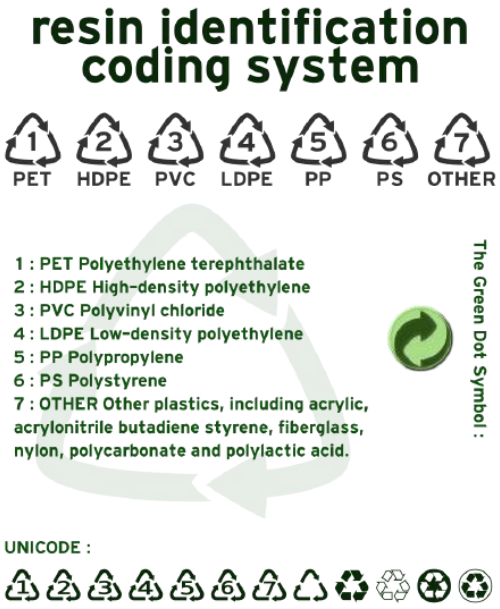



Figure 11.2.*c*): The resin identification coding system explains the chemical composition of various plastics. Some are commonly recycled, and others are not. The text in the image is, "1: PET Polyethylene terephthalate, 2: HDPE High-density polyethylene, 3: PVC Polyvinyl chloride, 4: LDPE Low-density polyethylene, 5: PP Polypropylene, 6: PS Polystyrene, 7: OTHER Other plastics, including acrylic, acrylonitrille butadiene styrene, fiberglass, nylon, polycarbonate and polylactic acid. Image by Filtre/OpenClipart Library (public domain).

## Biological Treatment

Landfill disposal of wastes containing significant organic fractions is increasingly discouraged in many countries, including the United States. Such disposal practices are even prohibited in several European countries. Since landfilling does not provide an attractive management option, other techniques have been identified. One option is to treat waste so that biodegradable materials are degraded and the remaining inorganic waste fraction (known as residuals) can be subsequently disposed or used for a beneficial purpose.

Biodegradation of wastes can be accomplished by using **aerobic** composting, **anaerobic digestion**, or **mechanical biological treatment** (MBT) methods. If the organic fraction can be separated from inorganic material, aerobic composting or anaerobic digestion can be used to degrade the waste and convert it into usable **compost** (figure **11.2.*d***). For example, organic wastes such as food waste, yard waste, and animal manure that consist of naturally degrading bacteria can be converted under controlled conditions into compost, which can then be utilized as natural fertilizer. Food scraps and yard waste together currently make up more than 30% of what we throw away and could be composted instead. Making compost keeps these materials out of landfills where they take up space and release methane, a potent greenhouse gas that contributes to climate change. Composting can be done at an industrial scale, such as with yard waste collected from homes in a city. It can also be done at home.



Figure **11.2.*d***): Compost is a nutrient-rich substance formed from decomposition of yard and food waste that enriches garden soil. Image by EPA (public domain).

Compost material may be used in a variety of applications. Compost enriches soil, helping retain moisture and suppress plant diseases and pests. Furthermore, It reduces the need for chemical fertilizers and encourages the production of beneficial bacteria and fungi. In addition to its use as a soil amendment for plant cultivation, compost can be used remediate soils, groundwater, and stormwater. Composting can be labor-intensive, and the quality of the compost is heavily dependent on proper control of the composting process. Inadequate control of the operating conditions can result in compost that is unsuitable for beneficial applications. Nevertheless, composting is becoming increasingly popular; composting diverted 82 million tons of waste material away the landfill waste stream in 2009, increased from 15 million tons in 1980. This diversion also prevented the release of approximately 178 million metric tons of carbon dioxide in 2009 – an amount equivalent to the yearly carbon dioxide emissions of 33 million automobiles.

All composting requires three basic ingredients: browns, greens, and water. "Browns" are materials such as dead leaves, branches, and twigs. "Greens" are materials such as grass clippings, vegetable waste, fruit scraps, and coffee grounds. The brown materials provide carbon for your compost, the green materials provide nitrogen, and the water provides moisture to help break down the organic matter. A compost pile should have an equal amount of browns to greens. Layers of these organic materials with different particle sizes should be alternated.

**Aerobic composting** is accomplished by placing selected proportions of organic waste into piles, rows or vessels, either in open conditions or within closed buildings fitted with gas collection and treatment systems. During the process, bulking agents such as wood chips are added to the waste material to enhance the aerobic degradation of organic materials. Finally, the material is allowed to stabilize and mature during a curing process where pathogens are concurrently destroyed. The end-products of the composting process include carbon dioxide, water, and the usable compost material.

In some cases, aerobic processes are not feasible. As an alternative, anaerobic processes may be utilized. **Anaerobic digestion** consists of degrading mixed or sorted organic wastes in vessels under anaerobic conditions. The anaerobic degradation process produces a combination of methane and carbon dioxide (biogas) and residuals (biosolids). Biogas can be used for heating and electricity production, while residuals can be used as fertilizers and soil amendments. Anaerobic digestion is a preferred degradation for wet wastes as compared to the preference of composting for dry wastes. The advantage of anaerobic digestion is

biogas collection; this collection and subsequent beneficial utilization makes it a preferred alternative to landfill disposal of wastes. Also, waste is degraded faster through anaerobic digestion as compared to landfill disposal.

**Backyard composting** requires a compost bin or pile to be placed in dry, shady spot near a water source (figure **11.2.*e***). Brown and green materials are added as they are collected, and larger pieces are chopped or shredded before they are added to the pile. Dry materials are moistened as they are added. Once the compost pile is established, grass clippings and green waste are mixed into the pile when added. When fruit and vegetable wastes are added, they are buried under 10 inches of compost material. Some homeowners cover the top of the compost with a tarp to keep it moist. When the material at the bottom is dark and rich in color, the compost is ready to use. This usually takes anywhere between two months to two years.



Figure **11.2.*e***): This compost bin contains food scraps and yard waste. Image by Ben_Kerckx/Pixabay (Pixabay license)

For those who do not have the space for an outdoor compost pile, materials can be composted indoors using a special type of bin, which can be purchased at a local hardware store, gardening supplies store, or made at home. Care must be taken to properly manage the compost pile and keep track of what is added such that it does not attract pests or rodents or smell bad. The compost should be ready in two to five weeks.

## Attribution

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

- Systems of Waste Management from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (CC-BY). Download for free at CNX.
- Composting at Home. Environmental Protection Agency. Accessed 01-18-2021 (public domain).
- 15.2 Waste Management Strategies from *Essentials of Environmental Science* by Kamala Dorsner and CK-12 (CC-BY)

11.2: Waste Reduction is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 11.3: Life Cycle Assessment

## Learning Objectives

After reading this module, students should be able to understand the basic elements of industrial ecology and life cycle analysi

## Problem Solving for Sustainability

It should be clear by now that making decisions and solving problems in support of greater sustainability of human-created systems and their impact on the natural environment is a complex undertaking. Often in modern life our decisions and designs are driven by a single goal or objective (e.g. greater monetary profitability, use of less energy, design for shorter travel times, generation of less waste, or reduction of risk), but in most cases solving problems sustainably requires a more holistic approach in which the functioning of many parts of the system must be assessed simultaneously, and multiple objectives must be integrated when possible. Furthermore, as noted in the Brundtland Report, often our decisions require the recognition of tradeoffs – there are many kinds of impacts on the environment and most decisions that we make create more than one impact at the same time. Of course choices must be made, but it is better if they are made with fuller knowledge of the array of impacts that will occur. The history of environmental degradation is littered with decisions and solutions that resulted in unintended consequences.

An illustrative example of the role of sustainability in solving problems is the issue of biofuels – turning plant matter into usable energy (mostly liquid hydrocarbon-based fuels). When viewed from afar and with a single goal, “energy independence,” using our considerable agricultural resources to turn solar energy, via photosynthesis, into usable fuels so that we can reduce our dependence on imported petroleum appears to be quite attractive. The United States is the largest producer of grain and forest products in the world. It has pioneered new technologies to maintain and even increase agricultural productivity, and it has vast processing capabilities to create artificial fertilizer and to convert biomass into agricultural products. And, after all, such a venture is both “domestic” and “natural” – attributes that incline many, initially at least, to be favorably disposed. However upon closer examination this direction is not quite as unequivocally positive as we might have thought. Yes it is possible to convert grain into ethanol and plant oils into diesel fuel, but the great majority of these resources have historically been used to feed Americans and the animals that they consume (and not just Americans; the United States is the world’s largest exporter of agricultural products). As demand has increased, the prices for many agricultural products have risen, meaning that some fraction of the world’s poor can no longer afford as much food. More marginal lands (which are better used for other crops, grazing, or other uses) have been brought under cultivation for fermentable grains, and there have been parallel “indirect” consequences globally – as the world price of agricultural commodities has risen, other countries have begun diverting land from existing uses to crops as well. Furthermore, agricultural runoff from artificial fertilizers has contributed to over 400 regional episodes of hypoxia (low oxygen) in estuaries around the world, including the U.S. Gulf Coast and Chesapeake Bay. In response to such problems, U.S. Congress passed the Energy Independence and Security Act in 2007, which limits the amount of grain that can be converted into biofuels in favor of using agriculturally-derived cellulose, the chief constituent of the cell walls of plants. This has given rise to a large scientific and technological research and development program to devise economical ways to process cellulosic materials into ethanol, and parallel efforts to investigate new cellulosic cropping systems that include, for example, native grasses. Thus, the seemingly simple decision to grow our biofuels industry in response to a political objective has had unintended political, financial, dietary, social, land use, environmental quality, and technological consequences.

With hindsight, the multiple impacts of biofuels have become clear, and there is always the hope that we can learn from examples like this. But we might also ask if there is a way to foresee all or at least some of these impacts in advance, and adjust our designs, processes, and policies to take them into account and make more informed decisions, not just for biofuels but also for complex societal problems of a similar nature. This approach is the realm of the field of **industrial ecology**, and the basis for the tool of **life cycle assessment (LCA)**, a methodology that has been designed to perform holistic analyses of complex systems.

### Industrial Ecology

Many systems designed by humans focus on maximizing profitability for the firm, business or corporation. In most cases this means increasing production to meet demand for the products or services being delivered. An unfortunate byproduct of this is the creation of large amounts of waste, many of which have significant impacts if they enter the environment. Figure **11.3.1** is a general-purpose diagram of a typical manufacturing process, showing the inputs of materials and energy, the manufacturing of

products, and the generation of wastes (the contents of the “manufacturing box” are generic and not meant to depict any particular industry—it could be a mine, a factory, a power plant, a city, or even a university).

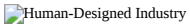



Figure **11.3.1**: **Human-Designed Industry** Generic representation of a human-designed industry.*Source: Theis, T.*

What many find surprising is the large disparity between the amounts of waste produced and the quantity of product delivered. Table **11.3.*a* Waste-to-Product Ratios for Selected Industries** provides such information, in the form of waste-to-product ratios, for a few common industries. The impacts of wastes on human health and the environment have historically been ignored, or steeply underpriced, so that little incentive has existed to limit waste production. More recently laws have been enacted that attempt to force those responsible for waste emissions into a more appropriate accounting (see **Environmental and Resource Economics** and **Modern Environmental Management** for a fuller treatment of the laws, regulations, and practices used to incorporate society’s costs into the production chain). Once realistic costs are assigned to the waste sector, manufacturers are quick to innovate and investigate ways to eliminate them.

Table **11.3.*a*:Waste-to-Product Ratios for Selected Industries:** Table shows the waste to product ratios for six common industries. *Source: Theis, T.*

| Industrial Sector | Waste-to-Product Ratio |
|---|---|
| Automobiles | 2/1 (up to 10/1 if consumer use is included) |
| Paper | 10/1 |
| Basic Metals (e.g. Steel and Aluminum) | 30-50/1 |
| Chemicals | 0.1-100/1 |
| Nanostructured materials (e.g. computer chips) | 700-1700/1 |
| Modern Agriculture | ~4/1 |

In 1989, Robert Frosch & Nicholas Gallopoulos, who worked in the General Motors Research Laboratory, published an important analysis of this problem in Scientific American (Frosch and Gallopoulos, 1989). Their paper was entitled “Strategies for Manufacturing”; in it they posed a critical question: Why is it that human-designed manufacturing systems are so wasteful, but systems in nature produce little, if any, waste? Although there had been many studies on ways to minimize or prevent wastes, this was the first to seek a systemic understanding of what was fundamentally different about human systems in distinction to natural systems. The paper is widely credited with spawning the new field of Industrial Ecology, an applied science that studies material and energy flows through industrial systems. Industrial Ecology is concerned with such things as closing material loops (recycling and reuse), process and energy efficiency, organizational behavior, system costs, and social impacts of goods and services. A principle tool of Industrial Ecology is life cycle assessment.

## Life Cycle Assessment Basics

Life Cycle Assessment (LCA) is a systems methodology for compiling and evaluating information on materials and energy as they flow through a product or service manufacturing chain. It grew out of the needs of industry, in the early 1960s, to understand

manufacturing systems, supply chains, and market behavior, and make choices among competing designs, processes, and products. It was also applied to the evaluation of the generation and emission of wastes from manufacturing activities. During the 1970s and 1980s general interest in LCA for environmental evaluation declined as the nation focused on the control of toxic substances and remediation of hazardous waste sites, but increasing concern about global impacts, particularly those associated with greenhouse gas emissions, saw renewed interest in the development of the LCA methodology and more widespread applications.

LCA is a good way to understand the totality of the environmental impacts and benefits of a product or service. The method enables researchers and practitioners to see where along the product chain material and energy are most intensively consumed and waste produced. It allows for comparisons with conventional products that may be displaced in commerce by new products, and helps to identify economic and environmental tradeoffs.

LCA can facilitate communication of risks and benefits to stakeholders and consumers (e.g. the "carbon footprint" of individual activities and life styles). Perhaps most importantly of all, LCA can help to prevent unintended consequences, such as creating solutions to problems that result in the transferal of environmental burdens from one area to another, or from one type of impact to another.

A complete LCA assessment defines a system as consisting of four general stages of the product or service chain, each of which can be further broken down into substages:

- Acquisition of materials (through resource extraction or recycled sources)
- Manufacturing, refining, and fabrication
- Use by consumers
- End-of-life disposition (incineration, landfilling, composting, recycling/reuse)

Each of these involves the transport of materials within or between stages, and transportation has its own set of impacts.

In most cases, the impacts contributed from each stage of the LCA are uneven, i.e. one or two of the stages may dominate the assessment. For example, in the manufacture of aluminum products it is acquisition of materials (mining), purification of the ore, and chemical reduction of the aluminum into metal that create environmental impacts. Subsequent usage of aluminum products by consumers contributes very few impacts, although the facilitation of recycling of aluminum is an important step in avoiding the consumption of primary materials and energy. In contrast, for internal combustion-powered automobiles, usage by consumers creates 70-80% of the life cycle impacts. Thus, it is not always necessary that the LCA include all stages of analysis; in many cases it is only a portion of the product/service chain that is of interest, and often there is not enough information to include all stages anyway. For this reason there are certain characteristic terminologies for various "scopes" of LCAs that have emerged:

- *Cradle-to-grave*: includes the entire material/energy cycle of the product/material, but excludes recycling/reuse.
- *Cradle-to-cradle*: includes the entire material cycle, including recycling/reuse.
- *Cradle-to-gate*: includes material acquisition, manufacturing/refining/fabrication (factory gate), but excludes product uses and end-of-life.
- *Gate-to-gate*: a partial LCA looking at a single added process or material in the product chain.
- *Well-to-wheel*: a special type of LCA involving the application of fuel cycles to transportation vehicles.
- *Embodied energy*: A cradle-to-gate analysis of the life cycle energy of a product, inclusive of the latent energy in the materials, the energy used during material acquisition, and the energy used in manufacturing intermediate and final products. Embodied energy is sometimes referred to as "emergy", or the cumulative energy demand (CED) of a product or service.

Over time the methodology for conducting Life Cycle Analyses (LCAs) has been refined and standardized; it is generally described as taking place in four steps: scoping, inventory, impact assessment, and interpretation (Figure **11.3.2**. The first three of these are consecutive, while the interpretation step is an ongoing process that takes place throughout the methodology.

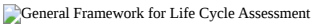



Figure \(\PageIndex{2}\): **General Framework for Life Cycle Assessment**
The four steps of life cycle assessment and their relationship to one another. *Source:* *Mr3641 via Wikipedia*

**Scoping** is arguably the most important step for conducting an LCA. It is here that the rationale for carrying out the assessment is made explicit, where the boundaries of the system are defined, where the data quantity, quality, and sources are specified, and where any assumptions that underlie the LCA are stated. This is critically important both for the quality of the resultant analysis, and for comparison among LCAs for competing or alternative products.

The **inventory analysis** step involves the collection of information on the use of energy and various materials used to make a product or service at each part of the manufacturing process. The inventory is probably the most tedious step in LCA since it involves locating, acquiring, and evaluating the quality of data and specifying the sources of uncertainties that may have arisen. For products that have been produced for a long time and for which manufacturing processes are well known, such as making steel, concrete, paper, most plastics, and many machines, data are readily available. But for newer products that are either under development or under patent protection, data are often considered proprietary and are generally not shared in open sources. Uncertainty can arise because of missing or poorly documented data, errors in measurement, or natural variations caused by external factors (e.g., weather patterns can cause considerable variation in the outputs of agricultural systems or the ways that consumers use products and services can cause variability in the emission of pollutants and the disposition of the product at end of life). Often the manufacturing chain of a process involves many steps resulting in a detailed inventory analysis.

The **life cycle impact assessment** (LCIA) takes the inventory data on material resources used, energy consumed, and wastes emitted by the system and estimates potential impacts on the environment. At first glance, given that an inventory may include thousands of substances, it may seem that the number of potential impacts is bewilderingly large, but the problem is made more tractable through the application of a system of impact classifications within which various inventory quantities can be grouped as having similar consequences on human health or the environment. Sometimes inventoried quantities in a common impact category originate in different parts of the life cycle and often possess very different chemical/biological/physical characteristics. The LCIA groups emissions based on their common impacts rather than on their chemical or physical properties, choosing a reference material for which health impacts are well known, as a basic unit of comparison. A key aspect is the conversion of impacts of various substances into the reference unit. This is done using characterization factors, some of which are well-known, such as global warming potential and ozone depletion potential, and $LC_{50}$ (the concentration of a substance at which fifty percent of an exposed population is killed), and others are still under development.

The **interpretation** step of LCA occurs throughout the analysis. As noted above, issues related to the rationale for conducting the LCA, defining the system and setting its boundaries, identifying data needs, sources, and quality, and choosing functional units, allocation procedures, and appropriate impact categories must all be addressed as the LCA unfolds. There are essentially two formal reasons for conducting an LCA: (a) identification of “hot spots” where material and/or energy use and waste emissions, both quantity and type, are greatest so that efforts can be focused on improving the product chain; and (b) comparison of results between and among other LCAs in order to gain insight into the preferable product, service, process, or pathway. In both cases, there are cautions that apply to the interpretation of results.

1. Typically a variety of assumption must be made in order to carry out the LCA. Sometimes these are minor, for example, exclusion of elements of the study that clearly have no appreciable impact on the results, and sometimes more critical, for

example choosing one set of system boundaries over another. These must be explicitly stated, and final results should be interpreted in light of assumptions made.

2. In the course of conducting an LCA it is usually the case that a variety of data sources will be used. In some cases these may be from the full-scale operation of a process, in others the source is from a small scale or even laboratory scale, in still other cases it may be necessary to simulate information from literature sources. Such heterogeneity inevitably leads to uncertainty in the final results; there are several statistical methods that can be applied to take these into account. An important aspect of the completed LCA is the degree of sensitivity the results display when key variables are perturbed. Highly sensitive steps in the chain have a greater need to narrow uncertainties before drawing conclusions with confidence.
3. Sometimes LCA impact categories overlap in the sense that the same pollutant may contribute to more than one category. For instance, if a given assessment comes up with high scores for both aquatic toxicity and human toxicity from, say, pesticide use then one might be justified in using both of these categories to draw conclusions and make choices based on LCA results. However, more typically elevated scores are found for categories that are not directly comparable. For instance, the extraction, refining, and use of petroleum generate a high score for global warming (due to GHG release), while the product chain for the biofuel ethanol has a high score for eutrophication (due to nitrogen release during the farming stage). Which problem is worse – climate change or coastal hypoxia? Society may well choose a course of action that favors one direction over another, but in this case the main value of the LCA is to identify the tradeoffs and inform us of the consequences, not tell us which course is "correct."
4. One of the inherent limits to LCA is its use for assessing risk. Risk assessment and management is a formal process that quantifies risks for a known population in a specific location exposed to a specific chemical for a defined period of time. It generates risk values in terms of the probability of a known consequence due to a sequence of events that are directly comparable, and upon which decisions on water, land, and air quality standards and their violation can be and are made. LCA is a method for evaluating the impacts of wastes on human health and the environment from the point of view of the product/service chain rather than a particular population. It can be used to identify the sources of contamination and general impacts on the environment – a sort of "where to look" guide for regulation, but its direct use in the environmental regulatory process has been, to date, rather limited. One application for LCA that has been suggested for regulatory use is for assessing the impacts of biofuel mandates on land use practices, in the United States and other regions, however no regulatory standards for land use have yet been proposed.

**Attribution**

Modified by Kyle Whittinghill from the following sources:

- Life Cycle Assessment from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (CC-BY). Download for free at CNX.

11.3: Life Cycle Assessment is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 11.4: Waste Disposal

There are three primary methods for waste disposal: open dumps, sanitary landfills, and incineration. Sanitary landfills and incineration prevent reuse, recycling, and proper decomposition. While open dumps promote decomposition better than other methods of waste disposal and allow discarded materials to be salvaged or recycled, they promote disease spread and cause water pollution. They are thus illegal in many countries.

## Open Dumps

**Open dumps** involve simply piling up trash in a designated area and is thus the easiest method of waste disposal (figure **11.4.*a***). Open dumps can support populations of organisms that house and transmit disease (reservoirs and vectors, respectively). Additionally, contaminants from the trash mix with rain water forming **leachate**, which infiltrates into the ground or runs off. This liquid leachate may contain toxic chemicals such as dioxin (a persistent organic pollutant), mercury, and pesticides.



Figure **11.4.*a***: An open dump in Vietnam. Image by Julien Belli (CC-BY).

## Sanitary Landfills

Despite advances in reuse and recycling, landfill disposal in **sanitary landfills** remains the primary waste disposal method in the United States. The rate of municipal solid waste (MSW) generation continues to increase, but overall landfill capacity is decreasing. Older existing landfills are increasingly reaching their authorized capacity. New regulations concerning proper waste disposal and the use of innovative liner systems to minimize the potential of groundwater contamination from leachate infiltration and migration have resulted in a substantial increase in the costs of landfill disposal. Finally, public opposition delays or, in many cases, prevents the approval of new landfills or expansion of existing facilities. Public opposition is partially inspired by memories of historic uncontrolled dumping practices the resulting undesirable side effects of uncontrolled vectors, contaminated groundwater, unmitigated odors, and subsequent diminished property values.

In sanitary landfills, trash is sealed from the top and the bottom to reduce contamination of surroundings (figure **11.4.*c***). Rainwater that percolates through a sanitary landfill is collected in the bottom liner, and this bottom layer thus prevents contamination of groundwater. The groundwater near the landfill is closely monitored for signs of contamination from the leachate. Layers of soil on top prevent disease spread. Each day after garbage is dumped in the landfill, it is covered with clay or plastic to prevent redistribution by animals or the wind.



Figure 11.4.*b*: A sanitary landfill in 1972. Image by Bill ShroutEPA (public domain).

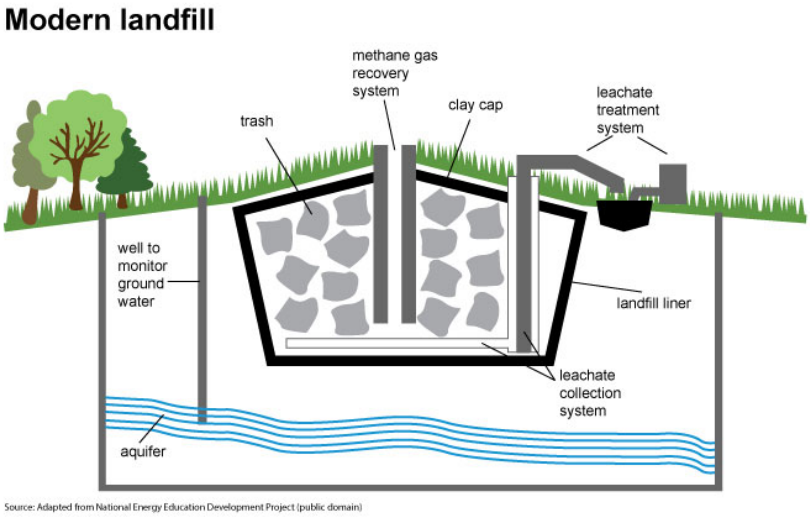



Figure 11.4.*c*: A sanitary landfill. Trash is compacted and stored underground. A landfill liner prevents leachate from draining into the aquifer and contaminating groundwater, which is monitored using a well. Additionally, leachate is collected and treated (leachate treatment system). A clay cap prevents animals or wind from accessing the trash. A methane gas recovery system captures the potent greenhouse gas, which can be used as a biofuel. Image by EPA/National Energy Education Development Project (public domain).

Several practices can reduce the environmental impact of sanitary landfills. Compacting in landfills reduces water and oxygen levels, slowing decomposition and promoting methane release. In the U.S., the Clear Air Act requires that landfills of a certain size collect landfill gas (biogas), which can be used as a biofuel for heating or electricity generation. Other gases such as ammonia and hydrogen sulfide may also be released by the landfill, contributing to air pollution. These gases are also monitored and, if necessary, collected for disposal. To address the often dry condition of wastes within landfills, the concept of bioreactor landfills has emerged. These recirculate leachate and/or inject other liquids to increase moisture and promote decomposition (and therefore increasing the rate of biogas production). Upon closure, many landfills undergo "land recycling" and can be redeveloped as golf courses, recreational parks, and other beneficial uses.

# Incineration

**Incineration** is simply burning trash. This has several advantages: it reduces volume and can be used to generate electricity (**waste-to-energy**). In fact, the sheer volume of the waste is reduced by about 85%. Incineration is costly, however, and it pollutes air and water. Air pollutants released by incineration include particulates, sulfur dioxide, nitrogen oxides, methane, heavy metals (such as lead and mercury), and dioxins. The byproduct of incineration, ash, is often toxic. Depending on its composition, ash

might require special disposal; other types of ash can be repurposed. Incineration is widely used in developed countries due to landfill space limitations. It is estimated that about 130 million tons of waste are annually combusted in more than 600 plants in 35 countries. Further, incineration is often used to effectively mitigate hazardous wastes such as chlorinated hydrocarbons, oils, solvents, medical wastes, and pesticides.

An incinerator processes trash and burns it in a **combustion chamber** (figure **11.4.*d* – *e***). The heat boils water, and the resultant steam is used to generate electricity. The smoke (called flue gases) goes through a **pollution-removal** before it is released, but it still contains some pollutants. The U.S. incinerated 11.8% of MSW in 2018.



Figure **11.4.*d***: Wheelabrator Technologies' Waste-to-Energy plant in Saugus, Massachusetts has been in service since 1975. Image and caption (modified) from Fletcher6 (CC-BY).

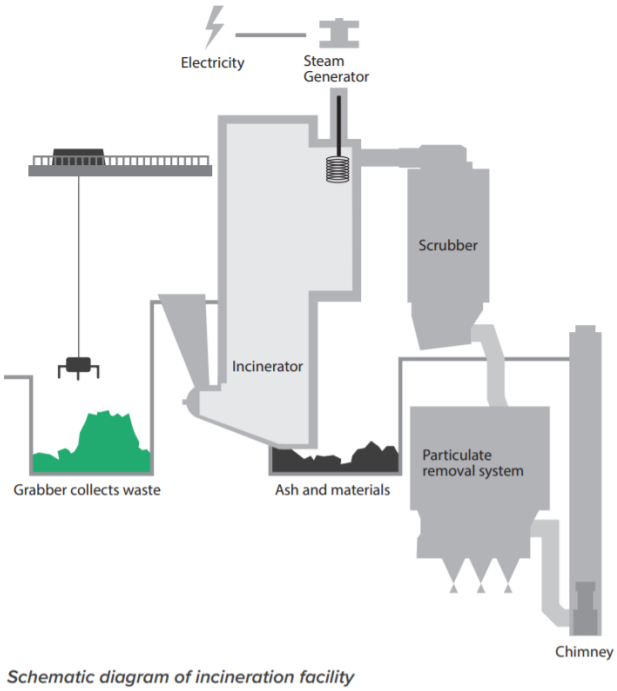



Figure **11.4.*e***: A schematic diagram of a waste incineration facility (waste-to energy system). A grabber collects the waste and transfers it into the incinerator. Heat from combusting the trash is used to produce steam and generate electricity. Emissions (flue gases) pass through a scrubber and particulate removal system to limit pollution before the gases are released through a chimney. Ash remains after incineration. Image by Kaza, Silpa; Bhada-Tata, Perinaz. 2018. Decision Maker's Guides for Solid Waste

Management Technologies. Urban Development Series Knowledge Papers;. World Bank, Washington, DC. © World Bank. (CC-BY)

There are two kinds of waste-to-energy systems: mass burn incinerators and refuse-derived incinerators. In **mass burn incinerators** all of the solid waste is incinerated. The heat from the incineration process is used to produce steam. This steam is used to drive electric power generators. Acid gases from the burning are removed by chemical scrubbers. Any particulates (small particles that remain suspended in the air) in the combustion gases are removed by **electrostatic precipitators**, which charge particulates and remove them with electrodes. The cleaned gases are then released into the atmosphere through a tall stack. The ashes from the combustion are sent to a landfill for disposal.

It is best if only combustible items (paper, wood products, and plastics) are burned. In a **refuse-derived incinerator**, non-combustible materials are separated from the waste. Items such as glass and metals may be recycled. The combustible wastes are then formed into fuel pellets which can be burned in standard steam boilers. This system has the advantage of removing potentially harmful materials from waste before it is burned. It also provides for some recycling of materials.

## Attribution

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

- Systems of Waste Management from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (CC-BY). Download for free at CNX.
- Solid Waste from *AP Environmental Science* by University of California College Prep (CC-BY). Download for free at CNX.
- Biomass explained: Landfill gas and biogas. 2020. EIA. Accessed 01-18-2021 (public domain).
- 15.2 Waste Management Strategies from *Essentials of Environmental Science* by Kamala Dorsner and CK-12 (CC-BY)

This page titled 11.4: Waste Disposal is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 11.5: Waste Regulations

## Regulatory Framework in the United States

During the course of the 20th century, especially following World War II, the United States experienced unprecedented economic growth. Much of the growth was fueled by rapid and increasingly complex industrialization. With advances in manufacturing and chemical applications also came increases in the volume, and in many cases the toxicity, of generated wastes. Furthermore, few if any controls or regulations were in place with respect to the handling of toxic materials or the disposal of waste products. Continued industrial activity led to several high-profile examples of detrimental consequences to the environment resulting from these uncontrolled activities. Finally, several forms of intervention, both in the form of government regulation and citizen action, occurred in the early 1970s.

Ultimately, several regulations were promulgated on the state and federal levels to ensure the safety of public health and the environment. With respect to waste materials, the Resource Conservation and Recovery Act (RCRA), enacted by the United States Congress, first in 1976 and then amended in 1984, provides a comprehensive framework for the proper management of hazardous and non-hazardous solid wastes in the United States. RCRA stipulates broad and general legal objectives while mandating the United States Environmental Protection Agency (USEPA) to develop specific regulations to implement and enforce the law. States and local governments can either adopt the federal regulations, or they may develop and enforce more stringent regulations than those specified in RCRA. Similar regulations have been developed or are being developed worldwide to manage wastes in a similar manner in other countries.

The broad goals of RCRA include: (1) the protection of public health and the environment from the hazards of waste disposal; (2) the conservation of energy and natural resources; (3) the reduction or elimination of waste; and (4) the assurance that wastes are managed in an environmentally-sound manner (e.g. the remediation of waste which may have spilled, leaked, or been improperly disposed). It should be noted here that the RCRA focuses only on active and future facilities and does not address abandoned or historical sites. These types of environmentally impacted sites are managed under a different regulatory framework, known as the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) of 1980, or more commonly known as "Superfund."

## Solid Waste Regulations

RCRA defines solid waste as any garbage or refuse, sludge from a wastewater treatment plant, water supply treatment plant, or air pollution control facility and other discarded material, including solid, liquid, semi-solid, or contained gaseous material resulting from industrial, commercial, mining, and agricultural operations, and from community activities. In general, solid waste can be categorized as either **non-hazardous waste** or **hazardous waste**.

Non-hazardous solid waste can be trash or garbage generated from residential households, offices and other sources. Generally, these materials are classified as **municipal solid waste** (MSW)**.** Alternatively, non-hazardous materials that result from the production of goods and products by various industries (e.g. coal combustion residues, mining wastes, cement kiln dust), are collectively known as industrial solid waste. Because they are classified as non-hazardous material, many components of municipal solid waste and industrial waste have potential for recycling and re-use. Significant efforts are underway by both government agencies and industry to advance these objectives.

Hazardous waste, generated by many industries and businesses (e.g. dry cleaners and auto repair shops), is constituted of materials that are dangerous or potentially harmful to human health and the environment. Waste is classified as hazardous if it exhibits at least one of these four characteristics:

- **Ignitability**, which refers to creation of fires under certain conditions; including materials that are spontaneously combustible or those that have a flash point less than 140 $^0$F.
- **Corrosivity**, which refers to capability to corrode metal containers; including materials with a pH less than or equal to 2 or greater than or equal to 12.5.
- **Reactivity**, which refers to materials susceptible to unstable conditions such as explosions, toxic fumes, gases, or vapors when heated, compressed, or mixed with water under normal conditions.
- **Toxicity**, which refers to substances that can induce harmful or fatal effects when ingested or absorbed, or inhaled.

## Radioactive Waste Regulations

Although non-hazardous waste and hazardous waste are regulated by RCRA, nuclear or radioactive waste is regulated in accordance with the Atomic Energy Act of 1954 by the Nuclear Regulatory Commission (NRC) in the United States.

Radioactive wastes are characterized according to four categories: (1) **High-level waste** (HLW), (2) **Transuranic waste** (TRU), (3) **Low-level waste** (LLW), and (4) **Mill tailings**. Various radioactive wastes decay at different rates, but health and environmental dangers due to radiation may persist for hundreds or thousands of years.

High-level waste is typically liquid or solid waste that results from government defense related activities or from nuclear power plants and spent fuel assemblies. These wastes are extremely dangerous due to their heavy concentrations of radionuclides, and humans must not come into contact with them.

Transuranic waste mainly results from the reprocessing of spent nuclear fuels and from the fabrication of nuclear weapons for defense projects. They are characterized by moderately penetrating radiation and a decay time of approximately twenty years until safe radionuclide levels are achieved. Following the passage of a reprocessing ban in 1977, most of this waste generation ended. Even though the ban was lifted in 1981, Transuranic waste continues to be rare because reprocessing of nuclear fuel is expensive. Further, because the extracted plutonium may be used to manufacture nuclear weapons, political and social pressures minimize these activities.

Low level wastes include much of the remainder of radioactive waste materials. They constitute over 80 percent of the volume of all nuclear wastes, but only about two percent of total radioactivity. Sources of Low level wastes include all of the previously cited sources of High-level waste and Transuranic waste, plus wastes generated by hospitals, industrial plants, universities, and commercial laboratories. Low level waste is much less dangerous than High-level waste, and NRC regulations allow some very low-level wastes to be released to the environment. Low level wastes may also be stored or buried until the isotopes decay to levels low enough such that it may be disposed of as non-hazardous waste. Low level wastes disposal is managed at the state level, but requirements for operation and disposal are established by the USEPA and NRC. The Occupational Health and Safety Administration (OSHA) is the agency in charge of setting the standards for workers that are exposed to radioactive materials.

## Attribution

Modified by Kyle Whittinghill the following sources:

- 15.3 Waste Policies from *Essentials of Environmental Science* by Kamala Dorsner and CK-12 (CC-BY)
- Systems of Waste Management from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (CC-BY). Download for free at CNX.

11.5: Waste Regulations is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 11.6: Solid Waste and Marine Life

Ocean dumping has also been a popular way for coastal communities to dispose of their solid wastes. In this method, large barges carry waste out to sea and dump it into the ocean (figure **11.6.*a***). That practice is now banned in the United States due to the pollution problems it created. However, much trash escapes to the ocean through littering and illegal dumping. Additionally, wind carries trash off of overfilled containers, landfills, or open dumps (where they are still used).



Figure **11.6.*a***: Left: A garbage barge leaves Manhattan. Right: Marine debris at the beach. Left image and caption by EPA (public domain), and right image and caption by NOAA (public domain).

**Garbage patches** form from waste that escapes to the ocean. Plastic breaks down into smaller pieces when exposed to sunlight and ample oxygen. These pieces of plastic are trapped in calm parts of the ocean that are surrounded by strong, circular currents (figure **11.6.*b***). Garbage patches may not be visible from the surface of the ocean, and the plastic pieces may be up to 20 meters deep.

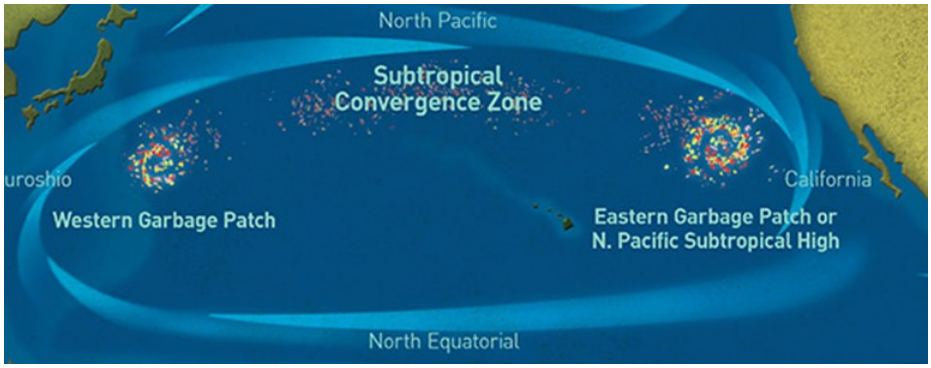



Figure **11.6.*b***: Locations of the largest garbage patches in the Pacific Ocean. The Western Garbage Patch is closer to Japan, and the Eastern Garbage Patch is closer to California. The Subtropical Convergence Zone is in the middle to the north. Ocean currents including the Kuroshio Current (left), North Pacific Current (top), California Current (right), and North Equatorial Current (bottom) are labeled. Image by Boyan Slat (public domain).

Plastic and other trash harms marine life in several ways. Old fishing nets and plastic six-pack rings trap wildlife. Additionally, organisms that ingest plastic risk choking. Ingested plastic may cut internal organs, limit space available for food, and carry toxins (figure **11.6.*c***).



Figure **11.6.c**: Much of this dead albatross's digestive system was filled with plastic waste. Image by Forest and Kim Starr (CC-BY).

## Attribution

Melissa Ha (CC-BY-NC) and Solid Waste from *AP Environmental Science* by University of California College Prep (CC-BY). Download for free at CNX.

---

This page titled 11.6: Solid Waste and Marine Life is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **19.3: Solid Waste and Marine Life** by Melissa Ha and Rachel Schleiger is licensed CC BY-NC 4.0.

# 11.7: Case Study- Electronic Waste and Extended Producer Responsibility

Electronic waste, commonly known as e-waste, refers to discarded electronic products such as televisions, computers and computer peripherals (e.g. monitors, keyboards, disk drives, and printers), telephones and cellular phones, audio and video equipment, video cameras, fax and copy machines, video game consoles, and others (Figure **11.7.1**)).

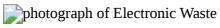

photograph of Electronic Waste

Figure **11.7.1**): Electronic Waste. Photograph shows many computers piled up in a parking lot as waste. Source: Bluedisk via Wikimedia Commons

In the United States, it is estimated that about 3 million tons of e-waste are generated each year. This waste quantity includes approximately 27 million units of televisions, 205 million units of computer products, and 140 million units of cell phones. Less than 15 to 20 percent of the e-waste is recycled or refurbished; the remaining percentage is commonly disposed of in landfills and/or incinerated. It should be noted that e-waste constitutes less than 4 percent of total solid waste generated in the United States. However, with tremendous growth in technological advancements in the electronics industry, many electronic products are becoming obsolete quickly, thus increasing the production of e-waste at a very rapid rate. The quantities of e-waste generated are also increasing rapidly in other countries such as India and China due to high demand for computers and cell phones.

In addition to the growing quantity of e-waste, the hazardous content of e-waste is a major environmental concern and poses risks to the environment if these wastes are improperly managed once they have reached the end of their useful life. Many e-waste components consist of toxic substances, including heavy metals such as lead, copper, zinc, cadmium, and mercury as well as organic contaminants, such as flame retardants (polybrominated biphenyls and polybrominated diphenylethers). The release of these substances into the environment and subsequent human exposure can lead to serious health and pollution issues. Concerns have also been raised with regards to the release of toxic constituents of e-waste into the environment if landfilling and/or incineration options are used to manage the e-waste.

Various regulatory and voluntary programs have been instituted to promote reuse, recycling and safe disposal of bulk e-waste. Reuse and refurbishing has been promoted to reduce raw material use energy consumption, and water consumption associated with the manufacture of new products. Recycling and recovery of elements such as lead, copper, gold, silver and platinum can yield valuable resources which otherwise may cause pollution if improperly released into the environment. The recycling and recovery operations have to be conducted with extreme care, as the exposure of e-waste components can result in adverse health impacts to the workers performing these operations. For economic reasons, recycled e-waste is often exported to other countries for recovery operations. However, lax regulatory environments in many of these countries can lead to unsafe practices or improper disposal of bulk residual e-waste, which in turn can adversely affect vulnerable populations.

In the United States, there are no specific federal laws dealing with e-waste, but many states have recently developed e-waste regulations that promote environmentally sound management. For example, the State of California passed the Electronic Waste Recycling Act in 2003 to foster recycling, reuse, and environmentally sound disposal of residual bulk e-waste. Yet, in spite of recent regulations and advances in reuse, recycling and proper disposal practices, additional sustainable strategies to manage e-waste are urgently needed.

One sustainable strategy used to manage e-waste is extended producer responsibility (EPR), also known as product stewardship. This concept holds manufacturers liable for the entire life-cycle costs associated with the electronic products, including disposal costs, and encourages the use of environmental-friendly manufacturing processes and products. Manufacturers can pursue EPR in multiple ways, including reuse/refurbishing, buy-back, recycling, and energy production or beneficial reuse applications. Life-cycle assessment and life-cycle cost methodologies may be used to compare the environmental impacts of these different waste management options. Incentives or financial support is also provided by some government and/or regulatory agencies to promote EPR. The use of non-toxic and easily recyclable materials in product fabrication is a major component of any EPR strategy. A growing number of companies (e.g. Dell, Sony, HP) are embracing EPR with various initiatives towards achieving sustainable e-waste management.

EPR is a preferred strategy because the manufacturer bears a financial and legal responsibility for their products; hence, they have an incentive to incorporate green design and manufacturing practices that incorporate easily recyclable and less toxic material

components while producing electronics with longer product lives. One obvious disadvantage of EPR is the higher manufacturing cost, which leads to increased cost of electronics to consumers.

There is no specific federal law requiring EPR for electronics, but the United States Environmental Protection Agency (USEPA) undertook several initiatives to promote EPR to achieve the following goals: (1) foster environmentally conscious design and manufacturing, (2) increase purchasing and use of more environmentally sustainable electronics, and (3) increase safe, environmentally sound reuse and recycling of used electronics. To achieve these goals, USEPA has been engaged in various activities, including the promotion of environmental considerations in product design, the development of evaluation tools for environmental attributes of electronic products, the encouragement of recycling (or *e*-cycling), and the support of programs to reduce e-waste, among others. More than 20 states in the United States and various organizations worldwide have already developed laws and/or policies requiring EPR in some form when dealing with electronic products. For instance, the New York State Wireless Recycling Act emphasizes that authorized retailers and service providers should be compelled to participate in take-back programs, thus allowing increased recycling and reuse of e-waste. Similarly, Maine is the first U.S. state to adopt a household e-waste law with EPR.

In Illinois, Electronic Products Recycling & Reuse Act requires the electronic manufacturers to participate in the management of discarded and unwanted electronic products from residences. The Illinois EPA has also compiled e-waste collection site locationswhere the residents can give away their discarded electronic products at no charge. Furthermore, USEPA compiled a list of local programs and manufacturers/retailers that can help consumers to properly donate or recycle e-waste.

Overall, the growing quantities and environmental hazards associated with electronic waste are of major concern to waste management professionals worldwide. Current management strategies, including recycling and refurbishing, have not been successful. As a result, EPR regulations are rapidly evolving throughout the world to promote sustainable management of e-waste. However, neither a consistent framework nor assessment tools to evaluate EPR have been fully developed.

## Attribution

Modified by Kyle Whittinghill from the following sources:

- Case Study: Electronic Waste and Extended Producer Responsibility from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (CC-BY). Download for free at CNX.
- 15.4 Case Study: Electronic Waste and Extended Producer Responsibility from *Essentials of Environmental Science* by Kamala Dorsner and CK-12 (CC-BY)

11.7: Case Study- Electronic Waste and Extended Producer Responsibility is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# CHAPTER OVERVIEW

## 12: Water Supply and Water Pollution


12.1: Fresh Water Supply

12.2: Case Study - The Aral Sea - Going, Going, Gone

12.3: Water Pollutants and Their Sources

12.4: Water and Wastewater Treatment

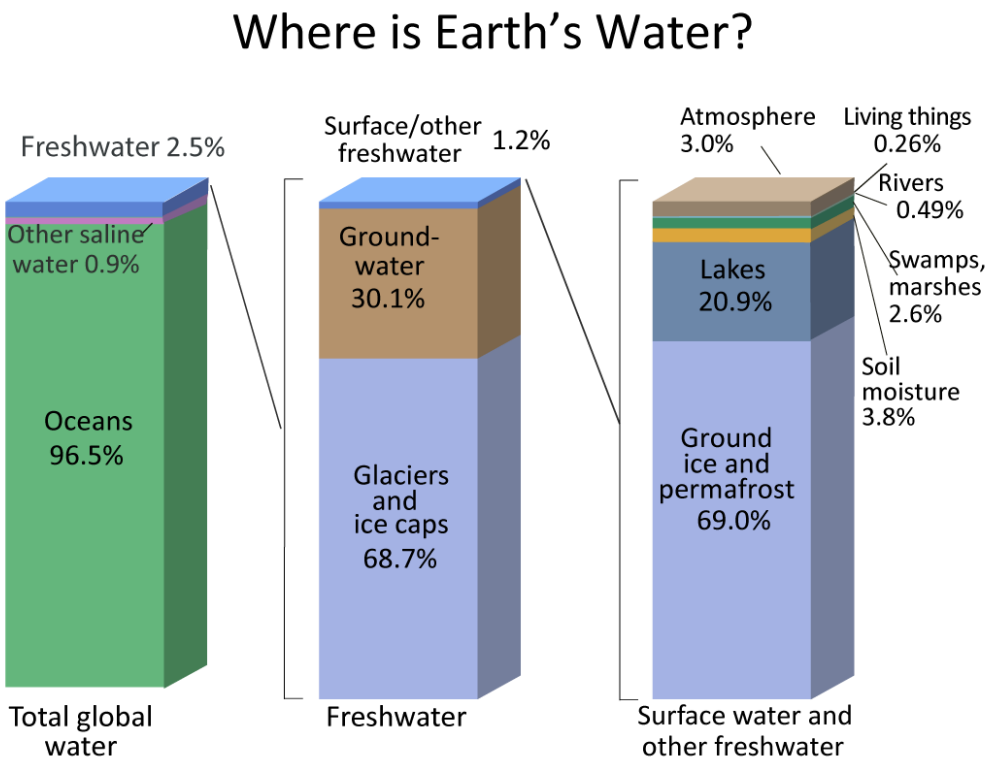
12.5: Mitigating Water Pollution


---


12: Water Supply and Water Pollution is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 12.1: Fresh Water Supply

Despite the vast volumes of water on Earth, only 2.5% is freshwater (figure **12.1.1**), 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 freshwater would be 2-3 tablespoons. Much of Earth's freshwater is trapped in glaciers and polar ice caps (figure **12.1.1**), 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 freshwater 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 **12.1.1**: 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 freshwater. Most of this freshwater (68.7%) is in glaciers and ice caps, 30.1% is groundwater, and 1.2% is surface and other freshwater. 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).

## Water Usage

Freshwater supply is one of the most important ecosystem services. In 2014, global water consumption was 3999 $km^3$ per 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 **12.1.2**).

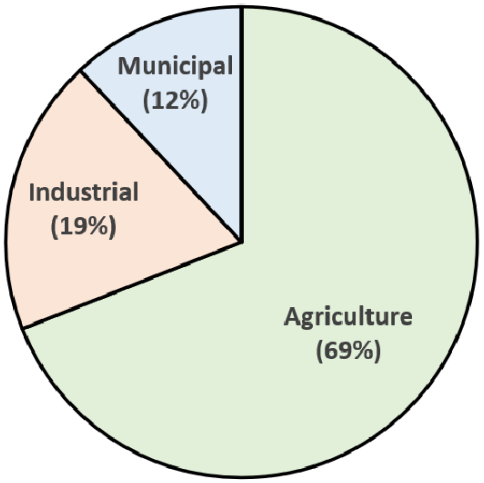



Figure 12.1.2: Global water usage by sector. Municipal refers to domestic use, such as by local businesses and households. Data from FAO. Image by Melissa Ha (CC-BY-NC).

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. There are several general categories of water use, including **offstream use**, which removes water from its source, e.g., irrigation, thermoelectric power generation (cooling electricity-producing equipment in fossil fuel, nuclear, and geothermal power plants), industry, and public supply; **consumptive use**, which is a type of offstream use where water does not return to the surface water or groundwater system immediately after use, e.g., irrigation water that evaporates or goes to plant growth; and **instream use**, which is water used but not removed from a river, mostly for hydroelectric power generation. The relative size of these three categories are instream use >> offstream use > consumptive use. In 2005, the U.S. used approximately 3,300 billion gallons per day for instream use, 410 billion gallons per day for offstream use, and 100 billion gallons per day for consumptive use. The major offstream uses of that water were thermoelectric (49%), irrigation (31%), public supply (11%), and industry (4%, see Figure 12.1.3). About 15% of the total water withdrawals in the U.S. in 2005 were saline water, which was used almost entirely for thermoelectric power generation. Almost all of the water used for thermoelectric power generation is returned to the river, lake, or ocean from where it came but about half of irrigation water does not return to the original source due to evaporation, plant transpiration, and loss during transport, e.g., leaking pipes. Total withdrawals of water in the U.S. actually decreased slightly from 1980 to 2005, despite a steadily increasing population. This is because the two largest categories of water use (thermoelectric and irrigation) stabilized or decreased over that time period due to better water management and conservation. In contrast, public supply water demand increased steadily from 1950 (when estimates began) through 2005. Approximately 77% of the water for offstream use in the U.S. in 2005 came from surface water and the rest was from groundwater (see Figure 12.1.4).

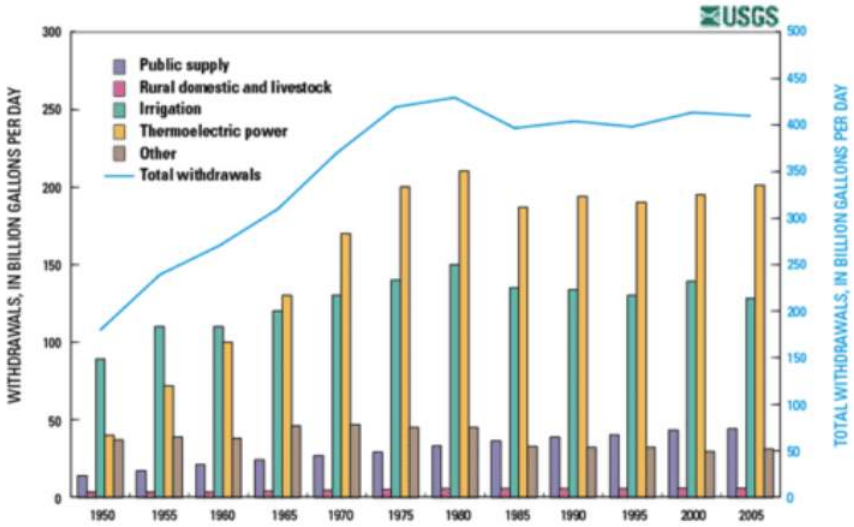



Figure **12.1.3**. Trends in Total Water Withdrawals by Water-use Category, 1950-2005 Trends in total water withdrawals in the U.S. from 1950 to 2005 by water use category, including bars for thermoelectric power, irrigation, public water supply, and rural domestic and livestock. Thin blue line represents total water withdrawals using vertical scale on right. Source: United States Geological Survey

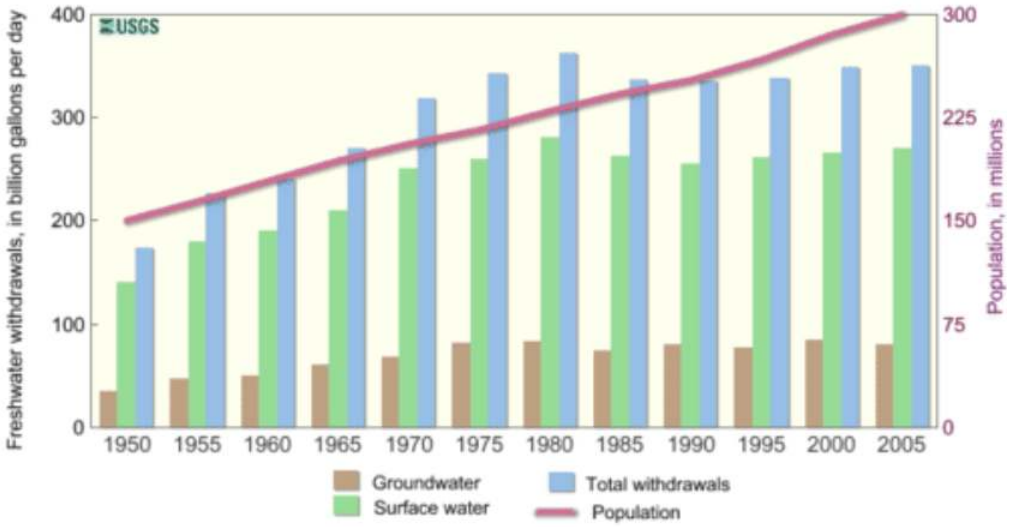



Figure **12.1.4**. Trends in Source of Fresh Water Withdrawals in the U.S. from 1950 to 2005 Trends in source of fresh water withdrawals in the U.S. from 1950 to 2005, including bars for surface water, groundwater, and total water. Red line gives U.S. population using vertical scale on right. Source: United States Geological Survey

Global total water use is steadily increasing at a rate greater than world population growth (Figure **12.1.5**). During the 20th century global population tripled and water demand grew by a factor of six. The increase in global water demand beyond the rate of population growth is due to improved standard of living without an offset by water conservation. Increased production of goods and energy entails a large increase in water demand. The major global water uses are irrigation (68%), public supply (21%), and industry (11%).

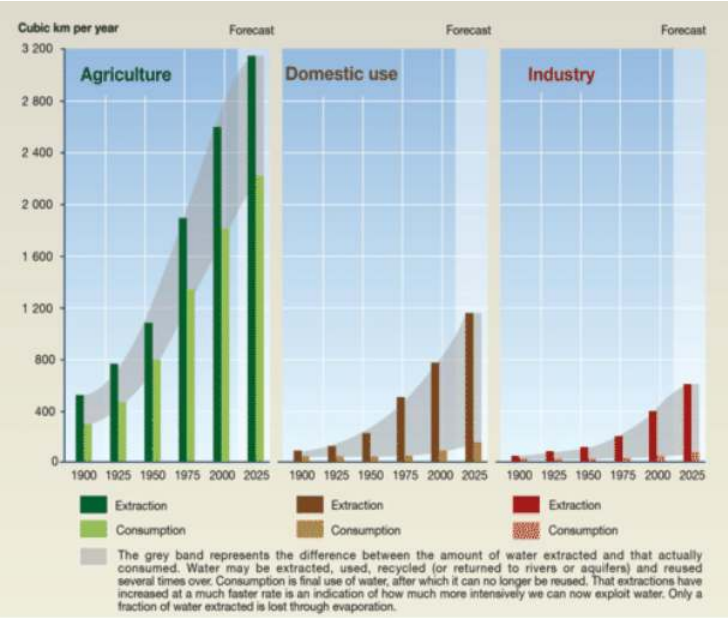



Figure 12.1.5. Trends in World Water Use from 1900 to 2000 and Projected to 2025 For each water major use category, including trends for agriculture, domestic use, and industry. Darker colored bar represents total water extracted for that use category and lighter colored bar represents water consumed (i.e., water that is not quickly returned to surface water or groundwater system) for that use category. Source: Igor A. Shiklomanow, State Hydrological Institute (SHI, St. Petersburg) and United Nations Educational, Scientific and Cultural Organisation (UNESCO, Paris), 1999

In the United States, 281 billion gallons of water were withdrawn per day in 2015, of which 82 billion gallons are fresh groundwater (figure 12.1.6). The state of California accounts for 9% of national water withdrawal (figure 12.1.7).

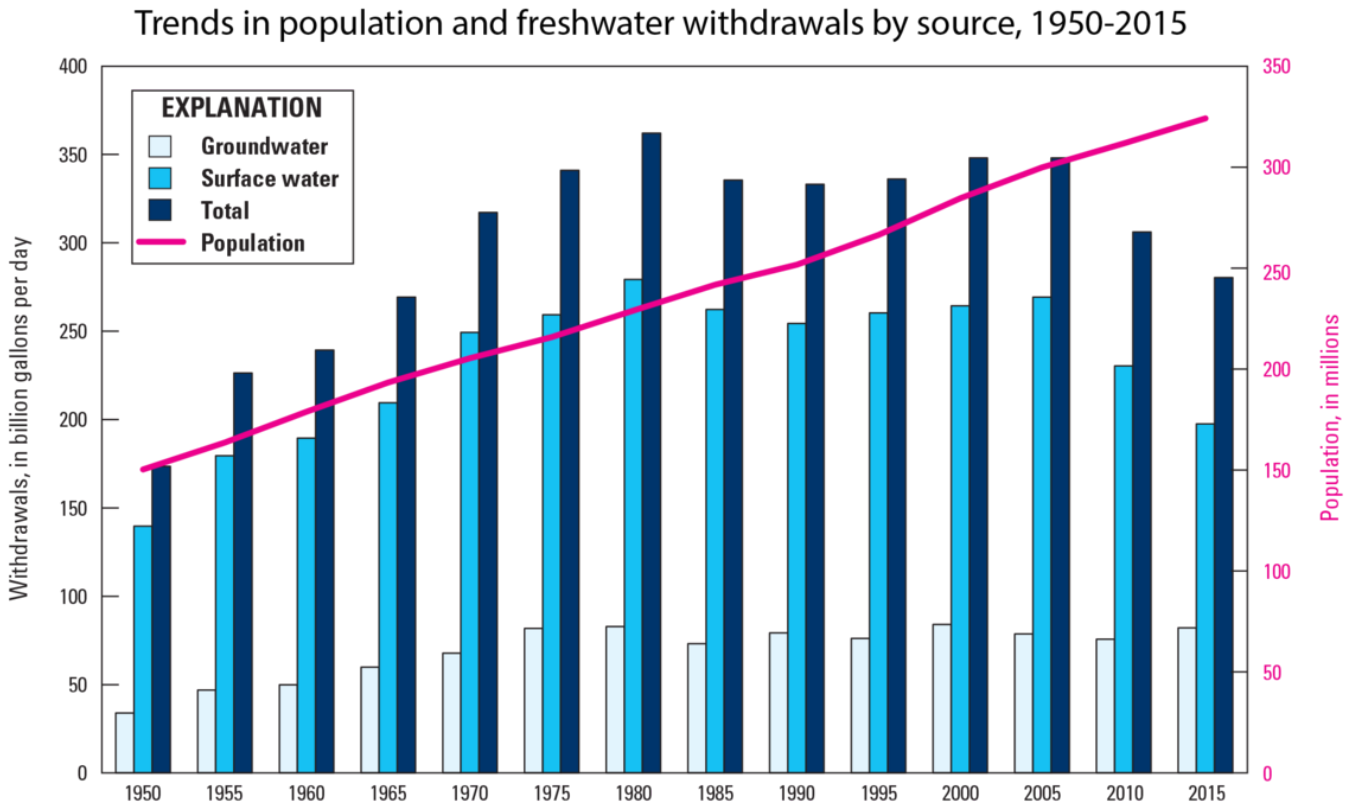



Figure 12.1.6: Average daily freshwater withdrawals (in billion gallons per day, left y-axis) over time (x-axis) in the United States. Light blue bars (left) represent groundwater withdrawals, medium blue (middle) represents withdrawals from surface water, and dark blue bars (right) represent total withdrawals 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).

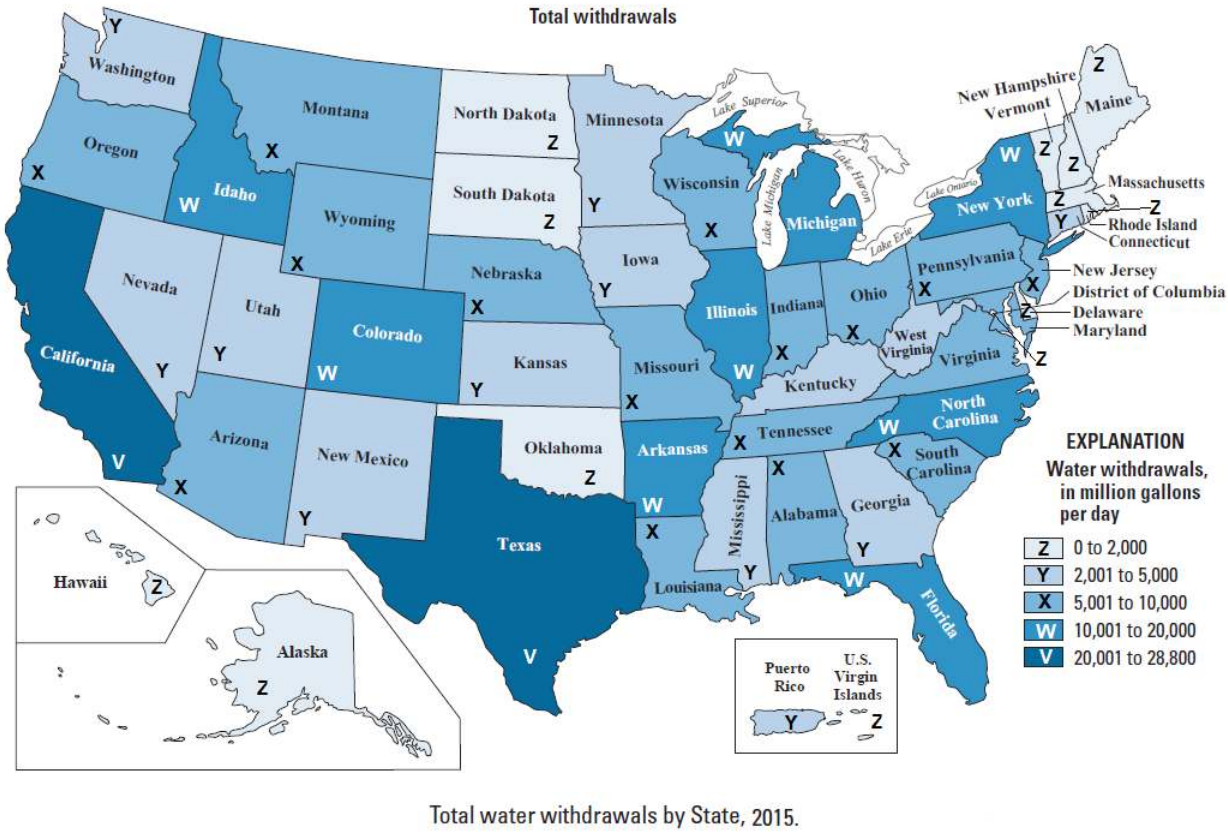



Figure 12.1.7: Total water withdrawals by state in 2015. The lightest blue (Z) indicates 0-2,000 million gallons per day, and the darkest blue (V) indicates 20,001-28,800 million gallons per day. California and Texas consume the most water, and Alaska, Hawaii, the Dakotas, Oklahoma, and some New England states consume the least. Image modified from USGS (public domain).

## Primary Freshwater Resources: Precipitation

Precipitation levels are unevenly distributed around the globe, affecting fresh water availability (figure 12.1.8). 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. 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”.

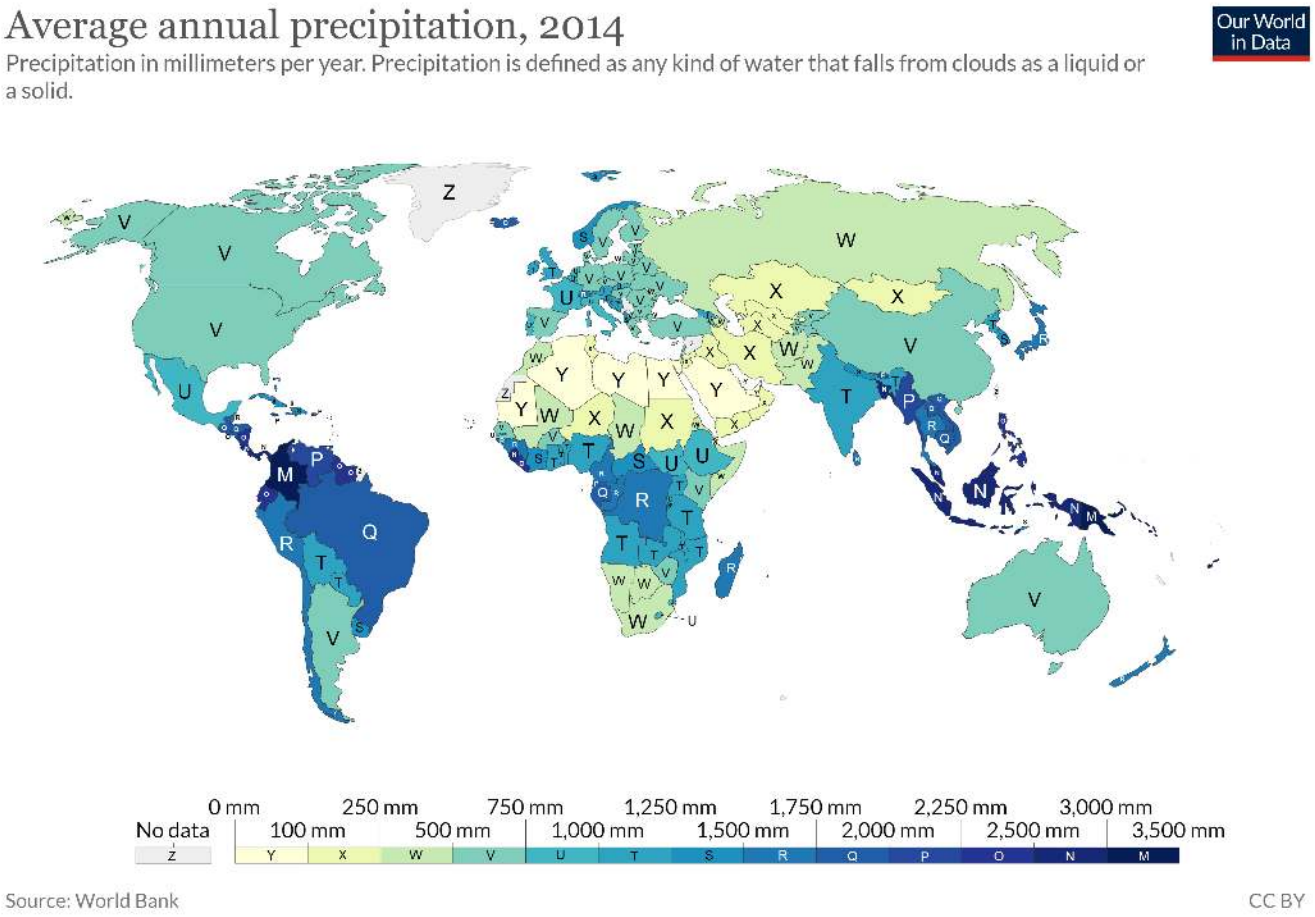



Figure 12.1.8: Annual average precipitation by country in 2014. Countries with the highest rainfall (3000-3500 mm) are Nicaragua, Costa Rica, Panama, Columbia, Sierra Leone (in West Africa), Bangladesh, and many of the islands is Southeast Asia. Countries with the least rainfall (0-100 mm) are Saudi 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 12.1.9). 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.

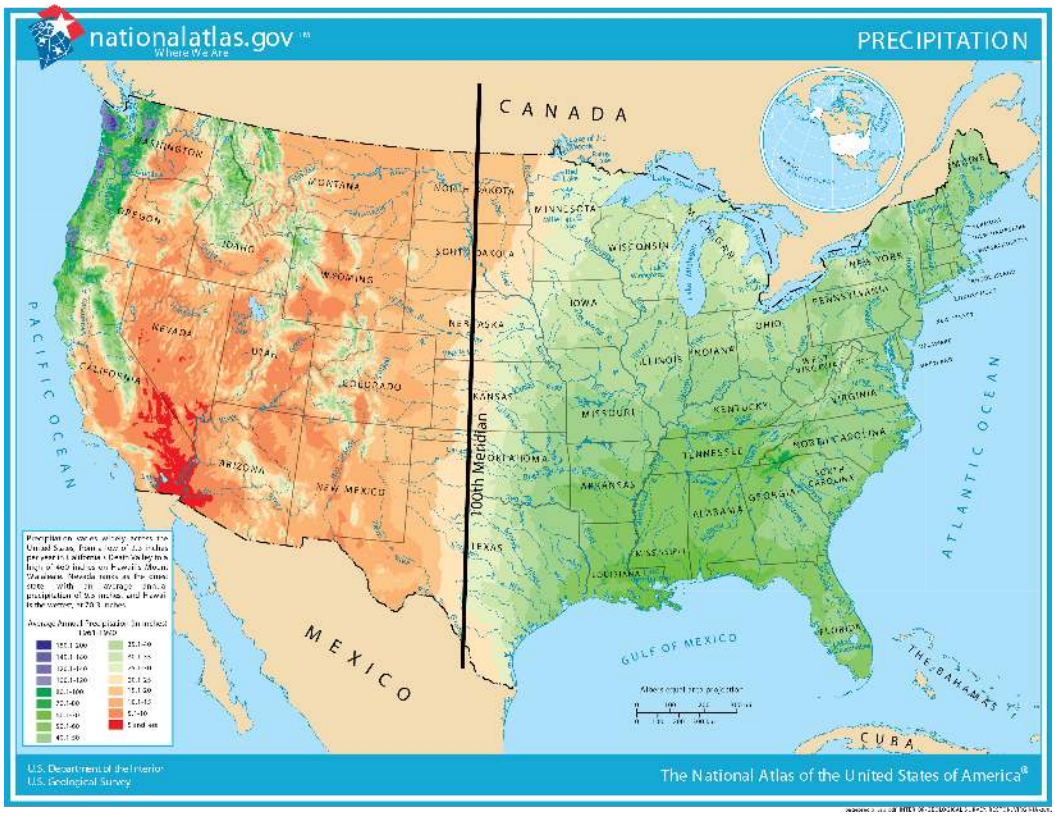



Figure 12.1.9: Distribution of precipitation in the United States. The 100th Meridian is approximately where the average precipitation transitions from relatively wet to dry. Eastern states are mostly green, indicating greater precipitation than most western states, which are mainly orange or red (indicating low precipitation). (Source: U.S. Geological Survey)

## Surface Water: Rivers and Lakes



Figure 12.1.10: Surface runoff, part of overland flow in the water cycle Source: James M. Pease at Wikimedia Commons

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 12.1.10) and seepage from surrounding land (figure 12.1.11).



Figure 12.1.11: **Groundwater Seepage** Groundwater seepage can be seen in Box Canyon in Idaho, where approximately 10 cubic meters per second of seepage emanates from its vertical headwall. *Source: NASA*

**River discharge** describes the volume of water moving through a river channel over time (see Figure 12.1.12). The relative contributions of surface runoff vs. groundwater seepage to river discharge depend on precipitation patterns, vegetation, topography, land use, and soil characteristics. Soon after a heavy rainstorm, river discharge increases due to surface runoff. The steady normal flow of river water is mainly from groundwater that discharges into the river. Gravity pulls river water downhill toward the ocean. Along the way the moving water of a river can erode soil particles and dissolve minerals, creating the river's load of moving sediment grains and dissolved ions. Groundwater also contributes a large amount of the dissolved ions in river water. The geographic area drained by a river and its tributaries is called a **watershed** or **drainage basin**. 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 12.1.12: **River Discharge** Colorado River, U.S.. Rivers are part of overland flow in the water cycle and an important surface water resource. *Source: Gonzo fan2007 at Wikimedia Commons*

The Colorado River, probably the most exploited river in the U.S., has many dams, some huge reservoirs, and several large aqueducts so that it can provide large amounts of fresh water to 7 states in the arid southwestern U.S. and Mexico. The primary use for the water is for a few large cities (Las Vegas, Phoenix, and Tuscon) and irrigation. Allocation of Colorado River water is strictly regulated. Fortunately, not all states use all of their water allocation because the total amount of allocated water is more than the typical Colorado River discharge. Colorado River water gets so saline due to evaporation along its course that the U.S. was forced to build a desalination plant near the border with Mexico so that it could be used for drinking and irrigation. The wetlands of the Colorado River delta and its associated ecosystem have been sadly degraded by the water overuse; some years, no river flow even reaches the ocean.

Lakes can also be an excellent source of fresh water for human use. They usually receive water from surface runoff and groundwater discharge. They tend to be short-lived on a geological time-scale because they are constantly filling in with sediment supplied by rivers. Lakes form in a variety of ways including glaciation (Great Lakes, North America, See Figure **12.1.13**:), recent tectonic uplift (Lake Tanganyika, Africa), and volcanic eruptions (Crater Lake, Oregon). People also create artificial lakes (**reservoirs**) by damming rivers. Large changes in climate can result in major changes in a lake's size. As Earth was coming out of the last Ice Age about fifteen thousand years ago, the climate in the western U.S. changed from cool and moist to warm and arid, which caused more than 100 large lakes to disappear. The Great Salt Lake in Utah is a remnant of a much larger lake called Lake Bonneville.



Figure **12.1.13**: **Great Lakes from Space** The Great Lakes hold 21% of the world's surface fresh water. Lakes are an important surface water resource. *Source:* *SeaWiFS Project, NASA/Goddard Space Flight Center, and ORBIMAGE*

## Surface Water: Glaciers

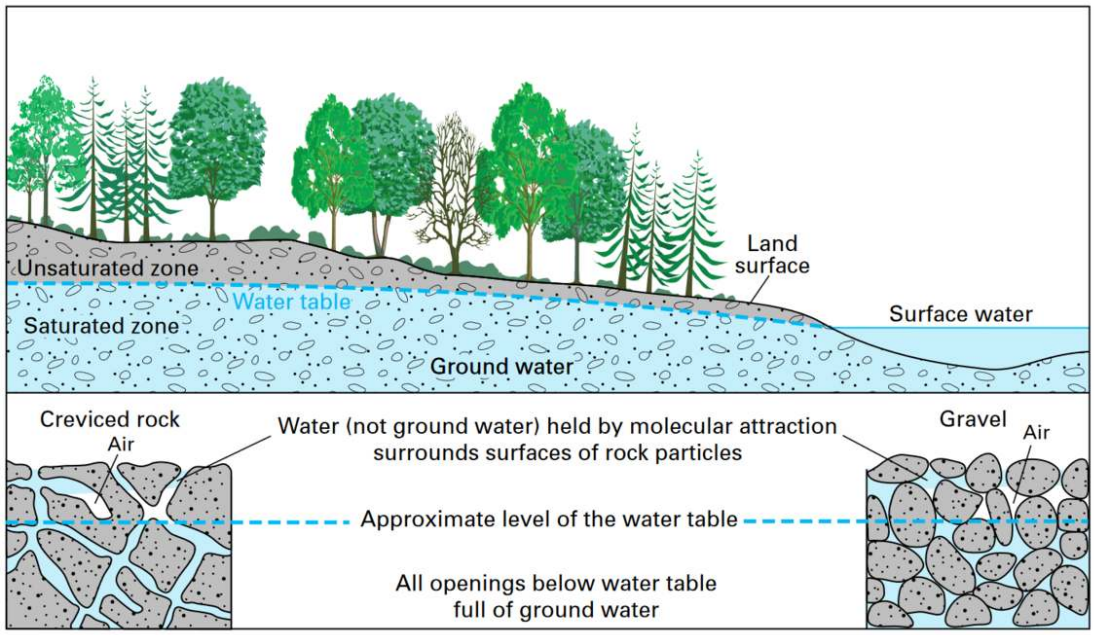
Although glaciers represent the largest reservoir of fresh water, they generally are not used as a water source because they are located too far from most people (see Figure **12.1.14**). Melting glaciers do provide a natural source of river water and groundwater. A large part of the world's population uses surface water which originally comes from melting snow and glaciers in mountain regions. 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 12.1.14: *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*

## Groundwater Resources

**Groundwater** is water located in small spaces, called **pore space**, between mineral grains and fractures in subsurface earth materials (rock or sediment, i.e., loose grains). Although most people in the world use surface water, groundwater is a much larger reservoir of usable freshwater, containing more than 30 times more water than rivers and lakes combined. 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. Groundwater is not located in underground rivers or lakes except where there are caves, which are relatively rare. A large area of subsurface, porous rock unit or sediment that holds extractable groundwater is an **aquifer**. Between the land surface and the depth where there is groundwater is the **unsaturated zone**, where pore spaces contain only air and water films on mineral grains (figure 12.1.15). The **saturated zone** of an aquifer is where groundwater completely fills pore spaces in earth materials (figure 12.1.15). The **water table** is the uppermost level at which the pores are fully saturated with water (figure 12.1.15).




How ground water occurs in rocks.

Figure 12.1.15: 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. Groundwater is in constant motion due to interconnection between pore spaces. Examples of aquifers are earth materials with abundant, large, well-connected pore spaces such as sand, gravel, uncemented sandstone, and any highly fractured rock. 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.

**Hydraulic conductivity** is another common measure of the connectedness of pore spaces and is a function of both permeability and fluid properties. Because it considers fluid properties, hydraulic conductivity is used by both petroleum geologists and hydrogeologists to describe the production capability of oil reservoirs and aquifers. A high hydraulic conductivity indicates a rapid transmission of fluid through an aquifer.

Most groundwater originates from rain or snowmelt, which infiltrates the ground and moves downward until it reaches the saturated zone. Other sources of groundwater include seepage from surface water (lakes, rivers, reservoirs, and swamps), surface water deliberately pumped into the ground, irrigation, and underground wastewater treatment systems, i.e., septic tanks. **Recharge areas** are locations where surface water infiltrates the ground rather than running off into rivers or evaporating (figure 12.1.16). 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.

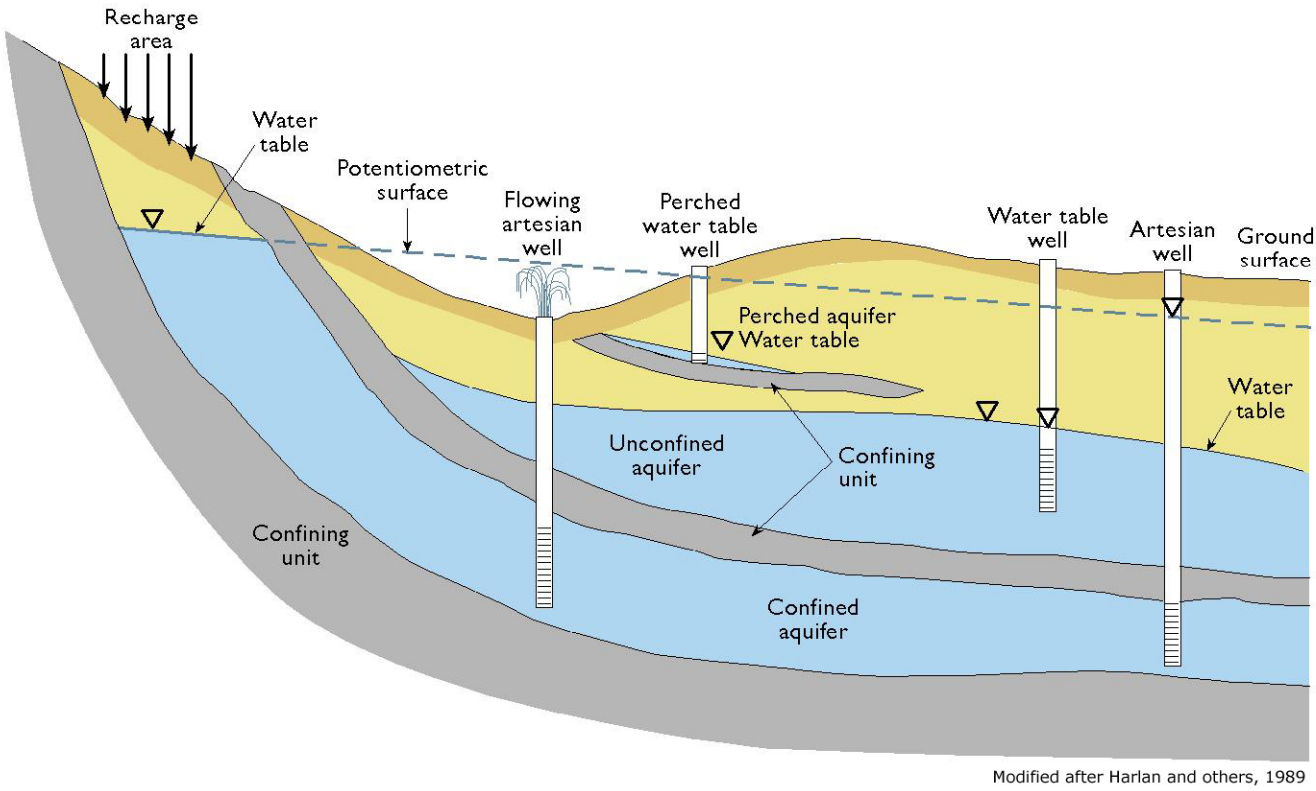



Figure 12.1.16): **Schematic Cross Section of Aquifer Types** This figure shows different types of aquifers and water wells, including unconfined aquifer, confined aquifer, water table well, artesian well, and flowing artesian well. Point of triangle is water level in each well and water table in other parts of figure. Water level in artesian well is at potentiometric surface and above local water table (dashed blue line) due to extra pressure on groundwater in confined aquifer. Water in flowing artesian well moves above land surface. *Source:* *Colorado Geological Survey*

Another reason for groundwater movement is that the surface of the water table commonly is not completely flat but mimics the topography of the land surface, especially in humid climates. There is "topography" to the water table because groundwater moves slowly through rock and soil, so it builds up in higher elevation areas. An earth material with low hydraulic conductivity is an **aquitard**. Examples of aquitards include clay, shale (sedimentary rock with abundant clay), and igneous and metamorphic rock, if they contain few fractures. In fact, when groundwater flows slowly through aquitards and deep underground, it can take many thousands of years to move relatively short distances. An **unconfined aquifer** has no aquitard above it and, therefore, it is exposed to the atmosphere and surface waters through interconnected pores (Figure 12.1.16). In an unconfined aquifer, groundwater flows because of gravity to lower water table levels, where it eventually may discharge or leave the groundwater flow system. Discharge areas include rivers, lakes, swamps, reservoirs, water wells, and **springs** (Figure 12.1.16). Springs are rivers that emerge from underground due to an abrupt intersection of the land surface and the water table caused by joints, caves, or faults that bring permeable earth materials to the surface. A **confined aquifer** is bounded by aquitards below and above, which prevents recharge from the surface immediately above. Instead, the major recharge occurs where the confined aquifer intercepts the land surface, which may be a long distance from water wells and discharge areas (Figure 12.1.16). Confined aquifers are commonly inclined away from recharge areas, so groundwater in a confined aquifer is under greater-than-atmospheric pressure due to the weight of water in the upslope direction. Similar to river discharge, groundwater discharge describes the volume of water moving through an aquifer over time. Total groundwater discharge depends on the permeability of the earth material, the pressure that drives groundwater flow, and the size of the aquifer. It is important to determine groundwater discharge to evaluate whether an aquifer can meet the water needs of an area.

## Human Impacts on Groundwater

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. Most shallow water wells are drilled into unconfined aquifers. These are called **water table wells** because the water level in the well coincides with the water table (Figure 12.1.16). 90% of all aquifers for water supply are unconfined aquifers composed of sand or gravel. To produce water from a well, you simply need to drill a hole that reaches the saturated zone and then pump water to the surface. Attempting to pump water from the unsaturated zone is like drinking root beer with a straw immersed only in the foam at the top.

To find a large aquifer for a city, hydrogeologists (geologists who specialize in groundwater) use a variety of information including knowledge of earth materials at the surface and sub-surface as well as test wells. Some people search for water by dowsing, where someone holds a forked stick or wire (called a divining rod) while walking over an area. The stick supposedly rotates or deflects downward when the dowser passes over water. Controlled tests show that a dowser's success is equal to or less than random chance. Nevertheless, in many areas water wells are still drilled on dowser's advice sometimes for considerable money. There is no scientific basis to dowsing.

Wells into confined aquifers typically are deeper than those into unconfined aquifers because they must penetrate a confining layer. The water level in a well drilled into a confined aquifer, which is an **artesian well**, (Figure 12.1.16), moves above the local water table to a level called the potentiometric surface because of the greater pressure on the groundwater. Water in a flowing well (Figure 12.1.17) moves all of the way to the land surface without pumping.



Figure 12.1.17: **A Flowing Well** Flowing artesian well where water moves above the land surface due to extra pressure on the groundwater in a confined aquifer. *Source: Environment Canada*

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 12.1.18). 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.

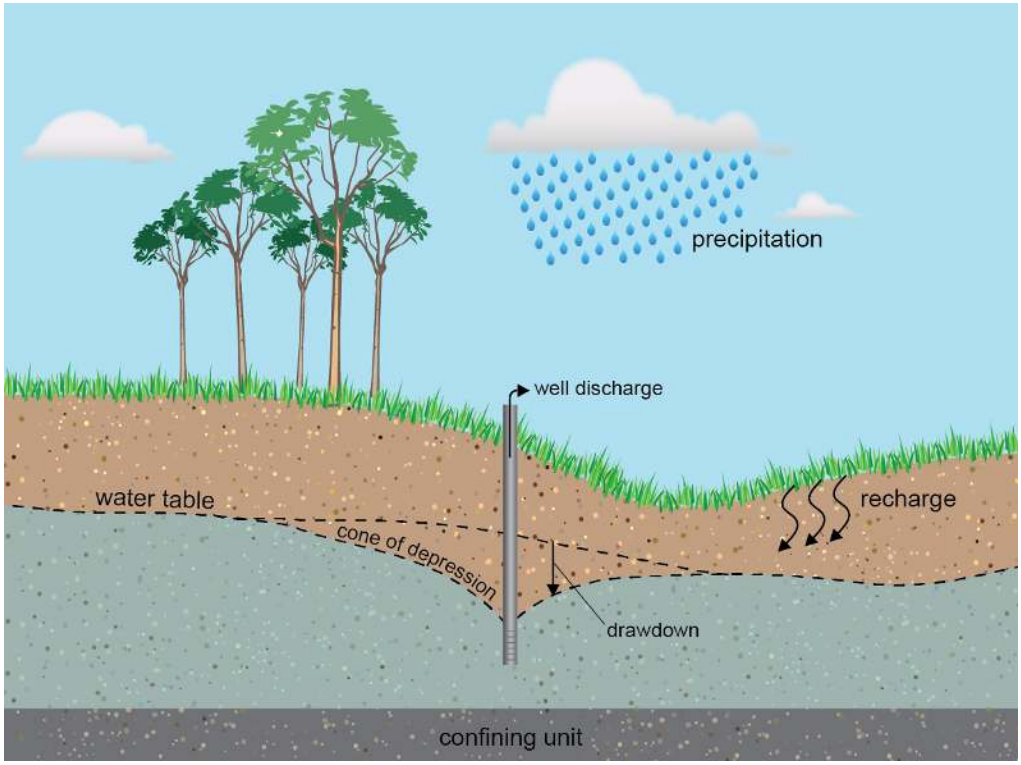



Figure 12.1.18: As groundwater is extracted from the well, the water table of the aquifer dips, forming a cone of depression. Drawdown is between the water table and the cone of depression. Precipitation is represented by water droplets falling from a cloud. It soaks into the ground, recharging the aquifer. At the bottom of the aquifer is the confining unit, which is impermeable to water. Image by Tara Gross/USGS (public domain).

A confined aquifer tends to be depleted from groundwater pumping more quickly than an unconfined aquifer, assuming similar aquifer properties and precipitation levels. This is because confined aquifers have smaller recharge areas, which may be far from the pumping well. Conversely, an unconfined aquifer tends to be more susceptible to pollution because it is hydrologically connected to the surface, which is the source of most pollution. Urban development usually worsens groundwater mining because natural recharge rates drop with the proliferation of impermeable pavement, buildings, and roads. Extensive groundwater pumping around Chicago has created a gigantic cone of depression there. Because the water table dropped up to 250 m (800 ft) in the area (Figure 12.1.19), many local public water suppliers have switched to Lake Michigan water. Chicago is fortunate to have a large alternate supply of fresh water; many arid locations don't have that luxury.




Figure 12.1.19: **Drop in Water Table in a Confined Aquifer in the Area of Chicago, Illinois and Milwaukee, Wisconsin, U.S. from 1864 - 1980** *Source: United States Geological Survey*

Another water resource problem associated with groundwater mining is **saltwater intrusion**, where overpumping of fresh water aquifers near ocean coastlines causes saltwater 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. Saltwater intrusion is a significant problem in many coastal areas of the U.S. including Long Island, New York; Cape Cod, Massachusetts; and southeastern and Gulf Coastal states.



Figure 12.1.20: Example of an earth fissure in Cedar City.

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 (Figure 12.1.20 – 21). Because the pores in the aquifer collapse as subsidence occurs, this permanently reduces the capacity of the aquifer to hold water in the future. In many places, water actually helps hold up the skeleton of the aquifer by the water pressure exerted on the grains in an aquifer. If pore pressure decreases because of groundwater mining, the aquifer can compact, causing the surface of the ground to sink. Areas especially susceptible to this effect are aquifers made of unconsolidated sediments. Unconsolidated sediments with multiple layers of clay and other fine-grained material are at higher risk because clay can compact considerably when drained of water [36, 37]. In many cases, the amount of compaction in one area will be greater than the amount of compaction in an adjacent area. The different amounts of compaction in areas that are next to each other can cause the land to offset and develop cracks and fissures.




Figure 12.1.21: The pole shows subsidence from groundwater pumping for irrigation in the San Joaquin Valley of California. The dates marked on the pole show land elevation in the past.

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 12.1.22). 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.

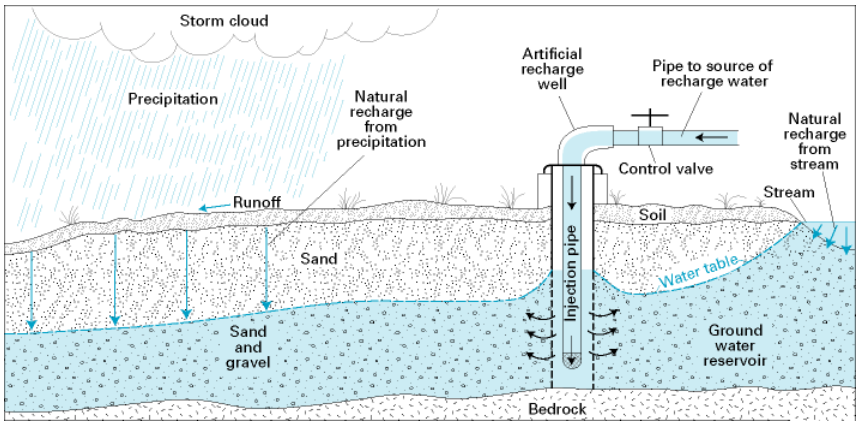



Figure 12.1.22: Groundwater can be replenished through natural recharge areas as well as artificial injection wells. In this diagram, precipitation emerges from a storm cloud. Some of the water washes away as runoff, and some water infiltrates through the sand (natural recharge from precipitation). The 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).

## Water Supply Crisis

The **water crisis** refers to a global situation where people in many areas lack access to sufficient water, clean water, or both. This section describes the global situation involving water shortages, also called **water stress**. In general, water stress is greatest in areas with very low precipitation (major deserts), large population density (e.g., India), or both. Future global warming could worsen the water crisis by shifting precipitation patterns away from humid areas and by melting mountain glaciers that recharge rivers downstream. Melting glaciers will also contribute to rising sea level, which will worsen saltwater intrusion in aquifers near ocean coastlines.

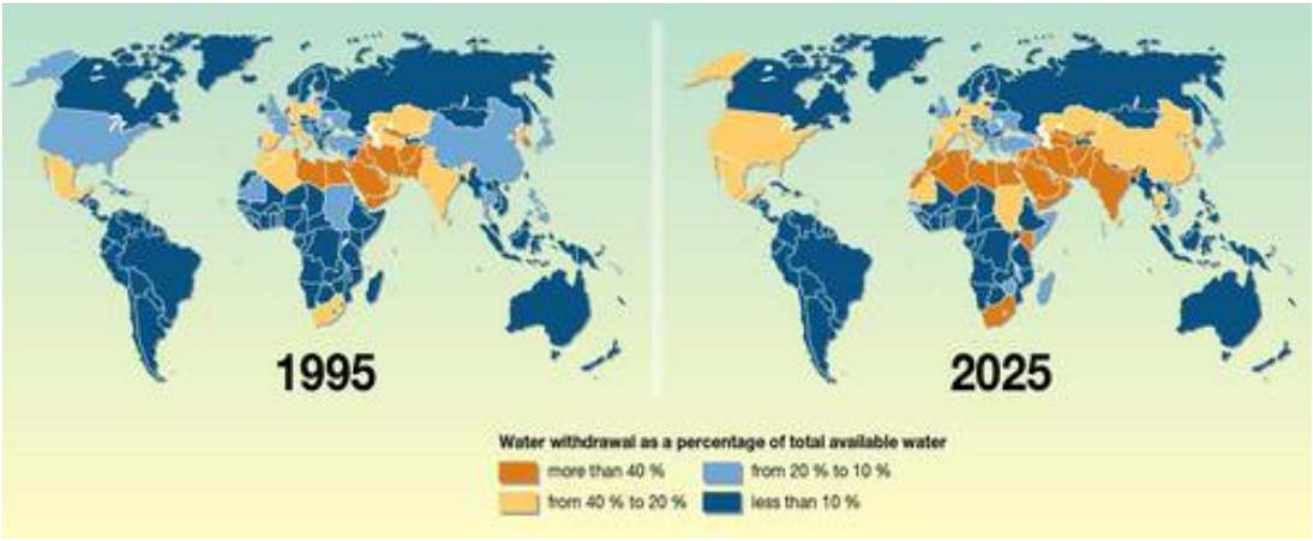



Figure **12.1.23**. Countries Facing Water Stress in 1995 and Projected in 2025 Water stress is defined as having a high percentage of water withdrawal compared to total available water in the area. Source: Philippe Rekacewicz (Le Monde diplomatique), February 2006

According to a 2006 report by the United Nations Development Programme, 700 million people (11% of the world's population) lived with water stress. Most of them live in the Middle East and North Africa. By 2025, the report projects that more than 3 billion people (about 40% of the world's population) will live in water-stressed areas with the large increase coming mainly from China and India. The water crisis will also impact food production and our ability to feed the ever-growing population. We can expect future global tension and even conflict associated with water shortages and pollution. Historic and future areas of water conflict include the Middle East (Euphrates and Tigris River conflict among Turkey, Syria, and Iraq; Jordan River conflict among Israel, Lebanon, Jordan, and the Palestinian territories), Africa (Nile River conflict among Egypt, Ethiopia, and Sudan), Central Asia (Aral Sea conflict among Kazakhstan, Uzbekistan, Turkmenistan, Tajikistan, and Kyrgyzstan), and south Asia (Ganges River conflict between India and Pakistan).

**Water scarcity** refers to water shortages, which can be physical or economic (Figure **12.1.24**). **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 and two billion people lack access to improved sanitation as simple as a pit latrine , and three billion people lack a facility to wash their hands. 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.

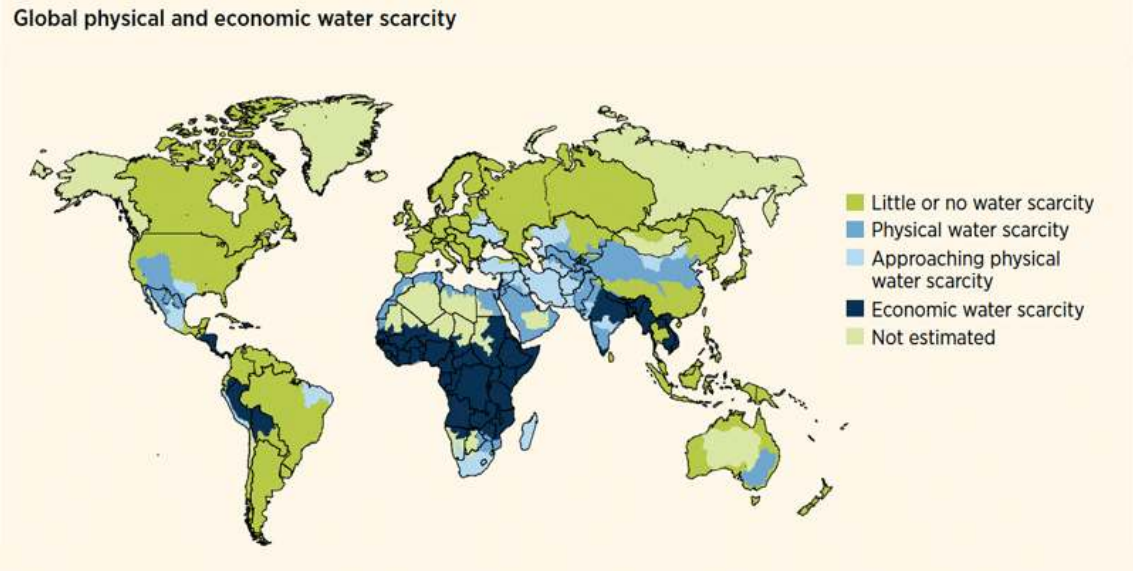



Figure 12.1.24: Regions facing or approaching water scarcity in 2012. Physical water stress occurs in the southwestern United States, the Middle East, North Africa, southern India, and southeast Australia. Economic water stress occurs in Central and East Africa, western India, and Southeast Asia. Image by World Water Assessment Programme (WWAP; public domain).

## Sustainable Solutions to the Water Supply Crisis?

The current and future water crisis described above requires multiple approaches to extending our fresh water supply and moving towards sustainability. Some of the longstanding traditional approaches include dams and aqueducts.

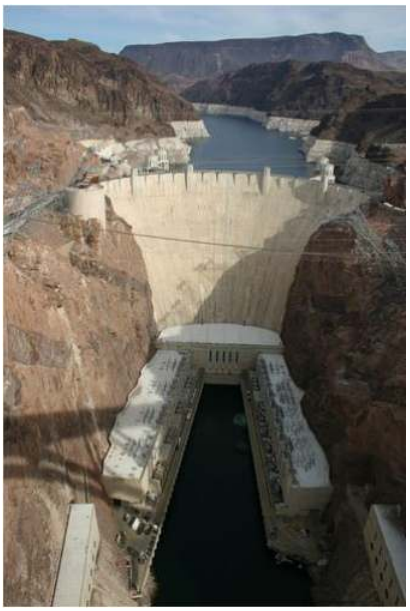


Figure 12.1.25. Hoover Dam, Nevada, U.S. Hoover Dam, Nevada, U.S.. Behind the dam is Lake Mead, the largest reservoir in U.S.. White band reflects the lowered water levels in the reservoir due to drought conditions from 2000 – 2010. Source: Cygnusloop99 at Wikimedia Commons

**Reservoirs** that form behind dams in rivers can collect water during wet times and store it for use during dry spells. 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, downstream river channel erosion, and impact on the ecosystem including a change from a river to lake habitat and interference with migration and spawning of fish.

**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. Owens Lake remains almost completely dry, but Mono Lake has recovered more significantly due to legal intervention.



Figure 12.1.26. The California Aqueduct California Aqueduct in southern California, U.S. Source: David Jordan at en.Wikipedia




Figure 12.1.27. **Map of California Aqueducts** Map of California aqueducts that bring water to southern California from central and northern California and from the Colorado River to the east. *Source: Central Basin Municipal Water District*

**Rainwater harvesting** involves catching and storing rainwater before it reaches the ground. Figure 12.1.28 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 12.1.29).

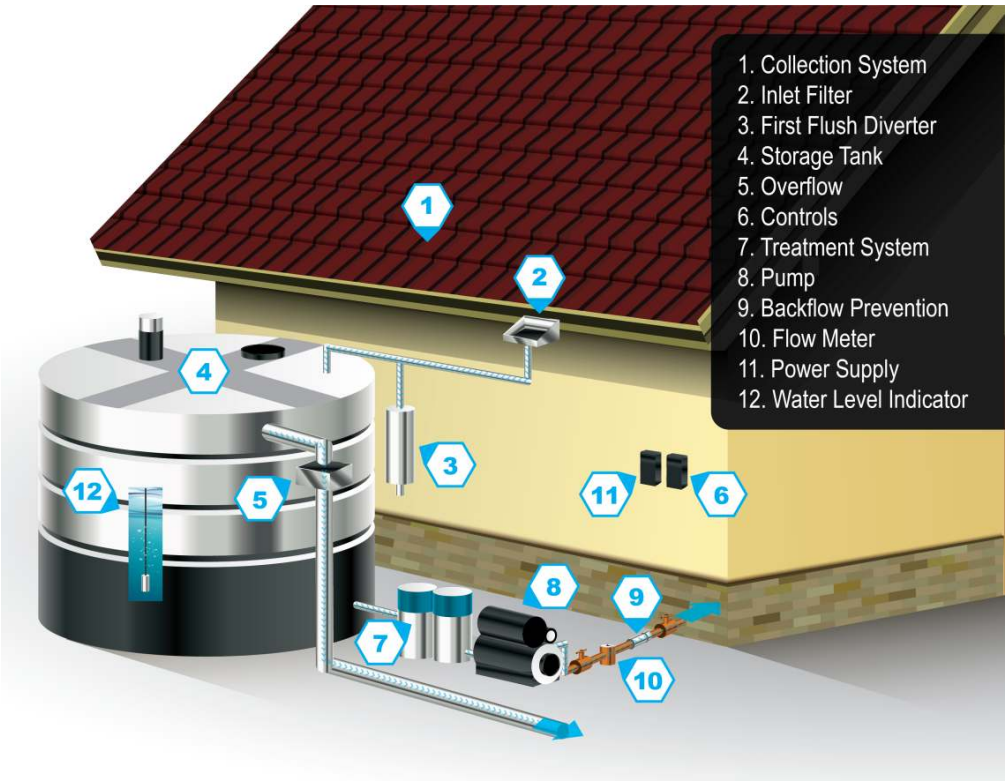



Figure **12.1.28**: The components of a rainwater harvesting system include (1) the collection system consisting of the roof surface and gutters that capture rain water and send it to (2) the inlet filter, a screen that catches large debris. (3) The first flush diverter removes debris not captured by 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 12.1.29: 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).

One method that can actually increase the amount of fresh water on Earth is **desalination**, which involves removing dissolved salt from seawater or saline groundwater (figure 12.1.30). 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, and electrodialysis. All of these procedures are moderately to very expensive and require considerable energy input, making the water produced much more expensive than fresh water 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 12.1.30: Port Stanvac Desalination Plant under construction in South Australia. Image by Vmenkov (CC-BY-SA).

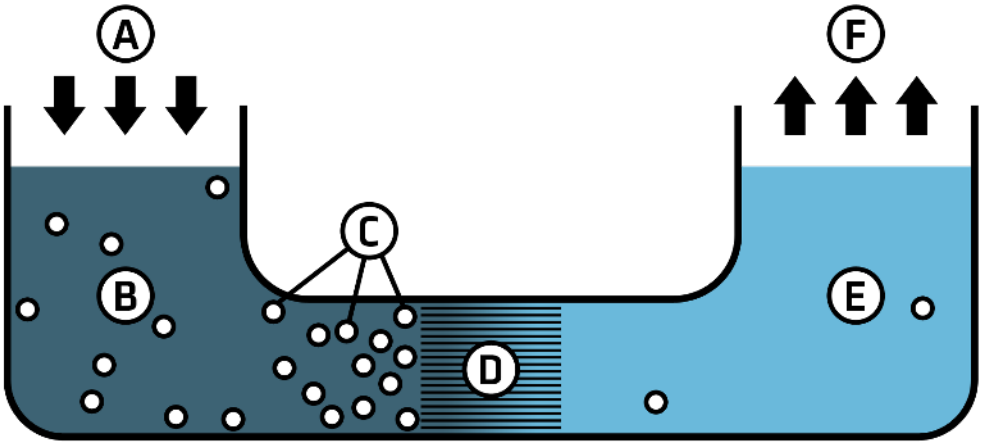



Figure 12.1.31: Reverse osmosis is one method of desalination. In this process, pressure (A) is applied to salt water, which contains ions (charged atoms or molecules) and other contaminants (C). The water is forced through a semipermeable membrane (D), through which the contaminants cannot pass. Clean freshwater that is safe to drink (E) is then distributed (F). Image by Colby Fisher (CC-BY-SA).

**Water recycling** refers to reusing water for appropriate purposes such as agriculture, municipal water supply, industrial processes, and environmental restoration (figure 12.1.32). 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.

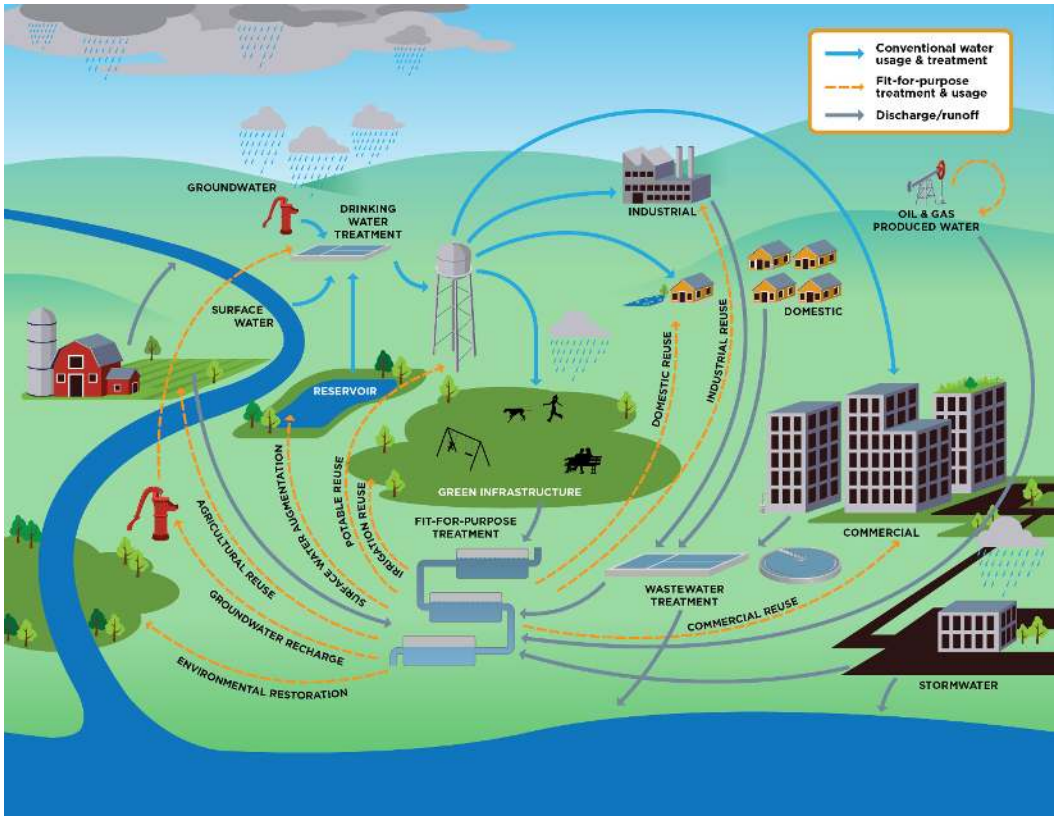


Figure 12.1.32: Flowchart of water reuse. Surface water, groundwater, and water from reservoirs undergoes drinking water treatment. From there, it serves commercial, industrial, and domestic purposes and supports green infrastructure. Used water is treated to the extent needed for its purpose (fit-for-purpose 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** 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 12.1.33). 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 12.1.33: 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).

Another important technique is **efficient irrigation,** which is extremely important because irrigation accounts for a much larger water demand than public water supply. 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 that minimize losses due to evaporation, no-till farming that reduces evaporative losses by covering the soil, and reusing treated wastewater from sewage treatment plants. Recycled wastewater has also been used to recharge aquifers.

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.)

### Suggested Supplementary Reading:

Weiss, K.R. 2018. Drying Lakes. *National Geographic.* March. p. 108-133.

*This article documents how many lakes across the globe are drying up, the reasons why, and the effect on humans. Overuse and a warming climate threaten lakes that provide sustenance and jobs for humans, while also providing critical habitat for animals.*

## 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.

8. Maupin MA, Kenny JF, Hutson SS, et al (2014) Estimated use of water in the United States in 2010. US Geological Survey

9. Gleick PH (1993) Water in Crisis: A Guide to the World's Fresh Water Resources, 1 edition. Oxford University Press

10. Hadley G (1735) Concerning the cause of the general trade-winds: By Geo. Hadley, Esq; FRS. Philosophical Transactions 39:58–62

11. Lorenz EN (1955) Available potential energy and the maintenance of the general circulation. Tell'Us 7:157–167

12. Halvorson SF, James Steenburgh W (1999) Climatology of lake-effect snowstorms of the Great Salt Lake. University of Utah

13. Powell JW (1879) Report on the lands of the arid region of the United States with a more detailed account of the land of Utah with maps. Monograph

14. U.S. Geological Survey (2012) Snowmelt - The Water Cycle, from USGS Water-Science School

34. Freeze AR, Cherry JA (1979) Groundwater, 1 edition. Prentice Hall

35. Reese RS (2014) Review of Aquifer Storage and Recovery in the Floridan Aquifer System of Southern Florida

36. Galloway D, Jones DR, Ingebritsen SE (1999) Land subsidence in the United States. U.S. Geological Survey

37. Knudsen T, Inkenbrandt P, Lund W, et al (2014) Investigation of land subsidence and earth fissures in Cedar Valley, Iron County, Utah: Utah Geological Survey


## Contributors and Attributions

Modified by Kyle Whittinghill and Melissa Ha from the following sources

- 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 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.
- Biogeochemical Cycles from *Biology 2e* by OpenStax (CC-BY). Access for free at openstax.org.
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.
- Basic Information about Water Reuse by United States Environmental Protection Agency (public domain)

---


This page titled 12.1: Fresh Water Supply is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **7.1: Water Cycle and Fresh Water Supply** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.
- **7.2: Water Supply Problems and Solutions** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# 12.2: Case Study - The Aral Sea - Going, Going, Gone

The Aral Sea is a lake located east of the Caspian Sea between Uzbekistan and Kazakhstan in central Asia (figure **12.2.*b***). This area is part of the Turkestan desert, which is the fourth largest desert in the world; it is produced from a rain shadow effect by Afghanistan's high mountains to the south. Due to the arid and seasonally hot climate there is extensive evaporation and limited surface waters in general. Summer temperatures can reach 60°C (140°F)! The water supply to the Aral Sea is mainly from two rivers, the Amu Darya and Syr Darya, which carry snow melt from mountainous areas.




Figure **12.2.2**. Map of Aral Sea Area Map shows lake size in 1960 and political boundaries of 2011. Countries in yellow are at least partially in Aral Sea drainage basin. Source: Wikimedia Commons

In the early 1960s the then-Soviet Union diverted the Amu Darya and Syr Darya Rivers for irrigation of one of the driest parts of Asia to produce rice, melons, cereals, and especially cotton. The Soviets wanted cotton or "white gold" to become a major export. They were successful and today Uzbekistan is one of the world's largest exporters of cotton. Unfortunately this action essentially eliminated any river inflow to the Aral Sea and caused it to disappear almost completely (Figure **12.2.2**).

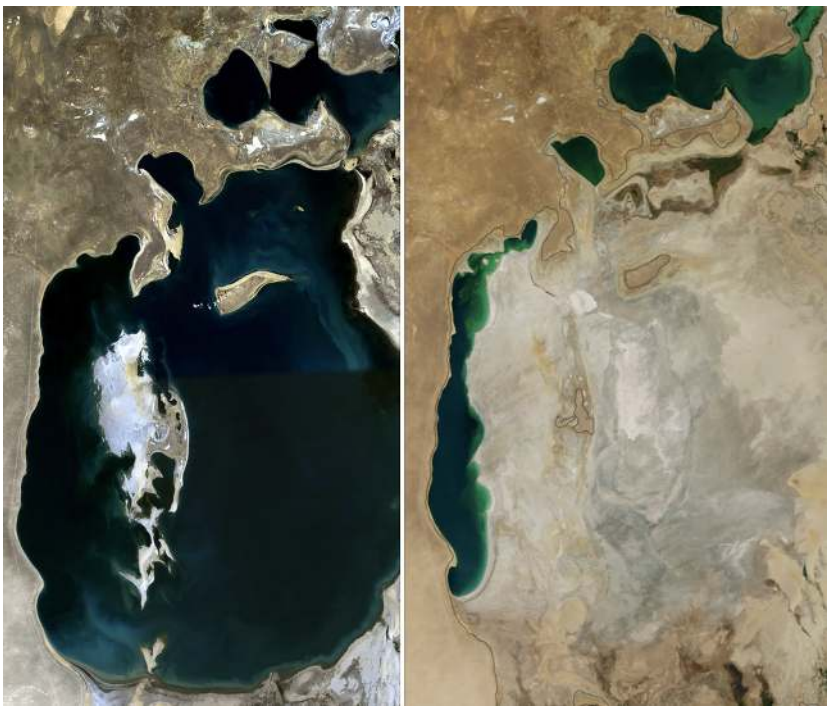


Figure **12.2.2** A comparison of the Aral Sea in 1989 (left) and 2014 (right). Credit: This work is in the Public Domain, CC0

In 1960, Aral Sea was the fourth largest inland water body; only the Caspian Sea, Lake Superior, and Lake Victoria were larger. Since then, it has progressively shrunk due to evaporation and lack of recharge by rivers. Before 1965, the Aral Sea received 2060 $km^3$ of fresh water per year from rivers and by the early 1980s it received none. By 2007, the Aral Sea shrank to about 10% of its original size and its salinity increased from about 1% dissolved salt to about 10% dissolved salt, which is 3 times more saline than seawater. These changes caused an enormous environmental impact. A once thriving fishing industry is dead as are the 24 species of fish that used to live there; the fish could not adapt to the more saline waters. The current shoreline is tens of kilometers from former fishing towns and commercial ports. Large fishing boats lie in the dried up lakebed of dust and salt (Figure **12.2.2**). A frustrating part of the river diversion project is that many of the irrigation canals were poorly built, allowing abundant water to leak or evaporate. An increasing number of dust storms blow salt, pesticides, and herbicides into nearby towns causing a variety of respiratory illnesses including tuberculosis.



Figure **12.2.3**. This abandoned ship lies in a dried up lake bed that was the Aral Sea near Aral, Kazakhstan Source: Staecker at Wikimedia Commons

The wetlands of the two river deltas and their associated ecosystems have disappeared. The regional climate is drier and has greater temperature extremes due to the absence of moisture and moderating influence from the lake. In 2003, some lake restoration work began on the northern part of the Aral Sea and it provided some relief by raising water levels and reducing salinity somewhat. The southern part of the Aral Sea has seen no relief and remains nearly completely dry. The destruction of the Aral Sea is one of the planet's biggest environmental disasters and it is caused entirely by humans. Lake Chad in Africa is another example of a massive lake that has nearly disappeared for the same reasons as the Aral Sea. Aral Sea and Lake Chad are the most extreme examples of large lakes destroyed by unsustainable diversions of river water. Other lakes that have shrunk significantly due to human diversions of water include the Dead Sea in the Middle East, Lake Manchar in Pakistan, and Owens Lake and Mono Lake, both in California.

## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.
- Case Study: The Aral Sea - Going, Going, Gone from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (licensed under CC-BY). Download for free at CNX.
- **Case Study - The Aral Sea - Going, Going, Gone** by Matthew R. Fisher, is licensed CC BY

---

This page titled 12.2: Case Study - The Aral Sea - Going, Going, Gone is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Matthew R. Fisher (OpenOregon) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.5: Case Study - The Aral Sea - Going, Going, Gone** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# 12.3: Water Pollutants and Their Sources

Although natural processes such as volcanic eruptions or evaporation sometimes can cause water pollution, most pollution is derived from human, land-based activities. Pollutants can spread through different water reservoirs, as the water carrying them progresses through stages of the water cycle (figure **12.3.1**). Water can be contaminated by various human activities or by existing natural features, like mineral-rich geologic formations. Agricultural activities, industrial operations, landfills, animal operations, and small and large scale sewage treatment processes, among many other things, all can potentially contribute to contamination. As water runs over the land or infiltrates into the ground, it dissolves material left behind by these potential contaminant sources. The risks and type of remediation for a contaminant depend on the type of chemicals present.

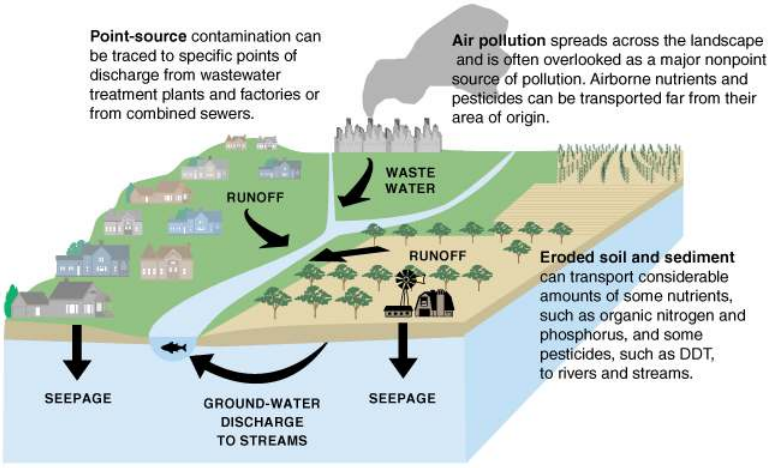



**Figure 12.3.1:** Sources of some water pollutants and movement of pollutants into different water reservoirs of the water cycle (such as rivers and aquifers). Point source contamination can be traced to specific points of discharge. Air pollution is an example of nonpoint source contamination because contaminants are throughout the air and landscape, ultimately entering the water from multiple locations. Eroded soil and sediment are another source of nonpoint source pollution. Image by USGS (public domain).

**Point source** pollution can be attributed to a single, definable origin. For example, animal factory farms (figure **12.3.2**) raise a large number and high density of livestock such as cows, pigs, and chickens.



**Figure 12.3.2:** Large animal farms are often referred to as concentrated feeding operations (CFOs). These farms are considered potential point sources of pollution because untreated animal waste may enter nearby waterbodies as untreated sewage. Credit: ehp.gov

Combined sewer systems that have a single set of underground pipes to collect both sewage and storm water runoff from streets for wastewater treatment can also be major point sources of pollutants. During heavy rain, storm water runoff may exceed sewer

capacity, causing it to back up. This spills untreated sewage directly into surface waters (figure **12.3.3**). Other examples include pipes from factories, waste disposal sites, storage tanks, and chemical spills.

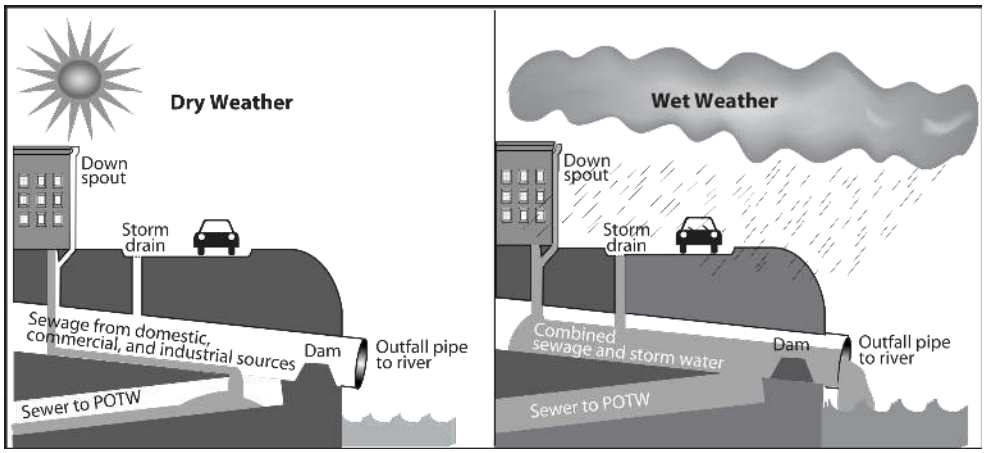



**Figure 12.3.3:** A combined sewer system is a possible major point source of water pollution during heavy rain due to overflow of untreated sewage. Sewage from domestic, commercial, and industrial sources enter the system through the down spout and storm water enters through the storm drain. During dry weather (and small storms), all sewage is handled by the publicly owned treatment works (POTW). A dam prevents raw sewage from flowing into a water body through the outfall pipe. During wet weather (large storms), the relief structure allows some of the combined sewage and storm water to be discharged untreated to an adjacent water body. Image by U.S. Environmental Protection Agency (public domain).

**Nonpoint source** pollution is from multiple dispersed sources. The whole of the contribution of pollutants is harmful, but the individual components may not reach harmful concentrations. Nonpoint sources of pollution include agricultural fields, cities, and abandoned mines. Rainfall runs over the land and through the ground, picking up pollutants from throughout the **watershed** (including areas of land and smaller streams that drain into a particular body of water). These pollutants might include herbicides, pesticides, and fertilizer from agricultural fields and lawns; oil, antifreeze, animal waste, and road salt from urban areas; and acid and toxic elements from abandoned mines. Then, this pollution is carried by runoff into surface water bodies and groundwater. Nonpoint source pollution, which is the leading cause of water pollution in the U.S., is usually much more difficult and expensive to control than point source pollution because of its low concentration, multiple sources, and much greater volume of water.

**Residence time** (the average time that a water molecule spends in a reservoir) is key to pollution problems because it affects pollution potential. Water in rivers has a relatively short residence time, so pollution usually is there only briefly. Of course, pollution in rivers may simply move to another reservoir, such as the ocean, where it can cause further problems. Groundwater is typically characterized by slow flow and longer residence time, which can make groundwater pollution particularly problematic. Finally, **pollution residence time** can be much greater than the water residence time because a pollutant may be taken up for a long time within the ecosystem or absorbed onto sediment.

According to a 2016 report, there were 20,912 cases of impaired water bodies in the United States, which means they could neither support a healthy ecosystem nor meet water quality standards (Table **12.3.1** and Figure **12.3.4**). Selected causes of impairment (water pollutants) are discussed below, categorized according to whether they arise from chemical, biological, or physical processes.

Table **12.3.1**: The leading causes of impaired bodies of water in the United States in 2016. Data from EPA (public domain).

| Cause of Impairment | Number of Impaired Bodies of Water |
|---|---|
| Polychlorinated biphenyls (PCBs) | 3,712 |
| Pathogens | 2,248 |
| Nutrients | 2,228 |
| Mercury | 2,138 |
| Metals (other than mercury) | 2,075 |
| Cause unknown- impaired biota | 1,852 |
| Organic enrichment/oxygen depletion | 1,281 |
| Turbidity | 1,175 |
| Pesticides | 795 |
| Salinity/total dissolved solids/chlorides/sulfates | 576 |
| pH/acidity/caustic conditions | 489 |
| Sediment | 453 |
| Temperature | 358 |
| Total toxics | 282 |
| Algal growth | 174 |
| Cause unknown | 159 |
| Dioxins | 136 |
| Toxic organics | 127 |
| Toxic inorganics | 99 |

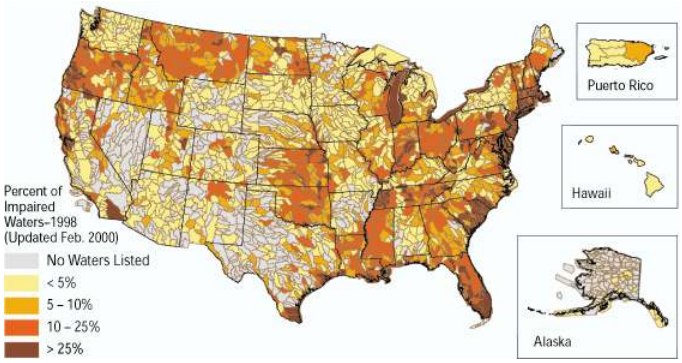



**Figure 12.3.4: Percentage of Impaired Water Bodies in a Watershed by State in USA Based on US EPA Data in 2000** Map of watersheds containing impaired water bodies from the U.S. Environmental Protection Agency's 1998 list of impaired waters

*Source: U.S. Geological Survey*

**Groundwater pollution** can occur from underground sources and all of the pollution sources that contaminate surface waters. Common sources of groundwater pollution are leaking underground storage tanks for fuel, septic tanks, agricultural activity, landfills, and fossil fuel extraction. Common groundwater pollutants include nitrate, pesticides, volatile organic compounds, and petroleum products. Polluted groundwater can be a more serious problem than polluted surface water because the pollution in groundwater may go undetected for a long time because usually it moves very slowly. As a result, the pollution in groundwater may

create a **contaminant plume**, a large body of flowing polluted groundwater (Figure 12.3.5), making cleanup very costly. By the time groundwater contamination is detected, the entity responsible for the pollution may be bankrupt or nonexistent. Another troublesome feature of groundwater pollution is that small amounts of certain pollutants, e.g., petroleum products and organic solvents, can contaminate large areas. In Denver, Colorado 80 liters of several organic solvents contaminated 4.5 trillion liters of groundwater and produced a 5 km long contaminant plume. A major threat to groundwater quality is from underground fuel storage tanks. Fuel tanks commonly are stored underground at gas stations to reduce explosion hazards. Before 1988 in the U.S. these storage tanks could be made of metal, which can corrode, leak, and quickly contaminate local groundwater. Now, leak detectors are required and the metal storage tanks are supposed to be protected from corrosion or replaced with fiberglass tanks. Currently there are around 600,000 underground fuel storage tanks in the U.S. and over 30% still do not comply with EPA regulations regarding either release prevention or leak detection.

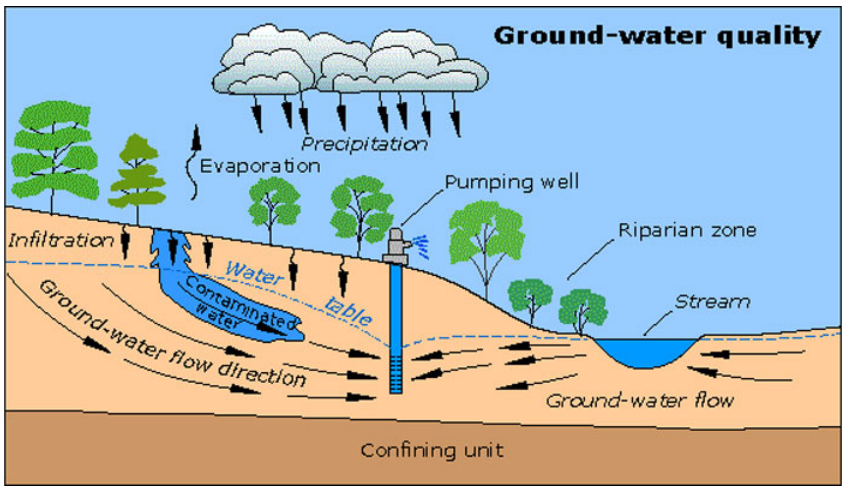



**Figure 12.3.5: Contaminant Plume in Groundwater** Mapping how a contaminant plume will migrate once it reaches groundwater requires understanding of the pollutant's chemical properties, local soil characteristics, and how permeable the aquifer is. *Source: United States Geological Survey*

## Biological Pollutants

**Pathogens** are disease-causing microorganisms, e.g., viruses, bacteria, parasitic worms, and protozoa, which cause a variety of intestinal diseases such as dysentery, typhoid fever, hepatitis, and cholera. Pathogens enter water primarily from human and animal fecal waste due to inadequate sewage treatment. In many underdeveloped countries, sewage is discharged into local waters either untreated or after only rudimentary treatment. In developed countries untreated sewage discharge can occur from overflows of combined sewer systems, poorly managed livestock factory farms, and leaky or broken sewage collection systems. Water with pathogens can be remediated by adding chlorine or ozone ($O_3$), by boiling, or by treating the sewage in the first place.

Unfortunately nearly a billion people around the world are exposed to waterborne pathogen pollution daily and around 1.5 million children mainly in underdeveloped countries die every year of waterborne diseases from pathogens. According to the World Health Organization, in 2008 approximately 880 million people in the world (or 13% of world population) did not have access to improved (safe) drinking water (World Health Statistics, 2010) (Figure 12.3.6). At the same time, about 2.6 billion people (or 40% of world population) lived without improved sanitation (Figure 12.3.7), which is defined as having access to a public sewage system, septic tank, or even a simple pit latrine. Each year approximately 1.7 million people die from diarrheal diseases associated with unsafe drinking water, inadequate sanitation, and poor hygiene, e.g., hand washing with soap. Almost all of these deaths are in developing countries, and around 90% of them occur among children under the age of 5 (Figure 12.3.8). Compounding the water crisis is the issue of social justice; poor people more commonly lack clean water and sanitation than wealthy people in similar areas. Globally, improving water, sanitation, and hygiene could prevent up to 9% of all disease and 6% of all deaths.

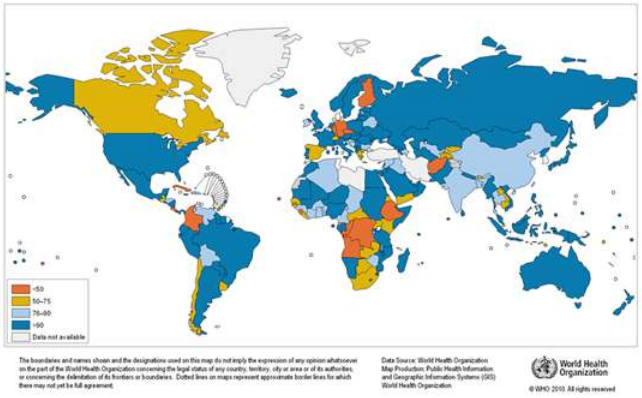



Figure 12.3.6: **Proportion of Population by Country Using Improved Drinking Water Sources in 2008** Improved drinking water sources, e.g., household connections, public standpipes, boreholes, protected dug wells and springs, and rainwater collections, are defined as those more likely to provide safe water than unimproved water sources, e.g., unprotected wells and springs, vendor-provided water, bottled water (unless water for other uses is available from an improved source), and tanker truck-provided water. *Source:* World Health Organization

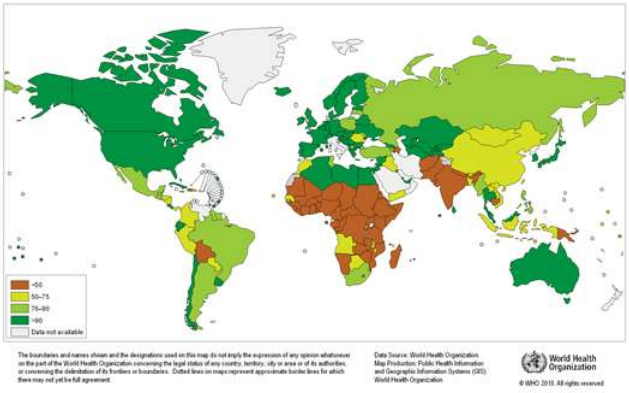



**Figure 12.3.7: Proportion of Population by Country Using Improved Sanitation Facilities in 2008** Improved sanitation facilities, e.g., connection to public sewers or septic systems, pour-flush latrines, pit latrines, and ventilated improved pit latrines, are defined as those more likely to be sanitary than unimproved facilities, e.g., bucket latrines, public latrines, and open pit latrines. *Source:* World Health Organization

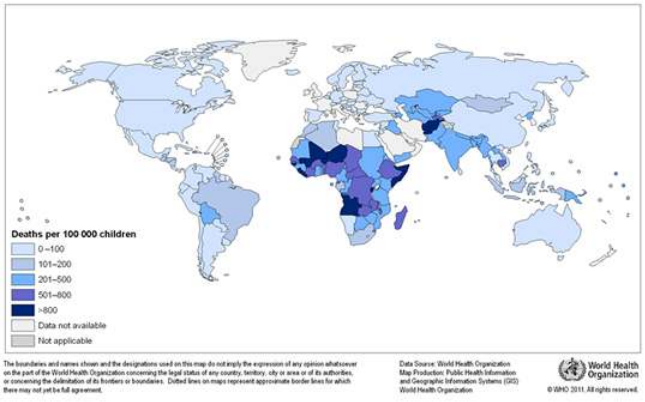



**Figure 12.3.8: Deaths by Country from Diarrhea Caused by Unsafe Water, Unimproved Sanitation, and Poor Hygiene in Children Less than 5 Years Old, 2004** *Source:* World Health Organization

## Chemical Pollutants

Chemical pollution from agriculture, industry, cities, and mining threatens global water quality. Air pollutants from these activities can also enter bodies of water (and become water pollutants) through dry deposition, precipitation, and runoff. Some chemical

pollutants have serious and well-known health effects, whereas many others have poorly known long-term health effects.

Any natural water contains dissolved chemicals, some of which are important human nutrients while others can be harmful to human health. The concentration of a water pollutant is commonly given in very small units such as parts per million (**ppm**) or even parts per billion (**ppb**). An arsenic concentration of 1 ppm means 1 part of arsenic per million parts of water. This is equivalent to one drop of arsenic in 50 liters of water. To give you a different perspective on appreciating small concentration units, converting 1 ppm to length units is 1 cm (0.4 in) in 10 km (6 miles) and converting 1 ppm to time units is 30 seconds in a year. **Total dissolved solids** (TDS) represent the total amount of dissolved material in water. Average TDS values for rainwater, river water, and seawater are about 4 ppm, 120 ppm, and 35,000 ppm, respectively.

Inorganic pollutants (figure **12.3.9**) include nutrients like nitrate ($NO_3^-$) and phosphate ($PO_4^{3-}$), heavy metals, chloride ($Cl^-$), and radioactive isotopes released from mining or nuclear accidents (such as cesium, iodine, uranium, and radon gas).



Figure **12.3.9**: This scientist demonstrates how to test water quality, and part of this involves measuring levels of inorganic pollutants. Image by Erin Gleeson (CC-BY).

## Toxic Heavy Metals

Examples of heavy metals include arsenic, mercury, lead, cadmium, and chromium, and they can bioaccumulate and biomagnify through the food chain (see *Biomagnification* below at the end of this section). Heavy metals are commonly produced by industry and at metallic ore mines. Arsenic and mercury are discussed in more detail below. The US EPA regulates 83 contaminants in drinking water to ensure a safe public water supply. Similarly, at the international level the World Health Organization has drinking water standards for a variety of contaminants.

Arsenic (As) has been famous as an agent of death for many centuries. In large doses arsenic causes cancer and can be fatal. Only recently have scientists recognized that health problems can be caused by drinking small arsenic concentrations in water over a long time. It attacks the central nervous system and can damage the respiratory system, bladder, lungs, liver, and kidneys. Arsenic enters the water supply naturally from weathering of arsenic-rich minerals and from human activities such as coal burning and smelting of metallic ores. The worst case of arsenic poisoning occurred in the densely populated impoverished country of Bangladesh, which had experienced 100,000s of deaths from diarrhea and cholera each year from drinking surface water contaminated with pathogens due to improper sewage treatment. In the 1970s the United Nations provided aid for millions of shallow water wells (20-100m deep), which resulted in a dramatic drop in pathogenic diseases. Unfortunately, many of the wells produced water naturally rich in arsenic (Figure **12.3.10**). Tragically, there are an estimated 77 million people (about half of the population) who inadvertently may have been exposed to toxic levels of arsenic in Bangladesh as a result. Arsenic poisoning leads to headaches, confusion, and diarrhea, and eventually to vomiting, stomach pain, and convulsions. If not treated, the final outcomes are heart disease, stroke, cancer, diabetes, coma, and death. There are ways to treat arsenic-rich groundwater, but it is a challenge in Bangladesh to implement the simple and effective technology that is available. The World Health Organization has called it the largest mass poisoning of a population in history.

Figure **12.3.10**: The distribution of arsenic in groundwater in Bangladesh. The WHO recommended safe level for arsenic is 10 µg/L. All of the green, orange, and red areas on the map exceed that limit.

Mercury enters the water supply naturally from weathering of mercury-rich minerals and, like mercury, from human activities such as coal burning and metal processing. Mercury (Hg) is used in a variety of electrical products, such as dry cell batteries, fluorescent light bulbs, and switches, as well as in the manufacture of paint, paper, vinyl chloride, and fungicides. Mercury acts on the central nervous system and can cause loss of sight, feeling, and hearing as well as nervousness, shakiness, and death. A famous mercury poisoning case in Minamata, Japan involved methylmercury-rich industrial discharge that caused high mercury levels in fish. People in the local fishing villages ate fish up to three times per day for over 30 years, which resulted in over 2,000 deaths. During that time the responsible company and national government did little to mitigate, help alleviate, or even acknowledge the problem.

## Biochemical Oxygen Demand

**Oxygen-demanding waste** is an extremely important pollutant to ecosystems. Most surface water in contact with the atmosphere has a small amount of dissolved oxygen, which is needed by aquatic organisms for cellular respiration. Decomposers, such as bacteria and fungi, also conduct cellular respiration and consume oxygen as they break down dead organic matter according to the following reaction:

$$\text{organic matter} + O_2 \rightarrow CO_2 + H_2O \tag{12.3.1}$$

Too much decaying organic matter in water is a pollutant because it removes oxygen from water, which can kill fish, shellfish, and aquatic insects. The amount of oxygen used by **aerobic** (in the presence of oxygen) bacterial decomposition of organic matter is called **biochemical oxygen demand** (BOD). The major source of dead organic matter in many natural waters is sewage; grass and leaves are smaller sources. An unpolluted water body with respect to BOD is a turbulent river that flows through a natural forest. Turbulence continually brings water in contact with the atmosphere where the $O_2$ content is restored. The dissolved oxygen content in such a river ranges from 10 to 14 ppm $O_2$, BOD is low, and clean-water fish such as trout thrive (figure **12.3.11**).



**Figure 12.3.11:** Left: Fish like rainbow trout thrive under low biological oxygen demand and ample levels of dissolved oxygen. Right: Fish like carp thrive under high biological oxygen demand and low dissolved oxygen levels. Left image by Engbretson Eric/U.S. Fish and Wildlife Service (public domain). Right image by X posid (public domain).

A polluted water body with respect to oxygen is a stagnant deep lake in an urban setting with a combined sewer system. This system favors a high input of dead organic carbon from sewage overflows and limited chance for water circulation and contact with the atmosphere. In such a lake, the dissolved $O_2$ content is ≤5 ppm $O_2$, BOD is high, and low $O_2$-tolerant fish, such as carp and catfish dominate (figure **12.3.11**).

## Plant Nutrients and Eutrophication

Excessive nutrients, particularly nitrogen (N) and phosphorous (P), are pollutants closely related to oxygen-demanding waste. Plants and algae require 15-20 nutrients for growth, most of which are plentiful in water. Nitrogen and phosphorus are called **limiting nutrients**, however, because they usually are present in water at low concentrations and therefore restrict the total amount of plant growth. This explains why nitrogen and phosphorus are major ingredients in most fertilizer.



**Figure 12.3.12:** A cyanobacteria (blue-green algae) bloom in Lake Erie resulting from fertilizer pollution. Image by NASA (public domain).

High concentrations of limiting nutrients, particularly nitrogen (N) and phosphorous (P), from human sources (mostly agricultural and urban runoff including fertilizer, sewage, and phosphorus-based detergent) can cause cultural **eutrophication**, which leads to the rapid growth of aquatic producers, particularly algae, called an *algal bloom*. Thick mats of floating algae or rooted plants lead to a form of water pollution that damages the ecosystem by clogging fish gills and blocking sunlight (figure **12.3.12**). When the prolific algal layer dies, it becomes oxygen-demanding waste, which can create very low $O_2$ water (<~2 ppm $O_2$), called **hypoxia** or dead zone because it causes death to organisms that are unable to leave that environment (figure **12.3.13**).

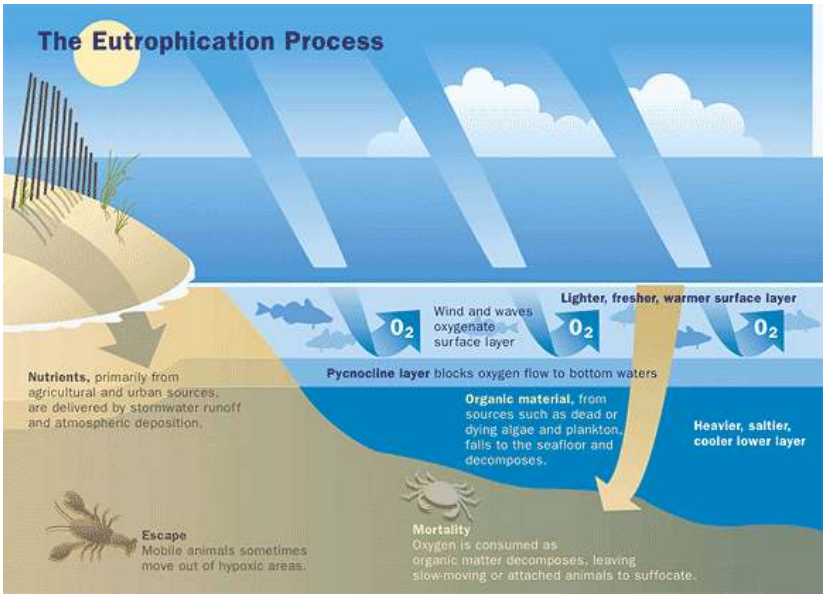



**Figure 12.3.13:** *The Eutrophication Process and how it leads to hypoxia*. The process of eutrophication. During eutrophication, increased nutrients causes the growth of photosynthetic microbes resulting in more biomass to decompose. More oxygen is used in decomposition, resulting in the death of larger organisms. The growth of photosynthetic microbes also leads to toxins (for example, those produced during red tides), which also leads to the death of larger organisms. Source: https://www.wri.org/our-work/project/eutrophication-and-hypoxia/about-eutrophication

A small percentage of algal species produce toxins that can kill fish, mammals, and birds, and may cause human illness. Exponential growths of these algae are called **toxic algal blooms** or **harful algal blooms** (figure **12.3.14**).



**Figure 12.3.14: Harmful Algal Bloom** Harmful algal bloom with deep red color. *Source: Kai Schumann via National Oceanic and Atmospheric Administration*

An estimated 50% of lakes in North America, Europe, and Asia are negatively impacted by cultural eutrophication. In addition, the size and number of marine hypoxic zones have grown dramatically over the past 50 years (Figure **12.3.15**), including a very large dead zone located offshore Louisiana in the Gulf of Mexico and measured 13,080 square kilometers in the summer of 2014. The eutrophication process in the Gulf of Mexico results from agricultural nutrient inputs in the Mississippi-Atchafalaya River Basin so the dead zone is cyclical and grows in the summer and shrinks during the winter due to decreased agriculture only to return the following summer. The Black Sea is another of the many dead zones that have been identified resulting from input from atmospheric deposition of NOx, intensive fertilizer use in farming practices, sewage input to water systems in the Danube River watershed in Germany. This resulted in the loss of fisheries and marine habits disrupted and reduced tourism. Cultural eutrophication and hypoxia are difficult to combat, because they are caused primarily by nonpoint source pollution, which is difficult to regulate, and N and P, which are difficult to remove from wastewater.

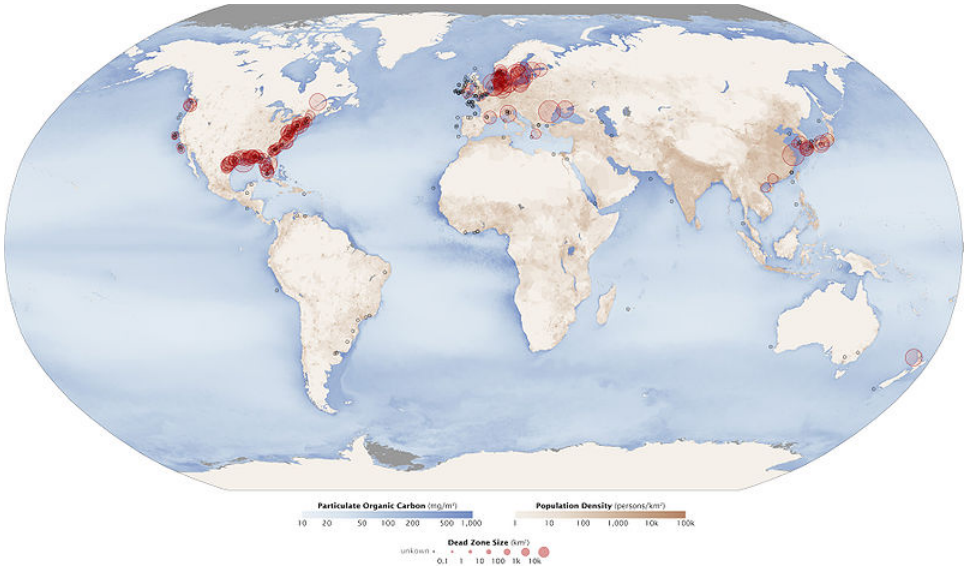



**Figure 12.3.15: Aquatic Dead Zones** Zones of hypoxia shown as red circles. Black dots show hypoxia zones of unknown size, brown shading shows population density, and blue shading shows density of particulate organic carbon, an indicator of organic productivity. *Source: Robert Simmon & Jesse Allen at NASA Earth Observatory via Wikimedia Commons*

For more information on Eutrophication see this video from the Khan Academy that explains the process of eutrophication and how dead zones are formed.

## Organic Chemical Pollutants

Organic pollutants include herbicides and pesticides, pharmaceuticals, fuel (such as oil spills), industrial solvents and cleansers, and synthetic hormones associated with pharmaceuticals. These synthetic hormones can act as **endocrine disruptors**. Many are

**persistent organic pollutants** (POPs), which are long-lived in the environment, biomagnify through the food chain, and can be toxic. As previously mentioned, DDT (pesticide), dioxin (herbicide by-product), and PCBs (polychlorinated biphenyls, which were used as a liquid insulator in electric transformers), are all POPs.

An example of organic chemical contamination is the Love Canal, in Niagara Falls, New York (Figure **12.3.16**). It is important because the pollution disaster at Love Canal, along with similar pollution calamities at that time (Times Beach, Missouri and Valley of Drums, Kentucky), helped to create **Superfund**, a federal program instituted in 1980 and designed to identify and clean up the worst of the hazardous chemical waste sites in the U.S.. Since the Love Canal tragedy, Superfund has analyzed tens of thousands of hazardous waste sites in the U.S. and cleaned up hundreds of the worst ones. Nevertheless, over 1,000 major hazardous waste sites with a significant risk to human health or the environment are still in the process of being cleaned.



**Figure 12.3.16:** The Love Canal in Niagra Falls, New York. Image by Environmental Protection Agency (public domain).

Love Canal is a neighborhood in Niagara Falls named after a large ditch (approximately 15 m wide, 3–12 m deep, and 1600 m long) that was dug in the 1890s for hydroelectric power. The ditch was abandoned before it actually generated any power and went mostly unused for decades, except for swimming by local residents. In the 1920s Niagara Falls began dumping urban waste into Love Canal, and in the 1940s the U.S. Army dumped waste from World War II there, including waste from the effort to build a nuclear bomb. Hooker Chemical Company purchased the land in 1942 and lined it with clay. From 1942 to 1952, the Hooker Chemical disposed of over 21,000 tons of chemical waste, including chlorinated hydrocarbons, benzene, and dioxin, into a canal and covered it with a layer of clay. Chlorinated hydrocarbons are a large group of organic chemicals that have chlorine functional groups, most of which are toxic and carcinogenic to humans (DDT and PCBs are an examples.) In 1953, Hooker sold the land to the Niagara Falls school board for $1, and included a clause in the sales contract that both described the land use (filled with chemical waste) and absolved them from any future damage claims from the buried waste. The school board promptly built a public school on the site and sold the surrounding land for a housing project that built 200 or so homes along the canal banks and another 1,000 in the neighborhood. During construction, the canal’s clay cap and walls were breached, damaging some of the metal barrels. Eventually, the chemical waste seeped into people's basements, and the metal barrels worked their way to the surface. Trees and gardens began to die; bicycle tires and the rubber soles of children's shoes disintegrated in noxious puddles. From the 1950s to the late 1970s, residents repeatedly complained of strange odors and substances that surfaced in their yards. Local residents allegedly experienced major health problems including high rates of miscarriages, birth defects, and chromosome damage, but studies by the New York State Health Department disputed that. Finally, in 1978 President Carter declared a state of emergency at

Love Canal, making it the first human-caused environmental problem to be designated that way, and the neighborhood had to be evacuated. Clean up of Love Canal, which was funded by Superfund and completely finished in 2004, involved removing contaminated soil, installing drainage pipes to capture contaminated groundwater for treatment, and covering it with clay and plastic. The total clean up cost was estimated to be $275 million.

**Oil spills** are another kind of organic pollution. Oil spills can result from supertanker accidents such as the Exxon Valdez in 1989, which spilled 10 million gallons of oil into the rich ecosystem of offshore south Alaska and killed massive numbers of animals. The largest marine oil spill was the Deepwater Horizon disaster, which began with a natural gas explosion (Figure **12.3.17**) at an oil well 65 km offshore of Louisiana and flowed for 3 months in 2010, releasing an estimated 200 million gallons of oil. The worst oil spill ever occurred during the Persian Gulf war of 1991, when Iraq deliberately dumped approximately 200 million gallons of oil in offshore Kuwait and set more than 700 oil well fires that released enormous clouds of smoke and acid rain for over nine months. During an oil spill on water, oil floats to the surface because it is less dense than water, and the lightest hydrocarbons evaporate, decreasing the size of the spill but polluting the air. Then, bacteria begin to decompose the remaining oil, in a process that can take many years. After several months only about 15% of the original volume may remain, but it is in thick asphalt lumps, a form that is particularly harmful to birds, fish, and shellfish. Cleanup operations can include *skimmer ships* that vacuum oil from the water surface (effective only for small spills), controlled *burning* (works only in early stages before the light, ignitable part evaporates but also pollutes the air), *dispersants* (detergents that break up oil to accelerate its decomposition, but some dispersants may be toxic to the ecosystem), and *bioremediation* (adding microorganisms that specialize in quickly decomposing oil, but this can disrupt the natural ecosystem).



**Figure 12.3.17: Deepwater Horizon Explosion** Boats fighting the fire from an explosion at the Deepwater Horizon drilling rig in Gulf of Mexico offshore Louisiana on April 20, 2010. *Source: United States Coast Guard via Wikimedia Commons*

## Physical Sources of Pollution

Trash, sediments, and thermal pollution arise from physical sources of pollution (figure **12.3.18**).



**Figure 12.3.18:** Water Pollution. Obvious water pollution in the form of floating debris; invisible water pollutants sometimes can be much more harmful than visible ones. Image by Stephen Codrington (CC-BY).

**Thermal pollution** occurs when water temperature exceeds its natural range. Many power plants (such as coal, natural gas, nuclear, etc.) rely on water from the environment for cooling. This water is released back into bodies of water at a higher temperature than usual. High temperature disrupt aquatic organisms for several reasons; one is that warmer waters cannot hold as much dissolved oxygen (see below; figure **12.3.19**). Dams can also raise water temperature to the detriment or the organisms that live there.

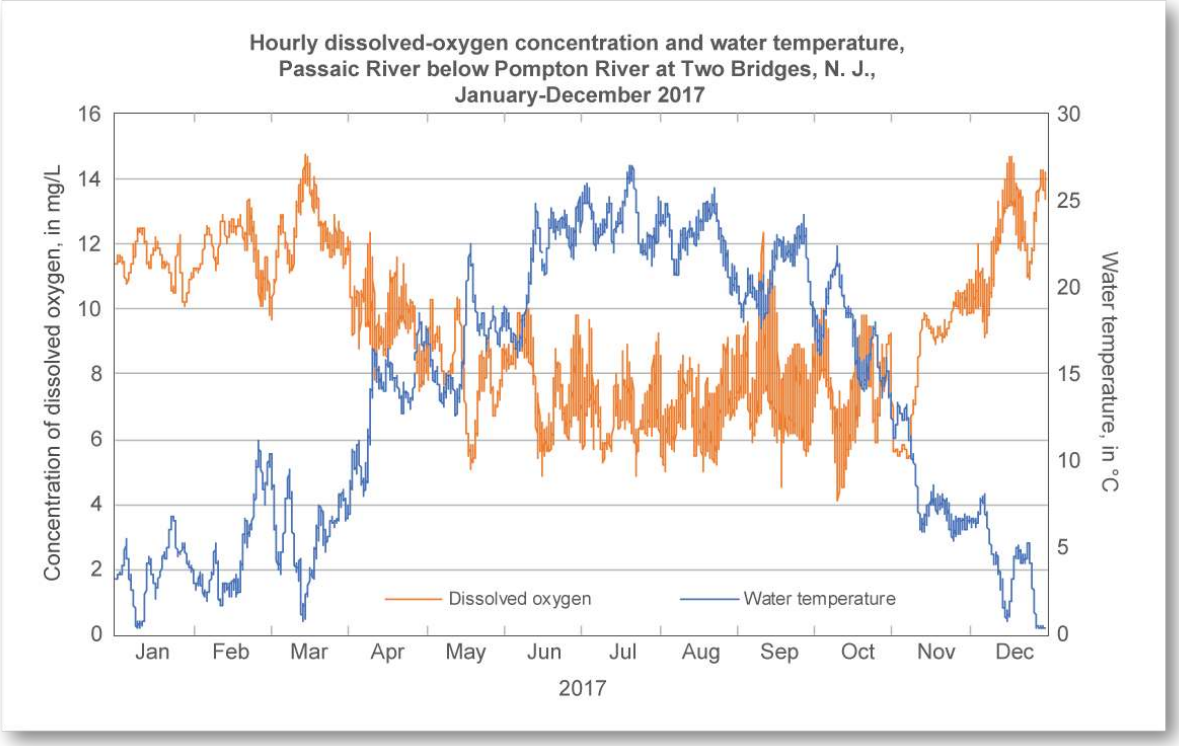



**Figure 12.3.19:** Water temperature affects dissolved oxygen concentrations in bodies of water. From January to December 2017, water temperature (blue) increases in summer months and then decreases. Dissolved oxygen levels (orange) dips in the summer

when temperature increases. Image and caption (modified) from USGS (public domain).

**Turbidity** is the clouding of water due to sediment. It can be caused by excessive phytoplankton, algae growth, urban runoff, or sediments from erosion. These suspended particles, in addition to making the water look dirty, also help promote the toxins in water as heavy metals and toxic organic compounds can attach easily to the suspended sediment. These suspended particles also absorb heat from the sun, making turbid waters warmer. This also reduces the oxygen content in the water, as more oxygen is dissolved in colder waters. The suspended particles also scatter light, decreasing the photosynthetic activity of plants and algae, which results in a positive feedback loop for decreasing oxygen even more. Excess sediments enter bodies of water when various land uses, such as mining, deforestation, and agriculture increases erosion.

## Other Water Quality Issues

Salt, typically sodium chloride, is a common inorganic contaminant. It can be introduced into groundwater from natural deposits or from anthropogenic sources like the salts applied to roads in the winter to keep ice from forming (figure **12.3.20**). Salt contamination can also occur from **saltwater intrusion**, where cones of depression around fresh groundwater pumping near ocean coasts induce the encroachment of saltwater into the freshwater body.



Figure **12.3.20**: A snow plow and road salt application truck that is used to clear snow and ice from roadways for improving traffic and pedestrian safety. The use of salt to deice pavement can be harmful to aquatic life in urban streams. Image and caption (modified) from USGS (public domain).

The acidity or alkalinity of a body of water can also impact its quality. **pH** is a measure of the concentration of hydrogen ions (protons) in a solution, which determines how acidic or basic (alkaline) a solution is. Acidic solutions have a high hydrogen ion concentration and a pH less than 7, and basic solutions have a pH of greater than 7. The pH of freshwater usually ranges from 5 to 9, and saltwater is slightly basic (pH = 8.2) in a healthy ecosystem. When conditions are too acidic, some aquatic animals cannot reproduce, and calcium carbonate structures (clams, snails, corals, etc.) dissolve. Acid deposition and ocean acidification are two examples of pH changes in water bodies caused by anthropogenic activity.

**Hard water** contains abundant calcium and magnesium, which reduces its ability to develop soapsuds and enhances scale (calcium and magnesium carbonate minerals) formation on hot water equipment. Water softeners remove calcium and magnesium, which allows the water to lather easily and prevent the minerals from precipitating on surfaces (figure **12.3.21**). Hard water develops naturally from the dissolution of calcium and magnesium carbonate minerals in soil; it does not have negative health effects in people.



Figure 12.3.21: Overly hard water (high calcium content) can cause film on glassware. Image by Liz Meimann/Iowa State Univ. (public domain).

## Biomagnification

Biomagnification represents the processes in an ecosystem that cause greater concentrations of a persistent, toxic substances, in organisms higher up the food chain. Many substances have been shown to bioaccumulate, including classical studies with the pesticide dichlorodiphenyltrichloroethane (DDT), which was published in the 1960s bestseller, *Silent Spring*, by Rachel Carson. Other chemicals that exhibit biomagnification are methylmercury, PCBs, cadmium, lead, and arsenic. Mercury and methylmercury are present in only very small concentrations in seawater; however, at the base of the food chain algae absorb methylmercury. Then, small fish eat the algae, large fish and other organisms higher in the food chain eat the small fish, and so on. Fish and other aquatic organisms absorb methylmercury rapidly but eliminate it slowly from the body. Therefore, each step up the food chain increases the concentration from the step below (Figure 12.3.22 – 23). Largemouth bass can concentrate methylmercury up to 10 million times over the water concentration and fish-eating birds can concentrate it even higher.

DDT was a commonly-used pesticide before its dangers became known. Figure 12.3.22 shows how DDT becomes concentrated in the tissues of organisms representing four successive trophic levels in a food chain. The concentration effect occurs because DDT is metabolized and excreted much more slowly than the nutrients that are passed from one trophic level to the next. So DDT accumulates in the bodies (especially in fat). Thus most of the DDT ingested as part of gross production is still present in the net production that remains at that trophic level. Thus, birds accumulated sufficient amounts of DDT to cause fragility in their eggshells which increased egg breakage during nesting, resulting in adverse effects on populations of predatory birds and scavengers (such as the California Condor). The use of DDT was banned in the United States in the 1970s.

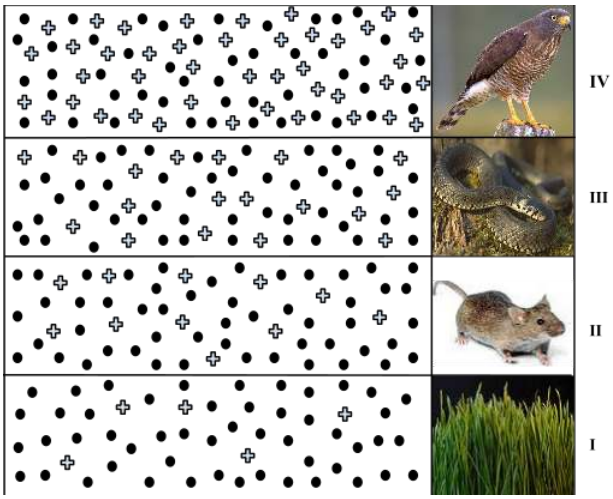



Figure 12.3.22: In biomagnification the concentration of the persistent toxins (crosses) increase higher up the food chain. An increase of toxin concentration as the food chain moves up to higher levels. Organisms at the top have a higher tissue concentration of toxins and pollutants than lower levels. The concentration system is due to persistence of the toxins, food chain energetics, and low rate of internal degradation or excretion of the substance.Trophic level I represents the primary producers. Trophic level II represents the primary consumers. Trophic level III represents the secondary consumers. Trophic level IV represents the tertiary consumers. (CC CC-SA 3.0).

Polychlorinated biphenyls (PCBs), which were used in coolant liquids in the United States until their use was banned in 1979. In a study performed by the National Oceanic and Atmospheric Administration (NOAA) in the Saginaw Bay of Lake Huron, PCB concentrations increased from the ecosystem's primary producers (phytoplankton) through the different trophic levels of fish species (Figure 12.3.23). The apex consumer (walleye) had more than four times the amount of PCBs compared to phytoplankton. Also, based on results from other studies, birds that eat these fish may have PCB levels at least one order of magnitude higher than those found in the lake fish.

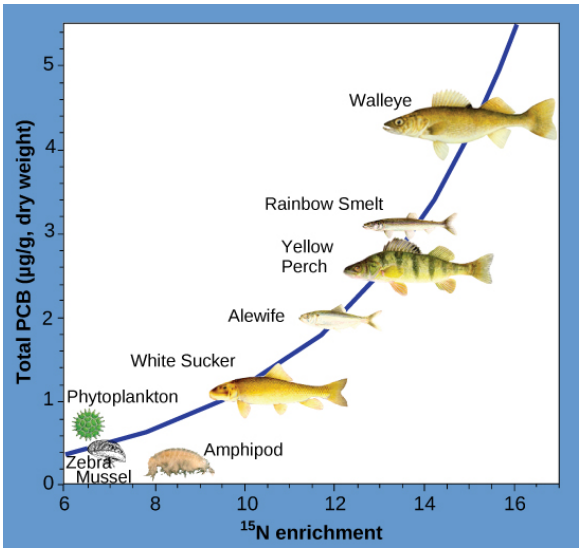



Figure 12.3.23: **PCB concentration in Lake Huron**: This chart shows the PCB concentrations found at the various trophic levels in the Saginaw Bay ecosystem of Lake Huron. Numbers on the x-axis reflect enrichment with heavy isotopes of nitrogen (15N), which is a marker for increasing trophic levels. Notice that the fish in the higher trophic levels accumulate more PCBs than those in lower trophic levels.

Other concerns have been raised by the accumulation of heavy metals, such as mercury and cadmium, in certain types of seafood. The United States Environmental Protection Agency (EPA) recommends that pregnant women and young children should not consume any swordfish, shark, king mackerel, or tilefish because of their high mercury content. These individuals are advised to eat fish low in mercury: salmon, sardines, tilapia, shrimp, pollock, and catfish.

## References


1. http://www.nature.com/scitable/knowledge/library/eutrophication-causes-consequences-and-controls-in-aquatic-102364466
2. www.globalpartnershipforocean...black-sea-blue
3. http://oceanservice.noaa.gov/educati...intsource.html
4. www2.epa.gov/nutrient-policy-data/cyanohabs
5. http://dujs.dartmouth.edu/winter-2012/eutrophication-in-the-gulf-of-mexico-how-midwestern-farming-practices-are-creating-a-%E2%80%98dead-zone%E2%80%99#.VObsBFPF9fw
6. http://water.epa.gov/type/watersheds/named/msbasin/zone.cfm
7. http://water.epa.gov/type/watersheds...hypoxia101.cfm
8. https://www.youtube.com/watch?<wbr/>v=rZFcLYqOxZQ
9. http://1.bp.blogspot.com/-HAO2PgXGXd...ia+diagram.png
10. http://www.nytimes.com/interactive/2...-map.html?_r=0
11. http://www.wri.org/resource/interact...cation-hypoxia
12. http://www.lenntech.com/eutrophicati...on-effects.htm
13. http://www.lenntech.com/turbidity.htm
14. Water for Life: Making it Happen (2005) World Health Organization and UNICEF. Retrieved from http://www.who.int/water_sanitation_health/waterforlife.pdf
15. World Health Statistics (2010) World Health Organization. Retrieved from http://www.who.int/whosis/whostat/EN_WHS10_Full.pdf


## Contributors and Attributions

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

- Water Contamination from *An Introduction to Geology* by Johnson et al. (licensed under CC-BY-NC-SA)
- Water Pollution from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- The Industrialization of Nature: A Modern History (1500 to the present), Water Pollution, and Case Study: The Love Canal Disaster from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors. Download for free at CNX. (licensed under CC-BY)
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0
- Eutrophication from *Oceanograpy* by Hill et. al. University of California, Davis
- **Biomagnification of Pesticides** by John W. Kimball, is licensed CC BY
- Biological Magnification from *General Microbiology* at Boundless.com. **License**: ***CC BY-SA: Attribution-ShareAlike***

---


This page titled 12.3: Water Pollutants and Their Sources is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **7.3: Water Pollution** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# 12.4: Water and Wastewater Treatment

Resolution of the global water pollution crisis requires multiple approaches to improve the quality of our fresh water and move towards sustainability. The most deadly form of water pollution, pathogenic microorganisms that cause waterborne diseases, kills almost 2 million people in underdeveloped countries every year. The best strategy for addressing this problem is proper sewage (wastewater) treatment. Untreated sewage is not only a major cause of pathogenic diseases, but also a major source of other pollutants, including oxygen-demanding waste, nutrients (N and P, particularly), and toxic heavy metals. Wastewater treatment is done at a sewage treatment plant in urban areas and through a septic tank system in rural areas.

The wastes generated by some 60% of the U.S. population are collected in sewer systems and carried along by some 14 billion gallons (~53 billion liters) of water a day. Of this enormous volume, some 10% is allowed to pass untreated into rivers, streams, and the ocean. The rest receives some form of treatment to improve the quality of the water (which makes up 99.9% of sewage) before it is released for reuse.

For many developing countries, financial aid is necessary to build adequate sewage treatment facilities. The World Health Organization estimates an estimated cost savings of between $3 and $34 for every $1 invested in clean water delivery and sanitation. The cost savings are from health care savings, gains in work and school productivity, and prevented deaths. Simple and inexpensive techniques for treating water at home include chlorination, filters, and solar disinfection. Another alternative is to use constructed wetlands technology (marshes built to treat contaminated water), which is simpler and cheaper than a conventional sewage treatment plant.

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. If you don’t like the taste of your tap water, then please use a water filter instead of bottled water!

## Sewage Treatment Plants

The main purpose of **sewage (wastewater) treatment** is to remove organic matter (oxygen-demanding waste) and kill bacteria. Special methods also can be used to remove nutrients and other pollutants. The numerous steps at a conventional sewage treatment plant include **pretreatment** (screening and removal of sand and gravel), **primary treatment** (settling or floatation to remove organic solids, fat, and grease), **secondary treatment** (aerobic bacterial decomposition of organic solids), **tertiary treatment** (bacterial decomposition of nutrients and filtration), **disinfection** (treatment with chlorine, ozone, ultraviolet light, or bleach to kill most microbes), and either **discharge** to surface waters (usually a local river) or reuse for some other purpose, such as irrigation, habitat preservation, and artificial groundwater recharge (Figure **12.4.1**).

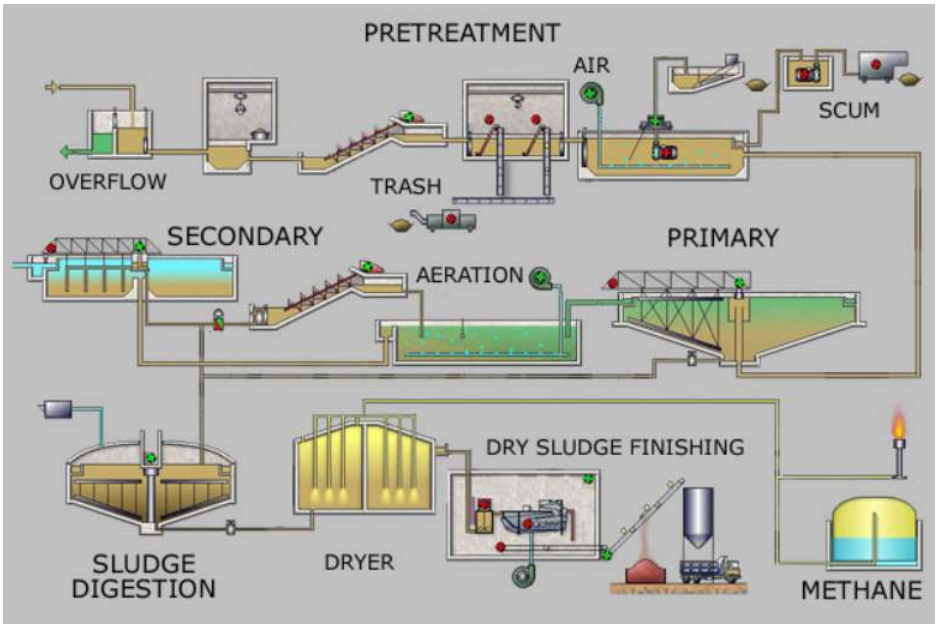



Figure **12.4.1**: Steps at a Sewage Treatment Plant The numerous processing steps at a conventional sewage treatment plant include pretreatment (screening and removal of sand and gravel), primary treatment (settling or floatation to remove organic solids, fat, and grease), secondary treatment (aerobic bacterial decomposition of organic solids), tertiary treatment (bacterial decomposition of nutrients and filtration), disinfection (treatment with chlorine, ozone, ultraviolet light, or bleach), and either discharge to surface waters (usually a local river) or reuse for some other purpose, such as irrigation, habitat preservation, and artificial groundwater recharge. Source: Leonard G.via Wikipedia

The concentrated organic solid produced during primary and secondary treatment is called **sludge**, which is treated in a variety of ways including landfill disposal, incineration, use as fertilizer, and anaerobic bacterial decomposition, which is done in the absence of oxygen. Anaerobic decomposition of sludge produces methane gas, which can be used as an energy source. To reduce water pollution problems, separate sewer systems (where street runoff goes to rivers and only wastewater goes to a wastewater treatment plant) are much better than combined sewer systems, which can overflow and release untreated sewage into surface waters during heavy rain. Some cities such as Chicago, Illinois have constructed large underground caverns and also use abandoned rock quarries to hold storm sewer overflow. After the rain stops, the stored water goes to the sewage treatment plant for processing.

## Septic Systems

Figure **12.4.10** A typical septic system.

In areas that are not served by sewage networks leading to a central sewage treatment plant, homeowners can use **septic systems** for disposal of sewage. The basic components of a septic tank system (Figure **12.4.2 – 3**) include a sewer line from the house, a septic tank (a large container where sludge settles to the bottom and microorganisms decompose the organic solids anaerobically), and the drain field (network of perforated pipes where the clarified water seeps into the soil and is further purified by bacteria). A typical septic tank is constructed of either concrete or plastic and has a volume of 5,000 litres to 10,000 litres (5 $m^3$ to 10 $m^3$). This forms the first treatment and is designed to be **anaerobic** (without oxygen) (Figure **12.4.3**). That promotes the activity of certain bacteria that help break down the waste. As the waste is degraded, some portions tend to sink to form sludge at the base of the tank, and others float to the surface, forming a scum layer. A septic tank may be divided into two parts to keep the sludge at the bottom and the scum on the top from draining out. The water then moves to the drainage field, which provides the right conditions for a different set of bacteria that operate in **aerobic** conditions. The drainage field includes an array of plastic pipes that are perforated to allow the effluent to drain out over a large area and seep slowly into the ground. In order to install a drainage field, it is first necessary to test the soil below, as it must be sufficiently permeable to allow the effluent to percolate away, but not so permeable that it flows too quickly and the soil is not able to filter out the pathogenic bacteria.

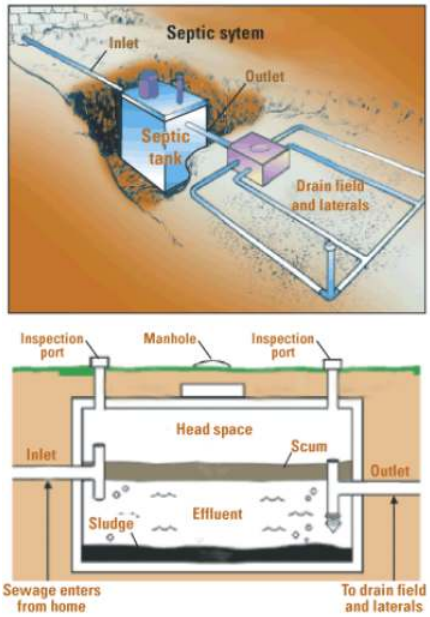



Figure 12.4.2. Septic System Septic tank system for sewage treatment. Source: United States Geological Survey

If they are properly installed and used, and if the sludge is periodically removed from the tank, a septic system should be effective in treating the sewage for decades. The anaerobic and aerobic bacteria should be able to break down the incoming waste and there should be little risk to the surface environment or groundwater. But many things can go wrong with a septic system, including the following:

- If inappropriate chemicals are added to the waste stream, they may interfere with the natural breakdown of the sewage. For example, bleach can kill the bacteria that break down sewage in the tank or drainage field
- If the tank is not periodically pumped out, solids can get into the drainage field and compromise the drainage, resulting in the flow of effluent toward the surface.
- If the soil is either not sufficiently permeable or too permeable, the effluent will not drain away (and will start to pool at the surface) or it will drain too quickly.
- If the drainage field is constructed in an area where the water table is close to surface, some of the effluent is likely to flow into the groundwater before it has been broken down by bacteria.

## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources:

- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.
- Water Treatment and Bioremediation from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Water Pollution from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors. Download for free at CNX. (licensed under CC-BY)
- **Sewage Treatment** by John W. Kimball, is licensed CC BY
- Groundwater Quality from *Physical Geology* by Steven Earle

---

This page titled 12.4: Water and Wastewater Treatment is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Matthew R. Fisher (OpenOregon) via source content that was edited to the style and standards of the LibreTexts platform.

- **7.4: Water Treatment** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# 12.5: Mitigating Water Pollution

Water pollution can be mitigated through regulation, bioremediation, and watershed management.

## Regulation

During the early 1900s, rapid industrialization in the U.S. resulted in widespread water pollution due to free discharge of waste into surface waters. The Cuyahoga River in northeast Ohio caught fire numerous times, including a famous fire in 1969 that caught the nation's attention (figure 12.5.1). In 1972 Congress passed one of the most important environmental laws in U.S. history, the Federal Water Pollution Control Act, which is more commonly called the **Clean Water Act** (CWA). The purpose of the Clean Water Act and later amendments is to maintain and restore water quality, or in simpler terms to make our water swimmable and fishable. It became illegal to dump pollution into surface water unless there was formal permission. The CWA regulates pollution from a single source such as that from industry or sewage treatment plants by setting pollution standards (maximum levels of each pollutant that can be in bodies of water or released at a time). United States water quality improved significantly as a result, but as there is still more work to be done. More progress is needed because currently the EPA considers over 40,000 U.S. water bodies as impaired, most commonly due to pathogens, metals, plant nutrients, and oxygen depletion. Another concern is protecting groundwater quality, which is not yet addressed sufficiently by federal law.



**Figure 12.5.1:** Cuyahoga River on fire in 1969. Image from USEPA Environmental-Protection-Agency (public domain).

Nonpoint sources of water pollution, e.g., agricultural runoff and urban runoff (Figure 12.5.2), are much harder to regulate because of their widespread, diffuse nature. There are many construction and agricultural practices that reduce polluted runoff including no-till farming and sediment traps. Artificial aeration or mechanical mixing can remediate lakes with oxygen depletion. Specific things that we can do to reduce urban runoff include the following: keep soil, leaves, and grass clippings off driveways, sidewalks, and streets; don't pour used motor oil, antifreeze, paints, pesticides, or any household hazardous chemical down the storm sewer or drain; recycle used motor oil; use hazardous waste disposal programs offered by the community; compost your organic waste; don't use fertilizers and herbicides on your lawn; and flush pet waste down the toilet.



**Figure 12.5.2: Storm Drain** Curbside storm drain receiving urban runoff. *Source: By Robert Lawton via Wikimedia Commons*

In the past, domestic and commercial refuse was commonly trucked to a “dump” (typically a hole in the ground), and when the hole was filled, it was covered with soil and forgotten. In situations like this, rain and melting snow can easily pass through the soil used to cover the refuse. This water passes into the waste itself, and the resulting landfill **leachate** that flows from the bottom of the landfill can seriously contaminate the surrounding groundwater and surface water. In the past few decades, regulations around refuse disposal have been significantly strengthened, and important steps have been taken to reduce the amount of landfill waste by diverting recyclable and compostable materials to other locations.

There are two fairly simple ways to significantly reduce the chance and degree of groundwater contamination from surface sources. One is to prevent rainwater from infiltrating down to the water table and picking up contaminants; this can be achieved by simply capping or roofing over the landfill, mine tailings, or spill site. The second is to provide an impermeable barrier beneath the contaminant. Modern landfills and mine tailings impoundments are all built using some combination of clay and engineered plastic barriers. Both of these solutions — caps and liners — are subject to failure due to leaks.

A modern engineered landfill has an impermeable liner (typically heavy plastic, although engineered clay liners or natural clay may be adequate in some cases), a plumbing system for draining leachate (the rainwater that flows through the refuse and becomes contaminated), and a network of monitoring wells both within and around the landfill (Figure **12.5.3**). Once part or all of a landfill site is full, it is sealed over with a plastic cover. The leachate must be treated, and that can be done in a normal sewage treatment plant. The monitoring wells are used to assess the level of the water table around the landfill and to collect groundwater samples so that any leakage can be detected. Because some leakage is almost inevitable, the ideal placement for landfills is in areas where the depth to the water table is significant (tens of meters if possible) and where the aquifer material is relatively impermeable. Landfills should also be situated far from streams, lakes, or wetlands so that contamination of aquatic habitats can be avoided.

Figure **12.5.3**: A cross-section of a typical modern landfill.

## Remediation

**Remediation** is the act of cleaning contamination. Biological remediation (**bioremediation**) is a waste management technique that involves the use of organisms such as plants, bacteria, and fungi to remove or neutralize pollutants from a contaminated site. According to the United States EPA, bioremediation is a “treatment that uses naturally occurring organisms to break down hazardous substances into less toxic or nontoxic substances”. This type of remediation is usually used on organic chemicals but also works on reducing or oxidizing inorganic chemicals like nitrate. **Phytoremediation** is a type of bioremediation that uses plants to absorb the chemicals over time.

Bioremediation is widely used to treat human sewage and has also been used to remove agricultural chemicals (pesticides and fertilizers) that leach from soil into groundwater. Certain toxic metals, such as selenium and arsenic compounds, can also be removed from water by bioremediation. Mercury is an example of a toxic metal that can be removed from an environment by bioremediation. Several species of bacteria can carry out the biotransformation of toxic mercury into nontoxic forms. These bacteria, such as *Pseudomonas aeruginosa*, can convert a charged form of mercury ($Hg^{2+}$) to an uncharged form (Hg), which is less toxic to humans.

Probably one of the most useful and interesting examples of the use of prokaryotes for bioremediation purposes is the cleanup of oil spills. To clean up these spills, bioremediation is promoted by adding inorganic nutrients that help bacteria already present in the environment to grow. Hydrocarbon-degrading bacteria feed on the hydrocarbons in the oil droplet, breaking them into inorganic compounds. Some species, such as *Alcanivorax borkumensis*, produce surfactants that break the oil into droplets, making it more accessible to the bacteria that degrade the oil. In the case of oil spills in the ocean, ongoing, natural bioremediation tends to occur, inasmuch as there are oil-consuming bacteria in the ocean prior to the spill. Under ideal conditions, it has been reported that up to 80 percent of the nonvolatile components (those that do not readily evaporate) in oil can be degraded within one year of the spill. Researchers have genetically engineered other bacteria to consume petroleum products; indeed, the first patent application for a bioremediation application in the U.S. was for a genetically modified oil-eating bacterium.



(a) (b)

**Figure 12.5.4:** (a) Cleaning up oil after the Exxon Valdez spill in Alaska (1989), the workers hosed oil from beaches and then used a floating boom to corral the oil, which was finally skimmed from the water surface. Some species of bacteria are able to solubilize and degrade the oil. (b) One of the most catastrophic consequences of oil spills is the damage to fauna. (credit a: modification of work by NOAA; credit b: modification of work by GOLUBENKOV, NGO: Saving Taman)

There are a number of cost/efficiency advantages to bioremediation, which can be employed in areas that are inaccessible without excavation. For example, hydrocarbon spills (specifically, oil spills) or certain chlorinated solvents may contaminate groundwater, which can be easier to treat using bioremediation than more conventional approaches. This is typically much less expensive than excavation followed by disposal elsewhere, incineration, or other off-site treatment strategies. It also reduces or eliminates the need for “pump and treat”, a practice common at sites where hydrocarbons have contaminated clean groundwater. Using microorganisms for bioremediation of hydrocarbons also has the advantage of breaking down contaminants at the molecular level, as opposed to simply chemically dispersing the contaminant.

**Chemical remediation** uses the introduction of chemicals to remove the contaminant or make it less harmful. One example is reactive barriers, a permeable wall in the ground or at a discharge point that chemically reacts with contaminants in the water. Reactive barriers made of limestone can increase the pH of acid mine drainage, making the water less acidic and more basic, which removes dissolved contaminants by precipitation into a solid form. **Physical remediation** consists of removing the contaminated water and either treating it (aka pump and treat) with filtration or disposing of it. All of these options are technically complex, expensive, and difficult, with physical remediation typically being the most costly.

Specific treatment methods for groundwater contamination depend on the geology, hydrology, and pollutant because some light contaminants flow on top of groundwater, others dissolve and flow with groundwater, and dense contaminants can sink below

groundwater. Sometimes slow flow through a soil can naturally purify groundwater because some pollutants, such as P, pesticides, and heavy metals, chemically bind with surfaces of soil clays and iron oxides. Other pollutants are not retained by soil particles: These include N, road salt, gasoline fuel, the herbicide atrazine, tetrachloroethylene (a carcinogenic cleaning solvent used in dry cleaning), and vinyl chloride. In other cases, slow groundwater flow can allow bacteria to decompose dead organic matter and certain pesticides. A common cleanup method called pump and treat involves pumping out the contaminated groundwater and treating it by oxidation, filtration, or biological methods. Sometimes soil must be excavated and sent to a landfill. In-situ treatment methods include adding chemicals to immobilize heavy metals, creating a permeable reaction zone with metallic iron that can destroy organic solvents, or using **bioremediation** by adding oxygen or nutrients to stimulate growth of microorganisms.

## Watershed Management

**Watershed management** involves reducing chemicals applied to land in the watershed (which will drain into a body of water) and runoff of those chemicals (figure $\mathbf{12.5.5-6}$). This strategy is more effective for nonpoint source pollution than setting pollution standards (as the CWA does) because it does not require each source of pollution be identified.



**Figure 12.5.5:** While some watersheds are relatively small, others encompass thousands of square miles and may contain streams, rivers, lakes, reservoirs, and underlying groundwater that are hundreds of miles inland. Shown here is an aerial view of Drakes Bay, part of California's Tomales-Drake watershed. Image and caption (modified) from Brian Cluer, NOAA Fisheries West Coast Region, California Coastal Office (public domain).

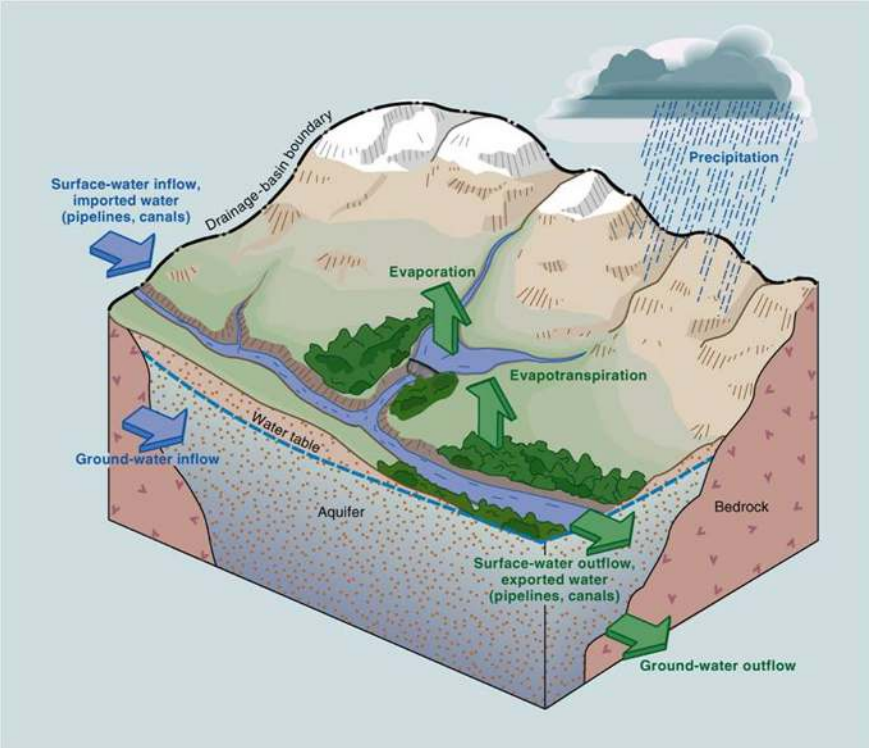



**Figure 12.5.6:** Diagram of a watershed, the area that drains into a body of water. The drainage-basin boundary defines the edges of the watershed. Precipitation falls on the watershed. Some reenters the atmosphere through evaporation or evapotranspiration. The water collects into smaller streams, which drain into a larger river. Additionally, surface water can enter the watershed (surface-water inflow) as water imported through pipelines or canals. Surface water exits the watershed (surface-water outflow) as water exported through pipelines canals. Similarly, ground-water inflow enters the system, and ground-water outflow exits the system. Ground water is stored in the aquifer underground, and the water table marks upper level of groundwater in the aquifer. Image from Healy, R.W., Winter, T.C., LaBaugh, J.W., and Franke, O.L., 2007, : Water budgets: Foundations for effective water-resources and environmental management: U.S. Geological Survey Circular 1308.

Maintaining or restoring **riparian areas** (riparian zones) is key to watershed management (figure **12.5.7**). These are areas of land close enough to a body of water to be affected by that body of water, for example, the lush region of vegetation surrounding a river. Riparian areas provide many ecosystem services that promote water quality and limit pollution. The vegetation absorbs nutrients that could otherwise cause eutrophication. It also provide shade, keeping the water cool and increasing its capacity to hold dissolved oxygen. The roots of the vegetation cling to soil, combating erosion. Additionally, roots and low-lying plants slow the flow of run-off, promoting infiltration. High infiltration rates have several benefits: (1) less runoff is available to bring pollution to the river, lake, or bay, (2) pollutants will be filtered from the runoff as it soaks into the ground, and (3) aquifers are replenished.



**Figure 12.5.7:** Riparian areas reduce water pollution by absorbing nutrients, keeping water cool, limiting erosion, and promoting infiltration. Image by BLM (public domain).

Watershed management plans typically leave undisturbed riparian vegetation directly adjacent to a body of water (figure **12.5.8**). Next (moving away from the water), could be managed forest followed by agriculture. The most intensive agriculture should be farthest from the body of water, particularly when pesticides and fertilizers are in use.

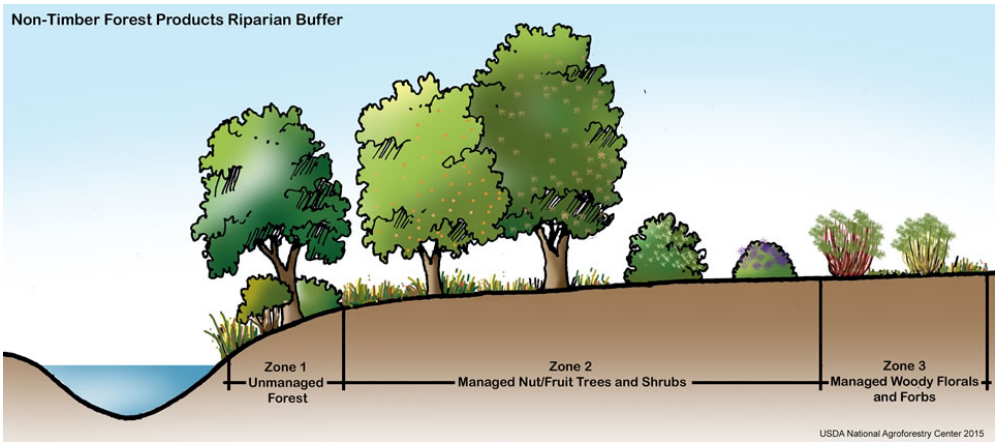



**Figure 12.5.8:** Although riparian forest buffers are often designed primarily for water quality benefits, these practices can also include woody species that provide products such as nuts, fruit, and decorative woody florals. These are examples of non-timber forest products. Zone 1, which is closest to the river, is unmanaged forest. Zone 2 contains managed nut or fruit trees and shrubs. Zone 3 contains managed woody florals and forbs. (Forbs are non-woody plants excluding grasses.) Image and caption by USDA National Agroforestry Center (public domain).

In urban areas can also be carefully structured to limit water pollution. Rain gardens next to buildings are areas with soil and vegetation that promote infiltration. Permeable pavement, which allows water to pass through it, also promotes infiltration, and reduces runoff, creating fewer opportunities for the water to acquire pollutants from washing over a dirty street.

# Contributors and Attribution

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

- Water Contamination from *An Introduction to Geology* by Johnson et al. (licensed under CC-BY-NC-SA)
- Water Treatment and Bioremediation from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Water Pollution from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors. Download for free at CNX. (licensed under CC-BY)
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0
- Groundwater Quality from *Physical Geology* by Steven Earle

This page titled 12.5: Mitigating Water Pollution is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# CHAPTER OVERVIEW

## 13: Agriculture


13.1: Food Security
13.2: Industrial Agriculture
13.3: Pests and Pesticides
13.4: Soil Degradation
13.5: Selective Breeding and Genetic Engineering
13.6: Pollination and Seed Dispersal
13.7: Sustainable Agriculture


---


13: Agriculture is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 13.1: Food Security

Progress continues in the fight against hunger, yet an unacceptably large number of people lack the food they need for an active and healthy life. In 2019, 690 million people in the world (8.9%) were undernourished (lacked sufficient calories; see The State of Food Security...), and 8.2% of the global population lived in poverty (see Goal 1...). Poverty is internationally defined as living on less than $1.90 per day. Poverty—not food availability—is the major driver of food insecurity. Improvements in agricultural productivity are necessary to increase rural household incomes and access to available food but are insufficient to ensure food security. Evidence indicates that poverty reduction and food security do not necessarily move in tandem. The main problem is lack of economic (social and physical) access to food at national and household levels and inadequate nutrition. Food security not only requires an adequate supply of food but also entails availability, access, and utilization by all—people of all ages, gender, ethnicity, religion, and socioeconomic levels.

## From Agriculture to Food Security

Agriculture and food security are inextricably linked. The agricultural sector in each country is dependent on the available natural resources, as well as the politics that govern those resources. **Staple food crops** are the main source of dietary energy in the human diet and include things such as rice, wheat, sweet potatoes, maize (corn), and cassava (figure **13.1.*a***).



Figure **13.1.*a***: Cassava tubers are a staple food crop in tropical regions. Image by Renatosjoao (CC-BY-SA).

### Food Security

An individual must have access to food in ample quantity and of sufficient nutritional quality at all times to be **food secure**. Those who never have sufficient quality food are **chronically food insecure**. Food security is determined by availability, access, and utilization.

**Food availability** refers to whether enough food is produced globally to feed the world's population. In fact, enough food is produced globally, but it is not accessible to everyone who needs it. **Food access** refers to the ability to obtain food in ample quantity and of nutritional quality. In the United States, residents of **food deserts** face limited access to nutritious foods, particularly fresh produce (figure **13.1.*b***). This problem is compounded when residents do not own a vehicle or have access to public transportation to obtain foods from surrounding areas or lack the funds to do so.

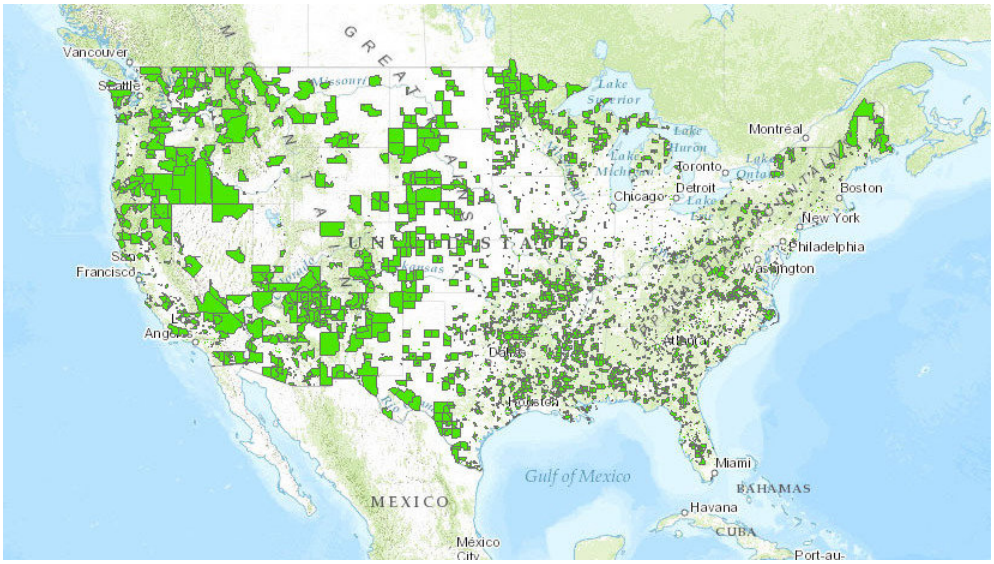



Figure **13.1.*b***: Food deserts in the contiguous United States are shaded in bright green. Food deserts are distributed over the United States. Examples of large food deserts are in southeastern Oregon and northeastern Arizona. Image by Economic Research Services, USDA (public domain).

**Food utilization** essentially translates the food available to a household into nutritional security for its members. One aspect of utilization is analyzed in terms of distribution according to need. Nutritional standards exist for the actual nutritional needs, which differ by gender, age, and life phase (for example, pregnancy), but these "needs" are often socially constructed based on culture. For example, in South Asia evidence shows that women eat after everyone else has eaten and are less likely than men in the same household to consume preferred foods such as meats and fish.

Food insecurity can result in **undernourishment** and **malnutrition** (a lack of essential nutrients; figure **13.1.*c***). The economic costs of undernourishment and malnutrition are substantial, potentially costing individuals 10% of their lifetime earnings and costing nations 2-3% of gross domestic product (GDP) in the worst-affected countries (Alderman 2005). This in part due to the reduced ability of under- and malnourished individuals to work. Additionally, without proper nutrition, children can develop disabilities that make it more difficult for them to join to workforce as adults. The economic costs reported here do not even account for the healthcare costs associated with under- and malnutrition.



Figure **13.1.*c***: A goiter is an enlarged thyroid gland that can result from deficiency of iodine, an essential nutrient. Iodized salt, commonly used in the United States, is one strategy to prevent iodine deficiencies. Image by Almazi/OpenStax (CC-BY). Access for free at openstax.org.

## Obesity

**Obesity** means having excess body fat resulting in a body mass index (BMI) of 30 or higher, and **overweight** individuals have a BMI of 25 or higher (figure **13.1.*d***). BMI is a common metric for obesity because it is determined based on weight and height; however, an individual with a big bone structure and/or high muscle mass could have a high BMI without facing the health risks associated with obesity. Likewise, an individual with excess body fat but little muscle mass might have a "healthy" BMI, but their excess body fat could still compromise their health. For these reasons, some experts rely on other metrics, such as relative fat mass (RFM). Relative fat mass is equal to height divided by waist circumference multiplied by 76 for females or 64 for males. An RFM of 32% or higher for females and 25% or higher for males is considered obese. Harvard Chan School of Public Health provides a thorough discussion of the advantages and disadvantages of various obesity metrics.

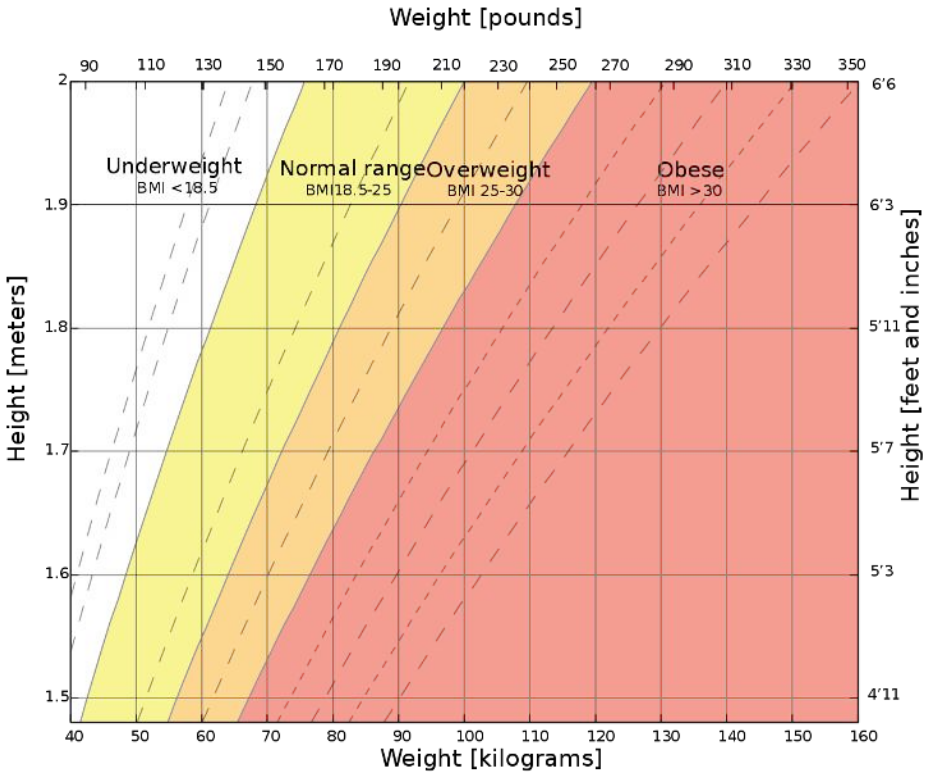



Figure **13.1.*d***: This graph shows BMI categories based on height and weight. Height in feet and inches is on the right axis, and weight in pounds is on the top axis. There are four BMI categories: underweight (< 18.5), normal (18.5-25), overweight (25-30), and obese (> 30). BMI is calculated as mass divided by height squared ($kg/m^2$). Image by InvictaHOG (public domain).

Obesity has become a significant global health challenge, yet is preventable and reversible. Since the 1970s, obesity rates have rapidly increased, resulting in a global obesity epidemic. Thirty-nine percent of the world's adults, or 1.9 billion people, are overweight in 2016. Of these, 650 million are obese (Obesity and Overweight). Obesity was linked to 4.7 million premature deaths in 2017 (about 8% of all deaths that year; Hannah/Our World in Data). The economic cost of obesity has been estimated at $2 trillion (Tremmel et al.).

Initially centered in developed countries, the obesity epidemic now affects countries of all economic statuses. In low-income countries, this results in a triple burden of undernourishment, malnutrition, and obesity. There is significant variation by region; some have very high rates of undernourishment and low rates of obesity, while in other regions the opposite is true (figure **13.1.*e***).

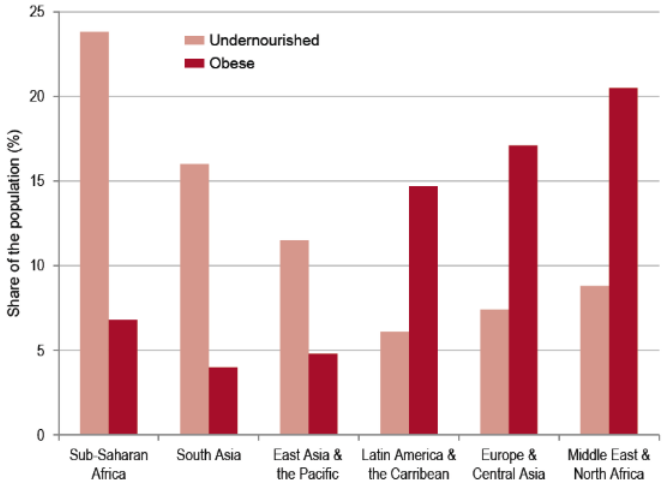



Figure **13.1.*e***: Obesity (red bars on the right) and undernourishment (pink bars on the left) by region. The share of the population (%) that is undernourished is approximately 24% in Sub-Saharan Africa while approximately 7% of the population in this region is obese. For South Asia 16% of the population is undernourished, and 4% is obese. These percentages 11% and 5% for East Asia and the Pacific, 5% and 15% for Latin America and the Caribbean, 7% and 17% for Europe and Central Asia, and 9% and 20% for the Middle East and North Africa, respectively.

Obesity accounts for a growing level and share of worldwide noncommunicable diseases, including diabetes, heart disease, and certain cancers that can reduce quality of life and increase public health costs of already under-resourced developing countries. Driven primarily by increasing availability of processed, affordable, and effectively marketed food, the global food system is falling short with rising obesity and related poor health outcomes. Due to established health implications and rapid increase in prevalence, obesity is now a recognized major global health challenge.

## Supplemental Reading

The United Nations explains global food supply and strategies for addressing world hunger here using a variety of informational graphics and videos.

## References

Alderman, H. 2005. Linkages between Poverty Reduction Strategies and Child Nutrition: An Asian Perspective. *Economic and Political Weekly*, *40*(46), 4837-4842. Accessed 2021-01-03.

Hannah Ritchie. 2017. Obesity. Our World in Data. Accessed 2021-01-03.

Goal 1: End poverty in all its forms everywhere. Sustainable Development. Department of Economic and Social Affairs. United Nations. Accessed 2021-01-03.

Obesity and Overweight. 2020. World Health Organization. Accessed 2021-01-03.

The State of Food Security and Nutrition in the World in 2020. Food and Agriculture Organization of the United Nations. Accessed 2021-01-03.

Tremmel, M., Gerdtham, U. G., Nilsson, P. M., & Saha, S. 2017. Economic Burden of Obesity: A Systematic Literature Review. *International journal of environmental research and public health*, *14*(4), 435. Accessed 2021-01-03.

## Attribution

Modified by Melissa Ha from Food Security from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)

This page titled 13.1: Food Security is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **14.1: Food Security** by Melissa Ha and Rachel Schleiger is licensed CC BY-SA 4.0.

# 13.2: Industrial Agriculture

Also known as conventional agriculture, **industrial agriculture** is a method of farming that involves use of synthetic fertilizer, synthetic pesticides, and machinery. Industrial agriculture is dependent on large investments in mechanized equipment powered mostly by fossil fuels. In the case of livestock, most production comes from systems where animals are highly concentrated and confined.

## Synthetic Pesticides

**Pests** are organisms that occur where they are not wanted or that cause damage to crops, humans or other animals (figure **13.2.*a***). Thus, the term "pest" is a highly subjective term. A **pesticide** is a term for any substance intended for preventing, destroying, repelling, or mitigating any pest. Though often misunderstood to refer only to insecticides, the term pesticide also applies to **herbicides** (weed killers), fungicides, and various other substances used to control pests. Chemical pesticides can be effective, fast acting, adaptable to all crops and situations. When first applied, pesticides can result in impressive production gains of crops. However, despite these initial gains, excessive use of pesticides can be ecologically unsound (see Disadvantages of Industrial Agriculture). By their very nature, most pesticides create some risk of harm—pesticides can cause harm to humans, animals, and/or the environment because they are designed to kill or otherwise adversely affect living things. At the same time, pesticides are useful to society because they can kill potential disease-causing organisms and control insects, weeds, worms, and fungi.



Figure **13.2.*a***: The cotton boll weevil is considered a major pest because of the damage it does to cotton plants. Image by ARS USDA (public domain).

## Monoculture Farming

Industrial agriculture employs **monoculture** farming, which involves growing only one species of crop in a large area (figure **13.2.*b***). The plants in a monoculture are evenly spaced and have the same planting, irrigation, fertilizer, harvesting, etc. requirements, and this uniformity results in efficient use of farming machinery. Often, the crop is a **high yield variety** (high yielding variety; HYV), which is produced through selectively breeding individual plants with desirable characteristics. Compared to traditional crops, HYVs can produce more per unit area of land. Monoculture farming compromises genetic and species diversity in agriculture, however, promoting the spread of pests and risking that all the plants for hundreds of acres could be susceptible to the same disease (see Genetic Diversity).



Figure **13.2.*b***: Monocultures like this potato farm (left) and cotton farm (right) make it easier to use farming equipment; however, this reduces genetic diversity and consumes fossil fuels. Left image by NightThree (CC-BY) and right image by Kimberly Vardeman (CC-BY).

## Advantages of Industrial Agriculture

Industrial agriculture has delivered tremendous gains in productivity and efficiency. Food production worldwide has overall risen since the 1940s (figure **13.2.*c***); the World Bank estimates that between 70 percent and 90 percent of the recent increases in food production are the result of industrial agriculture rather than greater acreage under cultivation (figure **13.2.*d***). United States consumers have come to expect abundant and inexpensive food. Additionally, the use of synthetic pesticides ensures that produce is relatively free of blemishes (figure **13.2.*e***).

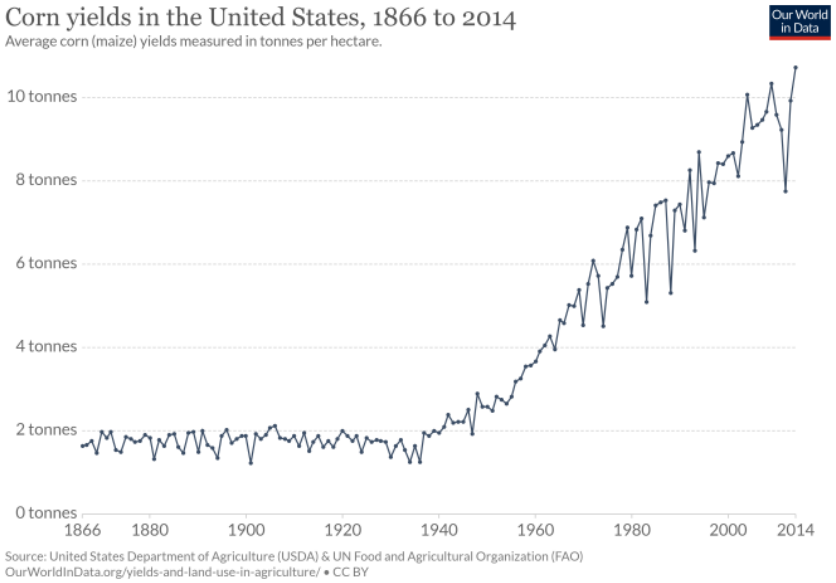



Figure **13.2.*c***: Corn yields in metric tons (tonnes; 1000 kilograms) per hectare over time. While yields have fluctuated from year to year, they have overall increased since the 1940s. This improving agricultural efficiency is partly due to industrial agricultural practices. Image by Hannah Ritchie and Max Roser/Our World in Data (CC-BY).

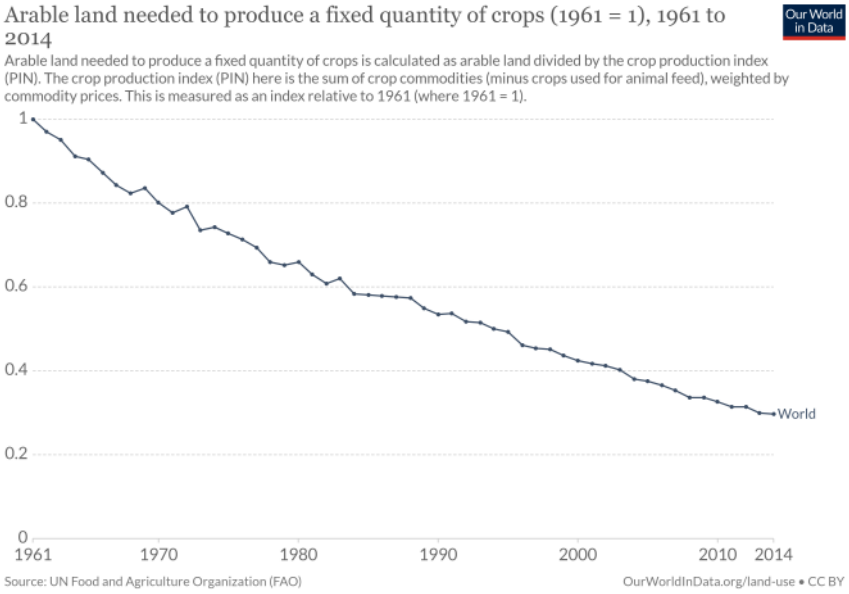



**Figure 13.2.*d*:** The arable (farmable) land required to produce a fixed amount of a crop has decreased over time due to improved agricultural efficiency. Image by Hannah Ritchie and Max Roser/Our World in Data (CC-BY).



Figure **13.2.*e***: Many of us carefully inspect produce at the grocery store and select those without blemishes. This visually appealing produce results from synthetic pesticides use. Image by Amanda Mills (public domain).

Industrial agriculture expands where crops can be cultivated. Applying synthetic fertilizers to poor soils can make them fertile. Additionally, irrigation systems promote high agricultural productivity in regions and during seasons that would otherwise only support the most drought-tolerant species. For example, about 25% of the food consumed in the United States is produced in California's Central Valley, which receives little rain in the summer months (figure **13.2.*f***).

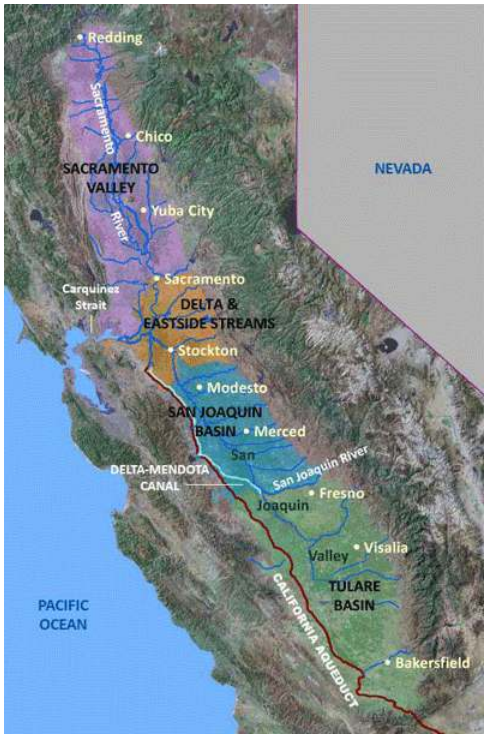



Figure **13.2.*f***: The four regions of California's Central Valley from north to south are the Sacramento Valley, Delta and Eastside Streams, San Joaquin Basin, and Tulare Basin. Image by USGS (public domain).

## Disadvantages of Industrial Agriculture

Economically, the U.S. agricultural sector includes a history of increasingly large federal expenditures. Also observed is a widening disparity among the income of farmers and the escalating concentration of **agribusiness**—industries involved with manufacture,

processing, and distribution of farm products—into fewer and fewer hands. Market competition is limited and farmers have little control over prices of their goods, and they continue to receive a smaller and smaller portion of consumer dollars spent on agricultural products.

Economic pressures have led to a tremendous loss to the number of farms, particularly small farms, and farmers during the past few decades. More than 155,000 farms were lost from 1987 to 1997. Economically, it is very difficult for potential farmers to enter the business today because of the high cost of doing business. Farming equipment is expensive, and the industrial agricultural system favors large corporate farms, which can more easily invest in this equipment. While this equipment promotes efficiency, it also reduces the number of jobs in agriculture (figure **13.2.*g***).

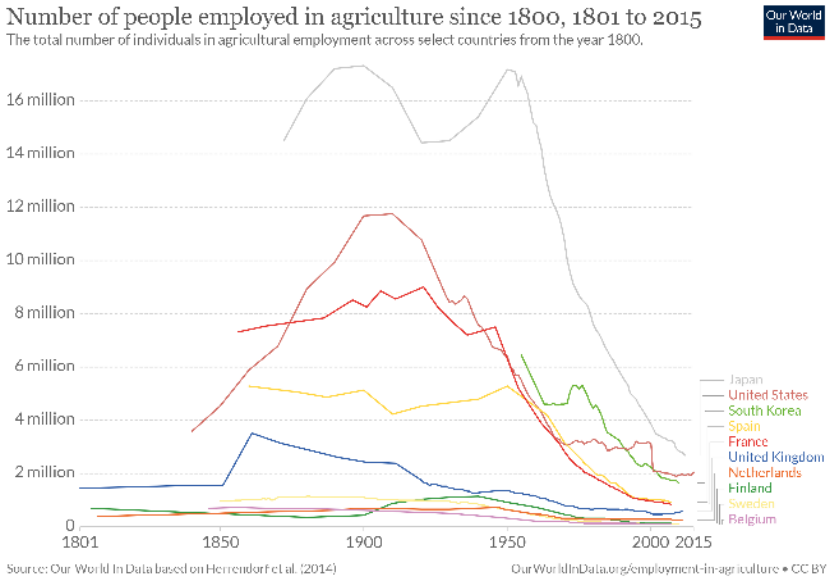



Figure **13.2.*g***: The number of people working in agriculture in selected countries over time (from 1801-2015). Generally, the number of agricultural workers has declined as farming practices increased in efficiency and as farming machinery completed tasks that were once manual. This is both an advantage because people are free to pursue other careers and a disadvantage because of the reduction in jobs. In Japan, the number of agricultural workers peaked above 16 million around 1900 and again around 1960. In the United States, it peaked to nearly 12 million around 1900. France followed a similar pattern, peaking around 1925 at approximately 9 million workers. There were over 6 million agricultural workers in South Korea in 1950, and this has overall declined (although there was a spike around 1980). In Spain the number of workers was over 5 million in 1950 and has since decreased. In the United Kingdom, the number of agricultural workers peaked around 1960 at about 3.5 million. The other countries shown (Netherlands, Finland, Sweden, and Belgium) have consistently had under 1 million agricultural workers. Image by Max Roser/Our World in Data (CC-BY).

The mechanical farming equipment required for industrial agriculture relies on fossil fuels. Burning fossil fuels releases air pollution, including greenhouse gases, which cause climate change. Additionally, the synthetic fertilizers and pesticides used in industrial agriculture are produced from fossil fuels. Furthermore, agriculture areas rarely contain as much biomass (organic matter that comprises living things) as biodiverse, intact ecosystems. Biomass is an important carbon sink because its organic molecules store carbon that could otherwise be in the atmosphere as carbon dioxide or methane, both greenhouse gases. Soils with a substantial organic component are also carbon sinks. Industrial agricultural practices that deplete these soils (see below) contribute to climate change, but alternative practices can promote this benefit of healthy soils (see Sustainable Agriculture).

While synthetic fertilizers provide crops with the necessary nutrients for high productivity, excess nutrients from fertilizers can enter bodies of water through runoff. This can lead to a bloom of algae or photosynthetic bacteria in lakes, rivers, and bays in a process called **eutrophication**. These photosynthetic microorganisms sometimes produce toxins that kill aquatic animals and can even harm the humans who consume these animals. Furthermore, they block aquatic plants from accessing light. Eventual decomposition of algal blooms requires oxygen, which results in **hypoxia** (low levels of dissolved oxygen), further harming aquatic species. Additionally, industrial agriculture pollutes water with pesticides and sediments. Pesticides are not only widespread in surface waters, but pesticides from every chemical class have been detected in groundwater.

Industrial agricultural practices often deplete soil quality. For example, wind and water erosion of exposed topsoil removes particles and nutrients from the soil (which also contributes to eutrophication and sediment pollution). Tilling (mixing the soil) and overgrazing of livestock exacerbates erosion. When livestock remove too much vegetation from an area, plant roots no longer anchor the soil in place, and erosion ensues. Farming equipment compacts the soil. This limits infiltration of water into the soil (decreases soil permeability), makes it more difficult for roots to penetrate the soil, and hinders gas exchange in roots. In extreme cases, soil has eroded and lost its ability to retain water that once arable (farmable) land becomes a desert-like (**desertification**; see Soil Degradation; figure **13.2.*h***).



Figure **13.2.*h***: Overgrazing depleted the soil in the Rukwa Plains of southwestern Tanzania. Image by Lichinga (CC-BY-SA).

Irrigation can lead to **salinization** (increased salinity). While the water itself eventually evaporates, is transpired, or drains from the soil, the minerals dissolved in the water may remain in the soil. Over time, these minerals can accumulate to levels that are toxic for most plants. As discussed in Water Usage, irrigation for agriculture accounts for 69% of global water usage. Withdrawal of sufficient water for irrigation contributes to water shortages and diverts water from ecosystems (see Aqueducts).

Repeated pesticide application exerts selective pressure on insects, fungi, and other crop pests to evolve pesticide resistance. By chance, some individuals in a pest population may carry gene versions that confer pesticide resistance. When these populations are exposed to pesticides, individuals with these gene versions are the most likely to survive and reproduce. They then pass these genes to their offspring, and pesticide resistance becomes more and more common in the population over time (figure **13.2.*i***). At this point, pesticides may need to be applied at greater concentrations or more frequently to achieve the same effect. If pesticide that was initially used is no longer effective at all, the farmer would be forced to find a different pesticide. Over 400 insects and mite pests and more than 70 fungal pathogens have become resistant to one or more pesticides.

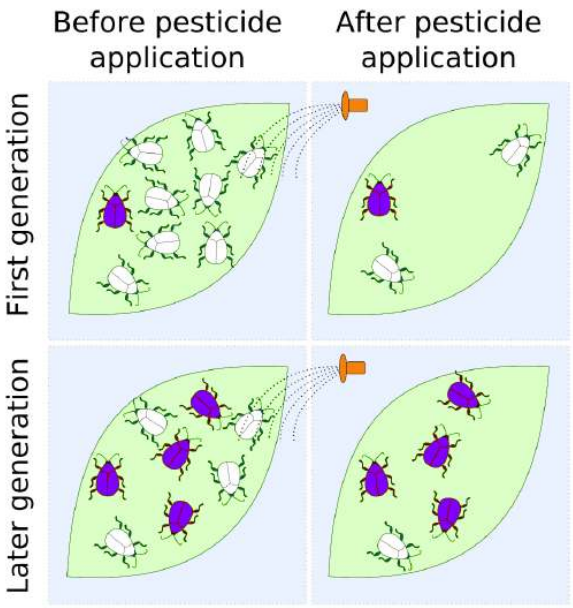



Figure **13.2.*i***: Pesticide application can exert selection for the evolution of pesticide resistance. In this hypothetical example, the purple beetle has a gene version that makes it resistant to pesticides (upper left). Pesticides are applied, killing most of the susceptible individuals (white beetles; upper right). In a later generation, purple, resistant beetles are more common because they had a better chance of surviving and reproducing (lower left). Upon another pesticide application, an even greater proportion of the population is resistant. Image modified from Delldot (CC-BY-SA).

The human health and environmental costs of pesticide use have been unevenly distributed. Despite the fact that the lion's share of chemical pesticides are applied in developed countries, 99 percent of all pesticide poisoning cases occur in developing countries where regulatory, health and education systems are weakest. Many farmers in developing countries overuse pesticides and do not take proper safety precautions because they do not understand the risks and fear smaller harvests. Making matters worse, developing countries seldom have strong regulatory systems for dangerous chemicals; pesticides banned or restricted in industrialized countries are used widely in developing countries. Farmers' perceptions of appropriate pesticide use vary by setting and culture. Prolonged exposure to pesticides has been associated with several chronic and acute health effects like non-Hodgkin's lymphoma, leukemia, as well as cardiopulmonary disorders, neurological and hematological symptoms, and skin diseases.

Pesticides have also placed stresses on pollinators and other beneficial insect species. They can lead to the destruction of natural enemies, which would regulate pest populations in an intact ecosystem. For example, once apple growers began controlling pests with the pesticide dichlorodiphenyltrichloroethane (DDT), which has since been banned, they quickly found their orchards being attacked by scale insects and mites. The reason: DDT had killed off their natural enemies. More generally, pollution from industrial agriculture along with habitat loss due to converting healthy ecosystems into agricultural fields is a major threat to biodiversity. As mentioned previously, monoculture farming limits biodiversity in crops.

With respect to animal agriculture, keeping livestock are kept in close quarters increases the risk of disease spread among them. For these reason, food animals are commonly given antibiotics, either preventatively or to treat existing infections. As discussed in the Infectious Diseases section, the overuse of antibiotics (whether in humans or farm animals) increases selection for the evolution of antibiotic-resistant strains of bacteria or other disease-causing organisms. Infections by antibiotic-resistant strains are more difficult to treat.

As with many industrial practices, potential health hazards are often tied to farming practices. Under research and investigation currently is the sub-therapeutic use of antibiotics in animal production, and pesticide and nitrate contamination of water and food. Farm worker health is also a consideration in all farming practices.

## Attributions

Modified by Melissa Ha from the following sources:

- Conventional Agriculture and Pests and Pesticides from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)

- Insecticides from *Biology* by John W. Kimball (licensed under CC-BY)
- Sustainable Agriculture: Definitions and Terms by Mary V. Gold, Alternative Farming Systems Information Center from National Agricultural Library, USDA (public domain)

---

This page titled 13.2: Industrial Agriculture is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **14.2: Industrial Agriculture** by Melissa Ha and Rachel Schleiger is licensed CC BY-SA 4.0.

# 13.3: Pests and Pesticides

**Pests** are organisms that occur where they are not wanted or that cause damage to crops or humans or other animals. Thus, the term "pest" is a highly subjective term. A **pesticide** is a term for any substance intended for preventing, destroying, repelling, or mitigating any pest. Though often misunderstood to refer only to insecticides, the term pesticide also applies to herbicides, fungicides, and various other substances used to control pests. By their very nature, most pesticides create some risk of harm—pesticides can cause harm to humans, animals, and/or the environment because they are designed to kill or otherwise adversely affect living things. At the same time, pesticides are useful to society because they can kill potential disease-causing organisms and control insects, weeds, worms, and fungi.



Figure **13.3.1**. The cotton boll weevil is considered a major pest because of the damage it does to cotton plants. *This work by Jimmy Smith is licensed under CC BY-NC-ND 4.0.*

## Pest Control is a Common Tool

The management of pests is an essential part of agriculture, public health, and maintenance of power lines and roads. Chemical pest management has helped to reduce losses in agriculture and to limit human exposure to disease vectors, such as mosquitoes, saving many lives. Chemical pesticides can be effective, fast acting, adaptable to all crops and situations. When first applied, pesticides can result in impressive production gains of crops. However, despite these initial gains, excessive use of pesticides can be ecologically unsound, leading to the destruction of natural enemies, the increase of pesticide resistance, and outbreaks of secondary pests.

These consequences have often resulted in higher production costs as well as environmental and human health costs– side-effects which have been unevenly distributed. Despite the fact that the lion's share of chemical pesticides are applied in developed countries, 99 percent of all pesticide poisoning cases occur in developing countries where regulatory, health and education systems are weakest. Many farmers in developing countries overuse pesticides and do not take proper safety precautions because they do not understand the risks and fear smaller harvests. Making matters worse, developing countries seldom have strong regulatory systems for dangerous chemicals; pesticides banned or restricted in industrialized countries are used widely in developing countries. Farmers' perceptions of appropriate pesticide use vary by setting and culture. Prolonged exposure to pesticides has been associated with several chronic and acute health effects like non-Hodgkin's lymphoma, leukemia, as well as cardiopulmonary disorders, neurological and hematological symptoms, and skin diseases.

### HUMAN HEALTH, ENVIRONMENTAL, AND ECONOMIC EFFECTS OF PESTICIDE USE IN POTATO PRODUCTION IN ECUADOR

The International Potato Center (CIP) conducted an interdisciplinary and inter-institutional research intervention project dealing with pesticide impacts on agricultural production, human health, and the environment in Carchi, Ecuador. Carchi is the most important potato-growing area in Ecuador, where smallholder farmers dominate production. They use tremendous amounts of pesticides for the control of the Andean potato weevil and the late blight fungus. Virtually all farmers apply class 1b highly toxic pesticides using hand pump backpack sprayers.

The study found that the health problems caused by pesticides are severe and are affecting a high percentage of the rural population. Despite the existence of technology and policy solutions, government policies continue to promote the use of pesticides. The study conclusions concurred with those by the pesticide industry, "that any company that could not ensure the safe use of highly toxic pesticides should remove them from the market and that it is almost impossible to achieve safe use of highly toxic pesticides among small farmers in developing countries."

Source: Yanggen et al. 2003.

## Persistent Organic Pollutants

**Persistent organic pollutants** (POPs) are a group of organic chemicals, such as DDT, that have been widely used as pesticides or industrial chemicals and pose risks to human health and ecosystems. POPs have been produced and released into the environment by human activity. They have the following three characteristics:

- **Persistent**: POPs are chemicals that last a long time in the environment. Some may resist breakdown for years and even decades while others could potentially break down into other toxic substances.
- **Bioaccumulative**: POPs can accumulate in animals and humans, usually in fatty tissues and largely from the food they consume. As these compounds move up the food chain, they concentrate to levels that could be thousands of times higher than acceptable limits.
- **Toxic**: POPs can cause a wide range of health effects in humans, wildlife and fish. They have been linked to effects on the nervous system, reproductive and developmental problems, suppression of the immune system, cancer, and endocrine disruption. The deliberate production and use of most POPs has been banned around the world, with some exemptions made for human health considerations (e.g., DDT for malaria control) and/or in very specific cases where alternative chemicals have not been identified. However, the unintended production and/or the current use of some POPs continue to be an issue of global concern. Even though most POPs have not been manufactured or used for decades, they continue to be present in the environment and thus potentially harmful. The same properties that originally made them so effective, particularly their stability, make them difficult to eradicate from the environment.

### POPs and Health

The relationship between exposure to environmental contaminants such as POPs and human health is complex. There is mounting evidence that these persistent, bioaccumulative and toxic chemicals (PBTs) cause long-term harm to human health and the environment. Drawing a direct link, however, between exposure to these chemicals and health effects is complicated, particularly since humans are exposed on a daily basis to many different environmental contaminants through the air they breathe, the water they drink, and the food they eat. Numerous studies link POPs with a number of adverse effects in humans. These include effects on the nervous system, problems related to reproduction and development, cancer, and genetic impacts. Moreover, there is mounting public concern over the environmental contaminants that mimic hormones in the human body (**endocrine disruptors**).

As with humans, animals are exposed to POPs in the environment through air, water and food. POPs can remain in sediments for years, where bottom-dwelling creatures consume them and who are then eaten by larger fish. Because tissue concentrations can increase or biomagnify at each level of the food chain, top predators (like largemouth bass or walleye) may have a million times greater concentrations of POPs than the water itself. The animals most exposed to PBT contaminants are those higher up the food web such as marine mammals including whales, seals, polar bears, and birds of prey in addition to fish species such as tuna, swordfish and bass (Figure **13.3.2**). Once POPs are released into the environment, they may be transported within a specific region and across international boundaries transferring among air, water, and land.

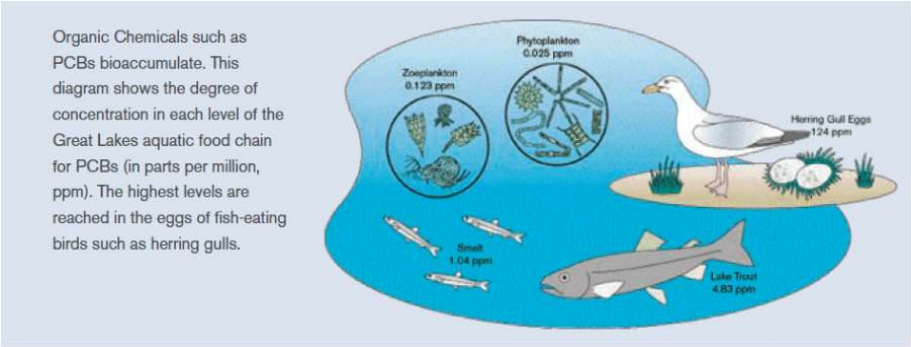



Figure **13.3.2**. Bioaccumulation and biomagnification. U.S. EPA. Great Lakes: The Great Lakes Atlas: Chapter Four the Great Lakes Today – Concerns. January 2009

## "Grasshopper Effect"

While generally banned or restricted, POPs make their way into and throughout the environment on a daily basis through a cycle of long-range air transport and deposition called the "**grasshopper effect.**"The "grasshopper" processes, illustrated in Figure **13.3.3**,

begin with the release of POPs into the environment. When POPs enter the atmosphere, they can be carried with wind currents, sometimes for long distances.

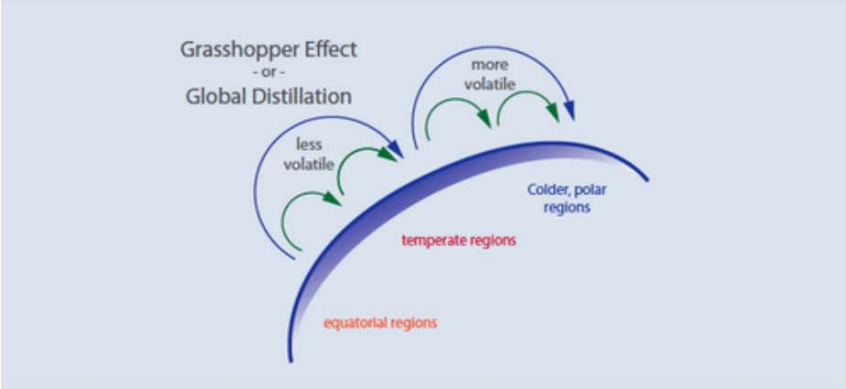



Figure **13.3.3**. How POPs move throughout the environment. Source: Environment Canada. The Science and the Environment Bulletin. May/June 1998

Through atmospheric processes, they are deposited onto land or into water ecosystems where they accumulate and potentially cause damage. From these ecosystems, they evaporate, again entering the atmosphere, typically traveling from warmer temperatures toward cooler regions. They condense out of the atmosphere whenever the temperature drops, eventually reaching highest concentrations in circumpolar countries. Through these processes, POPs can move thousands of kilometers from their original source of release in a cycle that may last decades.

## Contributors and Attributions

- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0. Modified from the original by Matthew R. Fisher.

---

This page titled 13.3: Pests and Pesticides is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Matthew R. Fisher (OpenOregon) via source content that was edited to the style and standards of the LibreTexts platform.

- **9.4: Pests and Pesticides** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# 13.4: Soil Degradation

Once fertile topsoil is lost, it is not easily replaced. **Soil degradation** refers to deterioration in the quality of the soil and the resultant reduction in its capacity to produce. Soils are degraded primarily by erosion, compaction, and salinization. Such processes often arise from poor soil management during agricultural activities. In extreme cases, soil degradation can lead to **desertification** (conversion of land to desert-like conditions) of croplands and rangelands in semi-arid regions. See the Sustainable Agriculture section for strategies to preserve soil quality (**soil conservation**).

## Erosion

**Erosion** is the biggest cause of soil degradation. Soil productivity is reduced as a result of losses of nutrients, water holding capacity, and organic matter. **Water holding capacity** is a measure of of the soils ability to retain water. The two agents of erosion are wind and water, which act to remove the finer particles from the soil. **Wind erosion** occurs mostly in flat, dry areas and moist, sandy areas along bodies of water. Wind not only removes soil, but also dries and degrades the soil structure. **Water erosion** is the most prevalent type of erosion. It occurs when raindrops splash on the ground and when water moves down slope as a thin film, small streams, or a large stream.

Some amount of soil erosion is a natural process along sloping areas and/or in areas with soft or materials that do not stick together and are susceptible to movement by water, wind, or gravity. For instance, soil material can be mobilized in strong windstorms, along the banks of rivers, in landslides, or by wave action along coastlines. However, human activities such as construction, logging, and off-road vehicle use promote erosion by removing the natural vegetation cover protecting the soil. Agricultural practices such as overgrazing and leaving plowed fields bare for extended periods contribute to farmland erosion. Each year, an estimated two billion metric tons of soil are eroded from farmlands in the United States alone. The soil transported by the erosion processes can also create problems elsewhere (for example, by clogging waterways and filling ditches and low-lying land areas). The areas most vulnerable to soil erosion include locations with thin organic (A and O) horizons and hilly terrains (figure **13.4.*a***).

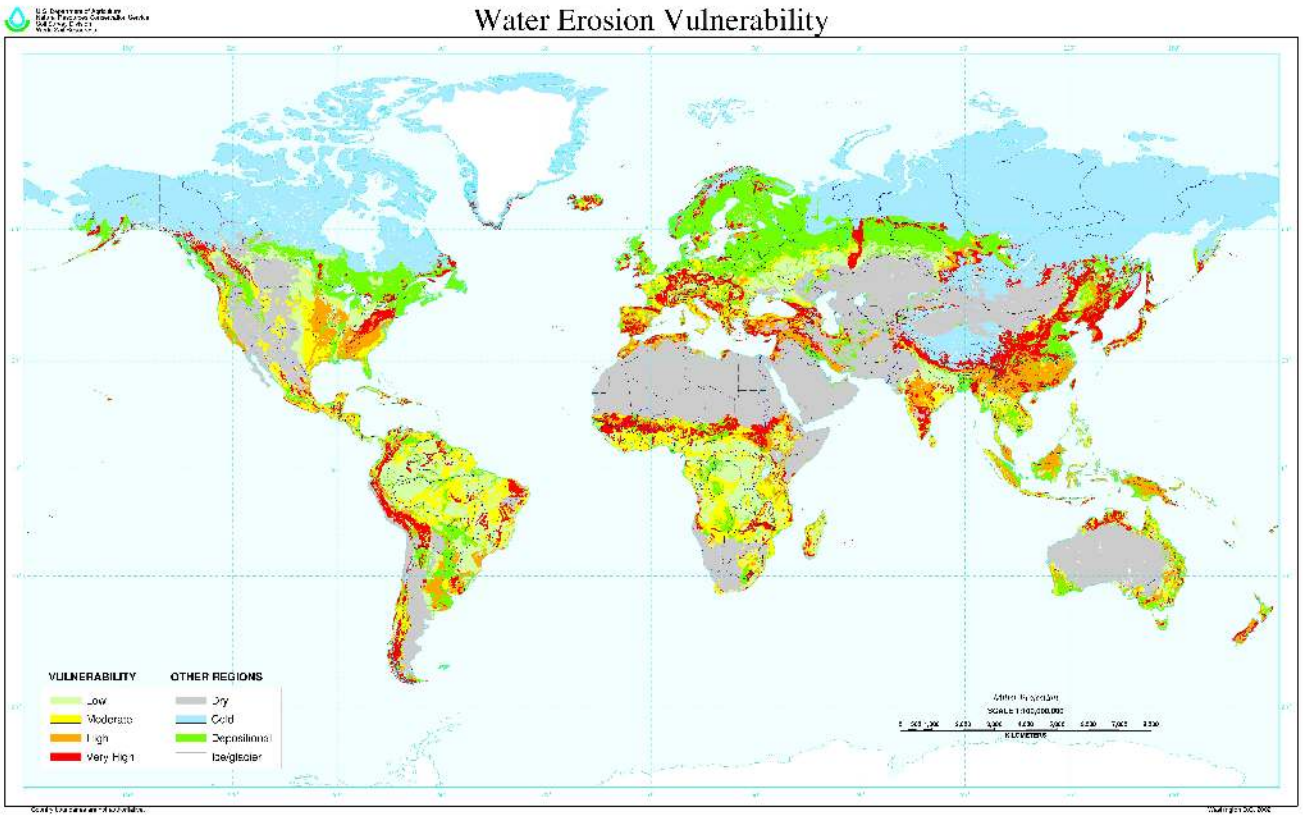



Figure 13.4.a: Global map of water erosion vulnerability. There are four levels of vulnerability: very high (red), high (orange), moderate (yellow), and low (light green). Examples of regions with very high vulnerability to erosion by water include the west coast of South America, parts of the eastern United States, regions around the Mediterranean Sea, a strip across western and central Africa, Korean and Japan, Northern Australia, and New Zealand. Examples of regions with low vulnerability include northern South America, the Democratic Republic of Congo, northern India and Nepal, and western Russia. Other regions are labeled as dry (gray), cold (light blue), depositional (bright green), or ice/glacier (white). For example, northern regions are cold and ice/glacier. Much of the western United States, Chile and Argentina, northern Africa, the Middle East, Kazakhstan, Mongolia, and much of Australia are considered dry. The largest depositional area (region where sediments accumulate) is northeastern Europe. Image by USDA-NRCS, Soil Science Division, World Soil Resources (public domain).

## Compaction

In modern agricultural practices, heavy machinery is used to prepare the seedbed, for planting, to control weeds, and to harvest the crop. The use of heavy equipment has many advantages in saving time and labor, but can cause compaction of soil and disruption of the natural soil biota. Much compaction is reversible and some is unavoidable with modern practices; however, serious compaction issues can occur when the equipment is used excessively during times when the soil has a high water content. The problem with soil compaction is that increased soil density limits root penetration depth and may inhibit proper plant growth.

## Salinization

When considerable quantities of salt accumulate in the soil in a process known as **salinization**, many plants are unable to grow properly or even survive. This is especially a problem in irrigated farmland. Groundwater used for irrigation contains small amounts of dissolved salts. Irrigation water that is not absorbed into the soil evaporates, leaving the salts behind. This process repeats itself and eventually severe salinization of the soil occurs. A related problem is water logging of the soil. When cropland is irrigated with excessive amounts of water in order to leach salts that have accumulated in the soil, the excess water is sometimes unable to drain away properly. In this case it accumulates underground and causes a rise in the subsurface water table. If the saline water rises to the level of the plant roots, plant growth is inhibited.

## Desertification

Land that was previously suited for growing crops may be turned into desert by climate change and the activities of humans, such as poor farming practices, livestock overgrazing, and overuse of available water. This process, called **desertification**, is a serious problem worldwide. Plants and soil types that are non-arid (not dry) specifically help water soak into the ground (infiltration) and water retention. When desertification begins, it leads to a reduction in vegetation and degraded soil quality, and this further increases the aridity and spreads the desert via a **positive feedback loop** (meaning that the processes feed on themselves promoting an increasing spiral).

Figure **13.4.*b*** shows areas of the world and their vulnerability to desertification. Note the red and orange areas in the western and midwestern United States. The Dust Bowl of the 1930s is a classic example of human-caused desertification (figure **13.4.*c***). Poor cultivation and grazing practices -- coupled with severe drought conditions -- led to severe wind erosion of soil in a region of the Great Plains that became known as the "Dust Bowl". Wind stripped large areas of farmlands of topsoil, and formed clouds of dust that traveled as far as the eastern United States.

Sometimes there is a conflict between what is known to prevent desertification and what an individual farmer feels is needed to do to make a living. Mitigating the desertification process includes both societal steps and individual education on alternatives.

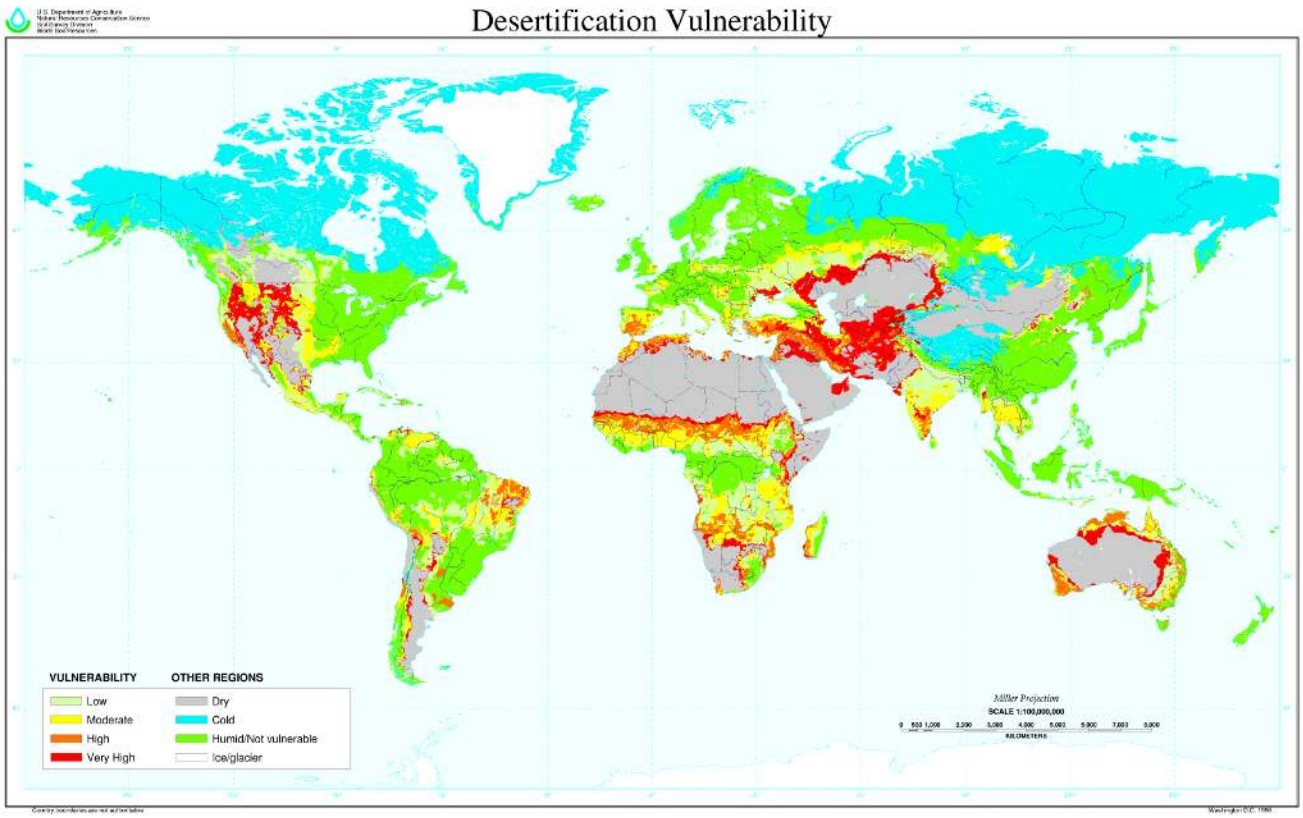



Figure **13.4.*b***: World map showing desertification vulnerability. There are four levels of vulnerability: very high (red), high (orange), moderate (yellow), and low (light green). Examples of regions with a very high vulnerability to desertification include the northwestern United States, a strip through western and central Africa, parts of the middle East, and parts of Australia. Examples of regions with low vulnerability include parts of Canada, parts of western Europe, and parts of Angola and the Democratic Republic of Congo. Other regions are labeled as dry (gray), cold (light blue), humid/not vulnerable (bright green), or ice/glacier (white). For example, northern regions are cold and ice/glacier. Some of southwestern United States, Chile and Argentina, northern Africa, the Middle East, Kazakhstan, Mongolia, and much of Australia are considered dry. Examples of humid/not vulnerable areas include the eastern United States, northern and eastern South America, western Europe, Central Africa, and East Asia. Image by USDA (public domain).



Figure **13.4.*c***: A cloud of dust in Kansas in 1935, part of the Dust Bowl. Image by NOAA Photo Library (public domain).

## Attribution

Modified by Melissa Ha from the following sources:

- The Great Basin and the Basin and the Range from *An Introduction to Geology* by Chris Johnson et al. (licensed under CC BY-NC-SA)
- Soils from *AP Environmental Science* by University of California College Prep, University of California (licensed under CC-BY). Download for free at CNX.
- Soil and Sustainability from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors (licensed under CC-BY). Download for free at CNX.

---

This page titled 13.4: Soil Degradation is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **7.5: Soil Degradation** by Melissa Ha and Rachel Schleiger is licensed CC BY-NC-SA 4.0.

# 13.5: Selective Breeding and Genetic Engineering

The development of a new crop variety is an example of **agricultural biotechnology**, a range of tools that include both traditional breeding techniques and more modern lab-based methods. Traditional methods date back thousands of years, whereas biotechnology uses the tools of genetic engineering developed over the last few decades.

## Selective Breeding (Artificial Selection)

Nearly all the fruits and vegetables found in your local market would not occur naturally. In fact, they exist only because of human intervention that began thousands of years ago. Humans created the vast majority of crop species by using traditional breeding practices on naturally-occurring, wild plants. These practices rely upon **selective breeding** (artificial selection), human-facilitated reproduction of individuals with desirable traits. For example, high yield varieties were produced through selective breeding. Traditional breeding practices, although low-tech and simple to perform, have the practical outcome of modifying an organism's genetic information, thus producing new traits.

Selective breeding is limited, however, by the life cycle of the plant and the genetic variants that are naturally present. For example, even the fastest flowering corn variety has a generation time of 60 days (the time required for a seed to germinate, produce a mature plant, get pollinated, and ultimately produce more seeds) in perfect conditions. Each generation provides an opportunity to selectively breed individual plants and generate seeds that are slightly closer to the desired outcome (for example, producing bigger, juicier kernels). Furthermore, if no individuals happen to possess gene variants that result in bigger, juicier kernels, it is not possible to artificially select this trait. Finally, traditional breeding shuffles all of the genes between the two individuals being bred, which can number into the tens of thousands (maize, for example, has 32,000 genes). When mixing such a large number of genes, the results can be unpredictable.

An interesting example is maize (corn). Biologists have discovered that maize was developed from a wild plant called teosinte. Through traditional breeding practices, humans living thousands of years ago in what is now Southern Mexico began selecting for desirable traits until they were able to transform the plant into what is now known as maize (figure **13.5.*a***). In doing so, they permanently (and unknowingly) altered its genetic instructions.

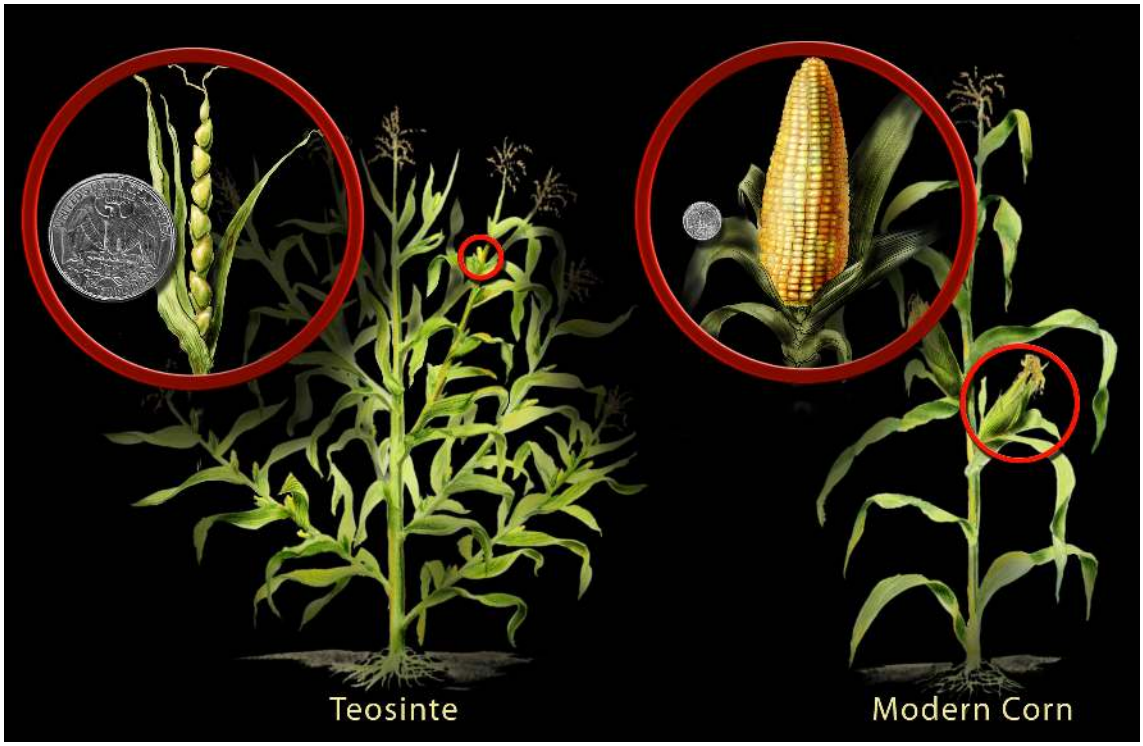



Figure **13.5.*a***: A wild grass called teosinte was genetically modified through selective breeding to produce what is now known as maize (corn). This process of transformation started thousands of years ago by indigenous people of what is now Mexico. Image by Nicolle Rager Fuller/National Science Foundation (public domain).

This history of genetic modification is common to nearly all crop species. For example, cabbage, broccoli, Brussel sprouts, cauliflower, and kale were all developed from a single species of wild mustard plant (figure **13.5.*b***). Wild nightshade was the source of tomatoes, eggplant, tobacco, and potatoes, the latter developed by humans 7,000 – 10,000 years ago in South America.



Figure **13.5.*b***: *Brassica oleracea* is a plant in the mustard family and is known as wild cabbage. From it were developed many familiar crops, such as cauliflower, broccoli, Brussel sprouts, and of course, cabbage. Image by Kulac (CC-BY-SA).

## Genetic Engineering

**Genetic engineering** is the process of directly altering an organism's DNA to produce the desired crops more rapidly than selective breeding. Because genes can be obtained from other species or even synthesized in the lab, scientists are not limited by existing genetic variation within a crop species (or closely related species with which they can be crossed). This broadens the possible traits that can be added to crops. Modern genetic engineering is more precise than selective breeding in the sense that biologists can modify just a single gene. Also, genetic engineering can introduce a gene between two distantly-related species, such as inserting a bacterial gene into a plant (figure **13.5.*c***).

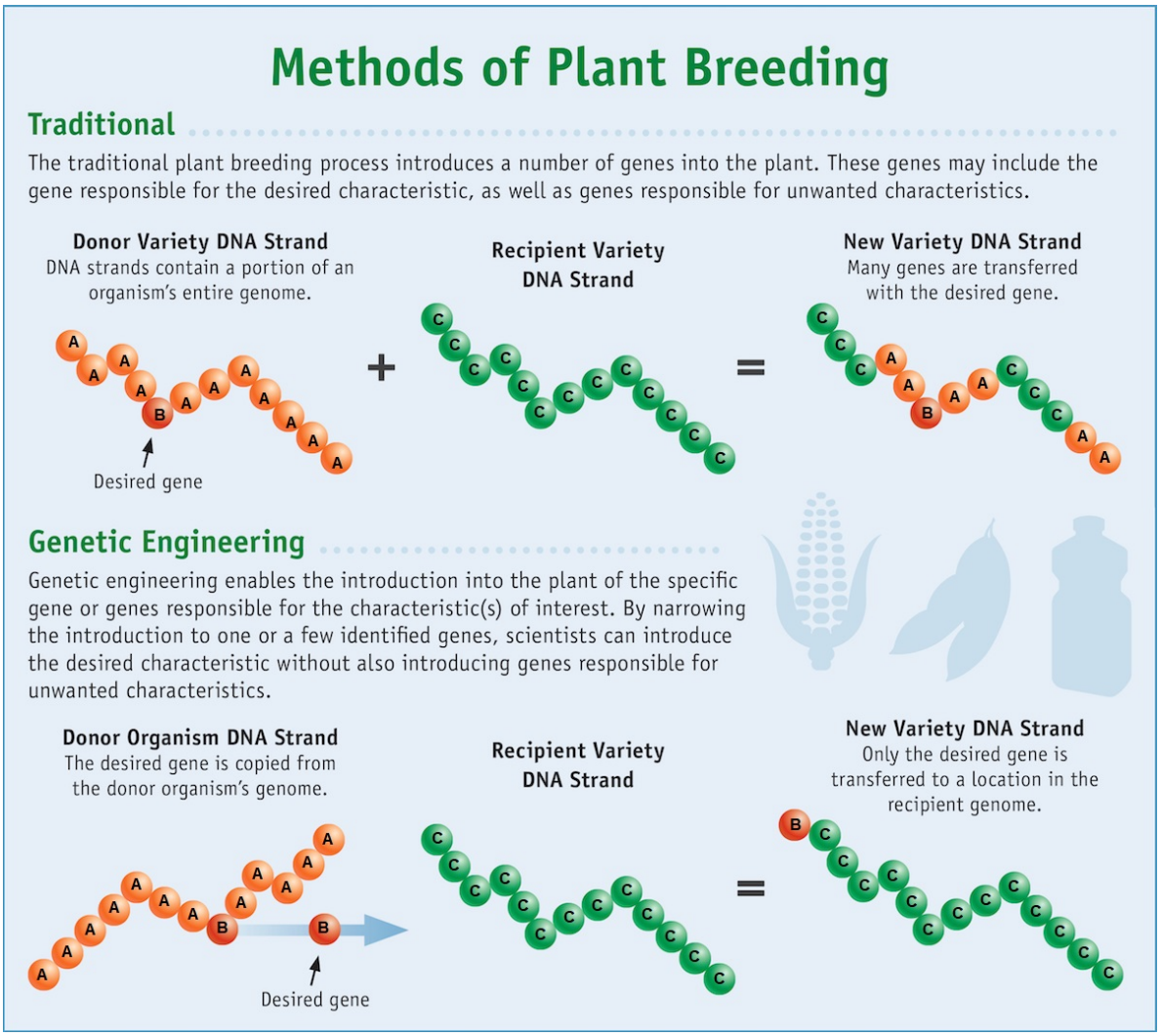



Figure **13.5.c**: Both selective (traditional) breeding and modern genetic engineering produce genetic modifications. Genetic engineering allows for fewer and more precise genetic modifications. The traditional plant breeding process introduces a number of genes into the plant. These genes may include the gene responsible for the desired characteristic, as well as genes responsible for unwanted characteristics. In traditional breeding, the donor variety DNA strand (A) and recipient variety DNA strand (C) are combined to produce the new variety DNA strand. The donor DNA strand contains a portion of an organism's entire genome, including the desired gene (B). In the new variety DNA strand, many genes are transferred with the desired gene. Genetic engineering enables the introduction into the plant of the specific gene or genes responsible for the characteristics of interest. By narrowing the introduction to one of a few identified genes, scientists can introduce the desired characteristic without also introducing genes responsible for unwanted characteristics. In genetic engineering, the desired gene (B) is copied from the donor organism's genome (A). The desired gene from the donor organism DNA strand and the recipient variety DNA strand (C) are combined to produce the new variety DNA strand. Only the desired gene is transferred to a location in the recipient genome. Modified from Michael J. Ermarth/FDA (public domain).

**Genetically modified organisms** (**GMOs**) are those that have had their DNA altered through genetic engineering. Genetically modified crops are sometimes called genetically engineered (GE) crops. **Transgenic organisms** are a type of genetically modified organism that contains genes from a different species. Because they contain unique combinations of genes and are not restricted to the laboratory, transgenic plants and other GMOs are closely monitored by government agencies to ensure that they are fit for human consumption and do not endanger other plant and animal life. Because these foreign genes (transgenes) can spread to other species in the environment, particularly in the pollen and seeds of plants, extensive testing is required to ensure ecological stability.

## How to Genetically Modify Plant Cells

DNA can be inserted into plant cells through various techniques. For example, a **gene gun** propels DNA bound to gold particles into plant cells. (DNA is negatively charge and clings to positively charged gold.) A more traditional approach employs the plant pathogen ***Agrobacterium tumefaciens*** (figure **13.5.*d***). Ordinarily, this bacterium causes crown gall disease in plants by inserting

a circular piece of DNA, called the **Ti plasmid**, into plant cells. This DNA incorporates into plant chromosomes, giving them genes to produce the gall (figure **13.5.*e***), which provides a home for the bacterial pathogen.

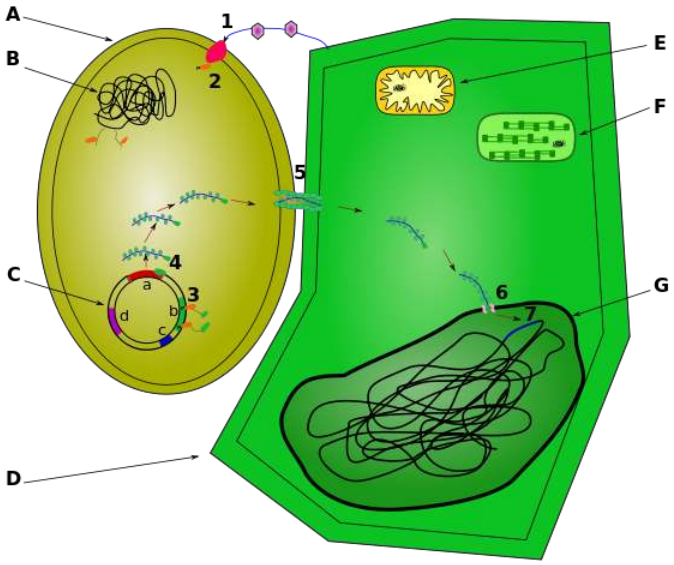



Figure **13.5.*d***: An *Agrobacterium tumefaciens* cell (A) inserts the Ti plasmid (C), which contains T-DNA (a) into a plant cell (D). The T-DNA eventually enters the plant nucleus (G), where the plant DNA is stored. Image by Chandres (CC-BY-SA).



Figure **13.5.*e***: A crown gall caused by *Agrobacterium tumefaciens*. This bacterium inserts DNA into plant cells, which causes them to produce the crown gall. This process can be adjusted by scientists to insert genes that code for desirable traits into crop species. Image by Scott Nelson (public domain).

Scientists alters the process by which *Agrobacterium* infects and genetically alter plant cells to produce genetically modified plants with agriculturally beneficial traits as follows (figure **13.5.*f***):

1. T-DNA, which codes for the crown gall is removed from the Ti plasmid, and genes for desired traits are added.
2. The modified plasmid is then added back to *Agrobacterium.*
3. *Agrobacterium* infects undifferentiated plant cells (stem cells that can develop into any part of the plant; figure **13.5.*g***).

4. The modified plant cells are given hormones to produce the entire plant.

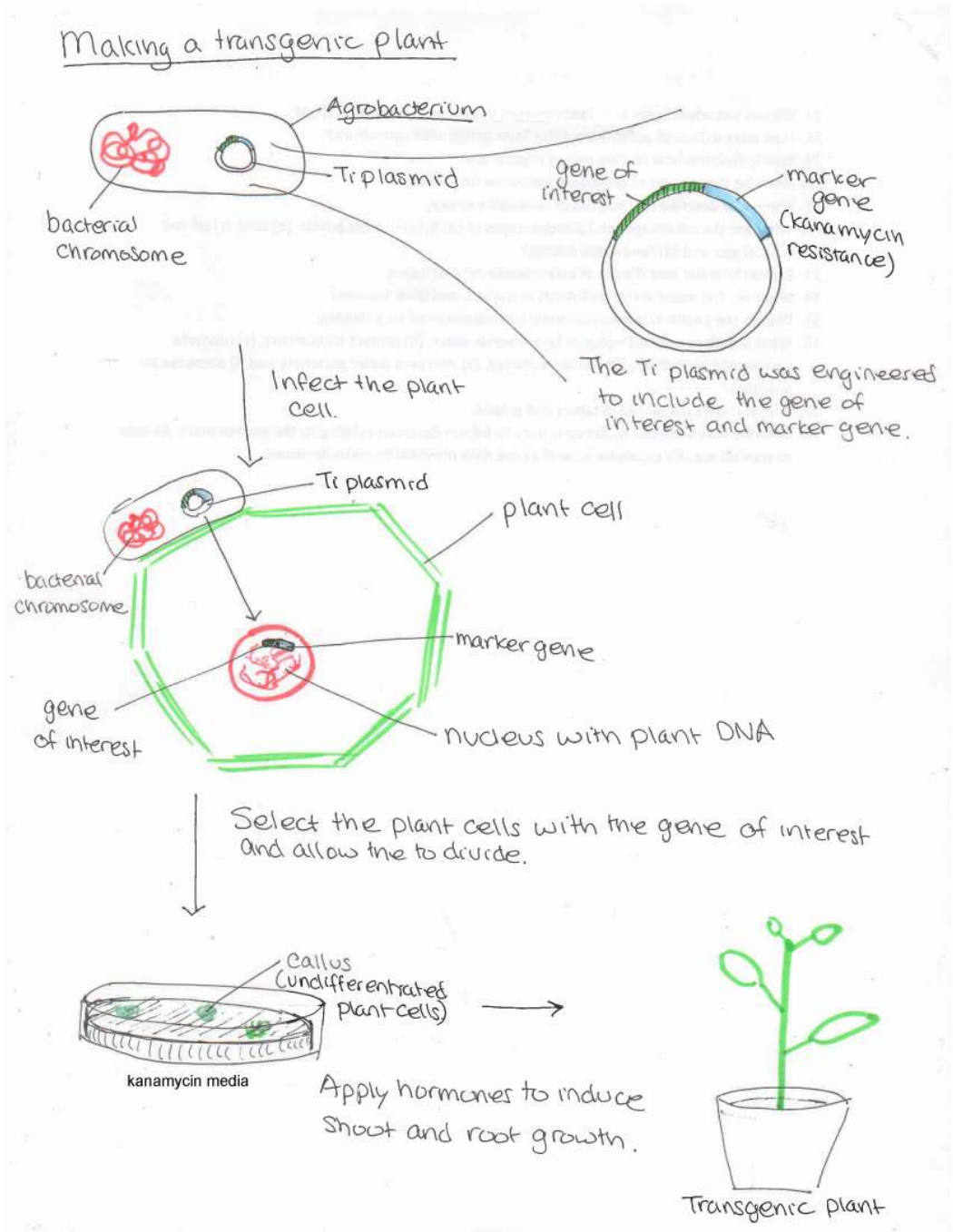



Figure **13.5.*f***: The process of producing a transgenic plant using *Agrobacterium*. The Ti plasmid was engineered to include the gene of interest and marker gene, which will later indicate which cells are transgenic. A rectangle represents *Agrobacterium*. It contains regular bacterial DNA and the Ti plasmid, represented by a circle. The T-DNA (which provides instructors for producing a crown gall) is removed, and a transgene (gene of interest) and maker gene are inserted instead. The gene of interest is colored green, and the marker gene (for resistance to the antibiotic kanamycin) is colored blue. Next, plant cells are infected with *Agrobacterium*. The bacterial cell inserts the Ti plasmid into a polyhedral plant cell. The transgene is transferred to the plant cell nucleus, which contains regular plant DNA. Next, the plant cells with the gene of interest are selected and allowed them to divide. This is done by culturing the plant cells on kanamycin media in a petri dish. Calluses, clusters of undifferentiated plant cells, are produced. Finally, hormones are applied to induce shoot and root growth. An arrow points from the petri dish to a potted transgenic plant. Image by Melissa Ha (CC-BY).

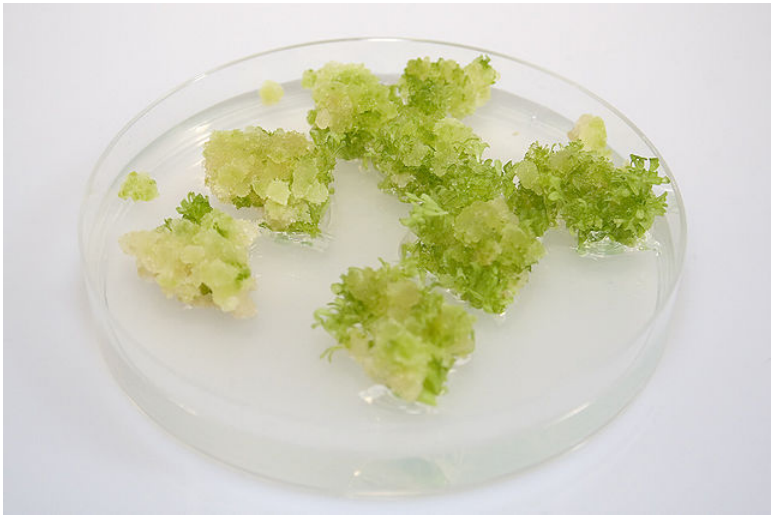


Figure **13.5.*g***: Individual plant cells are first genetically engineered with *Agrobacterium*. They then grow into **calluses** (blobs of undifferentiated plant cells) and are given hormones to induce root and shoot development. Because all the cells in the plant that eventually grows descended from a single genetically engineered cell, the entire plant is transgenic. Image by Igge (CC-BY-SA).

## Examples of Genetically Modified Crops

Many genetically modified crops have been approved in the U.S. and produce our foods. The first genetically modified organism approved by the U.S. Food and Drug Administration (FDA) in 1994 was **Flavr Savr™ tomatoes**, which have a longer shelf life (delayed rotting) because a gene responsible for breaking down cells in inhibited. Flavr Savr tomatoes are genetically modified (because their DNA has been altered) but not trasgenic (because they do not contain genes from another species). The Flavr Savr tomato did not successfully stay in the market because of problems maintaining and shipping the crop. **Golden rice** produces β-carotene, a precursor to vitamin A (figure **13.5.*h***; β-carotene is also in high concentrations in carrots, sweet potatoes, and cantaloupe, giving them their orange color.) **Roundup Ready®** corn, cotton, and soybeans are resistant to this common herbicide, making it easier to uniformly spray it in a field to kill the weeds without harming the crops (figure **13.5.*i***).



Figure **13.5.*h***: Golden rice has an orange-yellow color because it contains up to 35 µg β-carotene (a precursor to vitamin A) per gram of rice, which could prevent millions of cases of blindness worldwide. Image by International Rice Research Institute (CC-BY).




Figure **13.5.*i***: Roundup is a common herbicide. Roundup Ready® plants are genetically modified to resist roundup, meaning the herbicide does not kill them. This allows farmers to uniformly spray Roundup, killing weeds without harming the crops. Image by Mike Mozart (CC-BY).

Crops have also been engineered to produce insecticides. *Bacillus thuringiensis* (Bt) is a bacterium that produces protein crystals that are toxic to many insect species that feed on plants. Insects that have eaten Bt toxin stop feeding on the plants within a few hours. After the toxin is activated in the intestines of the insects, death occurs within a couple of days. The gene to produce **Bt toxin** has been added to many crops including corn (figure **13.5.*j***), potatoes, and cotton, providing plants with defense against insects.

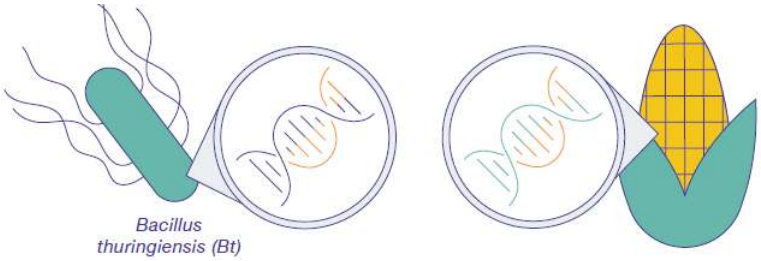



Figure **13.5.*j***: Genetic engineering to produce Bt corn. The gene from *Bacillus thuringiensis* that produces Bt toxin is removed and inserted into the corn plant. The Bt corn now produces the insecticide (Bt toxin) and can defend itself against pests. Image by FDA (public domain).

Genetically modified foods are widespread in the United States. For example, 94% of soy crops were genetically modified for herbicide resistance in 2020. Likewise, 8% of cotton and 10% of corn crops were modified for herbicide resistance in addition to the 83% of cotton and 79% of corn crops that were genetically modified in multiple ways.

Genetically modified animals have recently entered the market as well. **AquaAdvantage® salmon** are modified to grow more rapidly and were approved in November of 2015. However, as of March 2021, they have still not been sold due to legal challenges. In 2020, the FDA approved **GalSafe™ pigs** for medicine and food production. These pigs lack a molecule on the outside of their cells that cause allergies in some people.

## Advantages of Genetically Modified Crops

Advances in biotechnology may provide consumers with foods that are nutritionally-enriched, longer-lasting, or that contain lower levels of certain naturally occurring toxins present in some food plants. For example, researchers are using biotechnology to try to reduce saturated fats in cooking oils and reduce allergens in foods. Whether these benefits will reach the people who need them most remains to be seen. While cultivating golden rice could address vitamin A deficiency in millions of people, it has not

historically been accessible to these people because it is patented and expensive. Similarly, genetically modified seeds could increase the income of impoverished farmers if they were available at low or no cost, but this is not always the case.

Rainbow and SunUp papayas are a success story of how genetically modified crops can benefit small farmers and the economy in general. In the early 1990s, an emerging disease was destroying Hawaii's production of papaya and threatening to decimate the $11-million industry (figure **13.5.*k***). Fortunately, a man named Dennis Gonsalves (figure **13.5.*l***), who was raised on a sugar plantation and then became a plant physiologist at Cornell University, would develop papaya plants genetically engineered to resist the deadly virus. By the end of the decade, the Hawaiian papaya industry and the livelihoods of many farmers were saved thanks to the free distribution of Dr. Gonsalves's seeds.

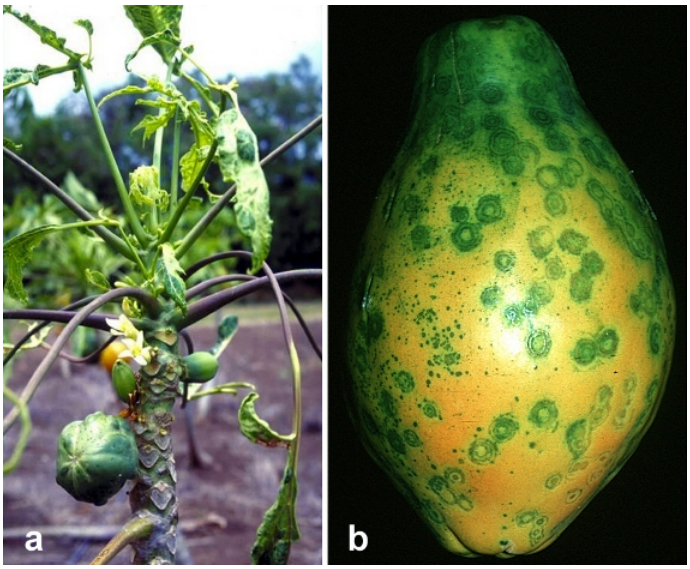



Figure **13.5.*k***: The symptoms of papaya ringspot virus are shown on the tree (a) and fruit (b). Image by APS (public domain).



Figure **13.5.*l***: Dennis Gonsalves genetically engineered papayas to resist the ringspot virus. Image by ARS USDA (public domain).

The effect of genetically modified crops on the environment depends on the specific genetic modification and which agricultural practices it promotes. For example, Bt crops produce their own insecticides such that external application of these chemicals is unnecessary, reducing the negative impacts of industrial agriculture. Ongoing research is exploring whether crops can be engineered to fix nitrogen in the atmosphere (as some bacteria do) rather than relying on ammonium, nitrites, and nitrates in the soil. If these crops were successfully engineered, they could reduce synthetic fertilizer application and minimize nutrient runoff that leads to eutrophication.

Genetically modified crops may have the potential to conserve natural resources, enable animals to more effectively use nutrients present in feed, and help meet the increasing world food and land demands. In practice, however, countries that use genetically modified crops compared to those that do not only enjoy a slight (or nonexistent) increase in yield.

## Disadvantages of Genetically Modified Crops

### Social Concerns

Intellectual property rights are one of the important factors in the current debate on genetically modified crops. Genetically modified crops can be patented by agribusinesses, which can lead to them controlling and potentially exploiting agricultural markets. Some accuse companies, such as Monsanto, of allegedly controlling seed production and pricing, much to the detriment of farmers (figure **13.5.*m***).




Figure **13.5.*m***: Protesters in Washington D.C. oppose genetically modified organisms and Monsanto specifically. Image cropped from Sarah Stierch (CC-BY).

### Environmental Concerns

Genetically modified crops present several environmental concerns. Monoculture farming already reduces biodiversity, and cultivating genetically modified crops, for which individual plants are quite similar genetically, exacerbates this. The use of Roundup Ready® crops naturally encourages widespread herbicide use, which could unintentionally kill nearby native plants. This practice would also increase herbicide residues on produce. While Bt crops are beneficial in the sense that they do not require external insecticide application, but Bt toxin is spread in their pollen. An early study found that Bt corn pollen may be harmful to monarch caterpillars (figure **13.5.*n***), but only at concentrations that are seldom reached in nature. Follow-up studies found that most of Bt corn grown did not harm monarchs; however, the one strain of Bt corn did was consequently removed from the market.



Figure **13.5.*n***: An early study suggested that pollen containing Bt toxin can harm beneficial and native insects, like this monarch caterpillar. However, without Bt crops, farmers are more likely to spray insecticides, circulating more harmful chemicals than pollen from Bt crops does. Image by Judy Gallagher (CC-BY).

Through interbreeding, or hybridization, genetically modified crops might share their transgenes with wild relatives. This could affect the genetics of those wild relatives and have unforeseen consequences on their populations and could even have implications for the larger ecosystem. For example, if a gene engineered to confer herbicide resistance were to pass from a genetically modified crop to a wild relative, it might transform the wild species into a **superweed** – a species that could not be controlled by herbicide. Its rampant growth could then displace other wild species and the wildlife that depends on it, thus inflecting ecological harm.

Not only could escaped genes alter weedy species, but they could also enter populations of native species. This could make some native species better competitors than they were previously, disrupting ecosystem dynamics. (They could potentially outcompete other native species with which they would otherwise coexist.)

While there is evidence of genetic transfer between genetically modified crops and wild relatives, there is not yet evidence of ecological harm from that transfer. Clearly, continued monitoring, especially for newly-developed crops, is warranted.

The escape of genetically modified animals has potential to disrupt ecosystems as well. For example, if AquaAdvantage salmon were to escape into natural ecosystem, as farmed fish often do, they could outcompete native salmon, including endangered species. Their genetic modification, which facilitates rapid growth, could result in a competitive advantage.

### Health Concerns

In addition to environmental risks, some people are concerned about potential health risks of genetically modified crops because they feel that genetic modification alters the intrinsic properties, or essence, of an organism. As discussed above, however, it is known that both traditional breeding practices and modern genetic engineering produce permanent genetic changes. Furthermore, selective breeding actually has a larger and more unpredictable impact on a species's genetics because of its comparably crude nature.

To address these concerns (and others), the US National Academies of Sciences, Engineering, and Medicine (NASEM) published a comprehensive, 500-page report in 2016 that summarized the current scientific knowledge regarding genetically modified crops. The report, titled *Genetically Engineered Crops: Experiences and Prospects*, reviewed more than 900 research articles, in addition to public comments and expert testimony. NASEM's GE Crop Report found "no substantiated evidence of a difference in risks to human health between current commercially available genetically engineered (GE) crops and conventionally bred crops, nor did it find conclusive cause-and-effect evidence of environmental problems from the GE crops." Additionally, the UN's Food and Agriculture Organization has concluded that risks to human and animal health from the use of GMOs are negligible. The scientific consensus on genetically modified crops is quite clear: they are safe for human consumption.

The potential of genetically modified crops to be allergenic is one of the potential adverse health effects, and it should continue to be studied, especially because some scientific evidence indicates that animals fed genetically modified crops have been harmed. NASEM's GE Crop Report concluded that when developing new crops, it is the product that should be studied for potential health and environmental risks, not the process that achieved that product. What this means is, because both traditional breeding practices and modern genetic engineering produce new traits through genetic modification, they both present potential risks. Thus, for the safety of the environment and human health, both should be adequately studied.

## Are Genetically Modified Crops the Solution We Need?

Significant resources, both financial and intellectual, have been allocated to answering the question: are genetically modified crops safe for human consumption? After many hundreds of scientific studies, the answer is yes. But a significant question still remains: are they necessary? Certainly, such as in instances like Hawaii's papaya, which were threatened with eradication due to an aggressive disease, genetic engineering was a quick and effective solution that would have been extremely difficult, if not impossible, to solve using traditional breeding practices.

However, in many cases, the early promises of genetically engineered crops – that they would improve nutritional quality of foods, confer disease resistance, and provide unparalleled advances in crop yields – have largely failed to come to fruition. NASEM's GE Crop Report states that while genetically modified crops have resulted in the reduction of agricultural loss from pests, reduced pesticide use, and reduced rates of injury from insecticides for farm workers, they have not increased the rate at which crop yields are advancing when compared to non-GE crops. Additionally, while there are some notable exceptions like golden rice or virus-resistant papayas, very few genetically engineered crops have been produced to increase nutritional capacity or to prevent plant disease that can devastate a farmer's income and reduce food security. The vast majority of genetically modified crops are

developed for only two purposes: to introduce herbicide resistance or pest resistance. Genetically modified crops are concentrated in developed countries, and their availability in developing countries, where they are perhaps most needed, is limited (figure **13.5.*o***).

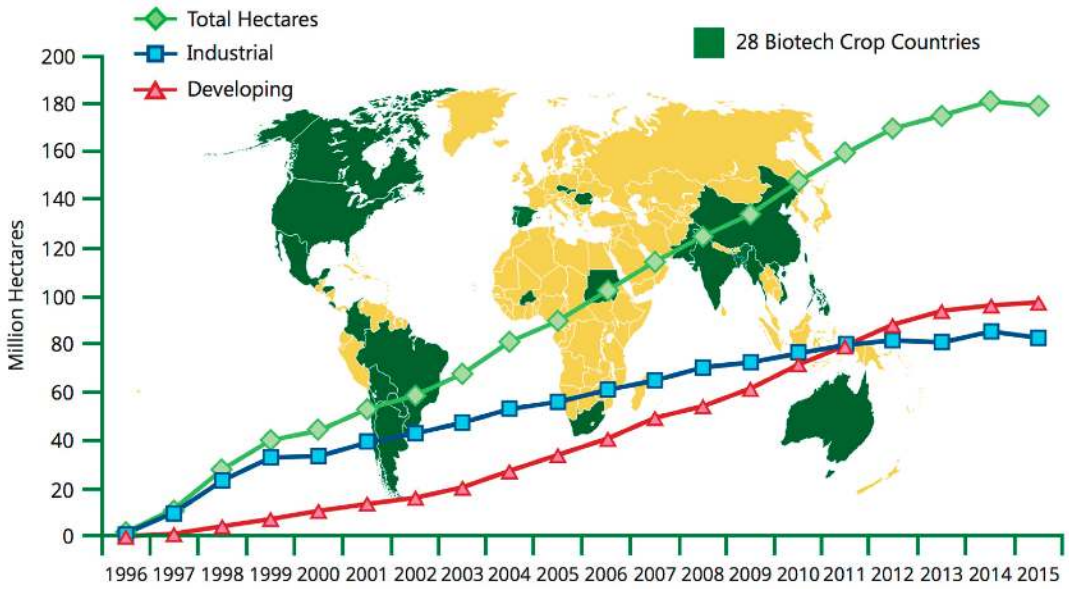



Figure **13.5.*o***: Global area of genetically modified crops in millions of hectares 1996–2015. Overall, the hectares of genetically modified crops has increased, and in 2011, developing countries surpassed industrial countries. By 2015, out of the 28 countries that grew biotech crops, 20 were developing, and only eight were developed countries. Latin American, African, and Asian farmers together grew 97.1 million hectares (54%) of the global 179.7 million biotech hectares, whereas industrial countries only planted 83 million hectares or 46%. Image and caption (modified) from Taheri, F., Azadi, H., & D'Haese, M. (2017). A World without Hunger: Organic or GM Crops? *Sustainability*, *9*(4), 580. doi:10.3390/su9040580. (CC-BY)

## Suggested Supplementary Reading

NASEM. 2016. Genetically Engineered Crops: Experiences and Prospects.

## Attributions

Modified by Melissa Ha from the following sources:

- Biotechnology and Genetic Engineering from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Biotechnology in Medicine and Agriculture from *Concepts in Biology* by OpenStax (licensed under CC-BY)

---

This page titled 13.5: Selective Breeding and Genetic Engineering is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **14.3: Selective Breeding and Genetic Engineering** by Melissa Ha and Rachel Schleiger is licensed CC BY-SA 4.0.

# 13.6: Pollination and Seed Dispersal

An estimated 90 percent of flowering plants depend on pollinators such as wasps, birds, bats, and bees, to reproduce. Plants and their pollinators are increasingly threatened around the world (Buchmann and Nabhan 1995; Kremen and Ricketts 2000). Pollination is critical to most major crops and virtually impossible to replace. For instance, imagine how costly fruit would be (and how little would be available) if its natural pollinators no longer existed and each developing flower had to be fertilized by hand.

Many animal species are important dispersers of plant seeds. It has been hypothesized that the loss of a seed disperser could cause a plant to become extinct. At present, there is no example where this has occurred. A famous example that has often been cited previously is the case of the dodo (Raphus cucullatus) and the tambalacoque (Sideroxylon grandiflorum). The dodo, a large flightless bird that inhabited the island of Mauritius in the Indian Ocean, became extinct due to overhunting in the late seventeenth century. It was once thought that the tambalacoque, a now endangered tree, depended upon the dodo to germinate its hard-cased seeds (Temple 1977). In the 1970s, only 13 trees remained and it was thought the tree had not reproduced for 300 years. The seeds of the tree have a very hard coat, as an experiment they were fed to a turkey; after passing through its gizzard the seeds were viable and germinated. This experiment led scientists to believe that the extinction of the dodo was coupled to the tambalacoque's inability to reproduce. However, this hypothesis has not stood up to further scrutiny, as there were several other species (including three now extinct species, a large-billed parrot, a giant tortoise, and a giant lizard) that were also capable of cracking the seed (Witmar and Cheke 1991; Catling 2001). Thus many factors, including the loss of the dodo, could have contributed to the decline of the tambalacoque. (For further details of causes of extinction see Historical Perspectives on Extinction and the Current Biodiversity Crisis). Unfortunately, declines and/or extinctions of species are often unobserved and thus it is difficult to tease out the cause of the end result, as multiple factors are often operating simultaneously. Similar problems exist today in understanding current population declines. For example, in a given species, population declines may be caused by loss of habitat, loss in prey species or loss of predators, a combination of these factors, or possibly some other yet unidentified cause, such as disease.

In the pine forests of western North America, corvids (including jays, magpies, and crows), squirrels, and bears play a role in seed dispersal. The Clark's nutcracker (Nucifraga columbiana) is particularly well adapted to dispersal of whitebark pine (Pinus albicaulis) seeds (Lanner 1996). The nutcracker removes the wingless seeds from the cones, which otherwise would not open on their own. Nutcrackers hide the seeds in clumps. When the uneaten seeds eventually grow, they are clustered, accounting for the typical distribution pattern of whitebark pine in the forest.

In tropical areas, large mammals and frugivorous birds play a key role in dispersing the seeds of trees and maintaining tree diversity over large areas. For example, three-wattled bellbirds (Procnias tricarunculata) are important dispersers of tree seeds of members of the Lauraceae family in Costa Rica. Because bellbirds return again and again to one or more favorite perches, they take the fruit and its seeds away from the parent tree, spreading Lauraceae trees throughout the forest (Wenny and Levy 1998).

13.6: Pollination and Seed Dispersal is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

- **1.5: Pollination and Seed Dispersal** by Nora Bynum is licensed CC BY 3.0. Original source: https://cnx.org/contents/AK5sUWpu1.1:GNNbYBSX1.

# 13.7: Sustainable Agriculture

**Sustainable agriculture** is a method of farming that does not deplete natural resources or degrade the environment and can therefore be continued indefinitely. Sustainable farms often rely on components of local ecosystems. For example, they might promote conditions for natural decomposition of wastes or employ natural enemies to control pests through predation or competition. More specifically, the 1977 and 1990 “Farm Bills” describe sustainable agriculture as "an integrated system of plant and animal production practices having a site-specific application that will, over the long term:

- satisfy human food and fiber needs;
- enhance environmental quality and the natural resource base upon which the agricultural economy depends;
- make the most efficient use of nonrenewable resources and on-farm resources and integrate, where appropriate, natural biological cycles and controls;
- sustain the economic viability of farm operations;
- enhance the quality of life for farmers and society as a whole."

Promoting biodiversity is key to sustainable agriculture, and high biodiversity results intact ecosystem services such as nutrient cycling and regulation of pest populations. This is consistent with the goal of sustainable agriculture: to mimic the processes found in natural ecosystems. In contrast to the monocultures of industrial agriculture, **polyculture** farming (growing different species together) is a common practice in sustainable agriculture. This assists with regulating pests and maintaining soil fertility (see below). **Seed banks**, locations at which many different types of seeds are stored, are key to conserving the genetic diversity of crops. These storage sites are cold enough to keep seeds frozen naturally (figure **13.7.*a***).



Figure **13.7.*a***: The Svalbard Global Seed Vault in Norway. Image by Miksu (CC-BY-SA).

## Integrated Pest Management

**Integrated Pest Management** (IPM) refers to a mix of farmer-driven, ecologically-based pest control practices that seek to reduce reliance on synthetic chemical pesticides. It uses several methods simultaneously to control pest populations. The steps of integrated pest management are to (1) identify true pests, (2) set thresholds and monitor (figure **13.7.*b***), and (3) develop an action plan. As many insects, microbes, and other organisms found in an agricultural area have a neutral or beneficial effect, it is not necessary to remove them. True pests are those that are causing economic harm, and they are managed (kept below economically damaging levels) rather than eradicated.



Figure **13.7.*b***: A retired Navy commander and landowner inspect an insect trap as part of an Integrated Pest Management (IPM) conservation practice. Image by NRCS (public domain).

An IPM Action Plan draws on four types of control: cultural, mechanical, biological, and chemical. These methods are listed in order of least to most environmental impact and are thus applied in this order. For example, cultural control is attempted first. If that is not effective, mechanical control is added to the plan, and so on. Chemical control is used as a last resort and is deemphasized in an IPM plan. When pesticides must be used, they are selected and applied in a way that minimizes adverse effects on beneficial organisms, humans, and the environment. It is commonly understood that applying an IPM approach does not necessarily mean eliminating pesticide use, although this is often the case because pesticides are often over-used for a variety of reasons.

### An Alternative to Spraying: Bollworm Control in Shandong

Farmers in Shandong (China) have been using innovative methods to control bollworm infestation in cotton when this insect became resistant to most pesticides. Among the control measures implemented were:

1. The use of pest resistant cultivars and interplanting of cotton with wheat or maize.
2. Use of lamps and poplar twigs to trap and kill adults to lessen the number of adults.
3. If pesticides were used, they were applied on parts of the cotton plant's stem rather than by spraying the whole field (to protect natural enemies of the bollworm).

These and some additional biological control tools have been effective in controlling insect populations and insect resistance, protecting surroundings and lowering costs.

## Cultural Control

**Cultural control** refers to minimizing the conditions that allow pests to thrive and spread. Examples include alternating which crops are planted each year (**crop rotation**), planting multiple types of crops near each other (**intercropping**; figure **13.7.*c***), selecting pest-resistant varieties, and planting pest-free rootstock (underground plant parts). Crop rotation prevents pests that are specialized to a particular type of crop from continuing year after year because their host plants are only available in certain years. Similarly, intercropping spatially limits the spread of pests. **Strip cropping** is a type of intercropping that involves growing different types of crops in alternating rows. Pests may infest one (or several) rows of their host species, but they would have to move past multiple rows of non-hosts to access additional host plants. Additionally, cultural control can involve optimizing irrigation and fertilizer application to promote plant defense and limit disease spread. Cultural control methods can be very effective and cost-efficient and present little to no risk to people or the environment.



Figure 13.7.*c*: *Momordica charantia* (bitter melon) intercropped with rice in the Philippines (cultural control). Bitter melon plants emerge from holes in plastic, which is used for mechanical control of weeds. Image by Judgefloro (CC-BY-SA).

## Mechanical Control

**Mechanical control** refers to physically removing pests or excluding them with barriers (figure 13.7.*d*). Mechanical controls that remove pests include sticky insect traps, mole traps, and removing weeds by hand. Netting to exclude birds, deer fencing, and weed cloth, plastic weed barriers (figure 13.7.*c*) or mulch are additional examples of mechanical control.



Figure 13.7.*d*: This solar insect trap is an example of mechanical control. It emits ultraviolet light to attract insects and is recharged using small solar panels. Image by MGK Solar Trap (CC-BY-SA).

## Biological Control

**Biological control** is the use of organisms to reduce pest populations (also see Invasive Species). One biological control strategy involves releasing natural enemies, such as predators, parasites, or parasitoids of the pest organisms (parasitoids are similar to parasites, but they consistently kill their hosts). Home gardeners can also rely on natural enemies by purchasing preying mantises, ladybugs (ladybird beetles), or lacewings for release (figure 13.7.*e*). Successful examples include ladybird beetles to depress aphid populations, parasitoid wasps to control whiteflies, and fungi, such as *Trichoderma*, to suppress fungal-caused plant diseases.



Figure **13.7.*e***: Lacewings are natural predators of garden and crop pests. Image by artsehn/Pixabay (Pixabay license).

If not used carefully, chemical control can have unintended negative consequences for biological control. In 1887, the cottony cushion scale (native to Australia) was devastating the citrus groves of California. A U.S. entomologist went to Australia to find a natural enemy and came back with the **vedalia beetle**, a species of ladybird beetle. Released in California, the beetle quickly brought the scale under control, at least until 1946. In that year, the pest made a dramatic comeback. This coincided with the first use of the now-banned pesticide dichlorodiphenyltrichloroethane (DDT) in the groves. Not only did DDT kill the target pest insects, but it killed the vedalia beetle as well. Only by altering spray procedures and reintroducing the beetle was the scale insect again controlled.

Another strategy for biological control involves releasing sterile males, which compete with fertile males for mates, ultimately decreasing pest population size. This technique was first applied against the screwworm fly, a serious pest of cattle (figure **13.7.*f***). The female flies lay their eggs in sores or other open wounds on the animals. After hatching, the larvae eat the tissues of their host. As they do so, they expose a still larger area to egg laying, often finally killing the host.



Figure **13.7.*f***: Thought to be eradicated in the United States by 1966, the screwworm fly, a cattle pest, reemerged in deer at Key Deer National Refuge in Florida in 2016. Sterile males were again released, and the pest was eradicated from Florida in 2017. Image by Judy Gallagher (CC-BY).

Prior to its eradication from the southeastern United States, the screwworm was causing huge annual livestock losses. The sterile male technique involves releasing factory-reared and sterilized flies into the natural population. Sterilization is done by exposing the factory flies to just enough gamma radiation to make them sterile but not enough to reduce their general vigor.

Starting in early 1958, up to 50 million sterilized flies were released each week from aircraft flying over Florida and parts of the adjoining states. Each time a fertile female in the natural population mated with a sterile male, the female layed sterile eggs. Since the females mate only once, her reproductive career was at an end. By early 1959, the pest was totally eliminated east of the Mississippi River. The problem in southwestern states was more challenging because the fly winters in Mexico and could move across the border with each new season. Even so, by expanding the program to include Mexico as well, the screwworm fly was finally eliminated from Mexico in 1991.

The sterile male technique also successfully controlled the Mediterranean fruit fly ("medfly"), a destructive fruit fly of citrus, peaches, pears, and apples in California.

### Chemical Control

**Chemical control** refers to the use of pesticides. If chemical control is needed, IPM favors highly targeted chemicals, such as pheromones to disrupt pest mating. **Pheromones** are chemical signals released by animals to communicate with other members of their species. Humans and many insect species alike release pheromones that function in attracting mates. Releasing the pheremones of insect pests can confuse males seeking mates and ultimately hinder them from reproducing (figure **13.7.*g***). This "male confusion" has been successful against the pink bollworm that infests cotton and reduced the need for conventional chemical insecticides by 90%. Pheromones have also been successful against pests that attack tomatoes, grapes, and peaches. If targeted chemicals are not effective, IPM may employ conventional pesticides, ideally only applying them to the spots that they are needed and at the lowest effective concentration. Broadcast spraying of non-specific pesticides is a last resort.



Figure **13.7.*g***: This pheromone trap is used to control the Asiatic rice borer. The caterpillars of this moth species infest rice stems. Image by Mehdi (CC-BY-SA).

## Sustainable Practices to Maintain Soil Fertility

A variety of sustainable practices can maintain soil quality. Many of these strategies have additional benefits such as regulating pests, limiting climate change, and preventing water pollution. These methods enrich the soil with nutrients, ensure proper water holding capacity (ability of the soil to retain water), and limit soil-degrading processes such as erosion and compaction.

### Crop Rotation

As previously mentioned, **crop rotations** are planned sequences of different crops over time on the same field (figure **13.7.*h***). Rotating crops provides productivity benefits by improving soil nutrient levels and breaking crop pest cycles. Farmers may also choose to rotate crops in order to reduce their production risk through diversification or to manage scarce resources, such as labor, during planting and harvesting. This strategy reduces the pesticide costs by naturally breaking the cycle of weeds, insects and diseases. Also, grass and legumes in a rotation protect water quality by preventing excess nutrients or chemicals from entering water supplies.



Figure **13.7.*h***: Crop rotation can be used at a large agricultural scale (left) or in a smaller garden (right). In each situation, which crop species are planted in each location changes year to year. Left image by USDA (public domain) and right image by Sten Porse (CC-BY-SA).

## Intercropping

**Intercropping** means growing two or more crops in close proximity to each other during part or all of their life cycles to promote soil improvement, biodiversity, and pest management. Incorporating intercropping principles into an agricultural operation increases diversity and interaction between plants, arthropods, mammals, birds and microorganisms resulting in a more stable crop-ecosystem and a more efficient use of space, water, sunlight, and nutrients (figure **13.7.*i***). This collaborative type of crop management mimics nature and is subject to fewer pest outbreaks, improved nutrient cycling and crop nutrient uptake, and increased water infiltration and moisture retention. Soil quality, water quality and wildlife habitat all benefit.



Figure **13.7.*i***: Intercropping of beans and maize (corn). Beans have root nodules that house nitrogen-fixing bacteria, enriching the soil. Image by AnnaJB (CC-BY-SA).

A common example of strip cropping (a type of intercropping; see above) involves alternating a row crop such as corn with a ground-covering crop such as alfalfa. The ground-covering crop helps reduce water runoff and traps soil eroded from the row crop. If this ground-covering crop is a member of the legume family such as alfalfa or soybeans and is associated with nitrogen-fixing bacteria, then alternating the strips from one planting to the next can also help maintain topsoil fertility.

## Cover Crops

**Cover crops** are those that are planted in the off-season to avoid leaving the soil bare. They can prevent soil and wind erosion, improve soil's physical and biological properties, supply nutrients, suppress weeds, improve the availability of soil water, and break pest cycles along with various other benefits. Cover crops are often members of the legume family and help enrich the soil with usable nitrogen. A single species or mix of cover crop species may be planted (figure **13.7.*j***).



Figure **13.7.*j***: A roller (crimper) on a tractor being used to roll down a rye and hairy vetch cover crop in April in the United States. While hairy vetch can limit soil erosion and add nitrogen to the soil, it is also an invasive species that must be carefully managed. Image and caption (modified) from Mr. 1032 (CC-BY-SA).

## Agroforestry

**Agroforestry** is the process of planting rows of trees interspersed with a cash crop (figure **13.7.*k* – *l***). Besides helping to prevent wind and water erosion of the soil, the trees provide shade which helps promote soil moisture retention. Decaying tree litter also provides some nutrients for the interplanted crops. The trees themselves may provide a cash crop. For example, fruit or nut trees may be planted with a grain crop. You can learn more about agroforestry using this interactive site by the United States Department of Agriculture and Forest Service.



Figure 13.7.*k*: Through agroforestry, high-value non-timber crops (food, medicinal plants, woody florals, and crafts) can be cultivated under the protection of a forest canopy that has been managed to provide a favorable crop environment. Image and caption (modified) from USDA/UFS (public domain).

## Contour and Terrace Farming

**Contour farming** involves plowing and planting crop rows along the natural contours of gently sloping land (figure 13.7.*l*). The lines of crop rows perpendicular to the slope help to slow water runoff, inhibit the formation of channels of water, and limit soil erosion (and the resultant loss of nutrients). **Terracing** is a common technique used to control water erosion on more steeply sloped hills and mountains (figure 13.7.*l*). Broad, level terraces are constructed along the contours of the slopes, and these act as dams trapping water for crops, reducing runoff, and limiting erosion.



Figure 13.7.*l*: Contour farming (left) and terracing (right). In contour farming, rows of crops wave perpendicular to the slopes of the land. This example (left) also employs agroforestry. Terrace farming involves making step-like structures (terraces) perpendicular to the incline of a slope to reduce erosion. Left image by Pixture2016 (CC-BY-SA). Right image by USDA (public domain).

## Minimal Tillage and No-till Agriculture

In modern agricultural practices, heavy machinery is used to prepare the seedbed for planting, to control weeds, and to harvest the crop. The use of heavy equipment has many advantages in saving time and labor, but can cause compaction of soil and disruption of the natural soil organisms. The problem with soil compaction is that increased soil density limits root penetration depth and may inhibit proper plant growth. Another aspect of soil tillage (mixing the soil) is that it may lead to more rapid decomposition of organic matter due to greater soil aeration. Over large areas of farmland, this has the unintended consequence of releasing more carbon and nitrous oxides (greenhouse gases) into the atmosphere, thereby contributing to climate change.

One of the easiest ways to prevent these problems is to minimize the amount of **tillage**, or turning over of the soil. In **minimal tillage** (conservation tillage) or **no-till agriculture**, the land is disturbed as little as possible by leaving crop residue in the fields (figure 13.7.*m*). Special seed drills inject new seeds and fertilizer into the unplowed soil. Tillage of fields does help to break up clods that were previously compacted, so best practices may vary at sites with different soil textures and composition. With proper planning, minimal tillage and no-till agriculture can simultaneously limit soil erosion and compaction, protect soil organisms, reduce costs (if performed correctly), and promote water infiltration. Furthermore, carbon can actually become sequestered into the soil with these methods, thus mitigating climate change. Minimal or no-till agriculture have proved a major success in Latin America and are being used in South Asia and Africa. However, a drawback of this method is that the crop residue can serve as a good habitat for insect pests and plant diseases.



Figure **13.7.*m***: No-till agriculture is an important tool to prevent loss of soil moisture. Image by USDA (public domain).

### Windbreaks

Creating **windbreaks** by planting tall trees along the perimeter of farm fields can help control the effects of wind erosion of soil (figure **13.7.*n***). Windbreaks reduce wind speed at ground level, an important factor in wind erosion. They also help trap snow in the winter months, leaving soil less exposed. As a side benefit, windbreaks also provide a habitat for birds and animals. One drawback is that windbreaks can be costly to farmers because they reduce the amount of available cropland.



Figure **13.7.*n***: Windbreaks are rows of trees and shrubs that reduce wind speed. They improve crop yields, reduce soil erosion, improve water-efficiency, protect livestock, and conserve energy. Image and caption (modified) from USDA/UFS (public domain).

## Organic Agriculture

**Organic agriculture** is often incorporated into sustainable agriculture (figure **13.7.*o***). To be certified as organic, farms must avoid using synthetic pesticides, synthetic fertilizers, and genetically modified organisms (figure **13.7.*p***). Organic meat, poultry, eggs, and dairy products come from animals that are given no antibiotics or growth hormones. Pests may instead be controlled by natural enemies or naturally produced substances such as neem oil or diatomaceous earth. Some organic alternatives to synthetic pesticides have relatively low environmental impact (such as salt spray), but there are many naturally produced compounds that are still toxic to people or cause ecological harm when widely used. Interestingly, external application of Bt toxin is approved for organic farming, but use of genetically engineered Bt crops is not. The latter results in lower concentrations of Bt toxin in the environment because it is locally produced directly by the plants themselves. In summary, although many practices of organic agriculture benefit the environment and meet sustainability goals, some organic farms are not sustainable and some sustainable farms are not organic.




Figure **13.7.*o***: United States Department of Agriculture (USDA) Conservationist Shemekia Mosley (right) works with Nuri Icgoren (left) who operates Urban Sprout Farms, a biodynamic, certified organic urban farm Lakewood Heights, Georgia. Image and caption (modified) by Preston Keres/USDA (public domain).




Figure **13.7.*p***: The USDA certified organic label. Modified from image by nikoretro (CC-BY-SA).

## Consumer Choices that Support Sustainable Agriculture

Even if you're not a farmer or lawmaker, you have the power to promote sustainable agriculture as a consumer. As discussed in Food Chains and Food Webs, it generally requires more land and more energy to produce meat compared to plant-based foods due to the inefficient transfer of energy from one trophic level to the next. For this reason, plant-based diets tend to be more sustainable, but this depends on the types of foods consumed and how they were produced.

**Local food** is not only fresher, but it requires fewer food miles (figure **13.7.*q***). In some ways, **organic food** causes less environmental degradation, but note that organic does not necessarily mean sustainable (see above). Because it limits machinery, synthetic pesticides, and synthetic fertilizers, it has a smaller **carbon footprint** (meaning it minimizes contribution to climate change). The Clean 15 is a list of produce that is low in pesticide residues. If you can't afford all organic foods, these are the best items to purchase non-organic. Examples include avocados, sweet corn, and pineapple. The Dirty Dozen lists produce that has the most pesticide residues. If you can only afford a few organic items, these are best to purchase organic. Examples include strawberries, spinach, and nectarines. Finally, some foods, such as beef, have higher **carbon footprint** and **water footprint** than others.



Figure **13.7.*q***: Farmers' markets are one option for purchasing locally grown foods. Image by NatalieMaynor (CC-BY).

## The Future of the Sustainable Agriculture Concept

Many in the agricultural community have adopted the sense of urgency and direction pointed to by the sustainable agriculture concept (figure **13.7.*r***). Sustainability has become an integral component of many government, commercial, and non-profit agriculture research efforts, and it is beginning to be woven into agricultural policy. Increasing numbers of farmers and ranchers have embarked on their own paths to sustainability, incorporating integrated and innovative approaches into their own enterprises.

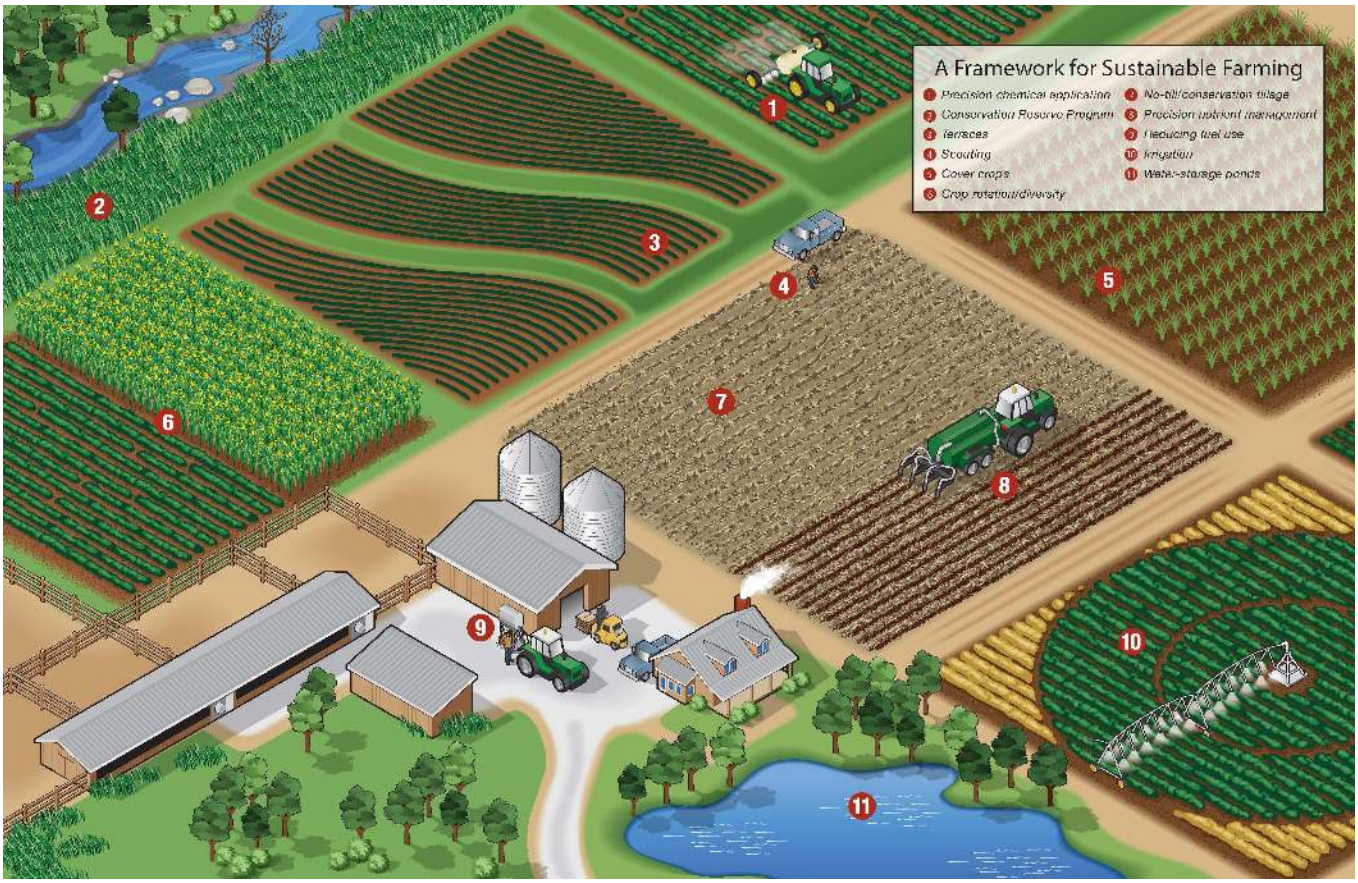



Figure 13.7.r: Sustainable agriculture combines many approaches. (1) If chemical are used, they are applied only where needed. (2) Through the Conservation Reserve Program, the most ecologically important patches are left intact. (3) Terraces on slopes reduce soil erosion. (4) Crops are scouted for pests as part of monitoring, a component of integrated pest management. (5) Cover crops prevent erosion and enrich the soil during off-seasons. (6) Different types of crops are included are planted in close proximity (intercropping) or in different years (crop rotation). (7) Land is either not tilled at all or tilled minimally (conservation tillage). (8) Through precision nutrient management, soil is sampled at different locations, and fertilizer is only applied where needed. (9) Heavy equipment is used only as needed and in such a way that conserves fuel. (10) Irrigation is precisely timed to conserve water. (11) Water storage ponds help recharge groundwater. Image by the United Soybean Board (CC-BY).

## Attributions

Modified by Melissa Ha from the following sources:

- Sustainable Agriculture from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Soils from *AP Environmental Science* by University of California College Prep, University of California (licensed under CC-BY). Download for free at CNX.
- Sustainability Definitions. USDA. Accessed 01-05-2021. (public domain)
- Biological Control from *Biology* by John W. Kimball (licensed under CC-BY)
- Integrated Pest Management (IPM) Principles. EPA. Accessed 01-05-2021 (public domain)
- Cover Crops and Soil Health. Natural Resources Conservation Service, USDA. Accessed 01-06-2021. (public domain)

This page titled 13.7: Sustainable Agriculture is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **14.4: Sustainable Agriculture** by Melissa Ha and Rachel Schleiger is licensed CC BY-SA 4.0.

# CHAPTER OVERVIEW

## 14: Population Ecology


14.1: Population Demography

14.2: Population Growth and Regulation

14.3: Human Population Growth


---


14: Population Ecology is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by LibreTexts.

# 14.1: Population Demography

Populations are dynamic entities. Populations consist all of the species living within a specific area, and populations fluctuate based on a number of factors: seasonal and yearly changes in the environment, natural disasters such as forest fires and volcanic eruptions, and competition for resources between and within species. The statistical study of population dynamics, demography, uses a series of mathematical tools to investigate how populations respond to changes in their biotic and abiotic environments. Many of these tools were originally designed to study human populations. For example, life tables, which detail the life expectancy of individuals within a population, were initially developed by life insurance companies to set insurance rates. In fact, while the term “demographics” is commonly used when discussing humans, all living populations can be studied using this approach.

## Population Size and Density

The study of any population usually begins by determining how many individuals of a particular species exist, and how closely associated they are with each other. Within a particular habitat, a population can be characterized by its population size (*N*), the total number of individuals, and its population density, the number of individuals within a specific area or volume. Population size and density are the two main characteristics used to describe and understand populations. For example, populations with more individuals may be more stable than smaller populations based on their genetic variability, and thus their potential to adapt to the environment. Alternatively, a member of a population with low population density (more spread out in the habitat), might have more difficulty finding a mate to reproduce compared to a population of higher density. As is shown in Figure **14.1.1**, smaller organisms tend to be more densely distributed than larger organisms.

Art Connection

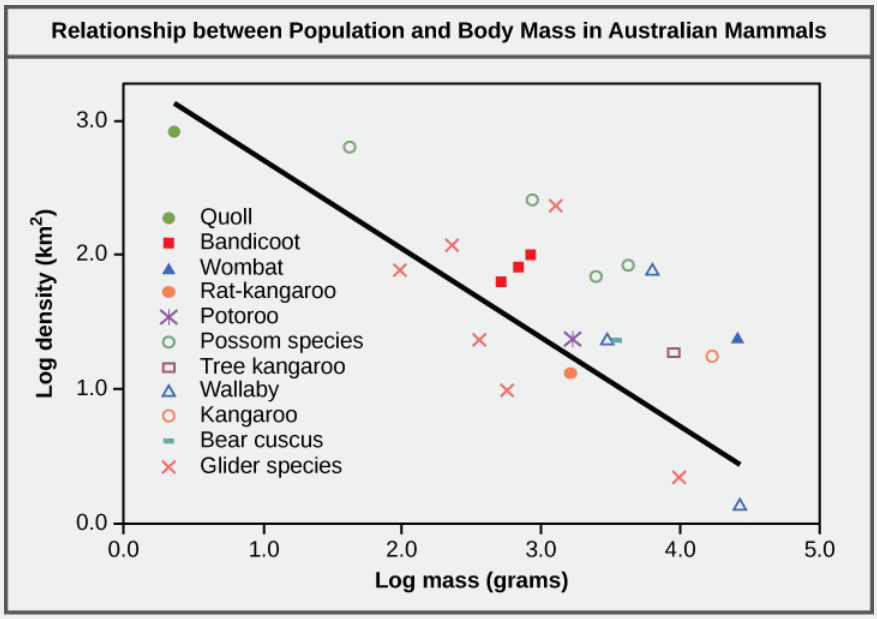



Figure **14.1.1**: Australian mammals show a typical inverse relationship between population density and body size.

As this graph shows, population density typically decreases with increasing body size. Why do you think this is the case?

## Population Research Methods

The most accurate way to determine population size is to simply count all of the individuals within the habitat. However, this method is often not logistically or economically feasible, especially when studying large habitats. Thus, scientists usually study populations by sampling a representative portion of each habitat and using this data to make inferences about the habitat as a whole. A variety of methods can be used to sample populations to determine their size and density. For immobile organisms such as plants, or for very small and slow-moving organisms, a quadrat may be used (Figure **14.1.2**). A quadrat is a way of marking off square areas within a habitat, either by staking out an area with sticks and string, or by the use of a wood, plastic, or metal square placed on the ground. After setting the quadrats, researchers then count the number of individuals that lie within their boundaries.

Multiple quadrat samples are performed throughout the habitat at several random locations. All of this data can then be used to estimate the population size and population density within the entire habitat. The number and size of quadrat samples depends on the type of organisms under study and other factors, including the density of the organism. For example, if sampling daffodils, a 1 $m^2$ quadrat might be used whereas with giant redwoods, which are larger and live much further apart from each other, a larger quadrat of 100 $m^2$ might be employed. This ensures that enough individuals of the species are counted to get an accurate sample that correlates with the habitat, including areas not sampled.

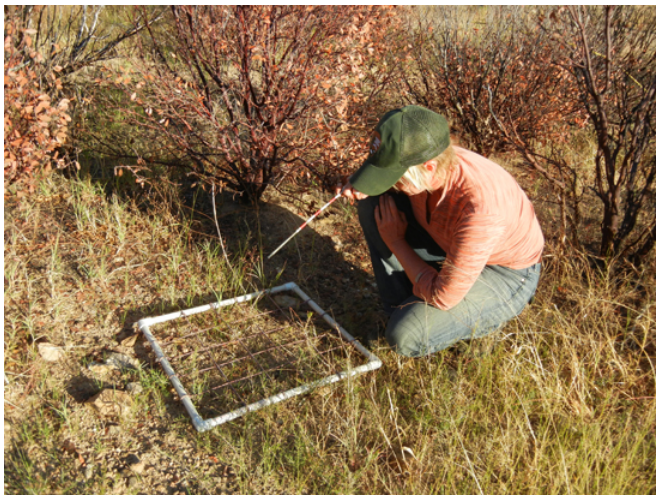
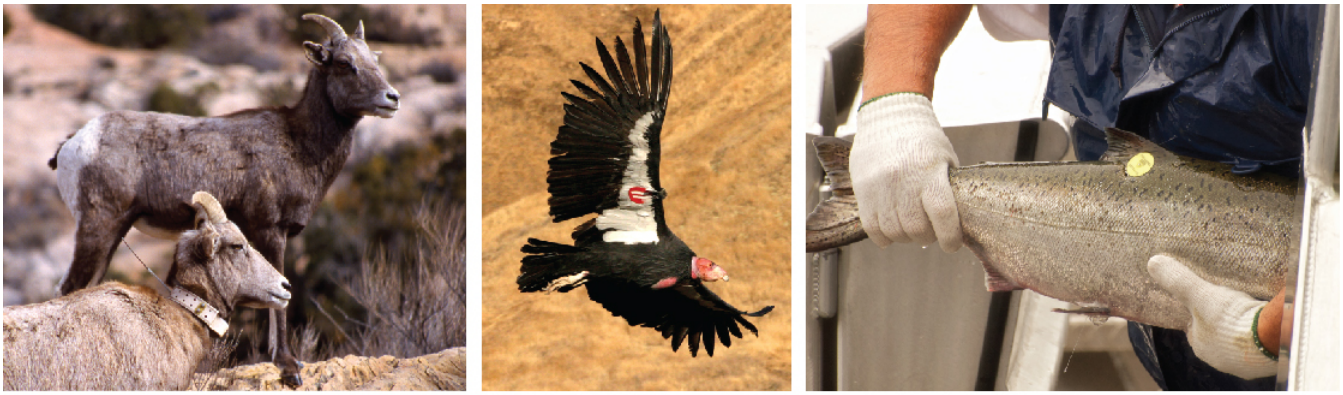


Figure **14.1.2**: A scientist uses a quadrat to measure population size and density. (credit: NPS Sonoran Desert Network)

For mobile organisms, such as mammals, birds, or fish, a technique called mark and recapture is often used. This method involves marking a sample of captured animals in some way (such as tags, bands, paint, or other body markings), and then releasing them back into the environment to allow them to mix with the rest of the population; later, a new sample is collected, including some individuals that are marked (recaptures) and some individuals that are unmarked (Figure **14.1.3**).



(a) (b) (c)

Figure **14.1.3**: Mark and recapture is used to measure the population size of mobile animals such as (a) bighorn sheep, (b) the California condor, and (c) salmon. (credit a: modification of work by Neal Herbert, NPS; credit b: modification of work by Pacific Southwest Region USFWS; credit c: modification of work by Ingrid Taylar)

Using the ratio of marked and unmarked individuals, scientists determine how many individuals are in the sample. From this, calculations are used to estimate the total population size. This method assumes that the larger the population, the lower the percentage of tagged organisms that will be recaptured since they will have mixed with more untagged individuals. For example, if 80 deer are captured, tagged, and released into the forest, and later 100 deer are captured and 20 of them are already marked, we can determine the population size ($N$) using the following equation:

$$\frac{\text{number marked first catch} * \text{total number of second catch}}{\text{number marked second catch}} = N$$

Using our example, the population size would be estimated at 400.

$$\frac{(80 * 100)}{20} = 400$$

Therefore, there are an estimated 400 total individuals in the original population.

There are some limitations to the mark and recapture method. Some animals from the first catch may learn to avoid capture in the second round, thus inflating population estimates. Alternatively, animals may preferentially be retrapped (especially if a food reward is offered), resulting in an underestimate of population size. Also, some species may be harmed by the marking technique, reducing their survival. A variety of other techniques have been developed, including the electronic tracking of animals tagged with radio transmitters and the use of data from commercial fishing and trapping operations to estimate the size and health of populations and communities.

## Species Distribution

In addition to measuring simple density, further information about a population can be obtained by looking at the distribution of the individuals. Species dispersion patterns (or distribution patterns) show the spatial relationship between members of a population within a habitat at a particular point in time. In other words, they show whether members of the species live close together or far apart, and what patterns are evident when they are spaced apart.

Individuals in a population can be more or less equally spaced apart, dispersed randomly with no predictable pattern, or clustered in groups. These are known as uniform, random, and clumped dispersion patterns, respectively (Figure **14.1.4**). Uniform dispersion is observed in plants that secrete substances inhibiting the growth of nearby individuals (such as the release of toxic chemicals by the sage plant *Salvia leucophylla*, a phenomenon called allelopathy) and in animals like the penguin that maintain a defined territory. An example of random dispersion occurs with dandelion and other plants that have wind-dispersed seeds that germinate wherever they happen to fall in a favorable environment. A clumped dispersion may be seen in plants that drop their seeds straight to the ground, such as oak trees, or animals that live in groups (schools of fish or herds of elephants). Clumped dispersions may also be a function of habitat heterogeneity. Thus, the dispersion of the individuals within a population provides more information about how they interact with each other than does a simple density measurement. Just as lower density species might have more difficulty finding a mate, solitary species with a random distribution might have a similar difficulty when compared to social species clumped together in groups.

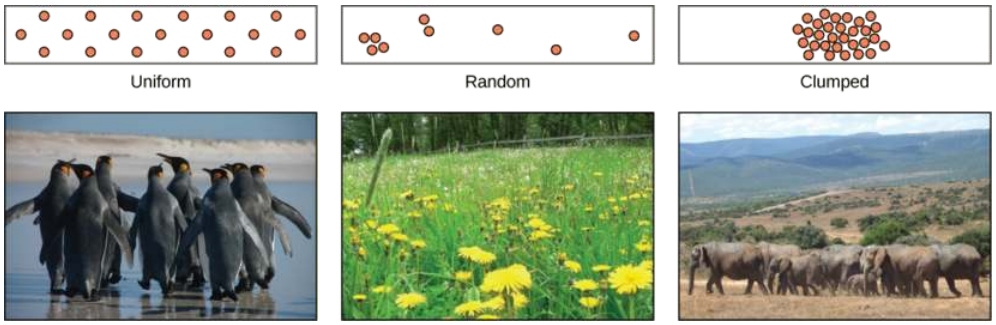



Figure **14.1.4**: Species may have uniform, random, or clumped distribution. Territorial birds such as penguins tend to have uniform distribution. Plants such as dandelions with wind-dispersed seeds tend to be randomly distributed. Animals such as elephants that travel in groups exhibit clumped distribution. (credit a: modification of work by Ben Tubby; credit b: modification of work by Rosendahl; credit c: modification of work by Rebecca Wood)

## Demography

While population size and density describe a population at one particular point in time, scientists must use demography to study the dynamics of a population. Demography is the statistical study of population changes over time: birth rates, death rates, and life expectancies. Each of these measures, especially birth rates, may be affected by the population characteristics described above. For example, a large population size results in a higher birth rate because more potentially reproductive individuals are present. In contrast, a large population size can also result in a higher death rate because of competition, disease, and the accumulation of waste. Similarly, a higher population density or a clumped dispersion pattern results in more potential reproductive encounters between individuals, which can increase birth rate. Lastly, a female-biased sex ratio (the ratio of males to females) or age structure

(the proportion of population members at specific age ranges) composed of many individuals of reproductive age can increase birth rates.

In addition, the demographic characteristics of a population can influence how the population grows or declines over time. If birth and death rates are equal, the population remains stable. However, the population size will increase if birth rates exceed death rates; the population will decrease if birth rates are less than death rates. Life expectancy is another important factor; the length of time individuals remain in the population impacts local resources, reproduction, and the overall health of the population. These demographic characteristics are often displayed in the form of a life table.

## Life Tables

Life tables provide important information about the life history of an organism. Life tables divide the population into age groups and often sexes, and show how long a member of that group is likely to live. They are modeled after actuarial tables used by the insurance industry for estimating human life expectancy. Life tables may include the probability of individuals dying before their next birthday (i.e., their mortality rate), the percentage of surviving individuals dying at a particular age interval, and their life expectancy at each interval. An example of a life table is shown in Table **14.1.1** from a study of Dall mountain sheep, a species native to northwestern North America. Notice that the population is divided into age intervals (column A). The mortality rate (per 1000), shown in column D, is based on the number of individuals dying during the age interval (column B) divided by the number of individuals surviving at the beginning of the interval (Column C), multiplied by 1000.

$$\text{mortality rate} = \frac{\text{number of individuals dying}}{\text{number of individuals surviving}} * 1000$$

For example, between ages three and four, 12 individuals die out of the 776 that were remaining from the original 1000 sheep. This number is then multiplied by 1000 to get the mortality rate per thousand.

$$\text{mortality rate} = \frac{12}{776} * 1000 \approx 15.5$$

As can be seen from the mortality rate data (column D), a high death rate occurred when the sheep were between 6 and 12 months old, and then increased even more from 8 to 12 years old, after which there were few survivors. The data indicate that if a sheep in this population were to survive to age one, it could be expected to live another 7.7 years on average, as shown by the life expectancy numbers in column E.

**Table 14.1.1:** Life Table of Dall Mountain Sheep[1]

| Age interval (years) | Number dying in age interval out of 1000 born | Number surviving at beginning of age interval out of 1000 born | Mortality rate per 1000 alive at beginning of age interval | Life expectancy or mean lifetime remaining to those attaining age interval |
|---|---|---|---|---|
| 0-0.5 | 54 | 1000 | 54.0 | 7.06 |
| 0.5-1 | 145 | 946 | 153.3 | -- |
| 1-2 | 12 | 801 | 15.0 | 7.7 |
| 2-3 | 13 | 789 | 16.5 | 6.8 |
| 3-4 | 12 | 776 | 15.5 | 5.9 |
| 4-5 | 30 | 764 | 39.3 | 5.0 |
| 5-6 | 46 | 734 | 62.7 | 4.2 |
| 6-7 | 48 | 688 | 69.8 | 3.4 |
| 7-8 | 69 | 640 | 107.8 | 2.6 |
| 8-9 | 132 | 571 | 231.2 | 1.9 |
| 9-10 | 187 | 439 | 426.0 | 1.3 |
| 10-11 | 156 | 252 | 619.0 | 0.9 |
| 11-12 | 90 | 96 | 937.5 | 0.6 |
| 12-13 | 3 | 6 | 500.0 | 1.2 |
| 13-14 | 3 | 3 | 1000 | 0.7 |

## Survivorship Curves

Another tool used by population ecologists is a survivorship curve, which is a graph of the number of individuals surviving at each age interval plotted versus time (usually with data compiled from a life table). These curves allow us to compare the life histories of different populations (Figure **14.1.5**). Humans and most primates exhibit a Type I survivorship curve because a high percentage of offspring survive their early and middle years—death occurs predominantly in older individuals. These types of species usually have small numbers of offspring at one time, and they give a high amount of parental care to them to ensure their survival. Birds are an example of an intermediate or Type II survivorship curve because birds die more or less equally at each age interval. These organisms also may have relatively few offspring and provide significant parental care. Trees, marine invertebrates, and most fishes exhibit a Type III survivorship curve because very few of these organisms survive their younger years; however, those that make it to an old age are more likely to survive for a relatively long period of time. Organisms in this category usually have a very large number of offspring, but once they are born, little parental care is provided. Thus these offspring are “on their own” and vulnerable to predation, but their sheer numbers assure the survival of enough individuals to perpetuate the species.

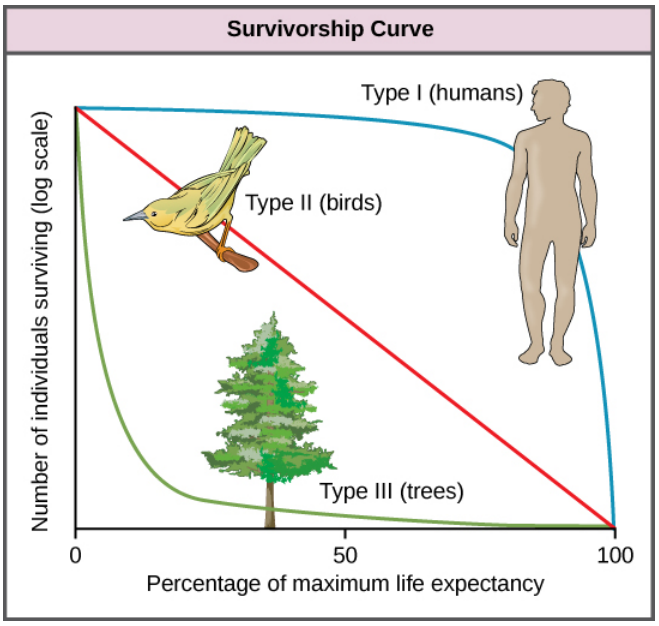



Figure **14.1.5**: Survivorship curves show the distribution of individuals in a population according to age. Humans and most mammals have a Type I survivorship curve because death primarily occurs in the older years. Birds have a Type II survivorship curve, as death at any age is equally probable. Trees have a Type III survivorship curve because very few survive the younger years, but after a certain age, individuals are much more likely to survive.

## Art Connections

Figure **14.1.1**: As this graph shows, population density typically decreases with increasing body size. Why do you think this is the case?

**Answer**

Smaller animals require less food and other resources, so the environment can support more of them.

## Life History Patterns and Energy Budgets

Energy is required by all living organisms for their growth, maintenance, and reproduction; at the same time, energy is often a major limiting factor in determining an organism's survival. Plants, for example, acquire energy from the sun via photosynthesis, but must expend this energy to grow, maintain health, and produce energy-rich seeds to produce the next generation. Animals have the additional burden of using some of their energy reserves to acquire food. Furthermore, some animals must expend energy caring for their offspring. Thus, all species have an energy budget: they must balance energy intake with their use of energy for metabolism, reproduction, parental care, and energy storage (such as bears building up body fat for winter hibernation).

### Parental Care and Fecundity

Fecundity is the potential reproductive capacity of an individual within a population. In other words, fecundity describes how many offspring could ideally be produced if an individual has as many offspring as possible, repeating the reproductive cycle as soon as possible after the birth of the offspring. In animals, fecundity is inversely related to the amount of parental care given to an individual offspring. Species, such as many marine invertebrates, that produce many offspring usually provide little if any care for the offspring (they would not have the energy or the ability to do so anyway). Most of their energy budget is used to produce many tiny offspring. Animals with this strategy are often self-sufficient at a very early age. This is because of the energy tradeoff these organisms have made to maximize their evolutionary fitness. Because their energy is used for producing offspring instead of parental care, it makes sense that these offspring have some ability to be able to move within their environment and find food and perhaps shelter. Even with these abilities, their small size makes them extremely vulnerable to predation, so the production of many offspring allows enough of them to survive to maintain the species.

Animal species that have few offspring during a reproductive event usually give extensive parental care, devoting much of their energy budget to these activities, sometimes at the expense of their own health. This is the case with many mammals, such as humans, kangaroos, and pandas. The offspring of these species are relatively helpless at birth and need to develop before they achieve self-sufficiency.

Plants with low fecundity produce few energy-rich seeds (such as coconuts and chestnuts) with each having a good chance to germinate into a new organism; plants with high fecundity usually have many small, energy-poor seeds (like orchids) that have a relatively poor chance of surviving. Although it may seem that coconuts and chestnuts have a better chance of surviving, the energy tradeoff of the orchid is also very effective. It is a matter of where the energy is used, for large numbers of seeds or for fewer seeds with more energy.

### Early versus Late Reproduction

The timing of reproduction in a life history also affects species survival. Organisms that reproduce at an early age have a greater chance of producing offspring, but this is usually at the expense of their growth and the maintenance of their health. Conversely, organisms that start reproducing later in life often have greater fecundity or are better able to provide parental care, but they risk that they will not survive to reproductive age. Examples of this can be seen in fishes. Small fish like guppies use their energy to reproduce rapidly, but never attain the size that would give them defense against some predators. Larger fish, like the bluegill or shark, use their energy to attain a large size, but do so with the risk that they will die before they can reproduce or at least reproduce to their maximum. These different energy strategies and tradeoffs are key to understanding the evolution of each species as it maximizes its fitness and fills its niche. In terms of energy budgeting, some species “blow it all” and use up most of their energy reserves to reproduce early before they die. Other species delay having reproduction to become stronger, more experienced individuals and to make sure that they are strong enough to provide parental care if necessary.

### Single versus Multiple Reproductive Events

Some life history traits, such as fecundity, timing of reproduction, and parental care, can be grouped together into general strategies that are used by multiple species. Semelparity occurs when a species reproduces only once during its lifetime and then dies. Such species use most of their resource budget during a single reproductive event, sacrificing their health to the point that they do not survive. Examples of semelparity are bamboo, which flowers once and then dies, and the Chinook salmon (Figure **14.1.5**a), which uses most of its energy reserves to migrate from the ocean to its freshwater nesting area, where it reproduces and then dies. Scientists have posited alternate explanations for the evolutionary advantage of the Chinook’s post-reproduction death: a programmed suicide caused by a massive release of corticosteroid hormones, presumably so the parents can become food for the offspring, or simple exhaustion caused by the energy demands of reproduction; these are still being debated.

Iteroparity describes species that reproduce repeatedly during their lives. Some animals are able to mate only once per year, but survive multiple mating seasons. The pronghorn antelope is an example of an animal that goes into a seasonal estrus cycle (“heat”): a hormonally induced physiological condition preparing the body for successful mating (Figure **14.1.5**b). Females of these species mate only during the estrus phase of the cycle. A different pattern is observed in primates, including humans and chimpanzees, which may attempt reproduction at any time during their reproductive years, even though their menstrual cycles make pregnancy likely only a few days per month during ovulation (Figure **14.1.5**c).



a b



c

**Figure 14.1.1:** The (a) Chinook salmon mates once and dies. The (b) pronghorn antelope mates during specific times of the year during its reproductive life. Primates, such as humans and (c) chimpanzees, may mate on any day, independent of ovulation. (credit a: modification of work by Roger Tabor, USFWS; credit b: modification of work by Mark Gocke, USDA; credit c: modification of work by "Shiny Things"/Flickr)

## Evolution Connection: Energy Budgets, Reproductive Costs, and Sexual Selection in *Drosophila*

Research into how animals allocate their energy resources for growth, maintenance, and reproduction has used a variety of experimental animal models. Some of this work has been done using the common fruit fly, *Drosophila melanogaster*. Studies have shown that not only does reproduction have a cost as far as how long male fruit flies live, but also fruit flies that have already mated several times have limited sperm remaining for reproduction. Fruit flies maximize their last chances at reproduction by selecting optimal mates.

In a 1981 study, male fruit flies were placed in enclosures with either virgin or inseminated females. The males that mated with virgin females had shorter life spans than those in contact with the same number of inseminated females with which they were unable to mate. This effect occurred regardless of how large (indicative of their age) the males were. Thus, males that did not mate lived longer, allowing them more opportunities to find mates in the future.

More recent studies, performed in 2006, show how males select the female with which they will mate and how this is affected by previous matings (Figure **14.1.2**).[1] Males were allowed to select between smaller and larger females. Findings showed that larger females had greater fecundity, producing twice as many offspring per mating as the smaller females did. Males that had previously mated, and thus had lower supplies of sperm, were termed "resource-depleted," while males that had not mated were termed "non-resource-depleted." The study showed that although non-resource-depleted males preferentially mated with larger females, this selection of partners was more pronounced in the resource-depleted males. Thus, males with depleted sperm supplies, which were limited in the number of times that they could mate before they replenished their sperm supply, selected larger, more fecund females, thus maximizing their chances for offspring. This study was one of the first to show that

the physiological state of the male affected its mating behavior in a way that clearly maximizes its use of limited reproductive resources.

| | Ratio large/small females mated |
|---|---|
| Non sperm-depleted | 8 ± 5 |
| Sperm-depleted | 15 ± 5 |

Figure **14.1.2**: Male fruit flies that had previously mated (sperm-depleted) picked larger, more fecund females more often than those that had not mated (non-sperm-depleted). This change in behavior causes an increase in the efficiency of a limited reproductive resource: sperm.

These studies demonstrate two ways in which the energy budget is a factor in reproduction. First, energy expended on mating may reduce an animal's lifespan, but by this time they have already reproduced, so in the context of natural selection this early death is not of much evolutionary importance. Second, when resources such as sperm (and the energy needed to replenish it) are low, an organism's behavior can change to give them the best chance of passing their genes on to the next generation. These changes in behavior, so important to evolution, are studied in a discipline known as behavioral biology, or ethology, at the interface between population biology and psychology.

## Footnotes

1. 1 Data Adapted from Edward S. Deevey, Jr., "Life Tables for Natural Populations of Animals," *The Quarterly Review of Biology* 22, no. 4 (December 1947): 283-314.
2. 1 Adapted from Phillip G. Byrne and William R. Rice, "Evidence for adaptive male mate choice in the fruit fly *Drosophila melanogaster*," Proc Biol Sci. 273, no. 1589 (2006): 917-922, doi: 10.1098/rspb.2005.3372.

## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- **Population Demography** and **Life Histories and Natural Selection** by OpenStax, is licensed CC BY by

  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).

- Population Dynamics and Regulation from *General Biology* by OpenStax (CC-BY)
- Population Demographics and Dynamics from *Environmental Biology* by Matthew R. Fisher (CC-BY)
- Principles of Population Growth from *Biology* by John W. Kimball (CC-BY)

---

This page titled 14.1: Population Demography is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **45.1: Population Demography** by OpenStax is licensed CC BY 4.0.
- **45.2: Life Histories and Natural Selection** by OpenStax is licensed CC BY 4.0.

# 14.2: Population Growth and Regulation

Although life histories describe the way many characteristics of a population (such as their age structure) change over time in a general way, population ecologists make use of a variety of methods to model population dynamics mathematically. These more precise models can then be used to accurately describe changes occurring in a population and better predict future changes. Certain models that have been accepted for decades are now being modified or even abandoned due to their lack of predictive ability, and scholars strive to create effective new models.

Population ecologists make use of a variety of methods to model population dynamics. An accurate model should be able to describe the changes occurring in a population and predict future changes. The two simplest models of population growth use deterministic equations (equations that do not account for random events) to describe the rate of change in the size of a population over time (Figure **14.2.1**). The first of these models, **exponential growth**, describes populations that increase in numbers without any limits to their growth. The second model, **logistic growth**, introduces limits to reproductive growth that become more intense as the population size increases. Neither model adequately describes natural populations, but they provide points of comparison.

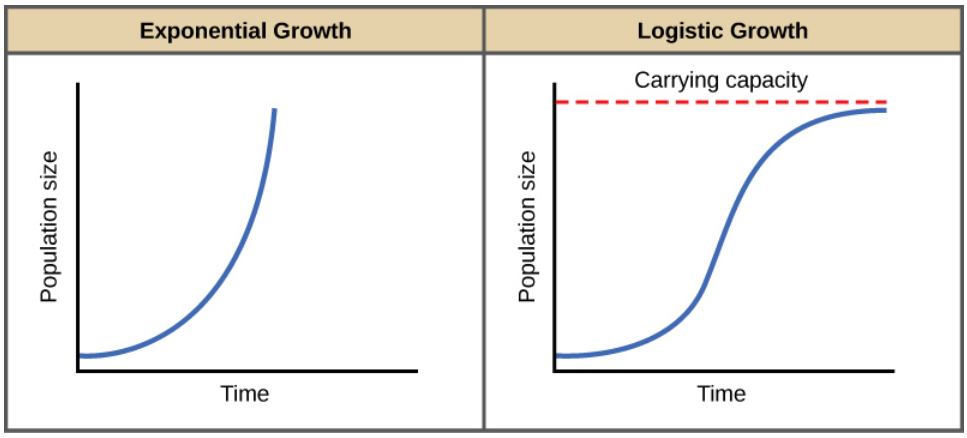



**Figure 14.2.1:** When resources are unlimited, populations exhibit exponential growth, resulting in a J-shaped curve. When resources are limited, populations exhibit logistic growth. In logistic growth, population expansion decreases as resources become scarce, and it levels off when the carrying capacity of the environment is reached, resulting in an S-shaped curve.

## The Population Growth Rate (*r* )

The **population growth rate** (sometimes called the rate of increase or per capita growth rate, *r*) equals the **birth rate** (*b*) minus the **death rate** (*d*) divided by the **initial population size** ($N_0$).

$$r = (b - d) \div N_0$$

Another method of calculating the population growth rate involves final and initial population size (figure **14.2.2**). In this case, **population growth rate** (*r*) equals the **final population size** (N) minus the **initial population size** ($N_0$) and divided by the **initial population size** ($N_0$).

$$r = (N - N_0) \div N_0$$



**Figure 14.2.2:** How could we calculate the growth rate of this caribou population per year if there were 200 individuals in 2016 and 300 individuals in 2018? Image by K. Joly/NPS (public domain).

## Exponential Growth

Charles Darwin, in his theory of natural selection, was greatly influenced by the English clergyman Thomas Malthus. Malthus published a book in 1798 stating that populations with unlimited natural resources grow very rapidly, and then population growth decreases as resources become depleted. This accelerating pattern of increasing population size is called exponential growth.

The best example of exponential growth is seen in bacteria. Bacteria are prokaryotes that reproduce quickly, about an hour for many species. If 1000 bacteria are placed in a large flask with an unlimited supply of nutrients (so the nutrients will not become depleted), after an hour, there is one round of division and each organism divides, resulting in 2000 organisms—an increase of 1000 (Figure **14.2.3**). In another hour, each of the 2000 organisms will double, producing 4000, an increase of 2000 organisms. After the third hour, there should be 8000 bacteria in the flask, an increase of 4000 organisms. The important concept of exponential growth is that the population growth rate—the number of organisms added in each reproductive generation—is accelerating; that is, it is increasing at a greater and greater rate. After 1 day and 24 of these cycles, the population would have increased from 1000 to more than 16 billion. When the population size, *N*, is plotted over time, a J-shaped growth curve is produced (Figure **14.2.1**).



**Figure 14.2.3:** Bacteria growing with abundant nutrients in a flask will exhibit exponential growth. Image by Soledad Mirand-Rottman (CC-BY-SA).

The bacteria-in-a-flask example is not truly representative of the real world where resources are usually limited. However, when a species is introduced into a new habitat that it finds suitable, it may show exponential growth for a while. In the case of the bacteria

in the flask, some bacteria will die during the experiment and thus not reproduce; therefore, the growth rate is lowered from a maximal rate in which there is no mortality. Additionally, ecologists are interested in the population at a particular point in time, an infinitely small time interval. For this reason, the terminology of differential calculus is used to obtain the "instantaneous" growth rate, replacing the *change* in number and time with an instant-specific measurement of number and time.

$$\frac{dN}{dT} = bN - dN = (b - d)N$$

Notice that the "*d*" associated with the first term refers to the derivative (as the term is used in calculus) and is different from the death rate, also called "$d$." The difference between birth and death rates is further simplified by substituting the term "*r*" (intrinsic rate of increase) for the relationship between birth and death rates:

$$\frac{dN}{dT} = rN$$

The value "$r$" can be positive, meaning the population is increasing in size; or negative, meaning the population is decreasing in size; or zero, where the population's size is unchanging, a condition known as zero population growth. A further refinement of the formula recognizes that different species have inherent differences in their intrinsic rate of increase (often thought of as the potential for reproduction), even under ideal conditions. Obviously, a bacterium can reproduce more rapidly and have a higher intrinsic rate of growth than a human. The maximal growth rate for a species is its biotic potential, or $r_{max}$, thus changing the equation to:

$$\frac{dN}{dT} = r_{\max}N$$

Populations of all species can, potentially, increase rapidly under conditions in which resource availability and other factors are not constraining. Examples of rapid population growth are illustrated in Figure **14.2.4**. However, unlimited growth cannot be sustained – in all of the cases in Figure **14.2.4**, the population sizes eventually leveled off, decreased, or crashed.

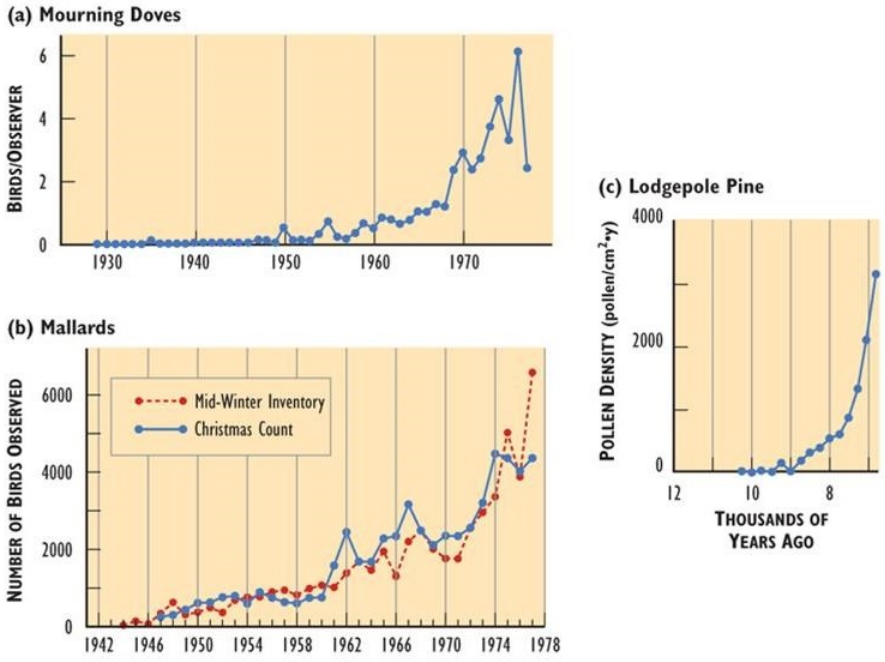



**Figure 14.2.4:** Rapid Growth of Some Natural Populations. (a) The population of mourning doves (Zenaida macroura) wintering in southern Ontario over 48 years. This used to be a rare bird, but it has apparently benefited from a warming climate, suburban habitat, and winter feeding. (b) The population of mallards (Anas platyrhynchos) wintering in southern Ontario over 35 years, illustrated with two independent sets of data. This duck has expanded its breeding and wintering ranges into eastern Canada, likely in response to habitat made available by the clearing of forest. (c) The population of lodgepole pine (Pinus contorta) near Snowshoe Lake, British Columbia, during natural afforestation following deglaciation 7000-9000 years ago. In this case, tree

populations are indicated by the amount of pollen in dated layers of lake sediment. Sources: Modified from (a) Freedman and Riley (1980); (b) Goodwin et al. (1977); (c) MacDonald and Cwynar (1991).

## Logistic Growth

Exponential growth is possible only when infinite natural resources are available; this is not the case in the real world. Charles Darwin recognized this fact in his description of the "struggle for existence," which states that individuals will compete (with members of their own or other species) for limited resources. The successful ones will survive to pass on their own characteristics and traits (which we know now are transferred by genes) to the next generation at a greater rate (natural selection). To model the reality of limited resources, population ecologists developed the logistic growth model.

In the real world, with its limited resources, exponential growth cannot continue indefinitely. Exponential growth may occur in environments where there are few individuals and plentiful resources, but when the number of individuals gets large enough, resources will be depleted, slowing the growth rate. Eventually, the growth rate will plateau or level off (Figure **14.2.1**). This population size, which represents the maximum population size that a particular environment can support, is called the carrying capacity, or $K$. There are three different sections to an S-shaped curve. Initially, growth is exponential because there are few individuals and ample resources available. Then, as resources begin to become limited, the growth rate decreases. Finally, growth levels off at the carrying capacity of the environment, with little change in population size over time.

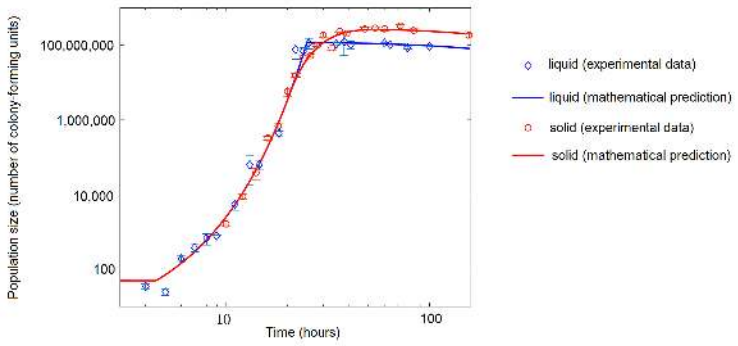



**Figure 14.2.4:** The growth of the bacterium *Escherichia coli* (*E. coli*) with limited nutrients over time in hours. Population size is in number of colony-forming units, which is a cell capable of dividing to produce a colony of cells. The bacteria were grown in liquid (blue diamonds) or on a semisolid medium (agar, red circles). The points represent actual measurements of populations size, and the lines represent a mathematical model, basically a prediction that is based on the data points. The data show logistic population growth, particularly in solid media. Initially, population size increases rapidly (concave part of curve), but the growth rate then stabilizes (straight line) before decreasing (convex part of curve) and leveling near the carrying capacity. Image modified from Shao X, Mugler A, Kim J, Jeong HJ, Levin BR, Nemenman I (2017) Growth of bacteria in 3-d colonies. PLoS Comput Biol 13(7): e1005679 (CC-BY).

In some populations, there are variations to the S-shaped curve. Examples in wild populations include sheep and harbor seals (figure **14.2.5**). In both examples, the population size exceeds the carrying capacity for short periods of time and then falls below the carrying capacity afterwards. This fluctuation in population size continues to occur as the population oscillates around its carrying capacity. Still, even with this oscillation the logistic model is confirmed.

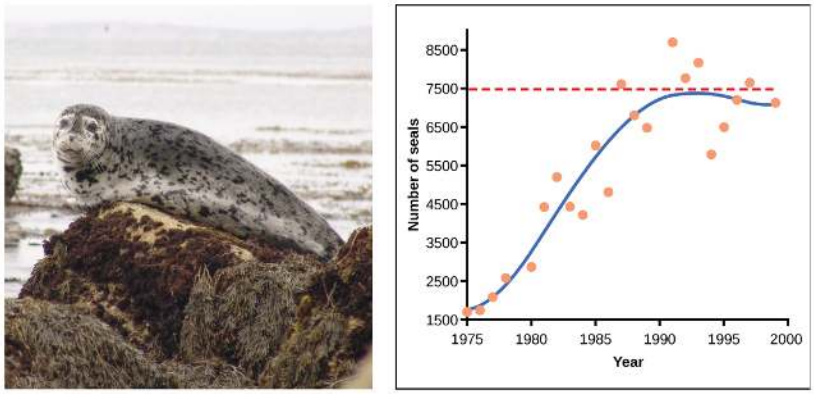



**Figure 14.2.5:** A natural population of seals shows logistic growth. The number of seals, ranging from 1500 to 8500, is on the y-axis, and the year, ranging from 1975-2000, is on the x-axis. Orange dots represent the observed population size each year, and the blue line represent the trend. From 1975 to 1983, the blue line curves upwards, representing an increase in both population size and population growth rate. From 1983 to 1992, the curve flattens, representing that population size is still increasing, but population growth rate is decreasing. After 1992, population fluctuates near the carrying capacity (dotted, horizontal line).

The formula we use to calculate logistic growth adds the carrying capacity as a moderating force in the growth rate. The expression "$K$–$N$" is indicative of how many individuals may be added to a population at a given stage, and "$K$–$N$" divided by "$K$" is the fraction of the carrying capacity available for further growth. Thus, the exponential growth model is restricted by this factor to generate the logistic growth equation:

$$\frac{dN}{dT} = r_{\max}\frac{dN}{dT}$$
$$= r_{\max}N\frac{(K-N)}{K}$$

Notice that when $N$ is very small, $(K-N)/K$ becomes close to $K/K$ or 1, and the right side of the equation reduces to $r_{max}N$, which means the population is growing exponentially and is not influenced by carrying capacity. On the other hand, when $N$ is large, $(K-N)/K$ come close to zero, which means that population growth will be slowed greatly or even stopped. Thus, population growth is greatly slowed in large populations by the carrying capacity $K$. This model also allows for the population of a negative population growth, or a population decline. This occurs when the number of individuals in the population exceeds the carrying capacity (because the value of $(K-N)/K$ is negative).

The logistic model assumes that every individual within a population will have equal access to resources and, thus, an equal chance for survival. For plants, the amount of water, sunlight, nutrients, and the space to grow are the important resources, whereas in animals, important resources include food, water, shelter, nesting space, and mates. In the real world, phenotypic variation among individuals within a population means that some individuals will be better adapted to their environment than others. The resulting competition between population members of the same species for resources is termed **intraspecific competition** (intra- = "within"; -specific = "species"). Intraspecific competition for resources may not affect populations that are well below their carrying capacity —resources are plentiful and all individuals can obtain what they need. However, as population size increases, this competition intensifies. In addition, the accumulation of waste products can reduce an environment's carrying capacity.

The logistic population growth model is not the only way that populations respond to limited resources. In some populations, growth is exponential until resources run low, wastes accumulate, or disease spreads (see limiting factors below), and the population then crashes (Figure **14.2.5**). Thus, population growth rate (and size) may plummet rapidly instead of tapering as it approaches the carrying capacity.

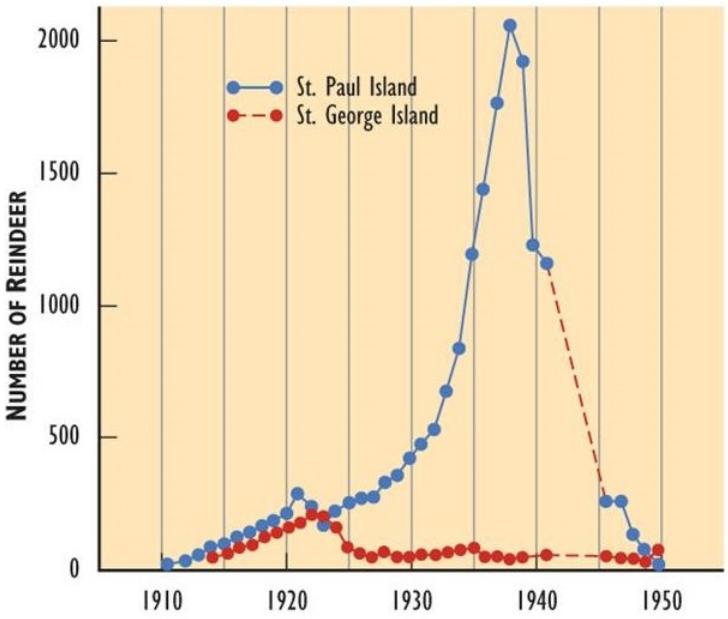



**Figure 14.2.5:** Population Growth and Crash. In 1910, reindeer (Rangifer tarandus tarandus; the Eurasian subspecies of caribou) were introduced to two islands in the Aleutian chain off Alaska in an attempt to establish a new food resource for local use. On both islands, the reindeer population increased rapidly. However, they exceeded the carrying capacity of the habitat and caused severe damage through overgrazing. The populations then crashed. Source: Modified from Krebs (1985).

## Factors that Regulate Population Growth

The logistic model of population growth, while valid in many natural populations and a useful model, is a simplification of real-world population dynamics. Implicit in the model is that the carrying capacity of the environment does not change, which is not the case. The carrying capacity varies annually: for example, some summers are hot and dry whereas others are cold and wet. In many areas, the carrying capacity during the winter is much lower than it is during the summer. Also, natural events such as earthquakes, volcanoes, and fires can alter an environment and hence its carrying capacity. Additionally, populations do not usually exist in isolation. They engage in interspecific competition: that is, they share the environment with other species, competing with them for the same resources. These factors are also important to understanding how a specific population will grow.

Nature regulates population growth in a variety of ways. These are grouped into density-dependent factors, in which the density of the population at a given time affects growth rate and mortality, and density-independent factors, which influence mortality in a population regardless of population density. Note that in the former, the effect of the factor on the population depends on the density of the population at onset. Conservation biologists want to understand both types because this helps them manage populations and prevent extinction or overpopulation.

### Density-dependent Regulation

Most **density-dependent factors** are biological in nature (biotic). Usually, the denser a population is, the greater its mortality rate. Density-dependent factors include predation, parasitism, herbivory, competition, and accumulation of waste. An example of density-dependent regulation is shown in figure 14.2.6 with results from a study focusing on the giant intestinal roundworm (*Ascaris lumbricoides*), a parasite of humans and other mammals. Denser populations of the parasite exhibited lower fecundity: they contained fewer eggs. One possible explanation for this is that females would be smaller in more dense populations (due to limited resources) and that smaller females would have fewer eggs. This hypothesis was tested and disproved in a 2009 study which showed that female weight had no influence. The actual cause of the density-dependence of fecundity in this organism is still unclear and awaiting further investigation.

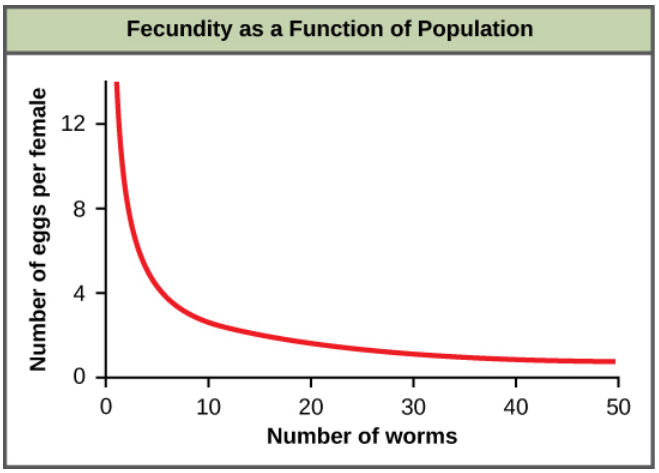



**Figure 14.2.6:** In this population of roundworms, fecundity (number of eggs) decreases with population density.[3]

As a population increases, its predators are able to harvest it more easily. Prey density also affects population growth rate of predators: low prey density increases the mortality of its predator because it has more difficulty locating its food sources. Parasites are able to pass from host to host more easily as the population density of the host increases. For this reason, epidemics among humans are particularly severe in cities. In fact, for most of the period since humans began living in cities, city populations have been maintained only through continual immigration from the countryside. Not until the development of community sanitation, immunization, and other public health measures did cities avoid periodic sharp drops in population as a result of epidemics. The recurrent epidemics of the "black death" in Europe that began in the fourteenth century caused a sharp decline in population. In just three years (1348–1350), at least one-quarter of the population of Europe died from the disease (probably plague). Similarly, herbivores can more easily spread between individual plants in a dense population. This is why strip cropping helps control pests. An herbivore or plant pathogen may infect one row of plants, but it is less likely to spread to more distant rows of that species.

Intraspecific competition (and interspecific competition) can be important density dependent factors. In the summer of 1980, much of southern New England was struck by an infestation of the gypsy moth (figure **14.2.6**). As the summer wore on, the larvae (caterpillars) pupated, the hatched adults mated, the females laid masses of eggs (each mass containing several hundred eggs) on virtually every tree in the region. In early May of 1981, the young caterpillars that hatched from these eggs began feeding and molting. In 72 hours, a 50-ft beech tree or a 25-ft white pine tree would be completely defoliated. Large patches of forest began to take on a winter appearance with their skeletons of bare branches. In fact, the infestation was so heavy that many trees were completely defoliated before the caterpillars could complete their larval development. The result: a massive die-off of the animals; very few succeeded in completing metamorphosis. Here, then, was a dramatic example of how competition among members of one species for a finite resource - in this case, food - caused a sharp drop in population. The effect was clearly density-dependent. The lower population densities of the previous summer had permitted most of the animals to complete their life cycle.



**Figure 14.2.6:** Gypsy moth caterpillars face intraspecific competition at high densities. Image by Editor at Large (CC-BY-SA).

## Density-independent Regulation

Many factors, typically physical or chemical in nature (abiotic), influence the mortality of a population regardless of its density, including weather, natural disasters, and pollution (Figure **14.2.7**). An individual deer may be killed in a forest fire regardless of how many deer happen to be in that area. Its chances of survival are the same whether the population density is high or low. The same holds true for cold winter weather. Not only do abiotic factors limit population growth but they often drive existing populations well below their previous level.



**Figure 14.2.7:** Severe weather, like Hurricane Florence of 2018, can reduce population size regardless of density. Image by NASA Goddard Space Flight Center (CC-BY)

These factors are described as density-independent because they exert their effect irrespective of the size of the population when the catastrophe struck. This graph (Figure **14.2.8**, from P. T. Boag and P.R. Grant in **Science**, **214:**82, 1981) shows the decline in the population of one of Darwin's finches (*Geospiza fortis*) on Daphne Major, a tiny (100 acres = 40 hectares) member of the Galapagos Islands. The decline (from 1400 to 200 individuals) occurred because of a severe drought that reduced the quantity of seeds on which this species feeds. The drought ended in 1978, but even with ample food once again available the finch population recovered only slowly.

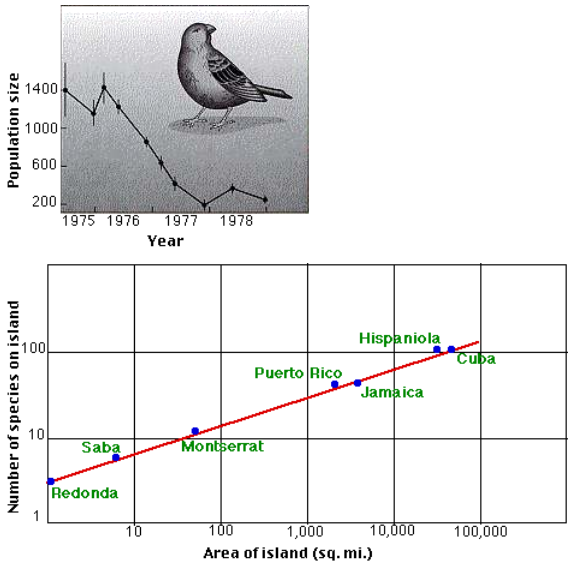



**Figure 14.2.7:** Left: This graph (from P. T. Boag and P.R. Grant in **Science**, **214:**82, 1981) shows the decline in the population of one of Darwin's finches (*Geospiza fortis*) on Daphne Major, a tiny (100 acres = 40 hectares) member of the Galapagos Islands. The decline (from 1400 to 200 individuals) occurred because of a severe drought that reduced the quantity of seeds on which this species feeds. Right: The graph (redrawn from R. H. MacArthur and E. O. Wilson, *The Theory of Island Biogeography*, Princeton University Press) shows the number of species of reptiles and amphibians on various islands in the West Indies.

In real-life situations, population regulation is very complicated and density-dependent and independent factors can interact. A dense population that is reduced in a density-independent manner by some environmental factor(s) will be able to recover differently than a sparse population. For example, a population of deer affected by a harsh winter will recover faster if there are more deer remaining to reproduce.

Catastrophic declines are particularly risky for populations living on islands. The smaller the island, the smaller the population of each species on it, and the greater the risk that a catastrophe will so decimate the population that it becomes extinct. This appears to be one reason for the clear relationship between size of island and the number of different species it contains. The graph (Figure **14.2.8**, redrawn from R. H. MacArthur and E. O. Wilson, *The Theory of Island Biogeography*, Princeton University Press) shows the number of species of reptiles and amphibians on various islands in the West Indies. In general, if one island has 10 times the area of another, it will contain approximately twice the number of species. The same principle applies to many **habitats**. In a sense, most habitats are islands. A series of ponds, a range of mountain tops, scattered groves of citrus trees, even individual trees within a grove, all are made up of patches of habitat separated by barriers to the free migration of their inhabitants.

This has practical as well as theoretical importance. As the human population grows, jungles are cleared for agriculture, farms are paved for shopping centers, rivers are dammed for hydroelectric power and irrigation, etc. Although wildlife sanctuaries are being established, they must be made large enough so that they can support populations large enough to survive density-independent checks when they strike.

### An example: Lake Guri

In 1986, the closing of a dam in Venezuela flooded over a thousand square miles (>2,500 $km^2$) turning hundreds of hilltops into islands. These ranged in size from less than 1 hectare (2.5 acres) to more than 150 hectares (370 acres).

Within 8 years:

- The tiniest islands (<1 hectare) lost 75% of the species that had lived there.
- The larger the island, the fewer species it lost.
- But all the islands - even the largest - lost their top predators; that is carnivores like pumas, jaguars (image), and eagles at the ends of food chains.
- Those animal species that did remain - mostly herbivores and small carnivores - greatly increased their populations because of

- a reduction in competition for resources
- no longer being eaten by predators

The intense grazing by the increased herbivore populations is degrading the variety of plant life on the smaller islands.

# Footnotes

1. 1 N.A. Croll et al., “The Population Biology and Control of *Ascaris lumbricoides* in a Rural Community in Iran.” *Transactions of the Royal Society of Tropical Medicine and Hygiene* 76, no. 2 (1982): 187-197, doi:10.1016/0035-9203(82)90272-3.
2. 2 Martin Walker et al., “Density-Dependent Effects on the Weight of Female *Ascaris lumbricoides* Infections of Humans and its Impact on Patterns of Egg Production.” *Parasites & Vectors* 2, no. 11 (February 2009), doi:10.1186/1756-3305-2-11.
3. 3 N.A. Croll et al., “The Population Biology and Control of *Ascaris lumbricoides* in a Rural Community in Iran.” *Transactions of the Royal Society of Tropical Medicine and Hygiene* 76, no. 2 (1982): 187-197, doi:10.1016/0035-9203(82)90272-3.
4. 4 David Nogués-Bravo et al., “Climate Change, Humans, and the Extinction of the Woolly Mammoth.” *PLoS Biol* 6 (April 2008): e79, doi:10.1371/journal.pbio.0060079.
5. 5 G.M. MacDonald et al., “Pattern of Extinction of the Woolly Mammoth in Beringia.” *Nature Communications* 3, no. 893 (June 2012), doi:10.1038/ncomms1881.

Freedman, B. and J. Riley. 1980. Population trends of various species of birds wintering in southern Ontario. Ontario Field Biologist, 34: 49-79.

Goodwin, C. E., B. Freedman, and S. M. McKay. 1977. Population trends in waterfowl wintering in the Toronto region, 1929–1976. Ontario Field Biologist, 31: 1-28.

Krebs, C.J. 2008. Ecology: The Experimental Analysis of Distribution and Abundance. 6th ed. Benjamin Cummings, New York, NY.

MacDonald, G.M. and L.C. Cwynar. 1991. Post-glacial population growth rates of Pinus contorta ssp. latifolia in western Canada. Journal of Ecology, 79: 417-429.

## Contributors and Attributions

Modified by Kyle Whittinghill and Melissa Ha from the following sources

- **45.4: Population Dynamics and Regulation** by OpenStax, is licensed CC BY by

  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).

- **45.3: Environmental Limits to Population Growth** by OpenStax, is licensed CC BY by

  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).

- **17.3B: Principles of Population Growth** by John W. Kimball, is licensed CC BY. This content is distributed under a Creative Commons Attribution 3.0 Unported (CC BY 3.0) license and made possible by funding from The Saylor Foundation
- Population Growth and Regulation from *Environmental Biology* by Matthew R. Fisher (CC-BY)
- Principles of Population Growth and The Human Population from *Biology* by John
- Ecology: From Individuals to the Biosphere in *Environmental Science: A Canadian Perspective* by Bill Freedman

This page titled 14.2: Population Growth and Regulation is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **45.3: Environmental Limits to Population Growth** by OpenStax is licensed CC BY 4.0.
- **45.4: Population Dynamics and Regulation** by OpenStax is licensed CC BY 4.0.

- **17.3B: Principles of Population Growth** by John W. Kimball is licensed CC BY 3.0. Original source: https://www.biology-pages.info/.

# 14.3: Human Population Growth

Concepts of animal population dynamics can be applied to human population growth. Humans are not unique in their ability to alter their environment. For example, beaver dams alter the stream environment where they are built. Humans, however, are unique in their ability to alter their environment with the conscious purpose of increasing its carrying capacity, sometimes to the detriment of other species (e.g., via artificial selection for crops that have a higher yield). This ability is a major factor responsible for human population growth and a way of overcoming density-dependent growth regulation. Much of this ability is related to human intelligence, society, and communication. Humans can construct shelter to protect them from the elements and have developed agriculture and domesticated animals to increase their food supplies. In addition, humans use language to communicate this technology to new generations, allowing them to improve upon previous accomplishments.

Although humans have increased the carrying capacity of their environment, the technologies used to achieve this transformation have caused unprecedented changes to Earth's environment, altering ecosystems to the point where some may be in danger of collapse. The depletion of the ozone layer, erosion due to acid rain, and damage from global climate change are caused by human activities. The ultimate effect of these changes on our carrying capacity is unknown. As some point out, it is likely that the negative effects of increasing carrying capacity will outweigh the positive ones—the carrying capacity of the world for human beings might actually decrease.

## History of Human Population Growth

The human population is growing rapidly. For most of human history, there were fewer than 1 billion people on the planet. During the time of the Agricultural Revolution, 10,000 B.C., there were only 5-10 million people on Earth - which is basically the population of New York City today. In 1800, when the Industrial Revolution began, there were approximately 1 billion people on Earth. Continued agricultural expansion and extraction of fossil fuels and minerals led to rapid global economic growth and, in turn, population growth in the 19th century. We've added over 6 billion people to the human population in just a little over 200 years (figure **14.3.1**). As of August 2020, the global human population is around 7.8 billion people.

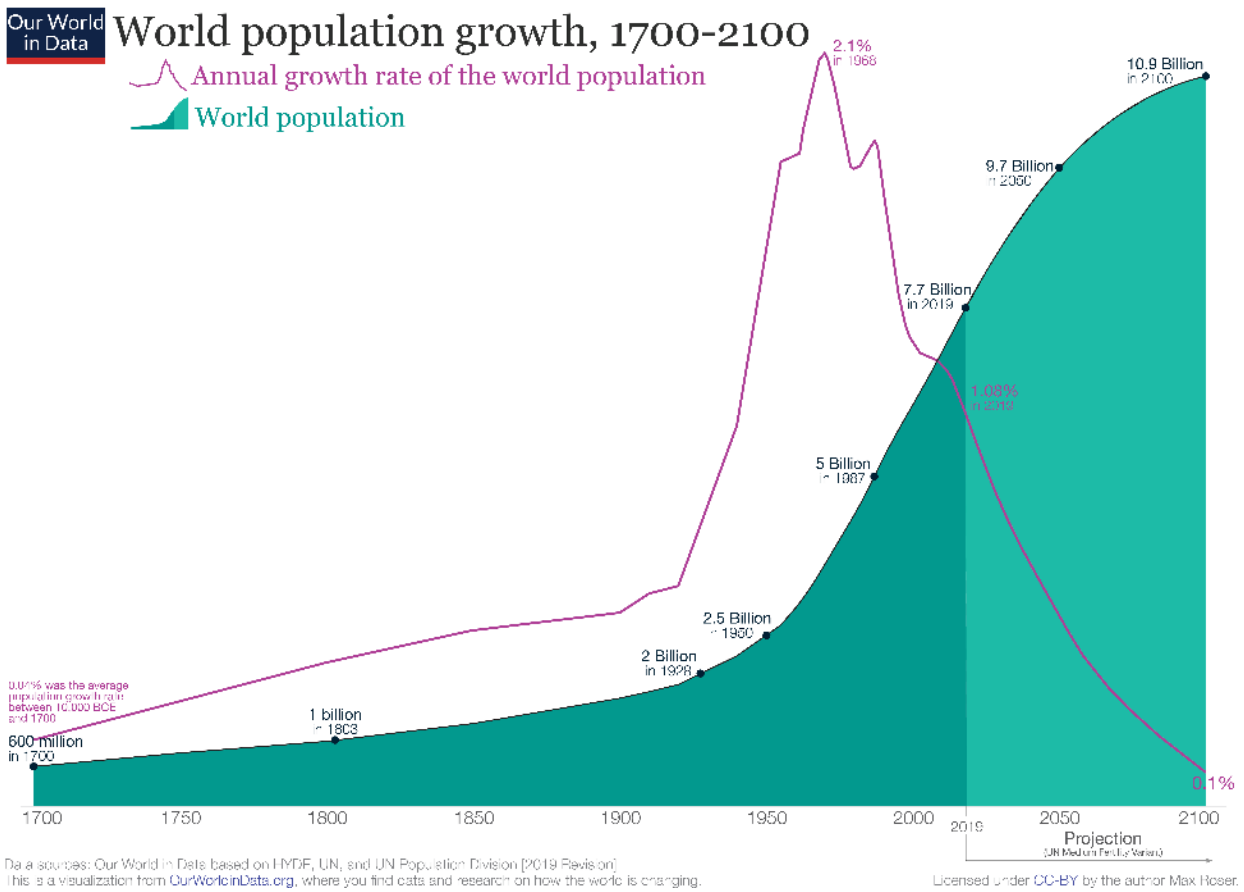


**Figure 14.3.1:** Human population growth. Since 1700, human population size (shaded in teal) has increased. It reached 1 billion in 1803, 2 billion in 1928, 2.5 billion in 1950, 5 billion in 1987, and 7.7 billion in 2019. It is projected to reach 10.9 billion in 2100. The population growth rate (pink line) was only 0.04% on average between 10,000 BCE and 1700. The population growth rate peaked in 1968 at 2.1%, and since then, it has slowed to 1.08% in 2019 and is projected to be at 0.1% in 2100. Image by Max Roser (2013) "Future Population Growth". Published online at OurWorldInData.org. (CC-BY)

The fundamental cause of the acceleration of growth rate for humans in the past 200 years has been the reduced death rate due to changes in public health and sanitation. Clean drinking water and proper disposal sewage has drastically improved health in developed nations. Also, medical innovations such as the use of antibiotics and vaccines have decreased the ability of infectious disease to limit human population growth. In the past, diseases such as the bubonic plaque of the fourteenth century killed between 30 and 60 percent of Europe's population and reduced the overall world population by as many as one hundred million people. Naturally, infectious disease continues to have an impact on human population growth, especially in poorer nations. For example, life expectancy in sub-Saharan Africa, which was increasing from 1950 to 1990, began to decline after 1985 largely as a result of HIV/AIDS mortality. According to a 2016 study by Marcus et al., The reduction in life expectancy caused by HIV/AIDS was estimated to be 8 years for 2016.

The world's human population is currently experiencing exponential growth even though human reproduction is far below its biotic potential (Figure **14.3.1**). To reach its biotic potential, all females would have to become pregnant every nine months or so during their reproductive years. Also, resources would have to be such that the environment would support such growth. Neither of these two conditions exists. In spite of this fact, human population is still growing exponentially. A consequence of exponential human population growth is the time that it takes to add a particular number of humans to the Earth is becoming shorter. Figure **14.3.2** shows that 123 years were necessary to add 1 billion humans in 1930, but it only took 12 years to add 1 billion people between 1975 and 1987.

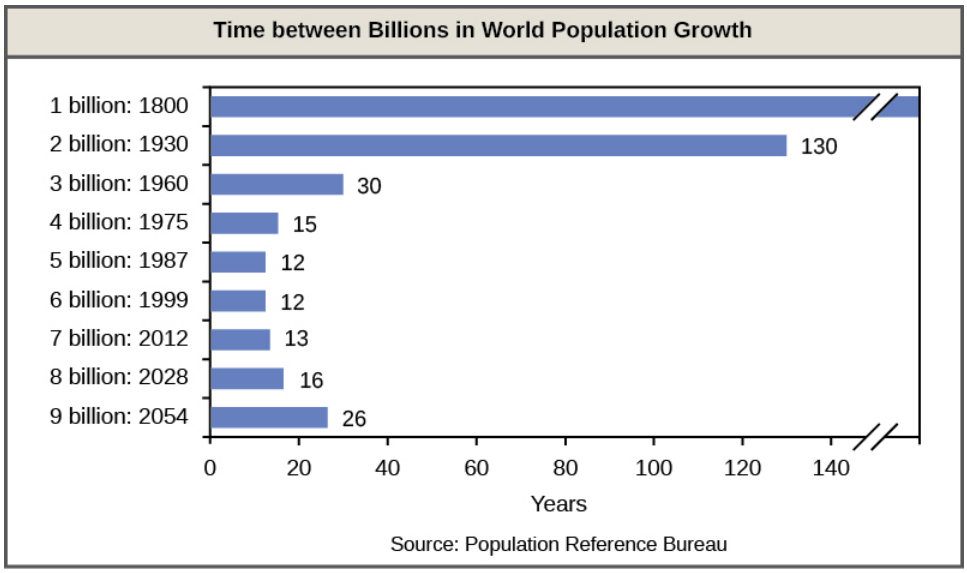



**Figure 14.3.2:** The time between the addition of each billion human beings to Earth decreases over time. (credit: modification of work by Ryan T. Cragun)

The actual growth rates in different countries are shown in Figure **14.3.3**, with the highest rates tending to be in the less economically developed countries of Africa and Asia.

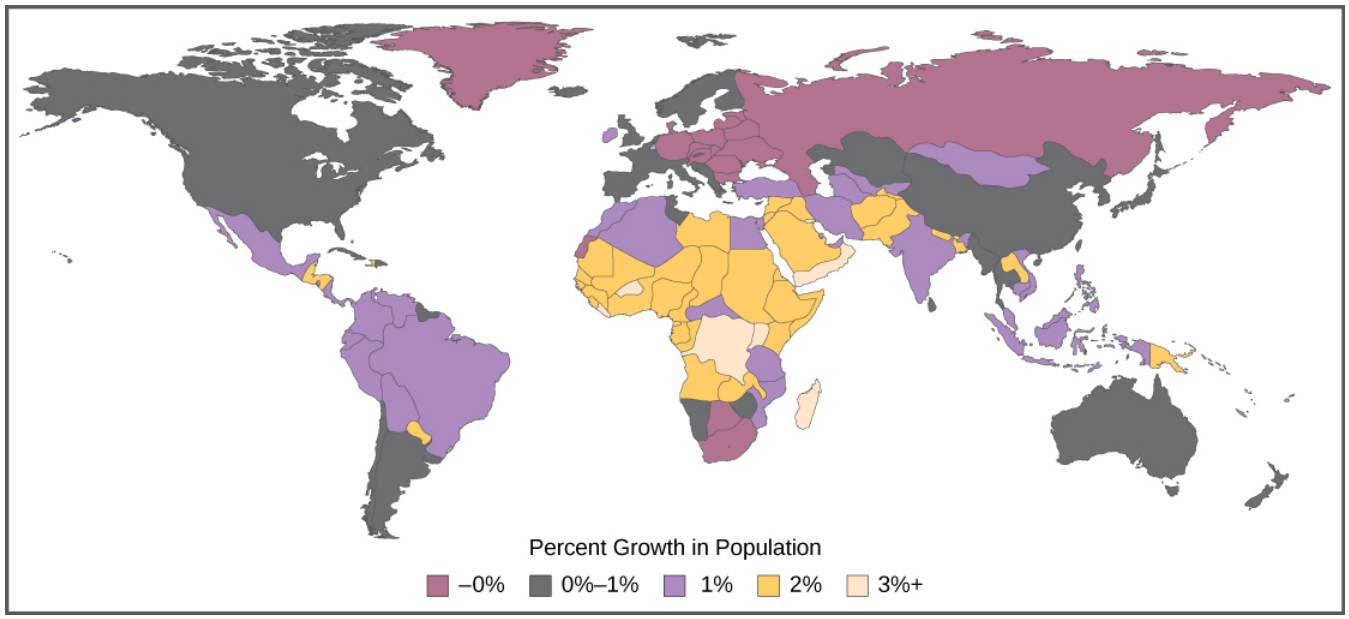



**Figure 14.3.3:** The percent growth rate of population in different countries is shown. Notice that the highest growth is occurring in less economically developed countries in Africa and Asia.



Click through this interactive view of how human populations have changed over time.

## Demographic Transition

Major changes in the human population first began during the 1700s in Europe and North America. First death rates fell, followed somewhat later by birth rates. Several advances in science and technology led to lower death rates, especially in children, in 18$^{th}$ century Europe and North America:

- New scientific knowledge of the causes of disease led to improved water supplies, sewers, and personal hygiene.
- Better farming techniques and machines increased the food supply.
- The Industrial Revolution of the 1800s led to new sources of energy, such as coal and electricity. This increased the efficiency of the new agricultural machines. It also led to train transport, which improved the distribution of food.

Because many more children survived to adulthood, birth rates also increased. As the gap between birth and death rates widened, the human population grew faster. It wasn't long before birth rates started to fall as well in Europe and North America. People started having fewer children because large families were no longer beneficial for several reasons.

- As child death rates fell and machines did more work, farming families no longer needed to have as many children to work in the fields.

- Laws were passed that required children to go to school. Therefore, they could no longer work and contribute to their own support. They became a drain on the family's income.

Eventually, birth rates fell to match death rates. As a result, population growth slowed to nearly zero. These changes in population that occurred in Europe and North America have been called the **demographic transition**. The transition can be summarized in the following four stages, which are illustrated in Figure **14.3.4** below:

- Stage 1—High birth and death rates lead to slow population growth.
- Stage 2—The death rate falls but the birth rate remains high, leading to faster population growth.
- Stage 3—The birth rate starts to fall, so population growth starts to slow.
- Stage 4—The birth rate reaches the same low level as the death rate, so population growth slows to zero.

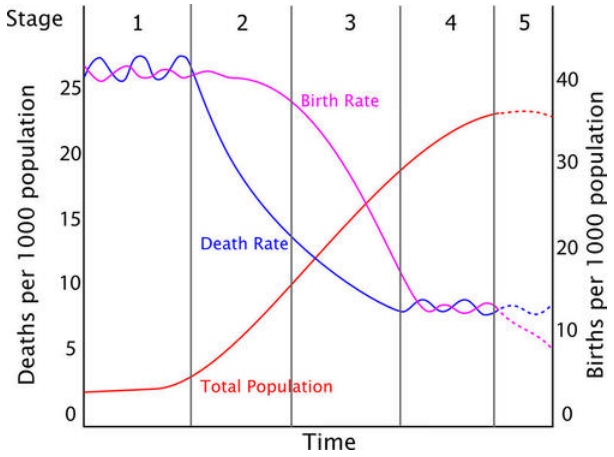



**Figure 14.3.4:** Stages of the Demographic Transition. In the demographic transition, the death rate falls first. After a lag, the birth rate also falls.

## Age Structure, Population Growth, and Economic Development

The age structure of a population is an important factor in population dynamics. Age structure is the proportion of a population at different age ranges. Age structure allows better prediction of population growth, plus the ability to associate this growth with the level of economic development in the region. Countries with rapid growth have a pyramidal shape in their age structure diagrams, showing a preponderance of younger individuals, many of whom are of reproductive age or will be soon (Figure **14.3.5**). This pattern is most often observed in underdeveloped countries where individuals do not live to old age because of less-than-optimal living conditions. Age structures of areas with slow growth, including developed countries such as the United States, still have a pyramidal structure, but with many fewer young and reproductive-aged individuals and a greater proportion of older individuals. Other developed countries, such as Italy, have zero population growth. The age structure of these populations is more conical, with an even greater percentage of middle-aged and older individuals.

Art Connection

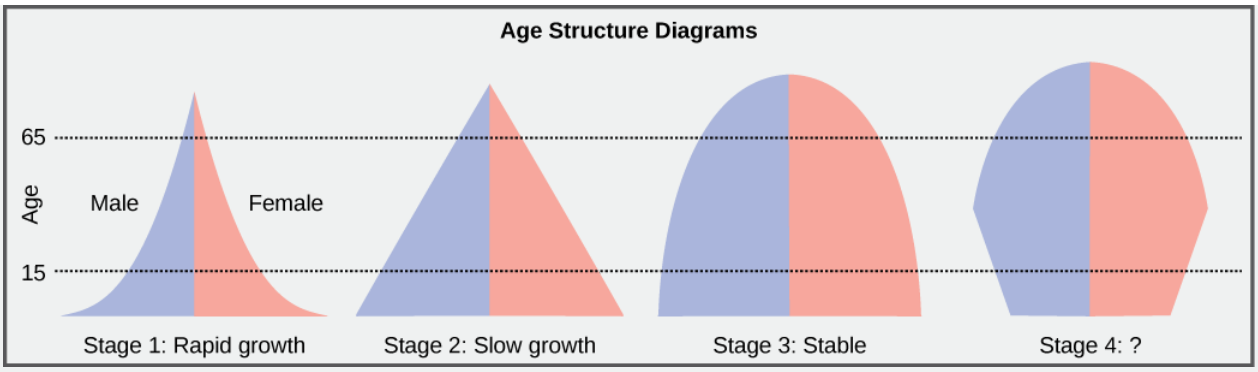
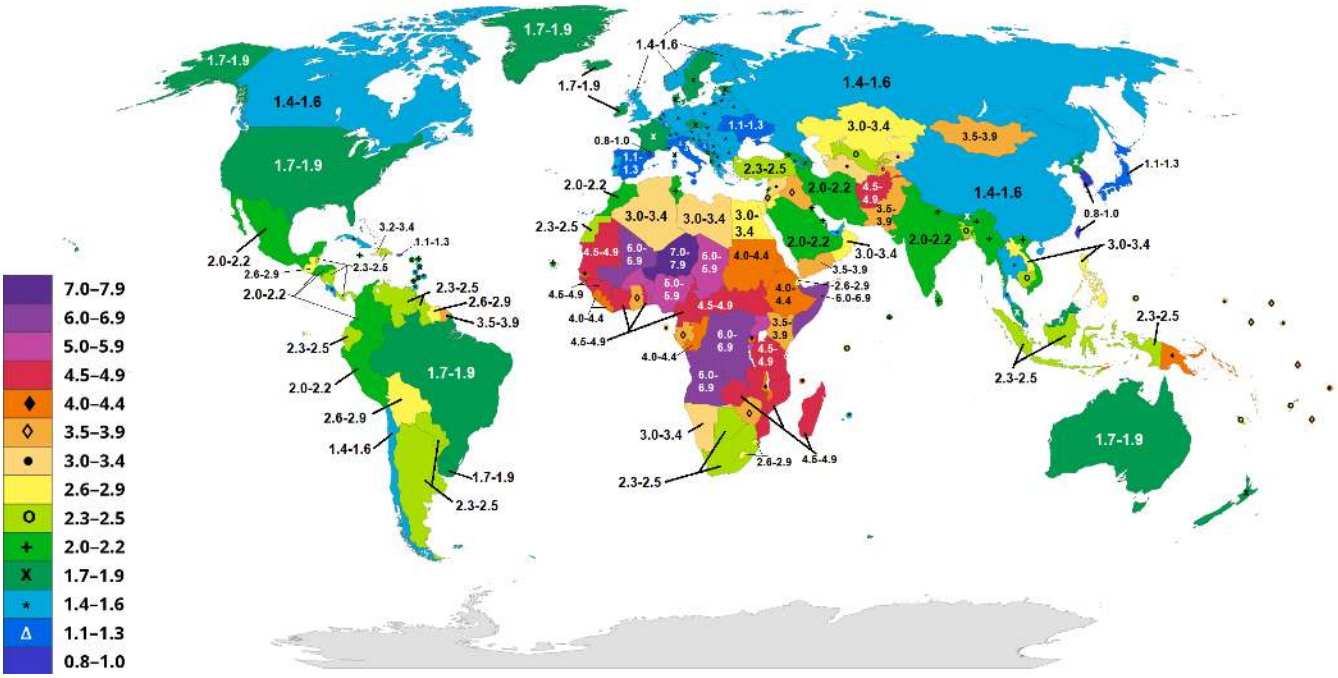



**Figure 14.3.5:** Typical age structure diagrams are shown. The rapid growth diagram narrows to a point, indicating that the number of individuals decreases rapidly with age. In the slow growth model, the number of individuals decreases steadily with age. Stable population diagrams are rounded on the top, showing that the number of individuals per age group decreases gradually, and then increases for the older part of the population.

Age structure diagrams for rapidly growing, slow growing and stable populations are shown in stages 1 through 3. What type of population change do you think stage 4 represents?

Although global population size continues to increase, the rate of human population growth has decreased. This means that the population size is not increasing as quickly as it did in the past (figure **14.3.1**). The **total fertility rate** (TFR) is the average number of children that each woman will have during her lifetime. The TFR is an average because, of course, some women will have more, some fewer, and some no children at all. Theoretically, when the TFR = 2, each pair of parents just replaces itself. Actually it takes a TFR of 2.1 or 2.2 to replace each generation - this number is called the **replacement fertility rate** - because some children will die before they grow up to have their own two children. Figure **14.3.6** compares he total fertility rate in different countries. Since the early 1960's the global total fertility rate (or per capita birth rate) has fallen due to several factors:

- Modern medicine raised survival rates, making large families unnecessary.
- In many regions birth control, became more available.
- Education slowed reproduction because people got married later.
- HIV decreased population growth rates in many countries.
- China's one-child policy had a noticeable effect on population growth in China (the country with the largest population).





https://bio.libretexts.org/@go/page/69500

**Figure 14.3.6:** Total fertility rate by country in 2020. Most countries in North and South America have a total fertility rate of 1.7-2.5 children per woman, but fertility rates are lower (for example, Canada and Chile) or higher (for example, Guatamala and Bolivia) in a few countries. Most of Europe and northern and Eastern Asia have low fertility rates (generally 1.4-1.6 children per woman). Many Middle Eastern countries and those in southern and southeastern Asia have slightly higher fertility rates (generally 1.7-2.5 children per woman). The fertility rate in Australia is 1.7-1.9 children per woman. High fertility rates are common in African countries. For example, the fertility rate in Niger is 7.0-7.9. Image modified from Korakys (CC-BY-SA).

In many nations, birth rates are presently below replacement rates. In fact, in all nations with a gross national income above $16,000 dollars per person, the birth rate is at or below the replacement rate of 2.1 lifetime births per female (Figure **14.3.7**).

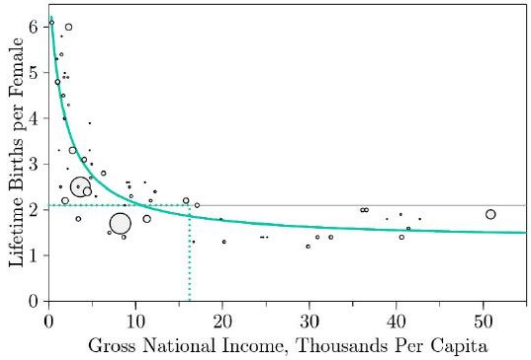



**Figure 14.3.7:** *Human fecundity (births per female) as a function of national per capita income.*

The TFR in the United States declined from more than 4 late in the nineteenth century to less than replacement in the early 1930s. However, when the small numbers of children born in the depression years reached adulthood, they went on a childbearing spree that produced the baby-boom generation (figure **14.3.8**). In 1957, more children were born in the United States than ever before (or since).

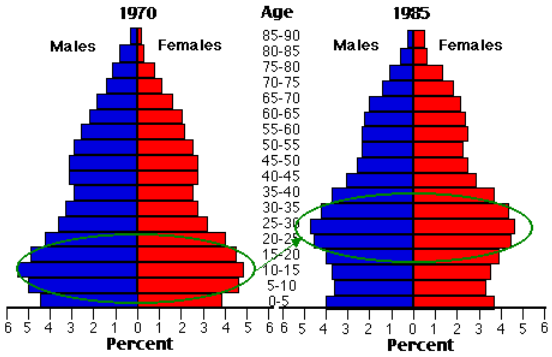



**Figure 14.3.8:** Age structure diagrams of the United States population in 1970 and 1985. The x-axis represents the percent of the population. The y-axis represents age groups in increments of five from 0-5 years all the way to 85-90 years. Blue bars to the left represent males, and red bars to the right represent females. Bars taper near the top of the diagram, showing that those in the oldest age categories represent a small percentage of the population. The 1970 age structure diagram bulges around the 10-15 age group, and this represents the baby boom (green oval). In 1985, this bulge has moved to the 25-30 age group.

## Long-Term Consequences of Exponential Human Population Growth

Many dire predictions have been made about the world's population leading to a major crisis called the "population explosion." In the 1968 book *The Population Bomb*, biologist Dr. Paul R. Ehrlich wrote, "The battle to feed all of humanity is over. In the 1970s hundreds of millions of people will starve to death in spite of any crash programs embarked upon now. At this late date nothing can prevent a substantial increase in the world death rate."[1] While many critics view this statement as an exaggeration, the laws of exponential population growth are still in effect, and unchecked human population growth cannot continue indefinitely.

Reducing the population growth rate will limit the consequences of overpopulation; however, regulation of the human population is a complex issue that presents ethical considerations. Efforts to moderate population control led to the 1979 One-Child Policy in China, which imposed fines on urban couples who had more than one child (figure **14.3.9**). Additionally, only a couple's first child was granted educational and health care benefits. The effectiveness of the policy in limiting overall population growth is controversial, as was the policy itself. Tragically, this policy combined with the desire of many couples to have a male heir led to female infanticide. Furthermore, enforcement of the policy led to human rights abuses including forced sterilizations and forced abortions. Over the years, some restrictions associated with the One-Child Policy were lifted. For example, in 2013, an adult who was an only child was allowed to have two children. The policy ended in 2016, and as of May 2021, each couple is now permitted to have up to three children.




**Figure 14.3.9:** A poster explaining and promoting the One-Child Policy. Image by Arian Zwegers (CC-BY).

**Zero population growth** occurs when the birth rate equals the death rate such that the rate of population growth is zero. Logically, this would occur when total fertility rate equals replacement fertility rate, but as described in the previous section, this depends on the age structure of a population.

All else being equal, reductions in total fertility rate will slow human population growth. Total fertility rate is negatively correlated with improved standard of living, access to health care, and gender equality. As countries develop economically, individuals become increasingly dependent on urban jobs rather than the family farm for income and provisions. While a large family would be needed to tend the family farm, supporting many children in an urban area is expensive. As a result, economic development is associated with couples choosing to have fewer children.

Access to healthcare decreases childhood mortality and provides the option of birth control. While decreased infant and childhood mortality alone would increase the population growth rate, people tend to have fewer children when they can be confident that all of their children are likely to survive; thus, as childhood mortality declines, total fertility rate also declines, lowering the population growth rate.

The education of women and girls is also associated with a decreased total fertility rate. Furthermore, family planning education programs have had highly positive effects in some countries on limiting population growth rates and increasing standards of living. When women have knowledge of family planning, power to make decisions about their family size, and a variety of career options, fertility rate declines. In regions that have made recent strides in educating girls, families more often choose to have a small number of children so that they can afford to pay for their education regardless of gender.

The impact of humans on our environment depends not only on population size but also on resource use. The **ecological footprint** is a measure of the area of land needed to support an individual or country. This includes agricultural lands used to

provide food, other lands used to provide goods (such as forests to provide timber) and fuels (such as oil fields to provide fossil fuels), and the land needed to dispose of wastes.

Countries differ in the average ecological footprint per person, and these differences are associated with region (figure **14.3.*b***). For example, in 2017, the ecological footprint per person in the United States was 8.1 global hectares compared to only 1.2 global hectares in India. (Visit footprintnetwork.org to see ecological footprints and related metrics for all countries.) Many factors affect the size of an ecological footprint from economic development to resource conservation to individual product choice. Reducing our ecological footprints can increase our carrying capacity and decrease environmental impact. So maybe we should update the question of, "What the Earths carrying capacity for humans," to, "What is the carrying capacity of people on Earth at which we can provide a fair and decent life for all humans while causing the minimum impact on the environment?

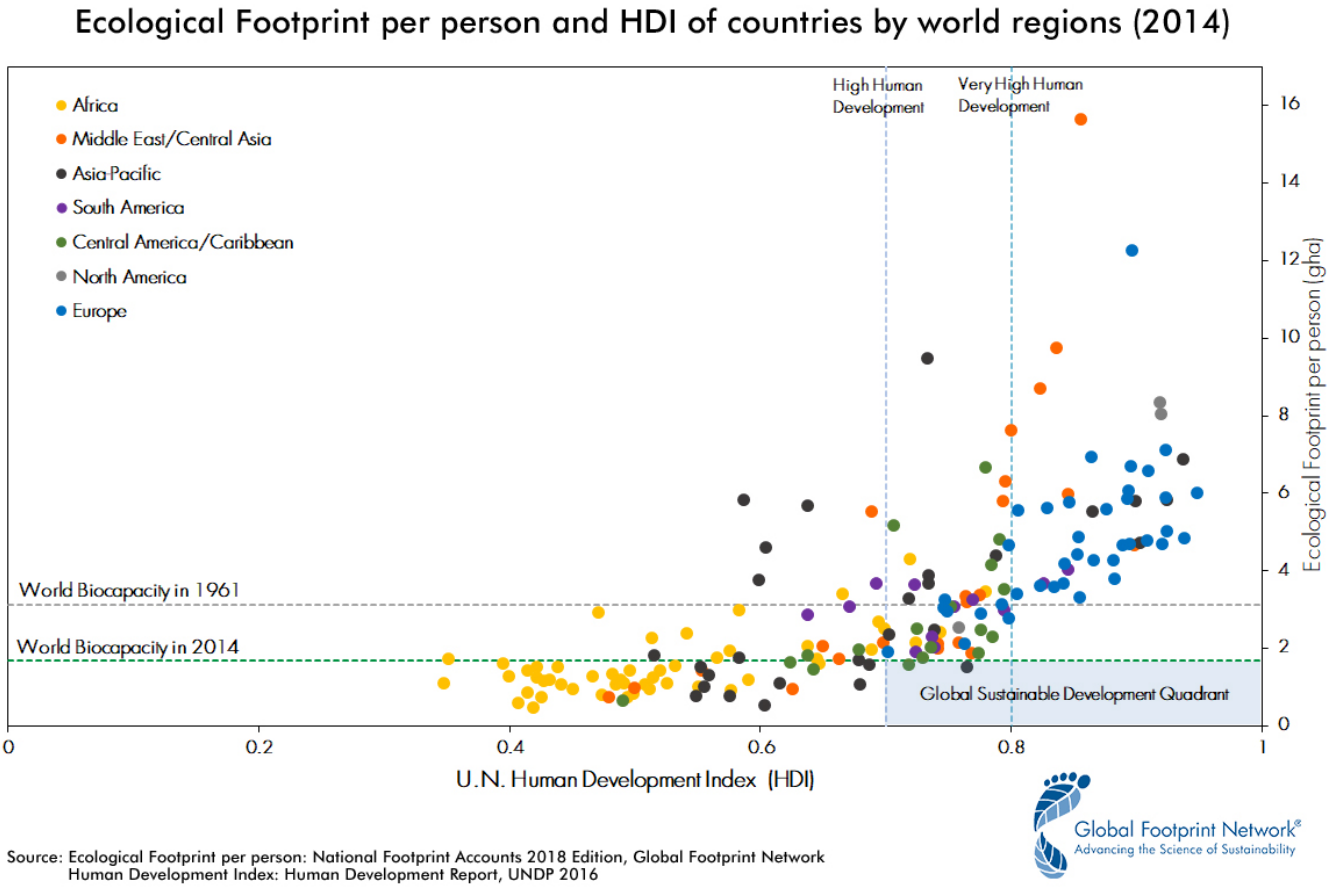



**Figure 14.3.*b*:** The ecological footprint per person in relation to the U.N. human development index for each country coded by region. Countries in some regions of the world have low ecological footprints per person, but this is because people are living in poverty (points in lower left of graph). This pattern is characteristic of many African countries. Other countries have large ecological footprints and live affluently (points in upper right of graph), including many European countries. To sustainably meet human needs, countries must achieve a high human development index and low ecological footprint (points in lower right of graph). Image by Footprint123 (CC-BY-SA).

Are you wondering about the size of your ecological footprint? You can estimate it here.

## References

Marcus, J. L., Chao, C. R., Leyden, W. A., Xu, L., Quesenberry, C. P., Jr, Klein, D. B., Towner, W. J., Horberg, M. A., & Silverberg, M. J. (2016). Narrowing the Gap in Life Expectancy Between HIV-Infected and HIV-Uninfected Individuals With Access to

Care. *Journal of acquired immune deficiency syndromes (1999)*, *73*(1), 39–46.

Population Growth (Annual %). 2019 World Bank Group. Accessed 29 August 2020.

Growing at a slower pace, world population is expected to reach 9.7 billion in 2050 and could peak at nearly 11 billion around 2100. 2019. United Nations Department of Economic and Social Affairs. Accessed 29 August 2020.

1. Paul R. Erlich, prologue to *The Population Bomb*, (1968; repr., New York: Ballantine, 1970).

## Contributors and Attributions

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

- **45.5: Human Population Growth** by OpenStax, is licensed CC BY by

  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).

- **6.4: A global transition** by Clarence Lehman, Shelby Loberg, & Adam Clark, is licensed CC BY-NC
- The Human Population from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Human Demography from *Introduction to Environmental Science*, 2nd edition by Caralyn Zehnder et al. (licensed under CC-BY-NC-SA).
- Human Population Growth from *General Biology* by OpenStax (licensed under CC-BY)
- The Industrialization of Nature: A Modern History (1500 to the present) from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors. Download for free at CNX. (licensed under CC-BY)
- The Human Population from *Biology* by John W. Kimball (licensed under CC-BY)

---

This page titled 14.3: Human Population Growth is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **45.5: Human Population Growth** by OpenStax is licensed CC BY 4.0.
- **6.22: Demographic Transition** by CK-12 Foundation is licensed CK-12. Original source: http://www.ck12.org/book/CK-12-Biology-Concepts.
- **6.4: A global transition** by Clarence Lehman, Shelby Loberg, & Adam Clark is licensed CC BY-NC 4.0. Original source: https://conservancy.umn.edu/handle/11299/204551.

# CHAPTER OVERVIEW

## 15: Community and Ecosystem Ecology


15.1: Community Ecology- Species Interactions

15.2: Community Structure

15.3: Community Dynamics

15.4: Ecology of Ecosystems

15.5: Energy Flow Through Ecosystems


---


This page titled 15: Community and Ecosystem Ecology is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

# 15.1: Community Ecology- Species Interactions

Populations rarely, if ever, live in isolation from populations of other species. In most cases, numerous species share a habitat. The interactions between these populations play a major role in regulating population growth and abundance. All populations occupying the same habitat form a community: populations inhabiting a specific area at the same time. The number of species occupying the same habitat and their relative abundance is known as species diversity. Areas with low diversity, such as the glaciers of Antarctica, still contain a wide variety of living things, whereas the diversity of tropical rainforests is so great that it cannot be counted. Ecology is studied at the community level to understand how species interact with each other and compete for the same resources.

## Competition

Resources are often limited within a habitat and multiple species may compete to obtain them. All species have an ecological **niche** in the a community, which describes how they acquire the resources they need, in what habitat they live, and how they interact with other species in the community (Figure 15.1.1). In othere words, the niche includes all the ways that the species interacts with the biotic and abiotic factors of the environment.

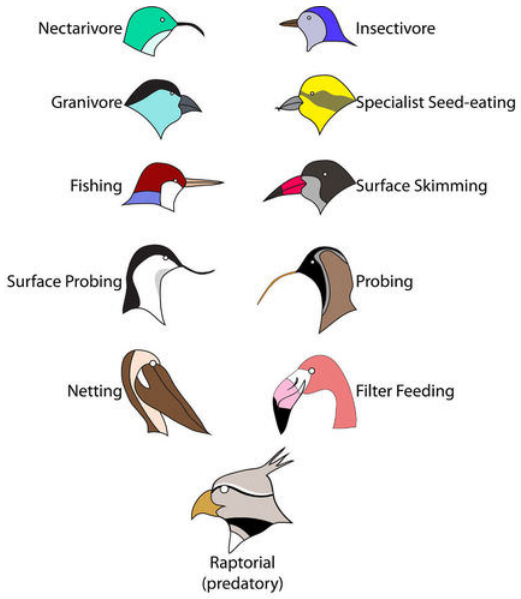



**Figure 15.1.1:** Bird Niches. Each of these species of birds has a beak that suits it for its niche. For example, the long slender beak of the nectarivore allows it to sip liquid nectar from flowers. The short sturdy beak of the granivore allows it to crush hard, tough grains.

**Competition** – occurs when the biological demand for an ecological resource exceeds the supply, causing organisms to interfere with each other. Plants, for example, often compete for access to limited supplies of sunlight, water, nutrients, and space. Animals may compete for food, nesting sites, mates, and other resources. **Intraspecific competition** occurs when individuals of the same species vie for access to resources, while **interspecific competition** occurs between species. Because use of resources by one individual takes those resources away from another individual we think of competition as an interaction that has a negative effect on both individuals or species in the interaction. For example, if flying squirrels and a species of bird are both using tree cavities for nests, having more squirrels is bad for the birds and having more birds is bad for the squirrels.

If a species is particularly effective at co-opting resources to its own benefit, it may displace other species, a phenomenon known as competitive displacement (or in extreme cases, competitive exclusion). This affects the presence and relative abundance of species in the community. For example, sugar maple (Acer saccharum) is a highly competitive tree in hardwood forests of eastern Canada. Where environmental conditions are well suited for this species, it can dominate mature stands. If large sugar maple trees are removed from a stand, perhaps by a selective timber harvest, other tree species (as well as small sugar maples) will benefit from the reduced competition and will grow more vigorously.

The **competitive exclusion principle** states that two species cannot occupy the same niche in a habitat indefinitely. In other words, different species cannot coexist forever in a community if they are competing for all the same resources. An example of this principle is shown in Figure **15.1.2**, with two protozoan species, *Paramecium aurelia* and *Paramecium caudatum*. When grown individually in the laboratory, they both thrive. But when they are placed together in the same test tube (habitat), *P. aurelia* outcompetes *P. caudatum* for food, leading to the latter's eventual extinction.

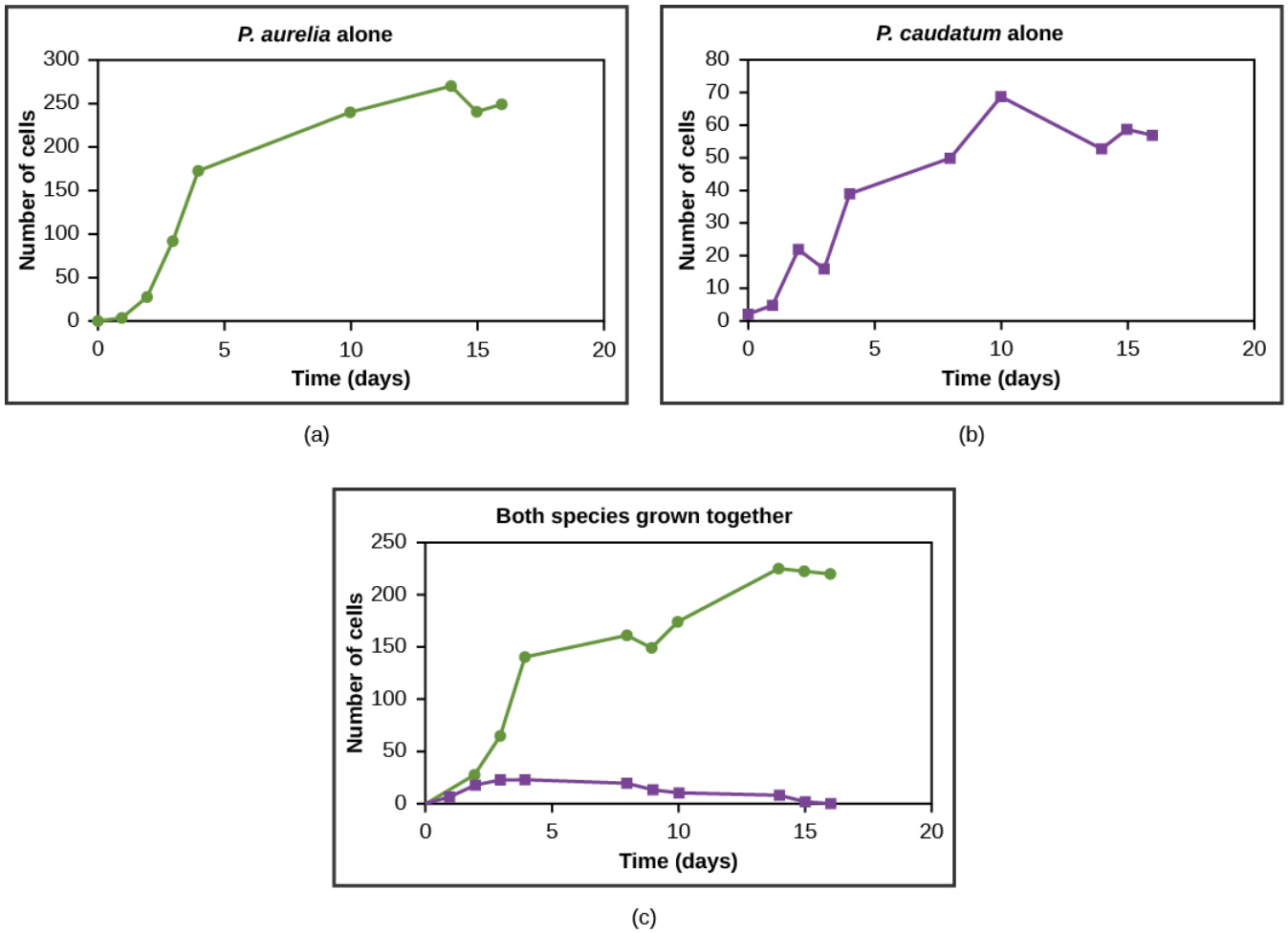



**Figure 15.1.2:** *Paramecium aurelia* and *Paramecium caudatum* grow well individually, but when they compete for the same resources, the *P. aurelia* outcompetes the *P. caudatum*.

This exclusion may be avoided if a population evolves to make use of a different resource, a different area of the habitat, or feeds during a different time of day, called resource partitioning. The two organisms are then said to occupy different microniches. These organisms coexist by minimizing direct competition.

## Facilitation or Symbiosis

Symbiotic relationships, or symbioses (plural), are close interactions between individuals of different species over an extended period of time which impact the abundance and distribution of the associating populations. Most scientists accept this definition, but some restrict the term to only those species that are mutualistic, where both individuals benefit from the interaction. In this discussion, the broader definition will be used.

A commensal relationship (**commensalism**) occurs when one species benefits from the close, prolonged interaction, while the other neither benefits nor is harmed. Many potential commensal relationships are difficult to identify because it is difficult to demonstrate that one partner is unaffected by the presence of the other. Birds nesting in trees provide an example of a commensal relationship (Figure **15.1.3**). The tree is not harmed by the presence of the nest among its branches. The nests are light and produce little strain on the structural integrity of the branch, and most of the leaves, which the tree uses to get energy by photosynthesis, are

above the nest so they are unaffected. The bird, on the other hand, benefits greatly. If the bird had to nest in the open, its eggs and young would be vulnerable to predators.



**Figure 15.1.3:** The southern masked-weaver bird is starting to make a nest in a tree in Zambezi Valley, Zambia. This is an example of a commensal relationship, in which one species (the bird) benefits, while the other (the tree) neither benefits nor is harmed. (credit: "Hanay"/Wikimedia Commons)

Another example or a commensal relationship involves the Little Blue Heron and the White Ibis, which are both wading birds. The Little Blue Heron catches more fish in the presence of the White Ibis, but the White Ibis is unaffected. Interestingly, Little Blue Herons attempt to catch fish more often in the presence of the species, but the success rate of their attempts does not change. Nevertheless, more frequent attempts still increases the total number of fish caught. The White Ibis may make fish more visible to Little Blue Herons, causing changes in their behavior (figure 15.1.4).



**Figure 15.1.4:** The Little Blue Heron (left) and the White Ibis (right) have a commensal relationship. Image by Russ (CC-BY).

A second type of symbiotic relationship is called **mutualism**, where two species benefit from their interaction. Some scientists believe that these are the only true examples of symbiosis. For example, termites have a mutualistic relationship with protozoa that live in the insect's gut (Figure 15.1.5a). The termite benefits from the ability of bacterial symbionts within the protozoa to digest cellulose. The termite itself cannot do this, and without the protozoa, it would not be able to obtain energy from its food (cellulose from the wood it chews and eats). The protozoa and the bacterial symbionts benefit by having a protective environment and a constant supply of food from the wood chewing actions of the termite.



a b

**Figure 15.1.5:** (a) Termites form a mutualistic relationship with symbiotic protozoa in their guts, which allow both organisms to obtain energy from the cellulose the termite consumes. (b) Lichen is a fungus that has symbiotic photosynthetic algae living inside its cells. (credit a: modification of work by Scott Bauer, USDA; credit b: modification of work by Cory Zanker)

Lichens have a mutualistic relationship between fungus and photosynthetic algae or bacteria (Figure 15.1.5b). As these symbionts grow together, the glucose produced by the algae provides nourishment for both organisms, whereas the physical structure of the lichen protects the algae from the elements and makes certain nutrients in the atmosphere more available to the algae. Another set of mutualisms are pollinators, such as bees, butterflies, and hummingbirds. Polinators benefit because they eat the collect pollen and/or nectar that they collect from flowers. The plants also benefit because their pollen is dispersed to other plants, allowing them to reproduce. Both the pollinators and the plants benefit, demonstrating a mutualism (figure 15.1.6).



**Figure 15.1.6:** A male Broad-Tailed Hummingbird visits a scarlet gilia flower at the Rocky Mountain Biological Laboratory. The hummingbird gains food (nectar) while aiding the gilia flower with reproduction. Image by David W. Inouye (CC-BY).

## Predation and Herbivory

Perhaps the classical example of species interaction is the predator-prey relationship. **Predation** occurs when one species (the **predator**) kills and eats multiple **prey** over its lifetime. Nature shows on television highlight the drama of one living organism killing another. Populations of predators and prey in a community are not constant over time: in most cases, they vary in cycles that appear to be related. Population sizes of predators and prey in a community are not constant over time, and they may vary in cycles that appear to be related.



**Figure 15.1.7:** The Canada lynx is an example of a predator. Image by Michael Zahra (CC-BY-SA).



**Figure 15.1.8:** The snowshoe hare is the prey of the Canada lynx. Image by the National Park Service (public domain).

The most often cited example of predator-prey dynamics is seen in the cycling of the lynx (predator, Figure 15.1.7) and the snowshoe hare (prey, Figure 15.1.8), using nearly 200 year-old trapping data from North American forests (Figure 15.1.8). This cycle of predator and prey lasts approximately 10 years, with the predator population lagging 1–2 years behind that of the prey population. As the hare numbers increase, there is more food available for the lynx, allowing the lynx population to increase as well. When the lynx population grows to a threshold level, however, they kill so many hares that hare population begins to decline, followed by a decline in the lynx population because of scarcity of food. When the lynx population is low, the hare population size begins to increase due, at least in part, to low predation pressure, starting the cycle anew.

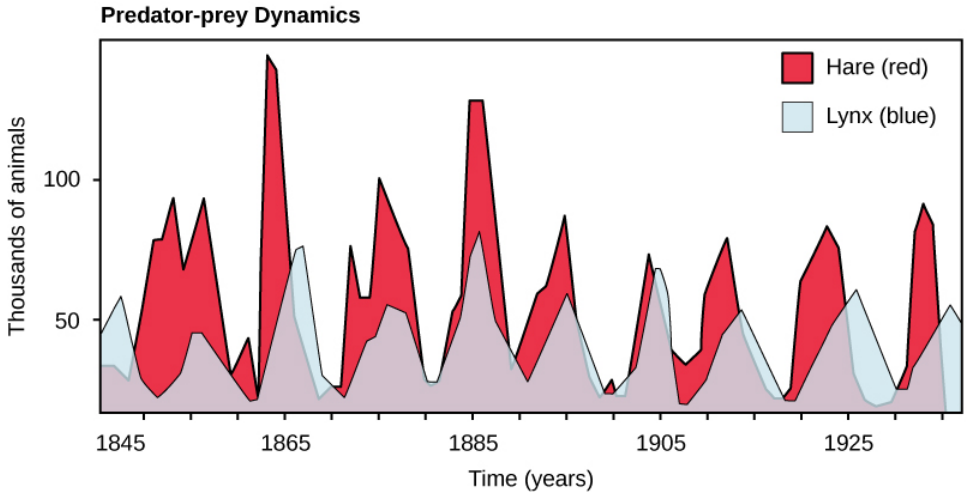



**Figure 15.1.8:** The cycling of lynx and snowshoe hare populations in Northern Ontario is an example of predator-prey dynamics.

The idea that the population cycling of the two species is entirely controlled by predation models has come under question. More recent studies have pointed to undefined density-dependent factors as being important in the cycling, in addition to predation. One possibility is that the cycling is inherent in the hare population due to density-dependent effects such as lower fecundity (maternal stress) caused by crowding when the hare population gets too dense. The hare cycling would then induce the cycling of the lynx because it is the lynxes' major food source. The more we study communities, the more complexities we find, allowing ecologists to derive more accurate and sophisticated models of population dynamics.

Herbivory describes the consumption of plants by insects and other animals, and it is another interspecific relationship that affects populations. Unlike animals, most plants cannot outrun predators or use mimicry to hide from hungry animals. Some plants have developed mechanisms to defend against herbivory. Other species have developed mutualistic relationships; for example, herbivory provides a mechanism of seed distribution that aids in plant reproduction.

## Adaptations of Predators

Predators have a variety of adaptations to catch and consume prey, and the specific adaptations depend on the predator. For example, raptors, such as owls and hawks, have hooked beaks for tearing flesh and talons for grabbing prey. Mammal predators often have sharp teeth and claws. Some predators can run quickly to chase prey, and others "sit and wait", lunging forward when prey pass them. Some predators, like rattlesnakes and tarantulas, subdue their prey by injecting them with venom (figure 15.1.9). Predators often have large eyes located forward (like a wolf) rather than eyes spaced far apart on the sides of the head (like a sheep). Forward-facing eyes allow for depth perception, which is key to tracking prey. In contrast, peripheral vision is expanded when eyes are located to the sides, and this helps prey identify threats.



**Figure 15.1.9:** This Asian birdeater tarantula uses its fangs to inject its prey with poisonous venom. Image by Matt Reinbold (CC-BY-SA).

## Defense Mechanisms against Predation and Herbivory

The study of communities must consider evolutionary forces that act on the members of the various populations contained within it. Species are not static, but slowly changing and adapting to their environment by natural selection and other evolutionary forces. Species have evolved numerous mechanisms to escape predation and herbivory. These defenses may be mechanical, chemical, physical, or behavioral.

Prey evolve mechanical, chemical, physical, or behavioral defenses against predators. Some prey have armor (such as a turtle shell or the bony plates protecting armadillos), a **mechanical defense** that reduces predation by discouraging physical contact. Many animals produce or obtain **chemical defenses** from plants and store them to prevent predation. Other species use their body shape and coloration as **camouflage** to avoid being detected by predators, an example of a **physical defense**. The tropical walking stick is an insect with the coloration and body shape of a twig, which makes it very hard to see when it is stationary against a background of real twigs (figure 15.1.10-a). In another example, the chameleon can change its color to match its surroundings (figure 15.1.10-b).




**Figure 15.1.10:** (a) The tropical walking stick and (b) the chameleon use their body shape and/or coloration to prevent detection by predators. (credit a: modification of work by Linda Tanner; credit b: modification of work by Frank Vassen)

**Mechanical defenses**, such as the presence of thorns on plants or the hard shell on turtles, discourage animal predation and herbivory by causing physical pain to the predator or by physically preventing the predator from being able to eat the prey (Figure 15.1.11. **Chemical defenses** are produced by many animals as well as plants, such as the foxglove which is extremely toxic when eaten. shows some organisms' defenses against predation and herbivory.



(a)



(b)



(c)



(d)

**Figure 15.1.11:** The (a) honey locust tree (*Gleditsia triacanthos*) uses thorns, a mechanical defense, against herbivores, while the (b) Florida red-bellied turtle (*Pseudemys nelsoni*) uses its shell as a mechanical defense against predators. (c) Foxglove (*Digitalis* sp.) uses a chemical defense: toxins produced by the plant can cause nausea, vomiting, hallucinations, convulsions, or death when consumed. (d) The North American millipede (*Narceus americanus*) uses both mechanical and chemical defenses: when threatened, the millipede curls into a defensive ball and produces a noxious substance that irritates eyes and skin. (credit a: modification of work by Huw Williams; credit b: modification of work by "JamieS93"/Flickr; credit c: modification of work by Philip Jägenstedt; credit d: modification of work by Cory Zanker)

Many species use their body shape and coloration to avoid being detected by predators, an example of a **physical defense**. The tropical walking stick is an insect with the coloration and body shape of a twig which makes it very hard to see when stationary against a background of real twigs (Figure 15.1.12a). In another example, the chameleon can change its color to match its surroundings (Figure 15.1.12b). Both of these are examples of **camouflage**, or avoiding detection by blending in with the background.



a b

**Figure 15.1.12:** (a) The tropical walking stick and (b) the chameleon use body shape and/or coloration to prevent detection by predators. (credit a: modification of work by Linda Tanner; credit b: modification of work by Frank Vassen)

Some species use coloration as a way of warning predators that they are not good to eat. For example, the cinnabar moth caterpillar, the fire-bellied toad, and many species of beetle have bright colors that warn of a foul taste, the presence of toxic chemical, and/or the ability to sting or bite, respectively. Predators that ignore this coloration and eat the organisms will experience their unpleasant taste or presence of toxic chemicals and learn not to eat them in the future. This type of defensive mechanism is called **aposematic coloration**, or warning coloration (Figure 15.1.13). Warning coloration only works if a predator uses eyesight to locate prey and can learn—a naïve predator must experience the negative consequences of eating one before it will avoid other similarly colored individuals. For example, the monarch butterfly caterpillar sequesters poisons from its food (plants and milkweeds) to make itself poisonous or distasteful to potential predators. The caterpillar is bright yellow and black to advertise its toxicity. The caterpillar is also able to pass the sequestered toxins on to the adult monarch, which is also dramatically colored black and red as a warning to potential predators.



(a) (b)

**Figure 15.1.13:** (a) The strawberry poison dart frog (*Oophaga pumilio*) uses aposematic coloration to warn predators that it is toxic, while the (b) striped skunk (*Mephitis mephitis*) uses aposematic coloration to warn predators of the unpleasant odor it produces. (credit a: modification of work by Jay Iwasaki; credit b: modification of work by Dan Dzurisin)

While some predators learn to avoid eating certain potential prey because of their coloration, other species have evolved mechanisms to mimic this coloration to avoid being eaten, even though they themselves may not be unpleasant to eat or contain toxic chemicals. In **Batesian mimicry**, a harmless species imitates the warning coloration of a harmful one. Assuming they share the same predators, this coloration then protects the harmless ones, even though they do not have the same level of physical or chemical defenses against predation as the organism they mimic. Many insect species mimic the coloration of wasps or bees, which are stinging, venomous insects, thereby discouraging predation (Figure 15.1.14).



a b

**Figure 15.1.14:** Batesian mimicry occurs when a harmless species mimics the coloration of a harmful species, as is seen with the (a) bumblebee and (b) bee-like robber fly. (credit a, b: modification of work by Cory Zanker)

In **Müllerian mimicry**, multiple species share the same warning coloration, but all of them actually have defenses. Figure 15.1.15 shows a variety of foul-tasting butterflies with similar coloration.



Figure 15.1.15: Several unpleasant-tasting *Heliconius* butterfly species share a similar color pattern with better-tasting varieties, an example of Müllerian mimicry. (credit: Joron M, Papa R, Beltrán M, Chamberlain N, Mavárez J, et al.)

Running from predators, hiding, and playing dead are examples of **behavioral defenses**. Some prey also exhibit behaviors that threaten predators. For example, snapping turtles stretch their legs to appear larger and snap aggressively at predators.

## Parasitism

A parasite is an organism that lives in or on another living organism and derives nutrients from it. In this relationship, the parasite benefits, but the organism being fed upon, the host, is harmed. The host is usually weakened by the parasite as it siphons resources the host would normally use to maintain itself. The parasite, however, is unlikely to kill the host, especially not quickly, because this would allow no time for the organism to complete its reproductive cycle by spreading to another host.

The reproductive cycles of parasites are often very complex, sometimes requiring more than one host species. A tapeworm is a parasite that causes disease in humans when contaminated, undercooked meat such as pork, fish, or beef is consumed (Figure

15.1.16). The tapeworm can live inside the intestine of the host for several years, benefiting from the food the host is bringing into its gut by eating, and may grow to be over 50 ft long by adding segments. The parasite moves from species to species in a cycle, making two hosts necessary to complete its life cycle. Another common parasite is *Plasmodium falciparum*, the protozoan cause of malaria, a significant disease in many parts of the world. Living in human liver and red blood cells, the organism reproduces asexually in the gut of blood-feeding mosquitoes to complete its life cycle. Thus malaria is spread from human to human by mosquitoes, one of many arthropod-borne infectious diseases.

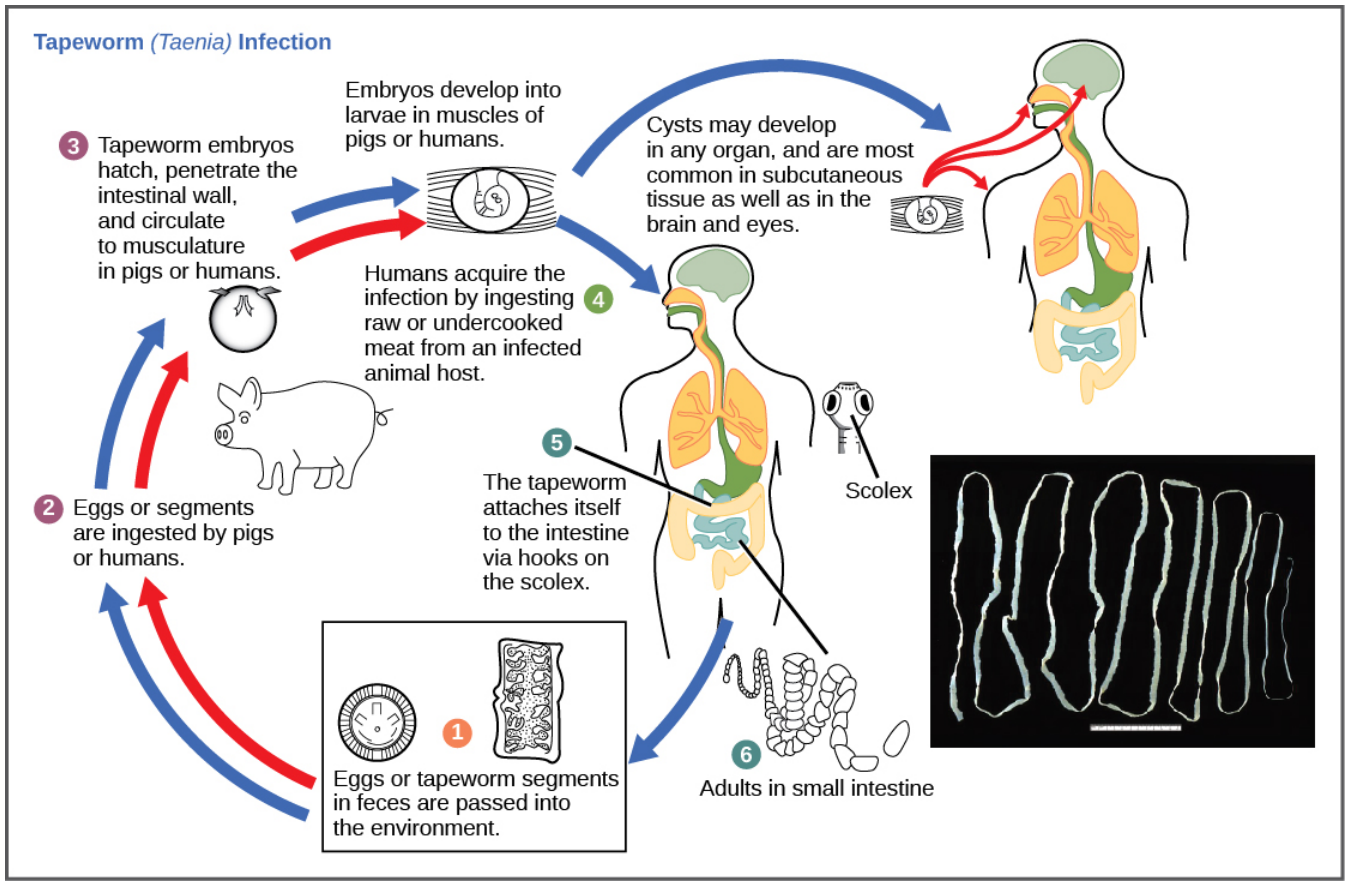



Figure 15.1.16: This diagram shows the life cycle of a pork tapeworm (*Taenia solium*), a human worm parasite. (credit: modification of work by CDC)

**Disease** – is a pathological relationship in which the health of plants or animals suffers from an infestation of another species, usually a microbe. Virulent diseases can cause enormous changes in the composition of ecological communities. In the early 1900s, the American chestnut (Castanea dentata) was afflicted by chestnut blight (Endothia parasitica), an introduced fungal pathogen. Because chestnuts have little immunity to this disease, the species was virtually eliminated from the forests of eastern North America by the 1950s. This change released other tree species from competition with the previously dominant chestnut, and they quickly filled gaps in the canopy created by its demise.

## Contributors and Attributions

Modified by Kyle Whittinghill and Melissa Ha from

- **45.6: Community Ecology** by OpenStax, is licensed CC BY by

  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).

- Community Ecology from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)

- **6.2: Ecosystems** by CK-12: Biology Concepts, is licensed CC BY-NC
- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.
- Ecology: From Individuals to the Biosphere from *Environmental Science: A Canadian Perspective* by Bill Freedman

---

This page titled 15.1: Community Ecology- Species Interactions is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **45.6: Community Ecology** by OpenStax is licensed CC BY 4.0.
- **6.2: Ecosystems** by CK-12 Foundation is licensed CK-12. Original source: http://www.ck12.org/book/CK-12-Biology-Concepts.

# 15.2: Community Structure

## Characteristics of Communities

Communities are complex entities that can be characterized by their structure (the types and numbers of species present) and dynamics (how communities change over time). Understanding community structure and dynamics enables community ecologists to manage ecosystems more effectively.

### Foundation Species

Foundation species are considered the “base” or “bedrock” of a community, having the greatest influence on its overall structure. They are usually the primary producers: organisms that bring most of the energy into the community. Kelp, brown algae, is a foundation species, forming the basis of the kelp forests off the coast of California.

Foundation species may physically modify the environment to produce and maintain habitats that benefit the other organisms that use them. An example is the photosynthetic corals of the coral reef (Figure **15.2.1**). Corals themselves are not photosynthetic, but harbor symbionts within their body tissues (dinoflagellates called zooxanthellae) that perform photosynthesis; this is another example of a mutualism. The exoskeletons of living and dead coral make up most of the reef structure, which protects many other species from waves and ocean currents.

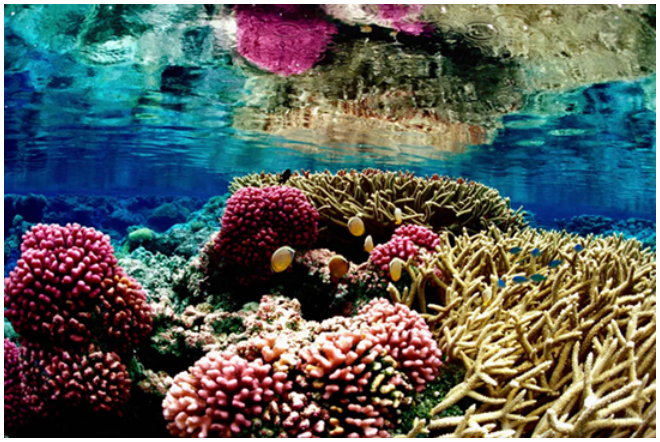


Figure **15.2.1**: Coral is the foundation species of coral reef ecosystems. (credit: Jim E. Maragos, USFWS)

### Biodiversity, Species Richness, and Relative Species Abundance

Biodiversity describes a community’s biological complexity: it is measured by the number of different species (species richness) in a particular area and their relative abundance (species evenness). The area in question could be a habitat, a biome, or the entire biosphere. Species richness is the term that is used to describe the number of species living in a habitat or biome. Species richness varies across the globe (Figure **15.2.2**). One factor in determining species richness is latitude, with the greatest species richness occurring in ecosystems near the equator, which often have warmer temperatures, large amounts of rainfall, and low seasonality. The lowest species richness occurs near the poles, which are much colder, drier, and thus less conducive to life in Geologic time (time since glaciations). The predictability of climate or productivity is also an important factor. Other factors influence species richness as well. For example, the study of island biogeography attempts to explain the relatively high species richness found in certain isolated island chains, including the Galápagos Islands that inspired the young Darwin. Relative species abundance is the number of individuals in a species relative to the total number of individuals in all species within a habitat, ecosystem, or biome. Foundation species often have the highest relative abundance of species.

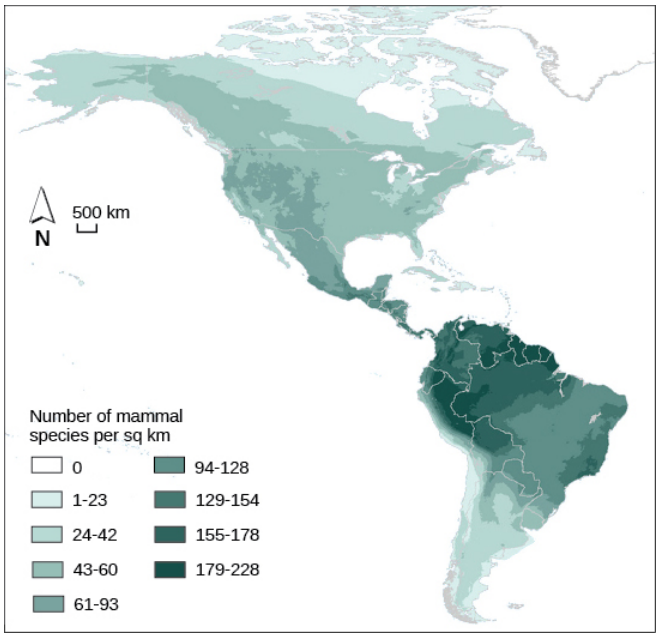



Figure **15.2.2**: The greatest species richness for mammals in North and South America is associated with the equatorial latitudes. (credit: modification of work by NASA, CIESIN, Columbia University)

## Keystone Species

A keystone species is one whose presence is key to maintaining biodiversity within an ecosystem and to upholding an ecological community's structure. The intertidal sea star, *Pisaster ochraceus*, of the northwestern United States is a keystone species (Figure **15.2.3**). Studies have shown that when this organism is removed from communities, populations of their natural prey (mussels) increase, completely altering the species composition and reducing biodiversity. Another keystone species is the banded tetra, a fish in tropical streams, which supplies nearly all of the phosphorus, a necessary inorganic nutrient, to the rest of the community. If these fish were to become extinct, the community would be greatly affected.



Figure **15.2.3**: The *Pisaster ochraceus* sea star is a keystone species. (credit: Jerry Kirkhart)

### Everyday Connection: Invasive Species

Invasive species are non-native organisms that, when introduced to an area out of their native range, threaten the ecosystem balance of that habitat. Many such species exist in the United States, as shown in Figure **15.2.4**. Whether enjoying a forest hike, taking a summer boat trip, or simply walking down an urban street, you have likely encountered an invasive species.

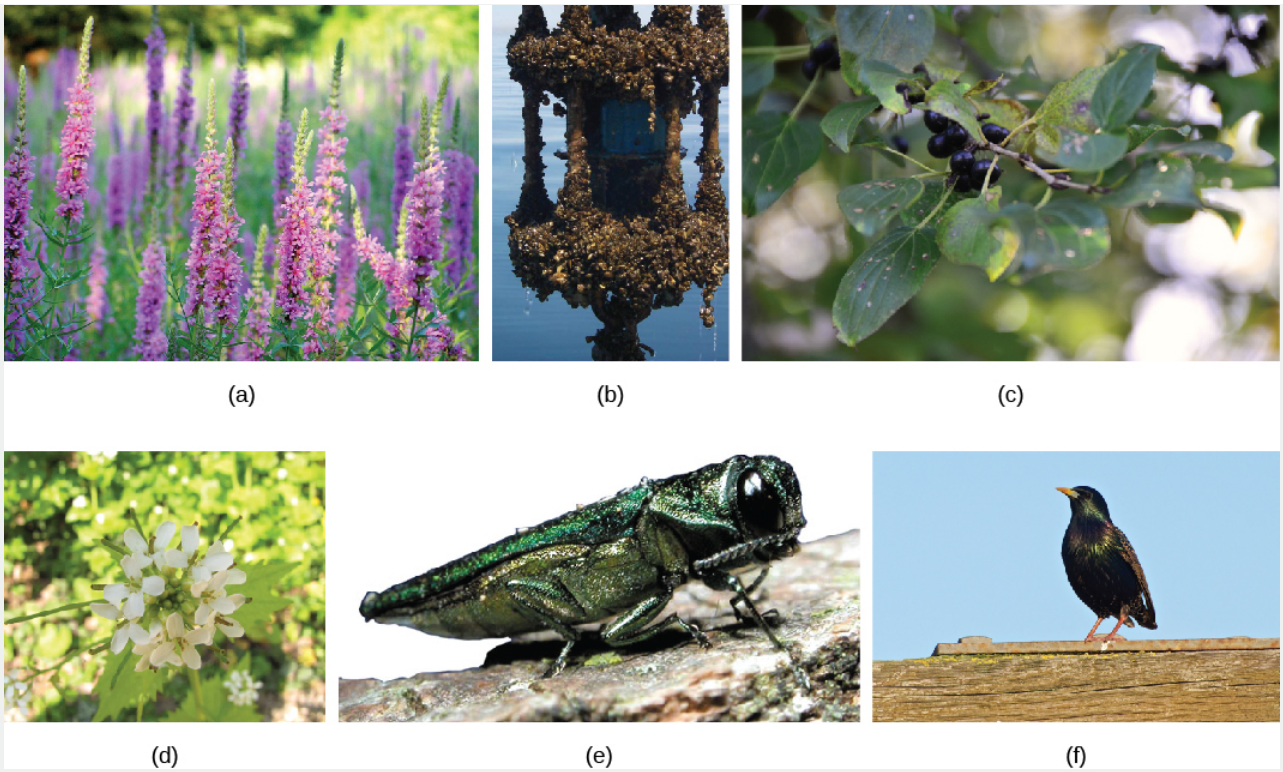



Figure **15.2.4**: In the United States, invasive species like (a) purple loosestrife (*Lythrum salicaria*) and the (b) zebra mussel (*Dreissena polymorpha*) threaten certain aquatic ecosystems. Some forests are threatened by the spread of (c) common buckthorn (*Rhamnus cathartica*), (d) garlic mustard (*Alliaria petiolata*), and (e) the emerald ash borer (*Agrilus planipennis*). The (f) European starling (*Sturnus vulgaris*) may compete with native bird species for nest holes. (credit a: modification of work by Liz West; credit b: modification of work by M. McCormick, NOAA; credit c: modification of work by E. Dronkert; credit d: modification of work by Dan Davison; credit e: modification of work by USDA; credit f: modification of work by Don DeBold)

One of the many recent proliferations of an invasive species concerns the growth of Asian carp populations. Asian carp were introduced to the United States in the 1970s by fisheries and sewage treatment facilities that used the fish's excellent filter feeding capabilities to clean their ponds of excess plankton. Some of the fish escaped, however, and by the 1980s they had colonized many waterways of the Mississippi River basin, including the Illinois and Missouri Rivers.

Voracious eaters and rapid reproducers, Asian carp may outcompete native species for food, potentially leading to their extinction. For example, black carp are voracious eaters of native mussels and snails, limiting this food source for native fish species. Silver carp eat plankton that native mussels and snails feed on, reducing this food source by a different alteration of the food web. In some areas of the Mississippi River, Asian carp species have become the most predominant, effectively outcompeting native fishes for habitat. In some parts of the Illinois River, Asian carp constitute 95 percent of the community's biomass. Although edible, the fish is bony and not a desired food in the United States. Moreover, their presence threatens the native fish and fisheries of the Great Lakes, which are important to local economies and recreational anglers. Asian carp have even injured humans. The fish, frightened by the sound of approaching motorboats, thrust themselves into the air, often landing in the boat or directly hitting the boaters.

The Great Lakes and their prized salmon and lake trout fisheries are also being threatened by these invasive fish. Asian carp have already colonized rivers and canals that lead into Lake Michigan. One infested waterway of particular importance is the Chicago Sanitary and Ship Channel, the major supply waterway linking the Great Lakes to the Mississippi River. To prevent the Asian carp from leaving the canal, a series of electric barriers have been successfully used to discourage their migration; however, the threat is significant enough that several states and Canada have sued to have the Chicago channel permanently cut off from Lake Michigan. Local and national politicians have weighed in on how to solve the problem, but no one knows whether the Asian carp will ultimately be considered a nuisance, like other invasive species such as the water hyacinth and zebra mussel, or whether it will be the destroyer of the largest freshwater fishery of the world.

The issues associated with Asian carp show how population and community ecology, fisheries management, and politics intersect on issues of vital importance to the human food supply and economy. Socio-political issues like this make extensive use of the sciences of population ecology (the study of members of a particular species occupying a particular area known as a habitat) and community ecology (the study of the interaction of all species within a habitat).

## Spatial Variation

**Spatial variation** of the environment reflects the fact that conditions are always changing from place to place, and sometimes extremely so. These spatial variations influence the character of ecological communities, in ways that may be gradual or more rapid:

- Gradual changes environmental conditions are associated with varying altitude on a mountain, differences of climate over large distances across continents, and other relatively continuous gradients. This type of spatial change is reflected in gradual variations of communities because individual species have different but overlapping tolerances and requirements of environmental conditions. These biological differences result in overlaps of the distributions of species, which can make it difficult for ecologists to determine the locations of boundaries (ecotones) between types of communities.
- Rapid changes in environmental conditions occur at sharp boundaries between different kinds of soil or bedrock, at interfaces between aquatic and terrestrial habitats, and in places affected by disturbance. The latter influence can occur, for instance, between a burned and unburned tract of forest, or between an ecological reserve and its surrounding area, which may be affected by agriculture or forestry. Relatively discrete changes in environmental conditions favour large differences in community types, with distinct boundaries between them.

## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- **45.6: Community Ecology** by OpenStax, is licensed CC BY by

  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).
- Community Ecology from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Ecology: From Individuals to the Biosphere from *Environmental Science: A Canadian Perspective* by Bill Freedman

---

This page titled 15.2: Community Structure is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **45.6: Community Ecology** by OpenStax is licensed CC BY 4.0.

# 15.3: Community Dynamics

## Community Dynamics

Community dynamics are the changes in community structure and composition over time. Sometimes these changes are induced by environmental disturbances such as volcanoes, earthquakes, storms, fires, and climate change. Communities with a stable structure are said to be at equilibrium. Following a disturbance, the community may or may not return to the equilibrium state.

Succession describes the sequential appearance and disappearance of species in a community over time. In primary succession, newly exposed or newly formed land is colonized by living things; in secondary succession, part of an ecosystem is disturbed and remnants of the previous community remain.

### Primary Succession and Pioneer Species

Primary succession occurs when new land is formed or rock is exposed: for example, following the eruption of volcanoes, such as those on the Big Island of Hawaii. As lava flows into the ocean, new land is continually being formed. On the Big Island, approximately 32 acres of land is added each year. First, weathering and other natural forces break down the substrate enough for the establishment of certain hearty plants and lichens with few soil requirements, known as pioneer species (Figure **15.3.1**). These species help to further break down the mineral rich lava into soil where other, less hardy species will grow and eventually replace the pioneer species. In addition, as these early species grow and die, they add to an ever-growing layer of decomposing organic material and contribute to soil formation. Over time the area will reach an equilibrium state, with a set of organisms quite different from the pioneer species.



Figure **15.3.1**: During primary succession in lava on Maui, Hawaii, succulent plants are the pioneer species. (credit: Forest and Kim Starr)

### Secondary succession

A classic example of secondary succession occurs in oak and hickory forests cleared by wildfire (Figure **15.3.2**). Wildfires will burn most vegetation and kill those animals unable to flee the area. Their nutrients, however, are returned to the ground in the form of ash. Thus, even when areas are devoid of life due to severe fires, the area will soon be ready for new life to take hold.

Before the fire, the vegetation was dominated by tall trees with access to the major plant energy resource: sunlight. Their height gave them access to sunlight while also shading the ground and other low-lying species. After the fire, though, these trees are no longer dominant. Thus, the first plants to grow back are usually annual plants followed within a few years by quickly growing and spreading grasses and other pioneer species. Due to, at least in part, changes in the environment brought on by the growth of the grasses and other species, over many years, shrubs will emerge along with small pine, oak, and hickory trees. These organisms are called intermediate species. Eventually, over 150 years, the forest will reach its equilibrium point where species composition is no longer changing and resembles the community before the fire. This equilibrium state is referred to as the climax community, which will remain stable until the next disturbance.

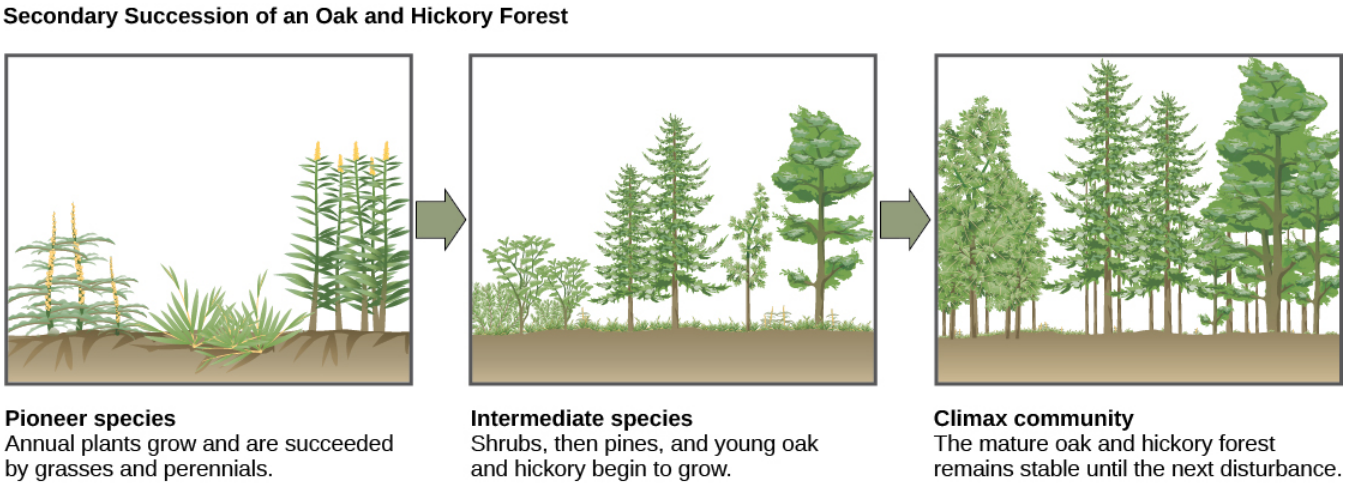



**Figure 15.3.2:** Secondary succession is shown in an oak and hickory forest after a forest fire.

## Disturbance

**Disturbance** is an event of destruction of some part of a community, an occurrence that is followed by a sometimes pronged period of ecological recovery called succession. All communities are dynamic, changing over time in their species composition and functional attributes (such as productivity, decomposition, and nutrient cycling). However, the rate of change depends on the stability of environmental conditions, which is greatest in communities that are close to the end-point of a succession. In contrast, the most dynamic communities are associated with the younger stages of succession. Disturbances can occur on two spatial scales.

- **Stand-replacing disturbances** are caused by wildfire, a disease epidemic, clear-cutting, and other cataclysmic events (Figure 15.3.3). This kind of disturbance is extensive and results in the immediate replacement of a community with a different one, followed by a period of successional recovery. Over time, succession may regenerate a community similar to what existed before the disruption, or a different one may result. The younger stages of a sere (successional sequence) are especially dynamic in terms of community change.
- **Microdisturbances** are local disruptions that affect small areas within an otherwise intact community. A microdisturbance may, for instance, be associated with the death of an individual large tree, which results in a gap in the canopy, below which community change is relatively dynamic as species compete to take advantage of the additional sunlight. Similarly, the death of an individual coral head represents a microdisturbance within a tropical reef community. Although ecological changes are dynamic within a gap created by a recent microdisturbance, at the stand level the community is relatively stable. Gap-phase community dynamics occur in all ecosystems but are especially important during later stages of succession, such as in older-growth forests.

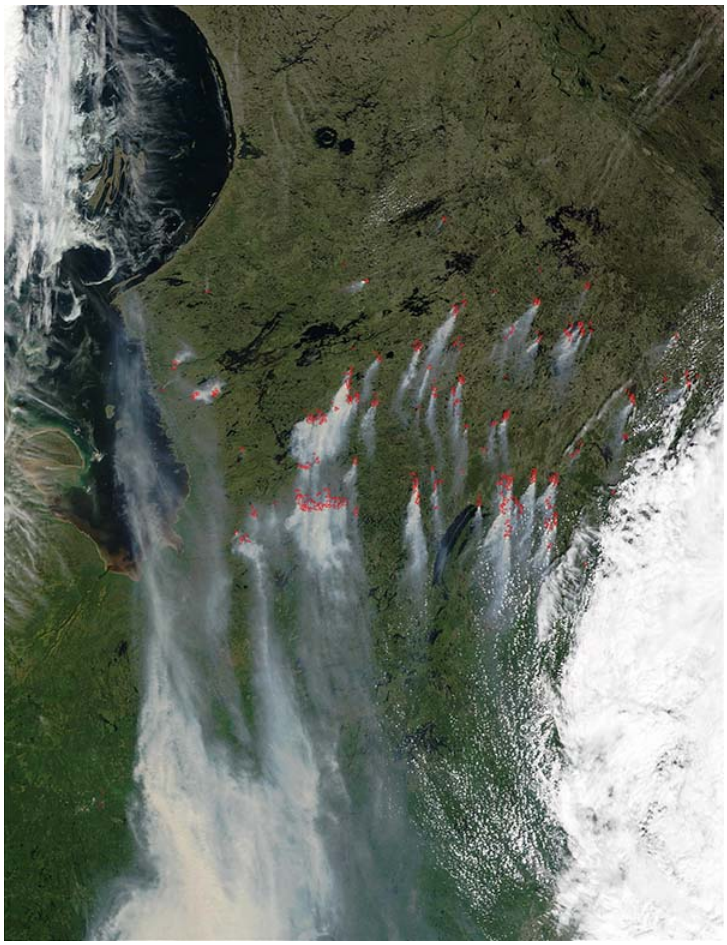


**Figure 15.3.3:** Ecosystems are occasionally subjected to catastrophic disturbances, such as these forest fires in 2002 in the boreal forest of northern Quebec. The individual fires are marked with a red dot, and their smoke plumes are blowing to the south. The large white mass at the bottom right is cloud cover. Source: NASA photo ID 751339; https://www.dvidshub.net/image/751339/fires-quebec-canada-send-smoke-us-natural-hazards#.VOS_vXUtHIU

# Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- **45.6: Community Ecology** by OpenStax, is licensed CC BY by

  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).

- Community Ecology from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Ecology: From Individuals to the Biosphere from *Environmental Science: A Canadian Perspective* by Bill Freedman

---

This page titled 15.3: Community Dynamics is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **45.6: Community Ecology** by OpenStax is licensed CC BY 4.0.

# 15.4: Ecology of Ecosystems

Life in an ecosystem is often about competition for limited resources, a characteristic of the theory of natural selection. Competition in communities (all living things within specific habitats) is observed both within species and among different species. The resources for which organisms compete include organic material from living or previously living organisms, sunlight, and mineral nutrients, which provide the energy for living processes and the matter to make up organisms' physical structures. Other critical factors influencing community dynamics are the components of its physical and geographic environment: a habitat's latitude, amount of rainfall, topography (elevation), and available species. These are all important environmental variables that determine which organisms can exist within a particular area.

An ecosystem is a community of living organisms and their interactions with their abiotic (non-living) environment. Ecosystems can be small, such as the tide pools found near the rocky shores of many oceans, or large, such as the Amazon Rainforest in Brazil (Figure 15.4.1).



(a) (b)

**Figure 15.4.1:** A (a) tidal pool ecosystem in Matinicus Island in Maine is a small ecosystem, while the (b) Amazon Rainforest in Brazil is a large ecosystem. (credit a: modification of work by "takomabibelot"/Flickr; credit b: modification of work by Ivan Mlinaric)

There are three broad categories of ecosystems based on their general environment: freshwater, ocean water, and terrestrial. Within these broad categories are individual ecosystem types based on the organisms present and the type of environmental habitat.

Ocean ecosystems are the most common, comprising 75 percent of the Earth's surface and consisting of three basic types: shallow ocean, deep ocean water, and deep ocean surfaces (the low depth areas of the deep oceans). The shallow ocean ecosystems include extremely biodiverse coral reef ecosystems, and the deep ocean surface is known for its large numbers of plankton and krill (small crustaceans) that support it. These two environments are especially important to aerobic respirators worldwide as the phytoplankton perform 40 percent of all photosynthesis on Earth. Although not as diverse as the other two, deep ocean ecosystems contain a wide variety of marine organisms. Such ecosystems exist even at the bottom of the ocean where light is unable to penetrate through the water.

Freshwater ecosystems are the rarest, occurring on only 1.8 percent of the Earth's surface. Lakes, rivers, streams, and springs comprise these systems; they are quite diverse, and they support a variety of fish, amphibians, reptiles, insects, phytoplankton, fungi, and bacteria.

Terrestrial ecosystems, also known for their diversity, are grouped into large categories called biomes, such as tropical rain forests, savannas, deserts, coniferous forests, deciduous forests, and tundra. Grouping these ecosystems into just a few biome categories obscures the great diversity of the individual ecosystems within them. For example, there is great variation in desert vegetation: the

saguaro cacti and other plant life in the Sonoran Desert, in the United States, are relatively abundant compared to the desolate rocky desert of Boa Vista, an island off the coast of Western Africa (Figure **15.4.2**).



(a) (b)

**Figure 15.4.2:** Desert ecosystems, like all ecosystems, can vary greatly. The desert in (a) Saguaro National Park, Arizona, has abundant plant life, while the rocky desert of (b) Boa Vista island, Cape Verde, Africa, is devoid of plant life. (credit a: modification of work by Jay Galvin; credit b: modification of work by Ingo Wölbern)

Ecosystems are complex with many interacting parts. They are routinely exposed to various disturbances, or changes in the environment that effect their compositions: yearly variations in rainfall and temperature and the slower processes of plant growth, which may take several years. Many of these disturbances are a result of natural processes. For example, when lightning causes a forest fire and destroys part of a forest ecosystem, the ground is eventually populated by grasses, then by bushes and shrubs, and later by mature trees, restoring the forest to its former state. The impact of environmental disturbances caused by human activities is as important as the changes wrought by natural processes. Human agricultural practices, air pollution, acid rain, global deforestation, overfishing, eutrophication, oil spills, and illegal dumping on land and into the ocean are all issues of concern to conservationists.

Equilibrium is the steady state of an ecosystem where all organisms are in balance with their environment and with each other. In ecology, two parameters are used to measure changes in ecosystems: resistance and resilience. The ability of an ecosystem to remain at equilibrium in spite of disturbances is called resistance. The speed at which an ecosystem recovers equilibrium after being disturbed, called its resilience. Ecosystem resistance and resilience are especially important when considering human impact. The nature of an ecosystem may change to such a degree that it can lose its resilience entirely. This process can lead to the complete destruction or irreversible altering of the ecosystem.

### Evolution Connection: Three-spined Stickleback

It is well established by the theory of natural selection that changes in the environment play a major role in the evolution of species within an ecosystem. However, little is known about how the evolution of species within an ecosystem can alter the ecosystem environment. In 2009, Dr. Luke Harmon, from the University of Idaho in Moscow, published a paper that for the first time showed that the evolution of organisms into subspecies can have direct effects on their ecosystem environment.[1]

The three-spines stickleback (*Gasterosteus aculeatus*) is a freshwater fish that evolved from a saltwater fish to live in freshwater lakes about 10,000 years ago, which is considered a recent development in evolutionary time (Figure **15.4.3**). Over the last 10,000 years, these freshwater fish then became isolated from each other in different lakes. Depending on which lake population was studied, findings showed that these sticklebacks then either remained as one species or evolved into two species. The divergence of species was made possible by their use of different areas of the pond for feeding called micro niches.

Dr. Harmon and his team created artificial pond microcosms in 250-gallon tanks and added muck from freshwater ponds as a source of zooplankton and other invertebrates to sustain the fish. In different experimental tanks they introduced one species of

stickleback from either a single-species or double-species lake.

Over time, the team observed that some of the tanks bloomed with algae while others did not. This puzzled the scientists, and they decided to measure the water's dissolved organic carbon (DOC), which consists of mostly large molecules of decaying organic matter that give pond-water its slightly brownish color. It turned out that the water from the tanks with two-species fish contained larger particles of DOC (and hence darker water) than water with single-species fish. This increase in DOC blocked the sunlight and prevented algal blooming. Conversely, the water from the single-species tank contained smaller DOC particles, allowing more sunlight penetration to fuel the algal blooms.

This change in the environment, which is due to the different feeding habits of the stickleback species in each lake type, probably has a great impact on the survival of other species in these ecosystems, especially other photosynthetic organisms. Thus, the study shows that, at least in these ecosystems, the environment and the evolution of populations have reciprocal effects that may now be factored into simulation models.



**Figure 15.4.3:** The three-spined stickleback evolved from a saltwater fish to freshwater fish. (credit: Barrett Paul, USFWS)

## Research into Ecosystem Dynamics: Ecosystem Experimentation and Modeling

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 **15.4.4** 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.

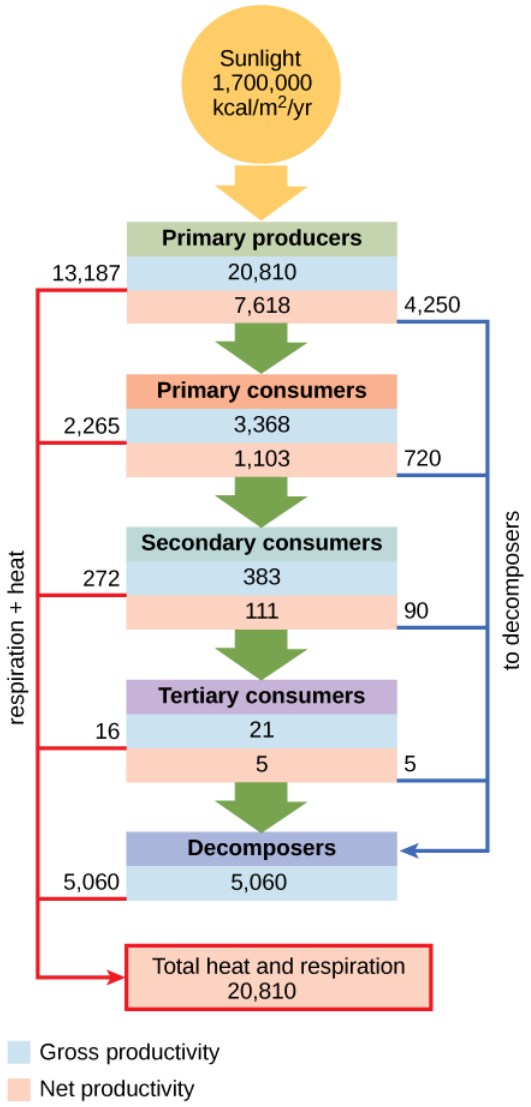



**Figure 15.4.4:** This conceptual model shows the flow of energy through a spring ecosystem in Silver Springs, Florida. Notice that the energy decreases with each increase in trophic level.

## 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.

Link to Learning

Visit The Darwin Project to view a variety of ecosystem models.

## Landscape Ecology

Landscapes (or seascapes in the marine context) are a mosaic of “patches”, each of which represents an ecological community. A landscape may contain various kinds of communities for the following reasons:

- each community reflects particular environmental conditions, such as different soil or bedrock types or variations of standing water (as in lakes, streams, or wetlands)
- the communities represent various stages in succession, such as patches of different age after wildfire or insect damage
- the communities may be related to land-use, as when parts of landscapes are affected by urbanization, agriculture, forestry, roads, or other human influences.

Over time, the spatial patterns of communities on landscapes are highly dynamic. This largely reflects the influence of disturbances and successional recovery. A patch that today is a pasture, a recent clear-cut, or a burn may be a mature forest after 50 years of succession. Similarly, a pond may in-fill over the centuries and become a wetland, which with further time may succeed into a forest. Ecologists use the term “shifting mosaic” to integrate the spatial and temporal variations of communities on landscapes. The following factors affect the shifting mosaic of communities.

- Patch size relates to the area of particular stands of communities (a stand is a community in a specific place). All species need some minimal area of habitat to support their populations, and small patches may not be adequate for that purpose. Relatively small patches may, however, help to support a population living in several stands on the landscape (an extensive population of this sort is known as a metapopulation). This can happen if the patches are connected by corridors to other suitable habitat, or if the species is capable of dispersing through surrounding inhospitable habitat (for this to occur, the landscape matrix must be permeable to movements of the species).
- The amount of edge is important because it influences the length of ecotone (transitional) habitat associated with a patch. A circular patch has the smallest ratio of edge to area, and smaller patches have higher ratios than larger ones of the same shape. An ecotone between patch types is a particular kind of habitat, and it may be selectively used by “edge species.” However, the greater the ratio of edge to area, the less “interior” habitat there is (this is uninfluenced by ecological conditions associated with an ecotone). Ecologists have identified “interior species” that are less successful if they try to use habitat close to an edge. Certain forest birds, for example, experience greater rates of predation and nest parasitism in small remnants of mature forest.
- Connectedness refers to the presence of links between otherwise discrete patches of similar habitat. These links may be used by a species as corridors to move among patches, allowing their metapopulation to function on the landscape. As was noted previously, connectedness is also related to the ability of a species to disperse among habitable patches through the surrounding habitat.
- Age-class adjacency is important in a landscape in which patches represent different stages of a successional sequence. This commonly occurs in landscapes affected by disturbances such as wildfire, insect epidemics, or clear-cutting. In general, patches of a similar post-disturbance age will be comparable in many aspects of habitat, while those of different age will be less similar. This can be an important consideration for movements of species among isolated patches that are suitable as habitat.

- Complex habitat requirements are characteristic of some larger animals, such as deer, bear, and wolf. These species need different kinds of habitat patches for specific purposes at various times of year. Because these animals participate in various kinds of communities, all of the habitat patches they need must be present on the landscape if a viable metapopulation is to be sustained.
- Landscape-level biodiversity is related to the richness of community types over a large area. A landscape that is uniformly covered by a single community has less biodiversity at this scale than one composed of a rich and dynamic mosaic of different kinds of communities.
- Landscape-level functions operate over extensive areas, and they may integrate the influences of many kinds of communities. A watershed, for example, is the expanse of terrain from which water drains into a stream, lake, or some other waterbody. Most watersheds contain various kinds of habitat patches, each with particular influences on hydrology and water chemistry. In general, watersheds covered with mature forest yield the cleanest flows of water. Other environmental services provided by well-vegetated landscapes include evapotranspiration, control of erosion, moderation of climatic extremes, and absorption of atmospheric carbon dioxide and release of oxygen.

Landscape ecology is an important subject area in environmental science. Humans commonly affect individual stands of particular kinds of communities, but many of the ecological effects must be managed at the scale of landscapes and seascapes.



**Figure 15.4.4:** A landscape is a mosaic of various kinds of communities, each stand of which represent a patch. In addition, landscapes are subjected to patch dynamics associated with natural disturbances, such as wildfire, windstorms, and insect outbreaks. However, the patch dynamics of many forested landscapes are being increasingly structured by forestry. In this aerial view of an area in New Brunswick, the natural forest is being harvested by clear-cutting (the lighter patches are snow in clear-cuts), which initiates a succession that restores a forest for another harvest in 60-80 years. Unless some areas are set aside for protection, this entire landscape may become used in this way. Source: M. Sullivan.

## 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.

## Contributors and Attributions

Modified by Kyle Whittinghill from the following sources

- **46.1: Ecology of Ecosystems** by OpenStax, is licensed CC BY by

  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).

- Essentials of Environmental Science by Kamala Doršner is licensed under CC BY 4.0.
- Ecology: From Individuals to the Biosphere from *Environmental Science: A Canadian Perspective* by Bill Freedman

---

This page titled 15.4: Ecology of Ecosystems is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **46.1: Ecology of Ecosystems** by OpenStax is licensed CC BY 4.0.

# 15.5: Energy Flow Through Ecosystems

Virtually every task performed by living organisms requires energy. In general, **energy** is defined as the ability to do work, or to create some kind of change. Energy exists in different forms. Examples include light energy, kinetic energy, heat energy, potential energy, and chemical energy.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.

It is important to understand how organisms acquire energy and how that energy is passed from one organism to another through **trophic interactions**. Trophic interactions in a community can be represented by diagrams called **food chains** and **food webs**. 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.

## How Organisms Acquire Energy in a Food Web

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** (or **producers**): organisms capable of synthesizing their own food (more specifically, capable of using inorganic carbon as a carbon source). Cells run on the chemical energy found mainly in carbohydrate molecules the majority of which are produced by autotrophs. 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. The energy that is harnessed by autotrophs enters the communities continuously and is transferred from one organism to another.

Photoautotrophs, such as plants, algae, and photosynthetic bacteria, serve as the energy source for a majority of the world's ecosystems. These ecosystems are often described by grazing food webs. Photoautotrophs harness the solar energy of the sun by converting it to chemical energy in the form of ATP (and NADP) (figure 15.5.1). The energy stored in ATP is used to synthesize complex organic molecules, such as glucose.

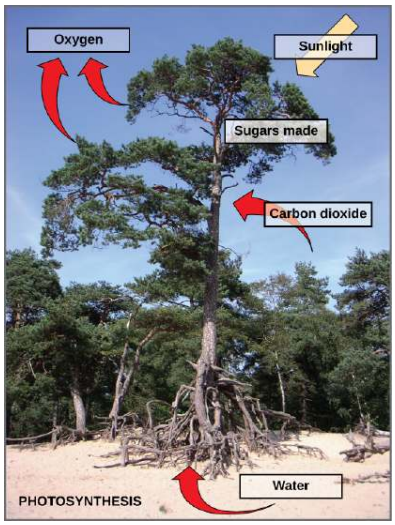



**Figure 15.5.1:** Photosynthesis uses solar energy, carbon dioxide, and water to release oxygen and to produce energy-storing sugar molecules. Image and caption by OpenStax (CC-BY). Access for free at openstax.org.

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 15.5.2). Many chemoautotrophs in hydrothermal vents use hydrogen sulfide ($H_2S$), 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.

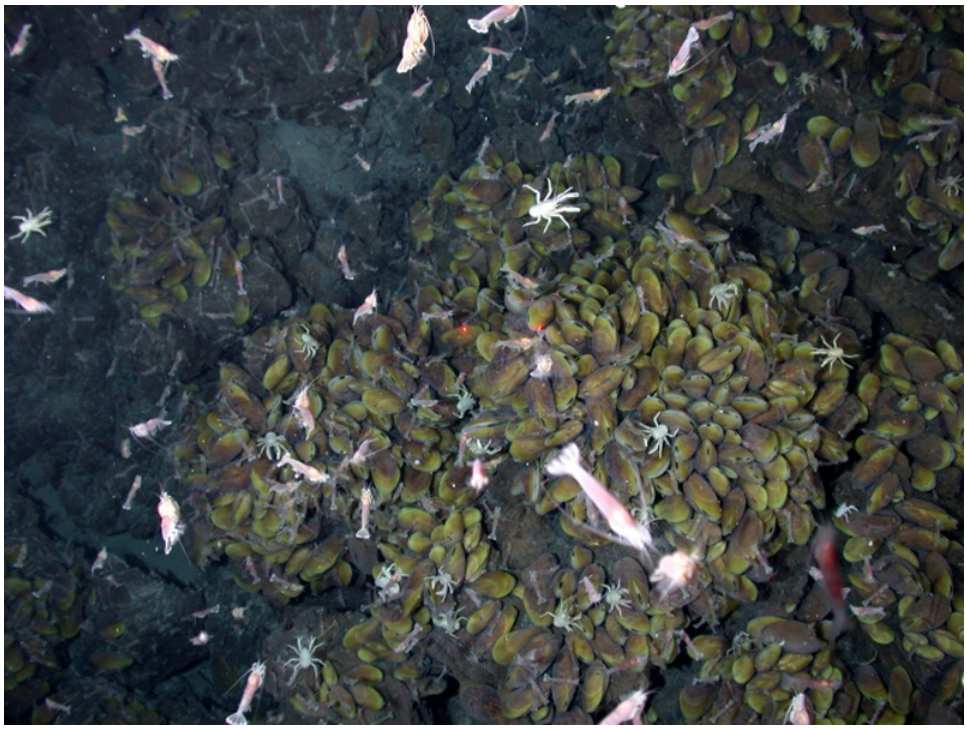


Figure 15.5.2: Swimming shrimp, a few squat lobsters, and hundreds of vent mussels are seen at a hydrothermal vent at the bottom of the ocean. As no sunlight penetrates to this depth, the ecosystem is supported by chemoautotrophic bacteria and organic material that sinks from the ocean's surface. This 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.

Heterotrophs **(consumers)**, like animals, fungi, and various microorganisms depend on producers, either directly or indirectly. For example, a deer obtains energy by eating plants (figure 15.5.3). A wolf eating a deer obtains energy that originally came from the plants eaten by that deer. Using this reasoning, all food eaten by humans can be traced back to producers that carry out photosynthesis (figure 15.5.4).



**Figure 15.5.3:** The energy stored in carbohydrate molecules from photosynthesis passes through the food chain. The predator that eats these deer is getting energy that originated in the photosynthetic vegetation that the deer consumed. (credit: Steve VanRiper, U.S. Fish and Wildlife Service)

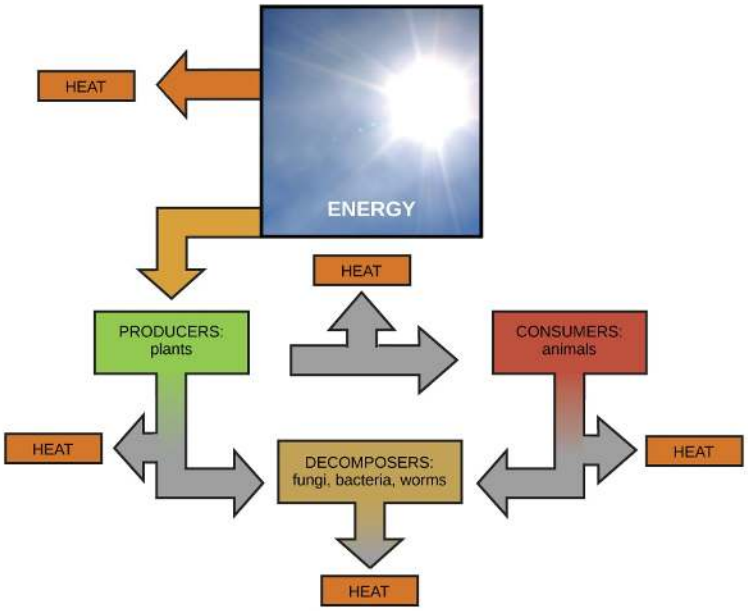



**Figure 15.5.4:** Ultimately, most life forms get their energy from the sun. This flow chart shows energy from the sun being captured by producers, such as plants, through photosynthesis. The energy is transferred to the consumers of the producers, such as animals. Energy can be obtained from producers directly (herbivores eat plants) or indirectly (carnivores eat herbivores). Decomposers eventually breakdown of dead organisms, including plant and animal material, and contribute to the nutrient pool. Fungi and bacteria are decomposers, and worms are detritivores (not shown). During each energy transfer, some of the energy in the system is lost as heat.

Consumers can be classified based on whether they eat animal or plant material (figure **15.5.5**). Consumers that feed exclusively on animals are called **carnivores**. Lions, tigers, snakes, sharks, sea stars, spiders, and ladybugs are all carnivores. **Herbivores** are consumers that feed exclusively on plant material, and examples include deer, koalas, some bird species, crickets, and caterpillars. Herbivores can be further classified into **frugivores** (fruit-eaters), **granivores** (seed eaters), **nectivores** (nectar feeders), and **folivores** (leaf eaters). Consumers that eat both plant and animal material are considered **omnivores**. Humans, bears, chickens, cockroaches, and crayfish are examples of omnivores.



**Figure 15.5.5:** Carnivores like the lion (top left) eat primarily meat. The ladybug (lower left) is also a carnivore that consumes small insects called aphids. Herbivores, like the mule deer (middle) eat primarily plant material. Omnivores like the bear (top right) and crayfish (bottom right) eat both plant and animal based food. Lion by Kevin Pluck; ladybug by Jon Sullivan; mule deer by Bill Ebbesen; bear by Dave Menke; crayfish by Jon Sullivan. All from OpenStax (CC-BY). Access for free at openstax.org.

Dead producers and consumers are eaten by **detritivores** (which ingest dead tissues) and **decomposers** (which further break down these tissues into simple molecules by secreting digestive enzymes). Invertebrate animals, such as worms and millipedes, are examples of detritivores, whereas fungi and certain bacteria are examples decomposers.

## Food Chains

A **food chain** is a linear sequence of organisms through which nutrients and energy pass as one organism eats another. Each organism in a food chain occupies a specific **trophic level** (energy level), its position in the food chain. The first trophic level in the food chain is the producers. The **primary consumers** (the herbivores that eat producers) are the second trophic level. Next are higher-level consumers. Higher-level consumers include **secondary consumers** (third trophic level), which are usually carnivores that eat the primary consumers, and **tertiary consumers** (fourth trophic level), which are carnivores that eat other carnivores. In the Lake Ontario food chain, shown in figure **15.5.6**, the Chinook salmon is the apex consumer at the top of this food chain. Some communities have additional trophic levels (quaternary consumers, fifth order consumers, etc.). Finally, detritivores and decomposers break down dead and decaying organisms from any trophic level. There is a single path through a food chain.

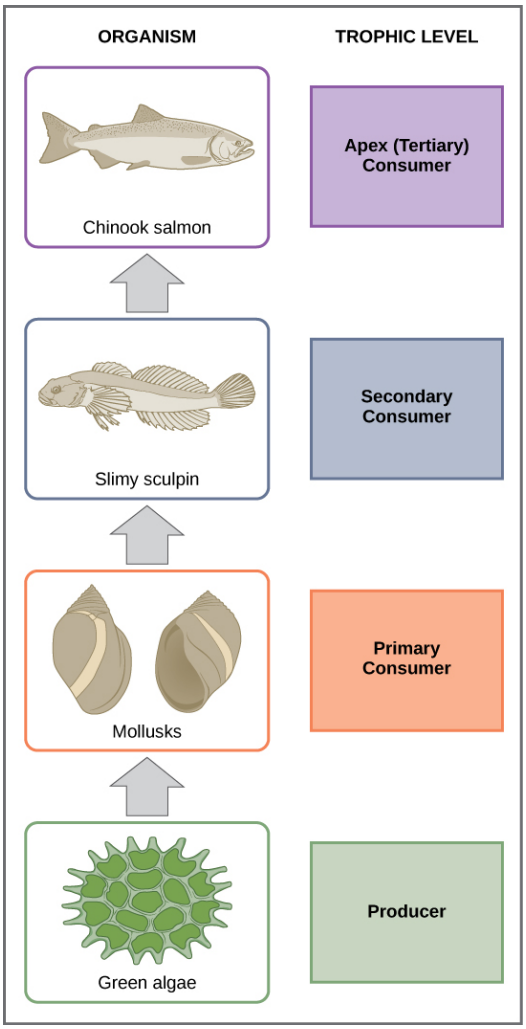



**Figure 15.5.6:** These are the trophic levels of a food chain in Lake Ontario at the United States–Canada border. Energy and nutrients flow from photosynthetic green algae (producers) at the base to the primary consumers, which are mollusks, or snails. The secondary consumers are small fish called slimy sculpin. The tertiary and apex consumer is Chinook salmon. Detritivores and decomposers are not shown. (credit: modification of work by National Oceanic and Atmospheric Administration/NOAA)

One major factor that limits the number of steps in a food chain is energy. Only about 10% of the energy in one trophic level is transferred to the next trophic level. This is because much energy is lost as heat during transfers between trophic levels or to decomposers due to the second law of thermodynamics. Thus, after four to six trophic energy transfers, the amount of energy remaining in the food chain may not be great enough to support viable populations at higher trophic levels.

Certain environmental toxins can become more concentrated as they move up the food chain, with the highest concentrations occurring in the top consumers, a process called **biomagnification**. Essentially, a top consumer ingests all the toxins that had previously accumulated in the bodies of the organisms at the lower trophic levels. This explains why frequently eating certain fish, like tuna or swordfish, increases your exposure to mercury, a toxic heavy metal.

## Food Webs

While food chains are simple and easy to analyze, there is a one problem when using food chains to describe most communities. Even when all organisms are grouped into appropriate trophic levels, some of these organisms can feed at more than one trophic level. In addition, species feed on and are eaten by more than one species. In other words, the linear model of trophic interactions, the food chain, is a hypothetical and overly simplistic representation of community structure. A holistic model—which includes 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. A **food web** is a concept that accounts for the multiple trophic interactions between each species (figure **15.5.7 – 9**).

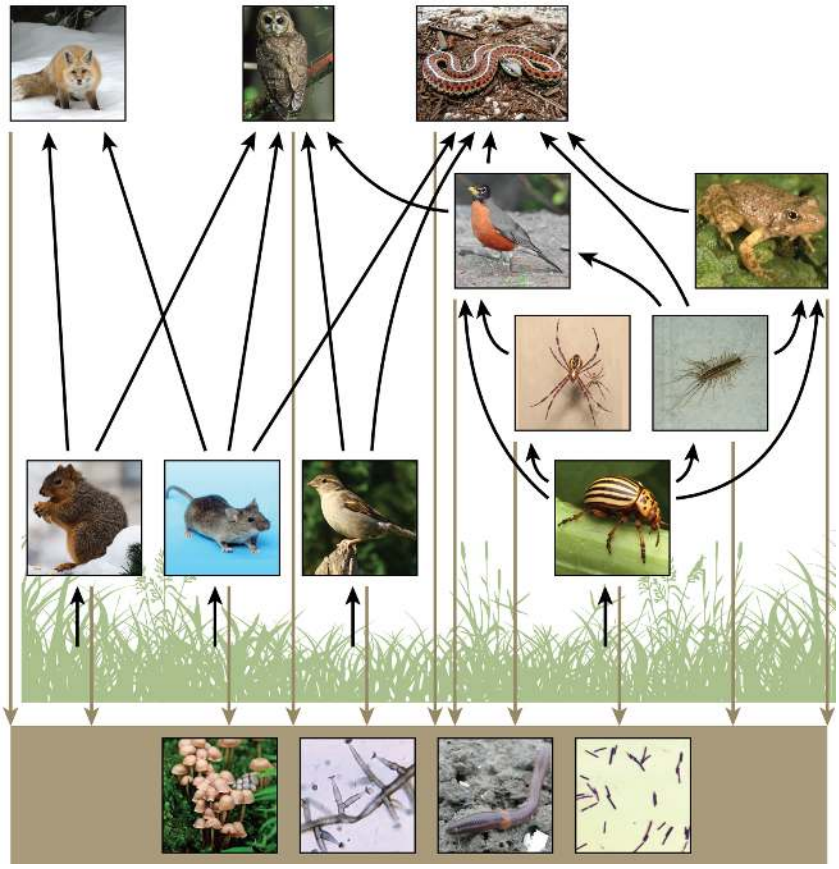


**Figure 15.5.7:** This food web shows the interactions among organisms across trophic levels. Arrows point from an organism that is consumed to the organism that consumes it and represent energy transfer. The producers (plants) capture energy from the sun. The level above the producers shows the primary consumers that eat the producers. Some examples are squirrels, mice, seed-eating birds, and beetles. Spiders and centipedes eat beetles. Robins eat beetles, spiders, and centipedes, and toads eat beetles and centipedes. Foxes eat squirrels and mice; owls eat squirrels, mice, seed-eating birds, and robins; and snakes eat mice, seed-eating birds, robins, centipedes, and toads. Some consumers are in between trophic levels because they eat a combination of primary, secondary, and/or tertiary consumers. All the producers and consumers eventually become nourishment for the decomposers (mushrooms, mold, and bacteria) and detritivores (earthworms) in the soil, which are depicted at the bottom of the illustration. (credit “fox”: modification of work by Kevin Bacher, NPS; credit “owl”: modification of work by John and Karen Hollingsworth, USFWS; credit “snake”: modification of work by Steve Jurvetson; credit “robin”: modification of work by Alan Vernon; credit “frog”: modification of work by Alessandro Catenazzi; credit “spider”: modification of work by “Sanba38″/Wikimedia Commons; credit “centipede”: modification of work by “Bauerph”/Wikimedia Commons; credit “squirrel”: modification of work by Dawn Huczek; credit “mouse”: modification of work by NIGMS, NIH; credit “sparrow”: modification of work by David Friel; credit “beetle”: modification of work by Scott Bauer, USDA Agricultural Research Service; credit “mushrooms”: modification of work by Chris Wee; credit “mold”: modification of work by Dr. Lucille Georg, CDC; credit “earthworm”: modification of work by Rob Hille; credit “bacteria”: modification of work by Don Stalons, CDC)

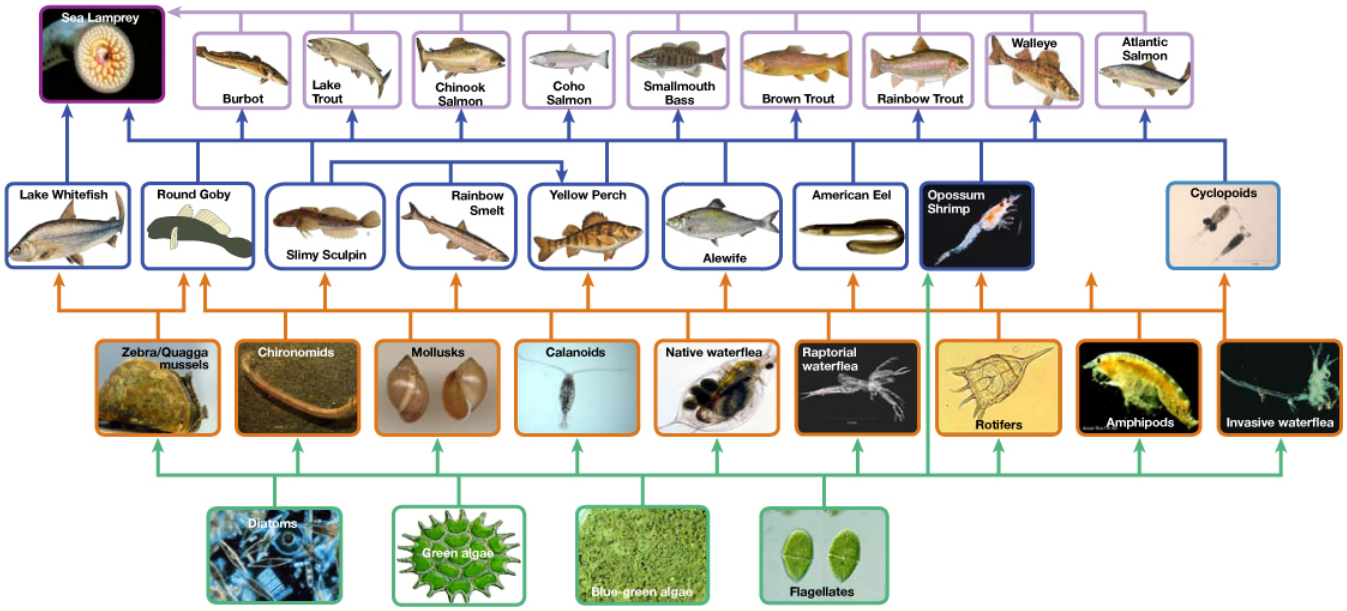



Figure 15.5.8: **Food web**: This food web shows the interactions between organisms across trophic levels in the Lake Ontario ecosystem. Primary producers are outlined in green, primary consumers in orange, secondary consumers in blue, and tertiary (apex) consumers in purple. Arrows point from an organism that is consumed to the organism that consumes it. Notice how some lines point to more than one trophic level. For example, the opossum shrimp eats both primary producers and primary consumers. (credit: NOAA, GLERL)

The trophic level of each species in a food web is not necessarily a whole number. In figure 15.5.9, phytoplankton are the primary producers (trophic level 1). Zooplankton only feed on phytoplankton, making them primary consumers (trophic level 2). Determining the trophic level of the other species is more complex. For example, krill eat both phytoplankton and zooplankton. If krill only ate phytoplankton they would primary consumers (trophic level 2). If they ate only zooplankton, they would be secondary consumers (trophic level 3). Since, krill consume both, their trophic level is 2.5.

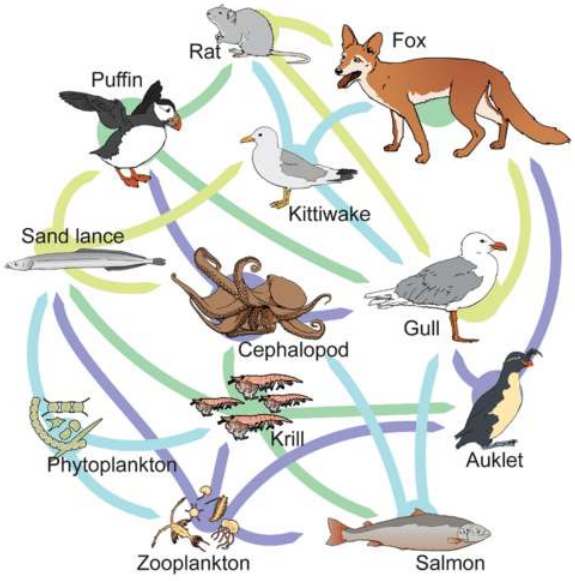



**Figure 15.5.9:** In this food web, each organism may have multiple food sources or be eaten by multiple species. Phytoplankton are the primary producers. They are consumed by sand lances, krill, and zooplankton. Sand lances are consumed by puffins, kittiwake, and cephalopods. Krill are consumed by the sand lance, cephalopods, auklets, and salmon. Zooplankton are consumed by the sand lance, krill, auklets, and salmon. Puffins are eaten by rats and gulls. Kittiwakes are eaten by rats, foxes, and gulls. Cephalopods are

consumed by puffins and gulls. Auklets are consumed by gulls and foxes. Salmon are consumed by cephalopods and gulls. Rats are consumed by foxes and gulls. Gulls are consumed by foxes. Detritivores and decomposers are not shown. Image by Mariana Ruiz Villarreal (LadyofHats) for CK-12 Foundation (CC-BY-NC).

## Ecosystem Productivity

Productivity within an ecosystem can be defined as the percentage of energy entering the ecosystem incorporated into biomass in a particular trophic level. Biomass is the total mass, in a unit area at the time of measurement, of living or previously living organisms within a trophic level. Ecosystems have characteristic amounts of biomass at each trophic level. For example, in the English Channel ecosystem the primary producers account for a biomass of 4 $g/m^2$ (grams per meter squared), while the primary consumers exhibit a biomass of 21 $g/m^2$.

The productivity of the primary producers is especially important in any ecosystem because these organisms bring energy to other living organisms by photoautotrophy or chemoautotrophy. The rate at which photosynthetic producers incorporate energy from the sun is called **gross primary productivity**. In a cattail marsh, plants only trap 2.2% of the energy from the sun that reaches them. Three percent of the energy is reflected, and another 94.8% is used to heat and evaporate water within and surrounding the plant. However, not all of the energy incorporated by producers is available to the other organisms in the food web because producers must also grow and reproduce, which consumes energy. At least half of the 2.2% trapped by cattail marsh plants is used to meet the plants own energy needs.

**Net primary productivity** is the energy that remains in the producers after accounting for the metabolic needs of the producers and heat loss. The net productivity is then available to the primary consumers at the next trophic level. One way to measure net primary productivity is to collect and weigh the plant material produced on a $m^2$ (about 10.7 $ft^2$) of land over a given interval. One gram of plant material (e.g., stems and leaves), which is largely the carbohydrate cellulose, yields about 4.25 kcal of energy when burned. Net primary productivity can range from 500 $kcal/m^2/yr$ in the desert to 15,000 $kcal/m^2/yr$ in a tropical rain forest.

The loss of energy between trophic levels is illustrated by the pioneering studies of Howard T. Odum in the Silver Springs, Florida, ecosystem in the 1940s (Figure **15.5.10**). the gross primary productivity (total energy accumulated by the primary producers) was 20,810 $kcal/m^2/yr$ (figure **15.5.*j***). The net primary productivity (energy available to consumers) was only 7,632 $kcal/m^2/yr$ after accounting for energy lost as heat and energy require to meet the producer's metabolic needs. . The primary consumers generated 3368 $kcal/m^2/yr$, the secondary consumers generated 383 $kcal/m^2/yr$, and the tertiary consumers only generated 21 $kcal/m^2/yr$. Thus, there is little energy remaining for another level of consumers in this ecosystem.

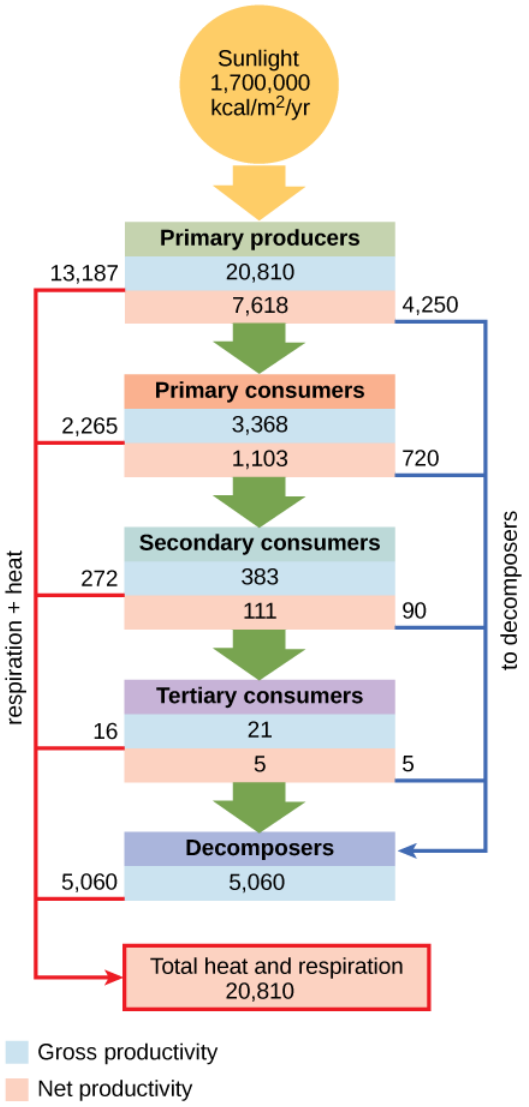



**Figure 15.5.10:** The flow of energy through a spring ecosystem in Silver Springs, Florida. The gross productivity (blue) for each trophic level is listed just above the net productivity (red). Notice that the energy decreases with each increase in trophic level. The energy content of primary producers (gross productivity) is 20,810 kcal/m$^2$/yr. The gross productivity of primary consumers is much smaller, about 3,368 kcal/m$^2$/yr. The gross productivity of secondary consumers is 383 kcal/m$^2$/yr, and the gross productivity of tertiary consumers is only 21 kcal/m$^2$/yr. The net productivity of each trophic level is less than the gross productivity because some energy is used to meet metabolic needs (respiration), and some energy is lost as heat. For example, the net productivity of primary consumers was 1,103 kcal/m$^2$/yr, only about a third of the gross productivity. Some of the energy in each trophic level (a total of 5,060 kcal/m$^2$/yr) is transferred to decomposers. All of the energy initially captured by primary producers (20,810 kcal/m$^2$/yr) is eventually released from the system.

## Ecological Efficiency: The Transfer of Energy between Trophic Levels

As illustrated in Figure **15.5.10**, large amounts of energy are lost from the ecosystem from one trophic level to the next level as energy flows from the primary producers through the various trophic levels of consumers and decomposers. The main reason for this loss is the second law of thermodynamics, which states that whenever energy is converted from one form to another, there is a tendency toward disorder (entropy) in the system. In biologic systems, this means a great deal of energy is lost as metabolic heat when the organisms from one trophic level consume the next level. In the Silver Springs ecosystem example (Figure **15.5.10**), we see that the primary consumers produced 1103 kcal/m$^2$/yr from the 7618 kcal/m$^2$/yr of energy available to them from the primary

producers. The measurement of energy transfer efficiency between two successive trophic levels is termed the **trophic level transfer efficiency** (TLTE) and is defined by the formula:

$$\text{TLTE} = \frac{\text{production at present trophic level}}{\text{production at previous trophic level}} * 100$$

In Silver Springs, the TLTE between the first two trophic levels was approximately 14.8 percent. The low efficiency of energy transfer between trophic levels is usually the major factor that limits the length of food chains observed in a food web. The fact is, after four to six energy transfers, there is not enough energy left to support another trophic level. In the Lake Ontario example shown in Figure 46.1.5, only three energy transfers occurred between the primary producer, (green algae), and the apex consumer (Chinook salmon).

Ecologists have many different methods of measuring energy transfers within ecosystems. Some transfers are easier or more difficult to measure depending on the complexity of the ecosystem and how much access scientists have to observe the ecosystem. In other words, some ecosystems are more difficult to study than others, and sometimes the quantification of energy transfers has to be estimated.

Another main parameter that is important in characterizing energy flow within an ecosystem is the **net production efficiency**. Net production efficiency (NPE) allows ecologists to quantify how efficiently organisms of a particular trophic level incorporate the energy they receive into biomass; it is calculated using the following formula:

$$\text{NPE} = \frac{\text{net consumer productivity}}{\text{assimilation}} * 100$$

**Net consumer productivity** is the energy content available to the organisms of the next trophic level. Assimilation is the biomass (energy content generated per unit area) of the present trophic level after accounting for the energy lost due to incomplete ingestion of food, energy used for respiration, and energy lost as waste. Incomplete ingestion refers to the fact that some consumers eat only a part of their food. For example, when a lion kills an antelope, it will eat everything except the hide and bones. The lion is missing the energy-rich bone marrow inside the bone, so the lion does not make use of all the calories its prey could provide.

Thus, NPE measures how efficiently each trophic level uses and incorporates the energy from its food into biomass to fuel the next trophic level. In general, cold-blooded animals (**ectotherms**), such as invertebrates, fish, amphibians, and reptiles, use less of the energy they obtain for respiration and heat than warm-blooded animals (**endotherms**), such as birds and mammals. The extra heat generated in endotherms, although an advantage in terms of the activity of these organisms in colder environments, is a major disadvantage in terms of NPE. Therefore, many endotherms have to eat more often than ectotherms to get the energy they need for survival. In general, NPE for ectotherms is an order of magnitude (10x) higher than for endotherms. For example, the NPE for a caterpillar eating leaves has been measured at 18 percent, whereas the NPE for a squirrel eating acorns may be as low as 1.6 percent.

## 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 15.5.11, 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 15.5.11), 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 **15.5.10**. 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 **15.5.11**).

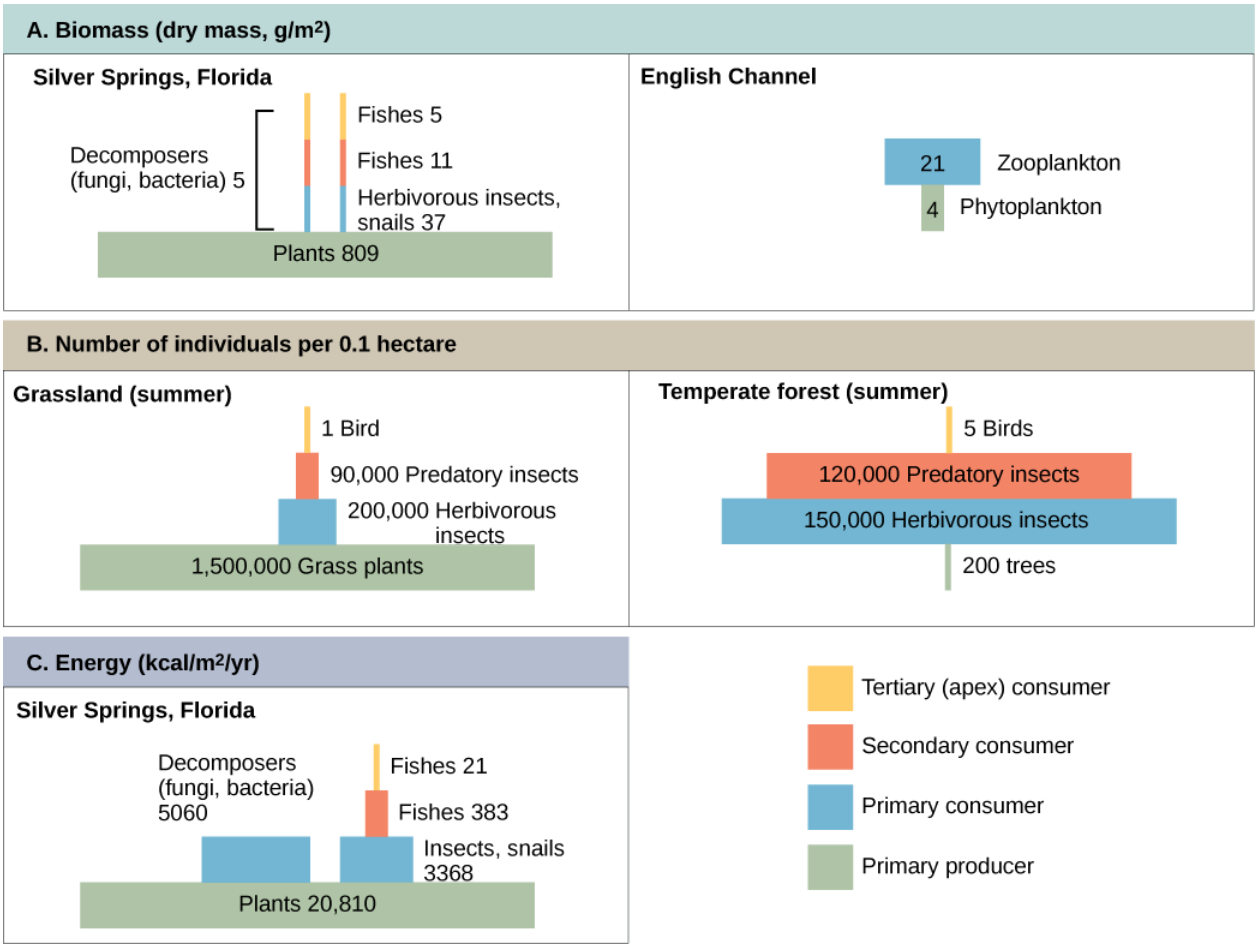



**Figure 15.5.11:** Ecological pyramids depict the (a) biomass, (b) number of organisms, and (c) energy in each trophic level.

## 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.

## Contributors and Attributions

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

- Ecology of Ecosystems, Energy Flow through Ecosystems, and Digestive Systems from *General Biology/Biology 2e* by OpenStax (licensed under CC-BY). Access for free at openstax.org. by

  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).

- Energy, Energy Enters Ecosystems Through Photosynthesis, and Energy Flow Through Ecosystems from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Ecosystem Productivity from *Biology* by John W. Kimball (licensed under CC-BY)

---

This page titled 15.5: Energy Flow Through Ecosystems is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **46.2: Energy Flow through Ecosystems** by OpenStax is licensed CC BY 4.0.
- **46.1: Ecology of Ecosystems** by OpenStax is licensed CC BY 4.0.

# CHAPTER OVERVIEW

## 16: Conservation Biology

**Conservation biology** is the management of Earth's ecosystems with the aim of protecting species, their communities, and ecosystems from escalated extinction rates and the destruction and degradation of their habitats. This field focuses on both evaluating past, current, and future trends and creating conservation action plans to reverse trends leading to extinction and destroyed habitats. More specifically, procedures have been put in place to protect threatened species, design habitat preserves, creating breeding programs, and reconciling conservation concerns with practical human needs.

As conservation biology needs to pull from so many fields, it is an interdisciplinary subject drawing on physical, life, and social sciences, as well as natural resource management procedures. The field seeks to integrate social science policy with theories from the fields of ecology, demography, taxonomy, and genetics. The principles underlying each of these disciplines have direct implications for the management of species and ecosystems, captive breeding and reintroduction, genetic analyses, and habitat restoration. Social science disciplines not only help with putting policies in place to legally protect species and habitats but also help fund conservation actions. Conservation action plans help direct research, monitoring, and education programs that engage concerns at both local and global scales.

The goal of this unit is to overview the essential conservation biology topics as we head into an uncertain future.



Figure 16.a: A road heading into our uncertain future. Image by Pikist (Public domain)

### Attributions

Rachel Schleiger (CC-BY-NC)

- 16.1: Introduction to Biodiversity
- 16.2: Measures of Biodiversity Loss
- 16.3: Threats to Biodiversity
- 16.4: Importance of Biodiversity
- 16.5: Preserving Biodiversity

Thumbnail image - "Environmental protection" is in the Public Domain

This page titled 16: Conservation Biology is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

# 16.1: Introduction to Biodiversity

Biodiversity is a broad term for biological variety, and it can be measured at a number of organizational levels. A common meaning of biodiversity is simply the number of species in a location or on Earth; for example, the American Ornithologists' Union lists 2078 species of birds in North and Central America. This is one measure of the bird biodiversity on the continent. More sophisticated measures or of diversity (diversity indexes) take into account the relative abundances of species (i.e. both the number of species and the number of individuals in each of those species). For example, a forest with 10 equally common species of trees is more diverse than a forest that has 10 species of trees wherein just one of those species makes up 95 percent of the trees rather than them being equally distributed. These estimation indexes, which came from information theory, are most useful as a first step in quantifying biodiversity between and within ecosystems; they are less useful when the main concern among conservation biologists is simply the loss of biodiversity. However, biologists recognize that measures of biodiversity, in terms of species diversity, may help focus efforts to preserve the biologically or technologically important elements of biodiversity. Biologists are also now using measures of biodiversity at other levels of biological organization (including genes, populations, and ecosystems).

The Lake Victoria cichlids provide an example through which we can begin to understand biodiversity (Figure **16.1.1**). The biologists studying cichlids in the 1980s discovered hundreds of cichlid species representing a variety of specializations to particular habitat types and specific feeding strategies: eating plankton floating in the water, scraping and then eating algae from rocks, eating insect larvae from the bottom, and eating the eggs of other species of cichlid. The cichlids of Lake Victoria are the product of an adaptive radiation. An adaptive radiation is a rapid (less than three million years in the case of the Lake Victoria cichlids) branching through speciation of a phylogenetic tree into many closely related species; typically, the species "radiate" into different habitats and niches. The Galápagos finches are an example of a modest adaptive radiation with 15 species. The cichlids of Lake Victoria are an example of a spectacular adaptive radiation that includes about 500 species.



**Figure 16.1.1:** Lake Victoria in Africa, shown in this satellite image, was the site of one of the most extraordinary evolutionary findings on the planet, as well as a casualty of devastating biodiversity loss. (credit: modification of work by Rishabh Tatiraju, using NASA World Wind software)

t the time biologists were making this discovery, some species began to quickly disappear. A culprit in these declines was a species of large fish that was introduced to Lake Victoria by fisheries to feed the people living around the lake. The Nile perch was introduced in 1963, but lay low until the 1980's when its populations began to surge. The Nile perch population grew by consuming cichlids, driving species after species to the point of extinction (the disappearance of a species). In fact, there were several factors that played a role in the extinction of perhaps 200 cichlid species in Lake Victoria: the Nile perch, declining lake water quality due to agriculture and land clearing on the shores of Lake Victoria, and increased fishing pressure. Scientists had not even cataloged all of the species present—so many were lost that were never named. The diversity is now a shadow of what it once was.

The cichlids of Lake Victoria are a thumbnail sketch of contemporary rapid species loss that occurs all over Earth and is caused by human activity. Extinction is a natural process of macroevolution that occurs at the rate of about one out of 1 million species becoming extinct per year. The fossil record reveals that there have been five periods of mass extinction in history with much higher rates of species loss, and the rate of species loss today is comparable to those periods of mass extinction. However, there is a major difference between the previous mass extinctions and the current extinction we are experiencing: human activity.

Specifically, three human activities have a major impact: destruction of habitat, introduction of exotic species, and over-harvesting. Predictions of species loss within the next century, a tiny amount of time on geological timescales, range from 10 percent to 50 percent. Extinctions on this scale have only happened five other times in the history of the planet, and they have been caused by cataclysmic events that changed the course of the history of life in each instance. Earth is now in one of those times.

## Species Diversity

**Species diversity** is the number of different species in a particular area and their relative abundance. The area in question could be a habitat, a biome, or the entire biosphere. Areas with low species diversity, such as the glaciers of Antarctica, still contain a wide variety of living organisms, whereas the diversity of tropical rainforests is so great that it cannot be accurately assessed. **Species richness**, the number of species living in a habitat or other unit, is one component of biodiversity. **Species evenness** is a component of species diversity based on **relative abundance** (the number individuals in a species relative to the total number of individuals in all species within a system). Foundation species often have the highest relative abundance of species. Two locations with the same richness do not necessarily have the same species evenness. For example, both communities in figure 16.1.2 have three different trees species and thus a species richness of three. However, there is a dominant species (represented by six individuals) in community #1. In community #2, there are three of individuals of each species. Therefore, community #2 has a greater species evenness and greater species diversity overall.

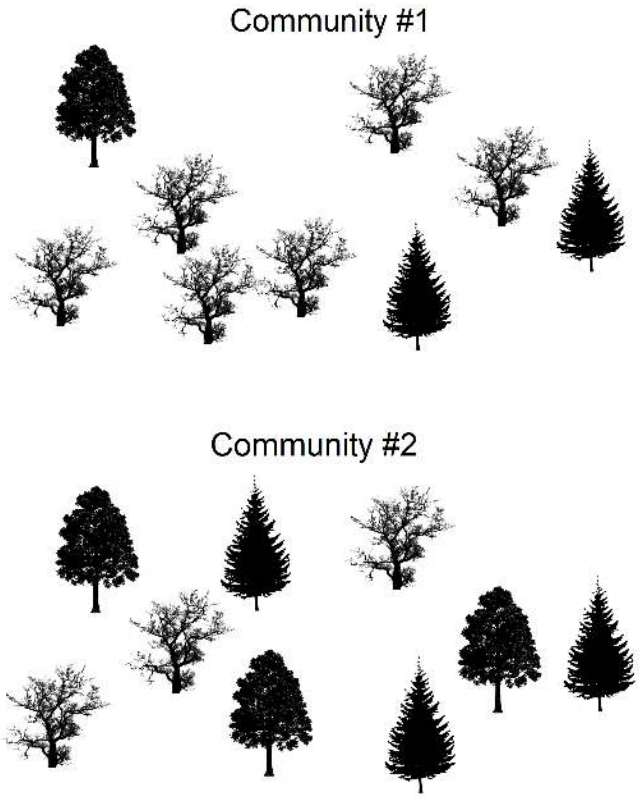



**Figure 16.1.2:** Two hypothetical tree communities have the same species richness, but community #2 (bottom) has a greater species evenness. Both communities have nine trees and three tree species. In community #1, one species is dominant, represented by six individuals. There are two individuals of a conifer species, and only one individual of the final species. In community #2, there are three individuals from each species. Images compiled by Melissa Ha from Alone tree George Hodan, Old Tree Silhouette, and Tree (all public domain).

## Other Types of Biodiversity

Scientists generally accept that the term biodiversity describes the number and kinds of species in a location or on the planet. Species can be difficult to define, but most biologists still feel comfortable with the concept and are able to identify and count

eukaryotic species in most contexts. Biologists have also identified alternate measures of biodiversity, some of which are important for planning how to preserve biodiversity.

**Genetic diversity** is one of those alternate concepts. Genetic diversity (or variation) is the raw material for adaptation in a species. A species' future potential for adaptation depends on the genetic diversity held in the genomes of the individuals in populations that make up the species. The same is true for higher taxonomic categories. A genus with very different types of species will have more genetic diversity than a genus with species that look alike and have similar ecologies. If there were a choice between one of these genera of species being preserved, the one with the greatest potential for subsequent evolution is the most genetically diverse one. It would be ideal not to have to make such choices, but increasingly this may be the norm.

Most genes code for proteins, which in turn carry out the metabolic processes that keep organisms alive and reproducing. Genetic diversity can be measured as **chemical diversity** in that different species produce a variety of chemicals in their cells, both the proteins as well as the products and byproducts of metabolism. This chemical diversity is important for humans because of the potential uses for these chemicals, such as medications. For example, the drug eptifibatide is derived from rattlesnake venom and is used to prevent heart attacks in individuals with certain heart conditions. At present, it is far cheaper to discover compounds made by an organism than to imagine them and then synthesize them in a laboratory. Chemical diversity is one way to measure diversity that is important to human health and welfare.

Humans have generated diversity in domestic animals, plants, and fungi through selective breeding. But even this diversity is suffering losses because of market forces and increasing globalism in human agriculture and migration, especially in heavily populated regions such as China, India, and Japan. For example, international seed companies produce only a very few varieties of a given crop and provide incentives around the world for farmers to buy these few varieties while abandoning their traditional varieties, which are far more diverse. The human population depends on crop diversity directly as a stable food source and its decline is troubling to biologists and agricultural scientists.

It is also useful to define **ecosystem diversity**, meaning the number of different ecosystems on the planet or in a given geographic area. Whole ecosystems can disappear even if some of the species might survive by adapting to other ecosystems. The loss of an ecosystem means the loss of interactions between species, the loss of unique features of coadaptation, and the loss of biological productivity that an ecosystem is able to create. An example of a largely extinct ecosystem in North America is the prairie ecosystem (Figure **16.1.3**). Prairies once spanned central North America from the boreal forest in northern Canada down into Mexico. They are now all but gone, replaced by crop fields, pasture lands, and suburban sprawl. Many of the species survive, but the hugely productive ecosystem that was responsible for creating the most productive agricultural soils is now gone. As a consequence, soils are disappearing or must be maintained at greater expense. The decline in soil productivity occurs because the interactions in the original ecosystem have been lost; this was a far more important loss to humans than the relatively few species that were driven extinct when the prairie ecosystem was destroyed.



**Figure 16.1.3:** The variety of ecosystems on Earth—from coral reef to prairie—enables a great diversity of species to exist. (credit "coral reef": modification of work by Jim Maragos, USFWS; credit: "prairie": modification of work by Jim Minnerath, USFWS)

## Current Species Diversity

Despite considerable effort, knowledge of the species that inhabit the planet is limited. About 1.5 million species have been described, but many more species are yet to be identified. Estimates of the total number of species on Earth range from 3 million to 100 million, with more recent estimates commonly ranging from 8 to 11 million species. A 2011 study suggests that only 13% of

**eukaryotic** species (such as plants, animals, fungi, and algae) have been named (Table **16.1.1** . Estimates of numbers of **prokaryotic** species (such as bacteria) are largely guesses, but biologists agree that science has only just begun to catalog their diversity. In fact, a 2017 study by Brendan Larsen and colleagues estimated that there are actually 1-6 billion species on Earth with at least 70% of them being bacteria. Given that Earth is losing species at an accelerating pace, science knows little about what is being lost.

**Table 16.1.1:** Estimates of the described and predicted species by Taxonomic group

| | Mora et al. 2011[1] | | Chapman 2009[2] | | Groombridge & Jenkins 2002[3] | |
|---|---|---|---|---|---|---|
| | Described | Predicted | Described | Predicted | Described | Predicted |
| Animalia | 1,124,516 | 9,920,000 | 1,424,153 | 6,836,330 | 1,225,500 | 10,820,000 |
| Chromista | 17,892 | 34,900 | 25,044 | 200,500 | — | — |
| Fungi | 44,368 | 616,320 | 98,998 | 1,500,000 | 72,000 | 1,500,000 |
| Plantae | 224,244 | 314,600 | 310,129 | 390,800 | 270,000 | 320,000 |
| Protozoa | 16,236 | 72,800 | 28,871 | 1,000,000 | 80,000 | 600,000 |
| Prokaryotes | — | — | 10,307 | 1,000,000 | 10,175 | — |
| Total | 1,438,769 | 10,960,000 | 1,897,502 | 10,897,630 | 1,657,675 | 13,240,000 |

There are various initiatives to catalog described species in accessible and more organized ways, and the internet is facilitating that effort. Nevertheless, at the current rate of species description, which according to the State of Observed Species reports is 17,000–20,000 new species a year, it would take close to 500 years to describe all of the species currently in existence. The task, however, is becoming increasingly impossible over time as extinction removes species from Earth faster than they can be described.

Naming and counting species may seem an unimportant pursuit given the other needs of humanity, but it is not simply an accounting. Describing species is a complex process by which biologists determine an organism's unique characteristics and whether or not that organism belongs to any other described species. It allows biologists to find and recognize the species after the initial discovery to follow up on questions about its biology. In addition, the unique characteristics of each species make it potentially valuable to humans or other species on which humans depend. Understanding these characteristics is the value of finding and naming species.

## DNA Barcoding

The technology of molecular genetics and data processing and storage are maturing to the point where cataloging the planet's species in an accessible way is close to feasible. **DNA barcoding** is one molecular genetic method, which takes advantage of a specialized structures inside of some cells called mitochondria (figure **16.1.4**). **Mitochondria** contain DNA that is separate from the rest of the cell, and one of the genes in mitochondrial DNA changes more quickly through the process of evolution than regular DNA. While plants contain mitochondria, DNA from their **chloroplasts**, the specialized structures in which photosynthesis occurs, are more often barcoded. Rapid DNA sequencing technology make the molecular genetics portion of the work relatively inexpensive and quick. Computer resources store and make available the large volumes of data. Projects are currently underway to use DNA barcoding to catalog museum specimens, which have already been named and studied, as well as testing the method on less studied groups. As of mid 2012, close to 150,000 named species had been barcoded. Early studies suggest there are significant numbers of undescribed species that looked too much like sibling species to previously be recognized as different. These now can be identified with DNA barcoding.

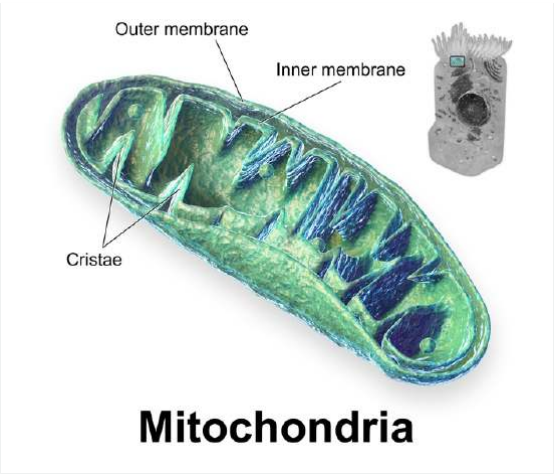



**Figure 16.1.4:** Mitochondria are specialized structures inside of eukaryotic cells. They contain separate DNA, and mitochondrial DNA contains a rapidly evolving gene used for DNA barcoding. Image by Blausen.com staff (2014). "Medical gallery of Blausen Medical 2014". *WikiJournal of Medicine* **1** (2). DOI:10.15347/wjm/2014.010. ISSN 2002-4436. (CC-BY)

## Patterns of Biodiversity

Biodiversity is not evenly distributed on Earth. Lake Victoria contained almost 500 species of cichlids alone, ignoring the other fish families present in the lake. All of these species were found only in Lake Victoria; therefore, the 500 species of cichlids were endemic. **Endemic species** are found in only one location. For example, the blue jay is endemic to North America, while the Barton Springs salamander is endemic to the mouth of one spring in Austin, Texas. Endemics with highly restricted distributions, like the Barton Springs salamander, are particularly vulnerable to extinction. Higher taxonomic levels, such as genera and families, can also be endemic. Lake Huron contains about 79 species of fish, all of which are found in many other lakes in North America. What accounts for the difference in fish diversity in these two lakes? Lake Victoria is a tropical lake, while Lake Huron is a temperate lake. Lake Huron in its present form is only about 7,000 years old, while Lake Victoria in its present form is about 15,000 years old. Biogeographers have suggested these two factors, latitude and age, are two of several hypotheses to explain biodiversity patterns on the planet.

### Career Connection: Biogeographer

Biogeography is the study of the distribution of the world's species—both in the past and in the present. The work of biogeographers is critical to understanding our physical environment, how the environment affects species, and how environmental changes impact the distribution of a species; it has also been critical to developing evolutionary theory. Biogeographers need to understand both biology and ecology. They also need to be well-versed in evolutionary studies, soil science, and climatology.

There are three main fields of study under the heading of biogeography: ecological biogeography, historical biogeography (called paleobiogeography), and conservation biogeography. Ecological biogeography studies the current factors affecting the distribution of plants and animals. Historical biogeography, as the name implies, studies the past distribution of species. Conservation biogeography, on the other hand, is focused on the protection and restoration of species based upon known historical and current ecological information. Each of these fields considers both zoogeography and phytogeography—the past and present distribution of animals and plants.

One of the oldest observed patterns in ecology is that species biodiversity in almost every taxonomic group increases as latitude declines. In other words, biodiversity increases closer to the equator (Figure **16.1.5**).

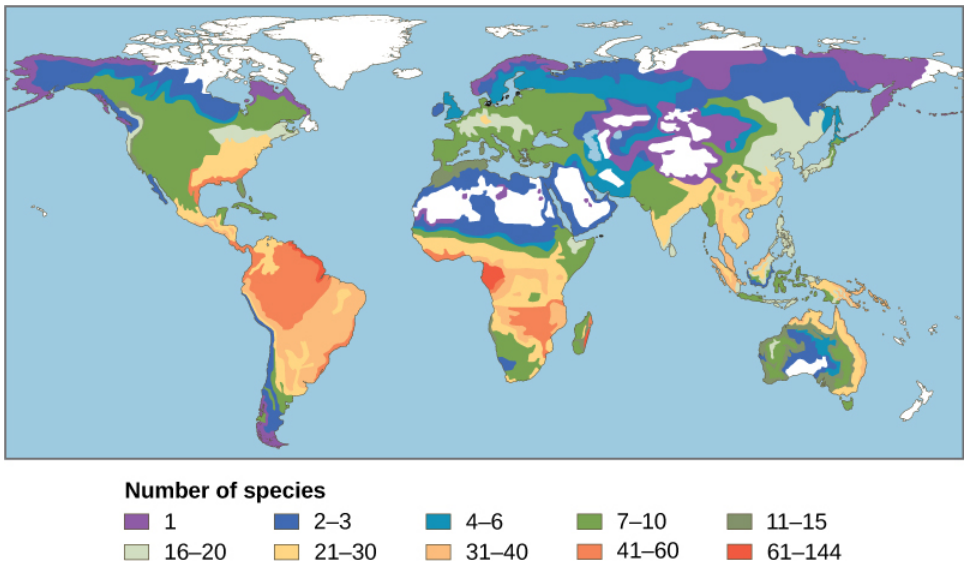



**Figure 16.1.5:** This map illustrates the number of amphibian species across the globe and shows the trend toward higher biodiversity at lower latitudes. A similar pattern is observed for most taxonomic groups.

It is not yet clear why biodiversity increases closer to the equator, but hypotheses include the greater age of the ecosystems in the tropics versus temperate regions that were largely devoid of life or drastically impoverished during the last glaciation. The idea is that greater age provides more time for **speciation**. Another possible explanation is the increased energy the tropics receive from the sun versus the decreased energy that temperate and polar regions receive. It is not entirely clear how greater energy input could translate into more species. The complexity of tropical ecosystems may promote speciation by increasing the **heterogeneity**, or number of **ecological niches**, in the tropics relative to higher latitudes. The greater heterogeneity provides more opportunities for coevolution, specialization, and perhaps greater selection pressures leading to population differentiation. However, this hypothesis suffers from some circularity—ecosystems with more species encourage speciation, but how did they get more species to begin with? The tropics have been perceived as being more stable than temperate regions, which have a pronounced climate and day-length seasonality. The tropics have their own forms of seasonality, such as rainfall, but they are generally assumed to be more stable environments and this stability might promote speciation.

Regardless of the mechanisms, it is certainly true that biodiversity is greatest in the tropics. The number of endemic species is higher in the tropics. The tropics also contain more biodiversity hotspots. At the same time, our knowledge of the species living in the tropics is lowest and because of recent, heavy human activity the potential for biodiversity loss is greatest.

## Biodiversity Hotspots

In 1988, British environmentalist Norman Myers developed a conservation concept to identify geographical areas rich in species and at significant risk for species loss: **biodiversity hotspots**. Biodiversity hotspots are geographical areas that contain high numbers of endemic species. **Endemic species** are found in only one location. For example, all of cichlid species found in Lake Victoria only occurred there. The blue jay is endemic to North America, while the Barton Springs salamander is endemic to the mouth of one spring in Austin, Texas. Endemic species with highly restricted distributions, like the Barton Springs salamander, are particularly vulnerable to extinction. If a population of a widespread species declines in one region, individuals from another region may be able to recolonize the first location, but this is not possible for endemic species. Endemic species are particularly common in isolated regions, such as mountaintops or islands. Endemism is especially likely on islands that are large and far from the mainland.

The purpose of the biodiversity hotspot concept was to identify important locations on the planet for conservation efforts, a kind of conservation triage. By protecting hotspots, governments are able to protect a larger number of species. The original criteria for a hotspot included the presence of 1500 or more endemic plant species and 70 percent of the area disturbed by human activity. There are now 34 biodiversity hotspots (Figure **16.1.6**) containing large numbers of endemic species, which include half of Earth's endemic plants.

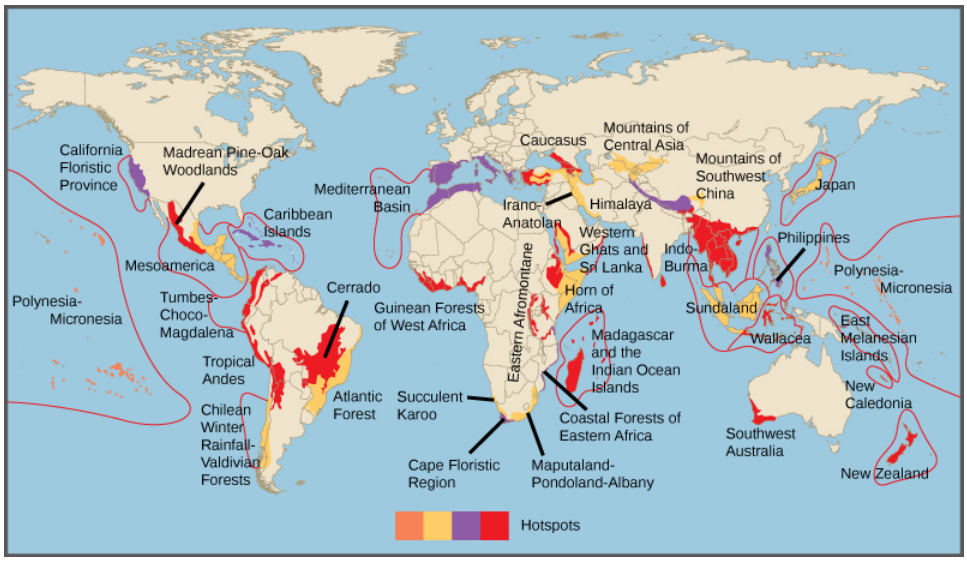



**Figure 16.1.6:** Conservation International has identified 34 biodiversity hotspots, which cover only 2.3 percent of the Earth's surface but have endemic to them 42 percent of the terrestrial vertebrate species and 50 percent of the world's plants.

## Case Study: Biodiversity Hotspots in South Africa

A rich species diversity is found in South Africa. With a land surface of approximately 148,000 $km^2$ (about 57,000 square miles), representing approximately 1% of the Earth's total surface, South Africa contains 10% of the world's total known bird, fish and plant species, and 6% of the world's mammal and reptile species.

This natural wealth is threatened by the expansion of the human population and the increasing demand this places on the environment. The Karoo (figure **16.1.7**) and the Cape are biodiversity hotspots in South Africa. South Africa has a wide range of climatic conditions and many variations in landscape, creating a variety of ecological niches, which can promote species diversity. These various landscapes give rise to the biomes which allow a wide variety of life to survive.




**Figure 16.1.7:** Map of South Africa, illustrating the Karoo, a dry region with hot summers and cold winters. Image by (WT-shared) NJR ZA at wts wikivoyage (CC-BY-SA).

More than 20,300 species of flowering plants occur in South Africa (figure **16.1.8**). One of the six most important areas of plant growth in the world is in the Western Cape. Most of the 2,000 threatened species of plants are found in the **fynbos** in South-West Cape (figure **16.1.9**).



**Figure 16.1.8:** Vegetation of the Karoo in near Willowmore, Eastern Cape, South Africa. Image by Winfried Bruenken (CC-BY-SA).



**Figure 16.1.9:** Vegetation of the fynbos in Western Cape, South Africa. Image by Joachim Huber (CC-BY-SA).

South Africa is home to 243 mammal species. Among the 17 threatened species in South Africa are the black rhino, pangolin, and giant golden mole. The blue antelope and quagga (figure **16.1.10**) have become extinct. Of more than 800 bird species, 26 are threatened, including the penguin, Cape vulture, martial eagle (figure **16.1.11**), and Cape parrot. In total 370 reptiles and amphibians occur in the region of which 21 are threatened. Six of these are endangered. Of the 220 freshwater fishes in South Africa, 21 are threatened. There are more than 2,000 marine fish species. While many insect species remain unidentified, 80,000 insects are known to occur.

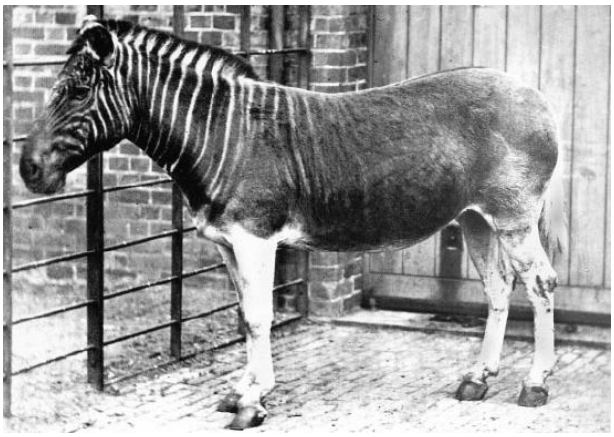


**Figure 16.1.10:** The quagga (*Equus quagga quagga*), now extinct, is technically a subspecies of the plains zebra. Image by Stanford Woods (public domain).



**Figure 16.1.11:** The Martial Eagle (*Polemaetus bellicosus*) is one of the threatened bird species found in South Africa. Image by Charles J. Sharp (CC-BY-SA).

## Footnotes

1. 1 Mora Camilo et al., "How Many Species Are There on Earth and in the Ocean?" *PLoS Biology* (2011), doi:10.1371/journal.pbio.1001127.
2. 2 Arthur D. Chapman, *Numbers of Living Species in Australia and the World*, 2nd ed. (Canberra, AU: Australian Biological Resources Study, 2009). www.environment.gov.au/biodiv...d-complete.pdf.
3. 3 Brian Groombridge and Martin D. Jenkins. *World Atlas of Biodiversity: Earth's Living Resources in the 21st Century*. Berkeley: University of California Press, 2002.
4. 4 International Institute for Species Exploration (IISE), *2011 State of Observed Species (SOS)*. Tempe, AZ: IISE, 2011. Accessed May, 20, 2012. species.asu.edu/SOS.
5. Brendan B. Larsen, Elizabeth C. Miller, Matthew K. Rhodes, and John J. Wiens, "Inordinate Fondness Multiplied and Redistributed: the Number of Species on Earth and the New Pie of Life," The Quarterly Review of Biology 92, no. 3 (September 2017): 229-265. DOI

## Contributors and Attributions

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

- Community Ecology and Importance of Biodiversity from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Preserving Biodiversity and **47.1: The Biodiversity Crisis** by OpenStax, is licensed CC BY by

  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).
- The Biodiverisity Crisis and Prelude to Conservation Biology and Biodiversity from *General Biology* by OpenStax (licensed under CC-BY)
- Biodiversity and Biomes from *Life Sciences Grade 10* by Siyavula (CC-BY)

This page titled 16.1: Introduction to Biodiversity is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **47.1: The Biodiversity Crisis** by OpenStax is licensed CC BY 4.0.
- **21.1: Importance of Biodiversity** by OpenStax is licensed CC BY 4.0.

# 16.2: Measures of Biodiversity Loss

## Extinct and Threatened Animals

The **International Union for the Conservation of Nature (IUCN)** coordinates efforts to catalog and preserve biodiversity worldwide. One way scientists gauge trends in biodiversity is by monitoring the fate of individual species. Since 1964, the IUCN has compiled information in the **Red List of Threatened Species**, which includes plants, animals, fungi, and selected brown algae species. The Red List is supported by scientific research. In 2011, the list contained 61,000 species, all with supporting documentation. Updates to the Red List are released every four years.

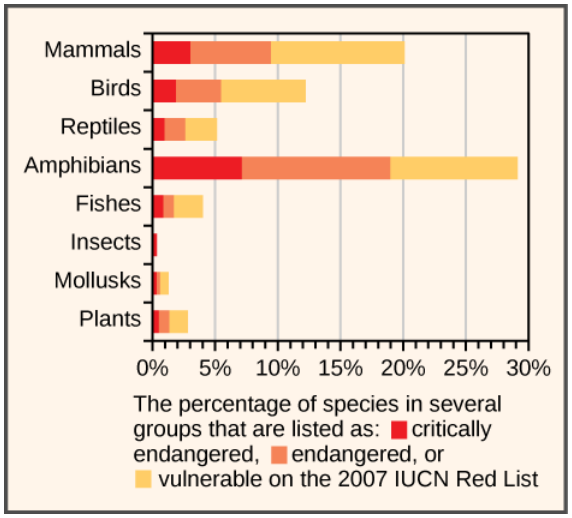



**Figure 16.2.1:** This chart shows the percentage of various animal species, by group, on the IUCN Red List as of 2007.

Species can be classified into nine Red List categories based on their extinction risk. Firstly, there are species that are already **extinct** and those that are **extinct in the wild**, meaning that the remaining individuals are only found in captivity.

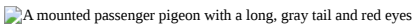



**Figure 16.2.2:** North American passenger pigeons lived in enormous flocks and were once the most numerous birds on earth. Market hunting on a massive scale and habitat destruction combined to extinguish them as a species in the early twentieth century.

Species at risk of extinction are called **threatened**. Species that are at risk of becoming threatened are called **near threatened**. The Emperor Penguin (*Aptenodytes forsteri*) is an example a near threatened species, mainly due to habitat loss and climate change (figure **16.2.3**).



**Figure 16.2.3:** The Emperor Penguin is an example of near threatened species. Image by Christopher Michel (CC-BY).

Those with little risk of extinction are designated **least concern**. Note that only a fraction of the 8-11 million species on Earth are identified (see The Number of Species on Earth). For many of the species that are identified, data still needs to be gathered before they can be assigned to a Red List category (**data deficient**). Less than 10% of the approximately 1.5 million identified species have been assessed for the Red List at all. Species that have not been assessed are deemed **not evaluated**.

There are three categories of threatened species: **vulnerable, endangered,** and **critically endangered**. Of these, critically endangered species have the greatest risk while vulnerable species are at the least risk out of the threatened categories. The African elephant (*Loxodonta africana*) is listed vulnerable species due to poaching and habitat loss (figure **16.2.4**). Interestingly, the forest subspecies (*L. africana cylotis*) is at much greater risk than the savanna subspecies (*L. africana africana*), but the Red List category is assigned at the species level in this case. The blue whale (*Balaenoptera musculus*) is endangered due to pollution, climate change, and poaching (figure **16.2.5**). Fortunately, blue whale population size is increasing. Several species of pitcher plants, which trap insects as a source of nitrogen, are critically endangered due to overexploitation and habitat loss (figure **16.2.6**).



**Figure 16.2.4:** The African forest elephant (*Loxodonta africana cylotis*) is at greater risk of extinction than the savanna subspecies. When both subspecies are grouped together, African elephants are considered vulnerable. Image by Peter H. Wrege (CC-BY-SA).



**Figure 16.2.5:** The blue whale (*Balaenoptera musculus*) is categorized as endangered according to the IUCN Red List. Image by Mike Baird (CC-BY).



**Figure 16.2.6:** *Nepenthes aristolochioides* is a critically endangered pitcher plant. Image by Alfindra Primaldhi (CC-BY).

Scientists know much more about the state of vertebrates—especially mammals, birds, and amphibians—than they do about other forms of animal life. Every one of the 5,488 species of mammals that have been described, for example, has been evaluated for purposes of the Red List. Of them, 76 species have become extinct since 1500, and two, Pere David's deer, which is native to China, and the scimitar oryx from Africa survive only in managed facilities. Another 29 of the mammal species listed as critically endangered are also tagged as "possibly extinct;" they are very likely gone, but the sort of exhaustive surveys required to confirm that fact have not been conducted. Overall, approximately 22% of mammal species worldwide are known to be threatened or extinct.

The Red List categorizes a smaller proportion of the world's 9,990 described bird species—14%—as threatened or extinct. But the raw number of species lost since 1500 is at least 134, and four more species persist only in zoos. Another 15 species of birds are considered possibly extinct. The fact that 86% of bird species are categorized as "not threatened" is good news in the context of the Red List.

Among the well-studied vertebrates, amphibians are faring especially poorly. Of the more than 6,000 known species of amphibians, 38 have become extinct worldwide since 1500, and another one, the Wyoming toad, survives only in captivity. Another 120 species are considered possibly extinct. Overall, 2,030, or one-third of the world's amphibian species are known to be threatened or extinct. More troubling still, many amphibian species—42.5%—are reported to be declining, and that number is probably low, since trend information is unavailable for 30.4% of species.

A Monteverde golden toad sits on a leaf.

**Figure 16.2.7:** The extinct golden toad of Monteverde, Costa Rica, was last seen in 1989. Its extinction is attributed to fungal infection combined with especially hot, dry weather. Source: U.S. Fish and Wildlife Service via Wikimedia Commons

Only small proportions of the world's species of reptiles and fish have been evaluated for purposes of the Red List. Among those, 34% of selected reptiles and 8% of selected bony fishes are threatened. Among those, the numbers of species that fall into the threatened category are very high: 1,275 of the 3,481 evaluated species, or 37%, for fish; and 423 of 1,385 evaluated species, or 31%, for reptiles. It should be noted, however, that these percentages are likely overestimates, since species of concern are more likely to be selected for evaluation than others.

The category "invertebrates" lumps together the vast majority of multi-cellular animals, an estimated 97% of all species. They include everything from insects and arachnids, to mollusks, crustaceans, corals, and more. Few of these groups have been assessed in a comprehensive way, but assessments within some groups call attention to disturbing, large-scale trends. For example, 33% of the world's reef-building corals are already considered threatened, and many more of them are experiencing rates of decline that move them toward threatened status (figure 16.2.8). The demise of reef-building corals has magnified ecological impacts, since so much other marine life depends on them.




**Figure 16.2.8:** Reef-formal corals, such this pink soft coral, are at high risk of extinction. Source: Linda Wade via National Oceanic & Atmospheric Administration (NOAA)

It should be understood that information about familiar creatures such as amphibians, mammals, and birds is just a beginning, and that even with the inclusion of some invertebrates the Red List does not provide a comprehensive picture of life on Earth. Scientists have described fewer than 2 million of the 8-9 million species of organisms thought to exist, most of which are insects. And of those 2 million, the status of only 44,838 has been assessed by IUCN.

In addition, it should be understood that among the species that have been assessed so far, there is a strong bias toward terrestrial vertebrates and plants, especially the ones that occur where biologists have visited frequently. Red List assessments also tend to focus on species that are likely to be threatened, since the effort also has the aim of enabling people to conserve species.

Whereas extinction is the global loss of a species, the elimination of species at a local level–known as **extirpation** – also poses threats to the integrity and sustainability of ecosystems. Widespread extirpation obviously leads to threatened or endangered status, but absence of species, even at a local scale, can affect ecosystem function. For example, by the mid-1920s wolves had been extirpated from Yellowstone National Park, although they continued to thrive elsewhere. When wolves were reintroduced to the park in the mid-1990s, numbers of elk (a main prey item) decreased significantly. This, in turn, reduced browsing pressure and had a significant effect on the vegetation and plant communities. What mattered for ecosystem function in Yellowstone was whether wolves were present *there*, not just whether the species survived somewhere.

## Biodiversity Change through Geological Time

The number of species on the planet, or in any geographical area, is the result of an equilibrium of two evolutionary processes that are ongoing: speciation and extinction. Both are natural “birth” and “death” processes of macroevolution. When speciation rates begin to outstrip extinction rates, the number of species will increase; likewise, the number of species will decrease when extinction rates begin to overtake speciation rates. Throughout Earth’s history, these two processes have fluctuated—sometimes leading to dramatic changes in the number of species on Earth as reflected in the fossil record (Figure **16.2.9**).

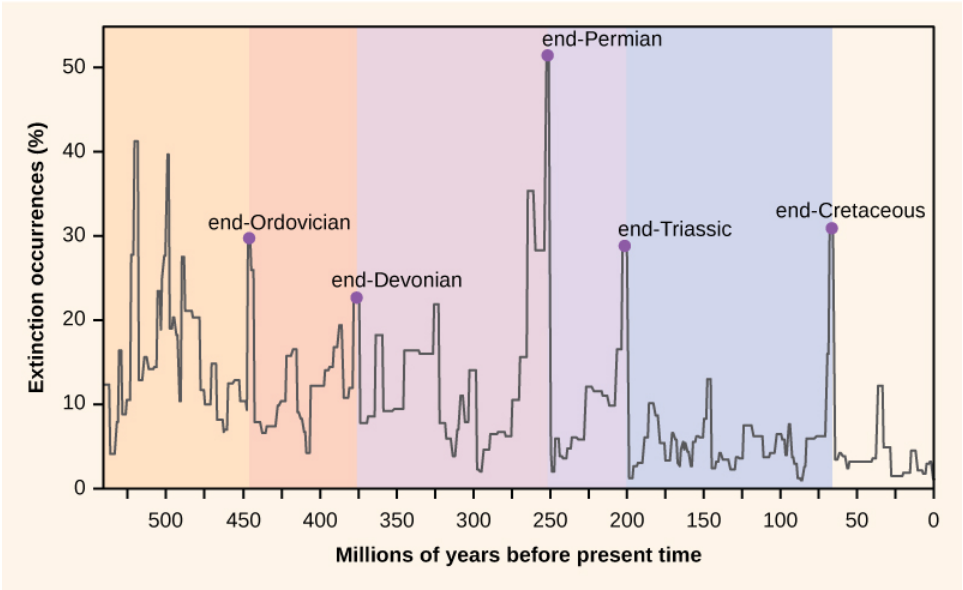



**Figure 16.2.9:** Percent extinction occurrences as reflected in the fossil record have fluctuated throughout Earth’s history. Sudden and dramatic losses of biodiversity, called mass extinctions, have occurred five times.

Paleontologists have identified five strata in the fossil record that appear to show sudden and dramatic (greater than half of all extant species disappearing from the fossil record) losses in biodiversity. These are called mass extinctions. There are many lesser, yet still dramatic, extinction events, but the five mass extinctions have attracted the most research. An argument can be made that the five mass extinctions are only the five most extreme events in a continuous series of large extinction events throughout the

Phanerozoic (since 542 million years ago). In most cases, the hypothesized causes are still controversial; however, the most recent event seems clear.

The fossil record of the mass extinctions was the basis for defining periods of geological history, so they typically occur at the transition point between geological periods. The transition in fossils from one period to another reflects the dramatic loss of species and the gradual origin of new species. These transitions can be seen in the rock strata. Table **16.2.2** provides data on the five mass extinctions.

**Table 16.2.2:** Mass extinctions: This table shows the names and dates for the five mass extinctions in Earth's history.

| Geological Period | Mass Extinction Name | Time (millions of years ago) |
|---|---|---|
| Ordovician–Silurian | end-Ordovician O–S | 450–440 |
| Late Devonian | end-Devonian | 375–360 |
| Permian–Triassic | end-Permian | 251 |
| Triassic–Jurassic | end-Triassic | 205 |
| Cretaceous–Paleogene | end-Cretaceous K–Pg (K–T) | 65.5 |

The Ordovician-Silurian extinction event is the first recorded mass extinction and the second largest. During this period, about 85 percent of marine species (few species lived outside the oceans) became extinct. The main hypothesis for its cause is a period of glaciation and then warming. The extinction event actually consists of two extinction events separated by about 1 million years. The first event was caused by cooling, and the second event was due to the subsequent warming. The climate changes affected temperatures and sea levels. Some researchers have suggested that a gamma-ray burst, caused by a nearby supernova, is a possible cause of the Ordovician-Silurian extinction. The gamma-ray burst would have stripped away the Earth's ozone layer causing intense ultraviolet radiation from the sun and may account for climate changes observed at the time. The hypothesis is speculative, but extraterrestrial influences on Earth's history are an active line of research. Recovery of biodiversity after the mass extinction took from 5 to 20 million years, depending on the location.

The late Devonian extinction may have occurred over a relatively long period of time. It appears to have affected marine species and not the plants or animals inhabiting terrestrial habitats. The causes of this extinction are poorly understood.

The end-Permian extinction was the largest in the history of life. Indeed, an argument could be made that Earth nearly became devoid of life during this extinction event. The planet looked very different before and after this event. Estimates are that 96 percent of all marine species and 70 percent of all terrestrial species were lost. It was at this time, for example, that the trilobites, a group that survived the Ordovician–Silurian extinction, became extinct. The causes for this mass extinction are not clear, but the leading suspect is extended and widespread volcanic activity that led to a runaway global-warming event. The oceans became largely anoxic, suffocating marine life. Terrestrial tetrapod diversity took 30 million years to recover after the end-Permian extinction. The Permian extinction dramatically altered Earth's biodiversity makeup and the course of evolution.

The causes of the Triassic–Jurassic extinction event are not clear and hypotheses of climate change, asteroid impact, and volcanic eruptions have been argued. The extinction event occurred just before the breakup of the supercontinent Pangaea, although recent scholarship suggests that the extinctions may have occurred more gradually throughout the Triassic.

The causes of the end-Cretaceous extinction event are the ones that are best understood. It was during this extinction event about 65 million years ago that the dinosaurs, the dominant vertebrate group for millions of years, disappeared from the planet (with the exception of a theropod clade that gave rise to birds). Indeed, every land animal that weighed more then 25 kg became extinct. The cause of this extinction is now understood to be the result of a cataclysmic impact of a large meteorite, or asteroid, off the coast of what is now the Yucatán Peninsula. This hypothesis, proposed first in 1980, was a radical explanation based on a sharp spike in the levels of iridium (which rains down from space in meteors at a fairly constant rate but is otherwise absent on Earth's surface) at the rock stratum that marks the boundary between the Cretaceous and Paleogene periods (Figure **16.2.10**). This boundary marked the disappearance of the dinosaurs in fossils as well as many other taxa. The researchers who discovered the iridium spike interpreted it as a rapid influx of iridium from space to the atmosphere (in the form of a large asteroid) rather than a slowing in the deposition of sediments during that period. It was a radical explanation, but the report of an appropriately aged and sized impact crater in 1991

made the hypothesis more believable. Now an abundance of geological evidence supports the theory. Recovery times for biodiversity after the end-Cretaceous extinction are shorter, in geological time, than for the end-Permian extinction, on the order of 10 million years.

Art Connection



**Figure 16.2.10:** In 1980, Luis and Walter Alvarez, Frank Asaro, and Helen Michels discovered, across the world, a spike in the concentration of iridium within the sedimentary layer at the K–Pg boundary. These researchers hypothesized that this iridium spike was caused by an asteroid impact that resulted in the K–Pg mass extinction. In the photo, the iridium layer is the light band. (credit: USGS)

Link to Learning

Explore this interactive website about mass extinctions.

## The Pleistocene Extinction

The Pleistocene Extinction is one of the lesser extinctions, and a recent one. It is well known that the North American, and to some degree Eurasian, megafauna, or large animals, disappeared toward the end of the last glaciation period. The extinction appears to have happened in a relatively restricted time period of 10,000–12,000 years ago. In North America, the losses were quite dramatic and included the woolly mammoths (last dated about 4,000 years ago in an isolated population), mastodon, giant beavers, giant ground sloths, saber-toothed cats, and the North American camel, just to name a few. The possibility that the rapid extinction of these large animals was caused by over-hunting was first suggested in the 1900s. Research into this hypothesis continues today. It seems likely that over-hunting caused many pre-written history extinctions in many regions of the world.

In general, the timing of the Pleistocene extinctions correlated with the arrival of humans and not with climate-change events, which is the main competing hypothesis for these extinctions. The extinctions began in Australia about 40,000 to 50,000 years ago, just after the arrival of humans in the area: a marsupial lion, a giant one-ton wombat, and several giant kangaroo species disappeared. In North America, the extinctions of almost all of the large mammals occurred 10,000–12,000 years ago. All that are left are the smaller mammals such as bears, elk, moose, and cougars. Finally, on many remote oceanic islands, the extinctions of many species occurred coincident with human arrivals. Not all of the islands had large animals, but when there were large animals, they were lost. Madagascar was colonized about 2,000 years ago and the large mammals that lived there became extinct. Eurasia and Africa do not show this pattern, but they also did not experience a recent arrival of humans. Humans arrived in Eurasia hundreds of thousands of years ago after the origin of the species in Africa. This topic remains an area of active research and hypothesizing. It seems clear that even if climate played a role, in most cases human hunting precipitated the extinctions.

## Present-Time Extinctions

The sixth, or Holocene, mass extinction appears to have begun earlier than previously believed and has mostly to do with the activities of *Homo sapiens*. Since the beginning of the Holocene period, there are numerous recent extinctions of individual species that are recorded in human writings. Most of these are coincident with the expansion of the European colonies since the 1500s.

One of the earlier and popularly known examples is the dodo bird. The dodo bird lived in the forests of Mauritius, an island in the Indian Ocean. The dodo bird became extinct around 1662. It was hunted for its meat by sailors and was easy prey because the dodo, which did not evolve with humans, would approach people without fear. Introduced pigs, rats, and dogs brought to the island by European ships also killed dodo young and eggs (Figure **16.2.11**).



**Figure 16.2.11.** The dodo bird was hunted to extinction around 1662. (credit: Ed Uthman, taken in Natural History Museum, London, England)

Steller's sea cow became extinct in 1768; it was related to the manatee and probably once lived along the northwest coast of North America. Steller's sea cow was first discovered by Europeans in 1741 and was hunted for meat and oil. The last sea cow was killed in 1768. That amounts to 27 years between the sea cow's first contact with Europeans and extinction of the species.

In 1914, the last living passenger pigeon died in a zoo in Cincinnati, Ohio. This species had once darkened the skies of North America during its migrations, but it was hunted and suffered from habitat loss through the clearing of forests for farmland. In 1918, the last living Carolina parakeet died in captivity. This species was once common in the eastern United States, but it suffered from habitat loss. The species was also hunted because it ate orchard fruit when its native foods were destroyed to make way for farmland. The Japanese sea lion, which inhabited a broad area around Japan and the coast of Korea, became extinct in the 1950s due to fishermen. The Caribbean monk seal was distributed throughout the Caribbean Sea but was driven to extinction via hunting by 1952.

These are only a few of the recorded extinctions in the past 500 years. The International Union for Conservation of Nature (IUCN) keeps a list of extinct and endangered species called the Red List. The list is not complete, but it describes 380 extinct species of vertebrates after 1500 AD, 86 of which were driven extinct by overhunting or overfishing.

## Estimates of Present-Time Extinction Rates

Estimates of extinction rates are hampered by the fact that most extinctions are probably happening without observation. The extinction of a bird or mammal is likely to be noticed by humans, especially if it has been hunted or used in some other way. But there are many organisms that are of less interest to humans (not necessarily of less value) and many that are undescribed.

The background extinction rate is estimated to be about one per million species per year (E/MSY). For example, assuming there are about ten million species in existence, the expectation is that ten species would become extinct each year (each year represents ten

million species per year).

One contemporary extinction rate estimate uses the extinctions in the written record since the year 1500. For birds alone this method yields an estimate of 26 E/MSY. However, this value may be underestimated for three reasons. First, many species would not have been described until much later in the time period, so their loss would have gone unnoticed. Second, the number of recently extinct species is increasing because extinct species now are being described from skeletal remains. And third, some species are probably already extinct even though conservationists are reluctant to name them as such. Taking these factors into account raises the estimated extinction rate closer to 100 E/MSY. The predicted rate by the end of the century is 1500 E/MSY.

A second approach to estimating present-time extinction rates is to correlate species loss with habitat loss by measuring forest-area loss and understanding species-area relationships. The species-area relationship is the rate at which new species are seen when the area surveyed is increased. Studies have shown that the number of species present increases as the size of the island increases. This phenomenon has also been shown to hold true in other habitats as well. Turning this relationship around, if the habitat area is reduced, the number of species living there will also decline. Estimates of extinction rates based on habitat loss and species-area relationships have suggested that with about 90 percent habitat loss an expected 50 percent of species would become extinct. Species-area estimates have led to species extinction rate calculations of about 1000 E/MSY and higher. In general, actual observations do not show this amount of loss and suggestions have been made that there is a delay in extinction. Recent work has also called into question the applicability of the species-area relationship when estimating the loss of species. This work argues that the species-area relationship leads to an overestimate of extinction rates. A better relationship to use may be the endemics-area relationship. Using this method would bring estimates down to around 500 E/MSY in the coming century. Note that this value is still 500 times the background rate.

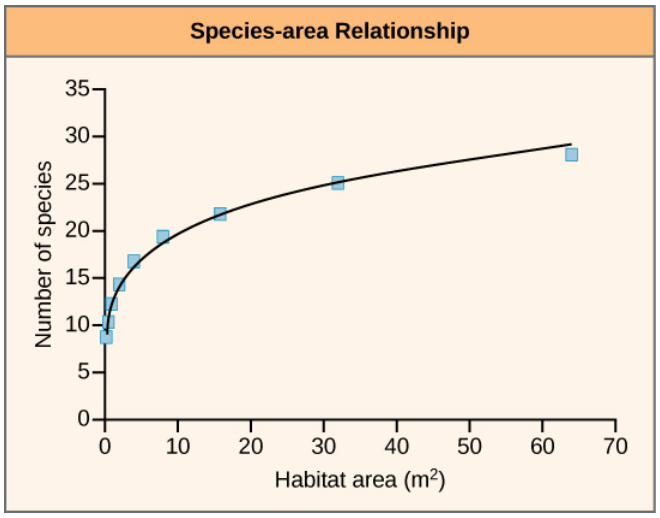



**Figure 16.2.12:** Studies have shown that the number of species present increases with the size of the habitat. (credit: modification of work by Adam B. Smith)

## References


Birds of the World. 2020. The Cornell Lab of Ornithology. Accessed 2020-07-29.

The IUCN Red List of Threatened Species. 2020. IUCN. Accessed 2020-07-29.

Mammal Diversity Database. 2020. American Society of Mammalogists. Accessed 2020-07-29.

1. 1 Mora Camilo et al., “How Many Species Are There on Earth and in the Ocean?” *PLoS Biology* (2011), doi:10.1371/journal.pbio.1001127.
2. 2 Arthur D. Chapman, *Numbers of Living Species in Australia and the World*, 2nd ed. (Canberra, AU: Australian Biological Resources Study, 2009). www.environment.gov.au/biodiv...d-complete.pdf.
3. 3 Brian Groombridge and Martin D. Jenkins. *World Atlas of Biodiversity: Earth’s Living Resources in the 21st Century*. Berkeley: University of California Press, 2002.
4. 4 International Institute for Species Exploration (IISE), *2011 State of Observed Species (SOS)*. Tempe, AZ: IISE, 2011. Accessed May, 20, 2012. species.asu.edu/SOS.

5. Brendan B. Larsen, Elizabeth C. Miller, Matthew K. Rhodes, and John J. Wiens, "Inordinate Fondness Multiplied and Redistributed: the Number of Species on Earth and the New Pie of Life," The Quarterly Review of Biology 92, no. 3 (September 2017): 229-265. DOI

6. Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., et al. (2011, March). Has the Earth's sixth mass extinction already arrived? *Nature*, *471*, 51-57. doi:10.1038/nature09678

   Vie, J-C, Hilton-Taylor, C. & Stuart S.N. (Eds.). (2009). *Wildlife in a Changing World: An Analysis of the 2008 IUCN Red List of Threatened Species*™. Gland, Switzerland: IUCN. Retrieved from http://data.iucn.org/dbtw-wpd/edocs/RL-2009-001.pdf.

## Contributors and Attributions

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

- **5.5: Preserving Biodiversity**, Community Ecology and Importance of Biodiversity from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Preserving Biodiversity and **47.1: The Biodiversity Crisis** by OpenStax, is licensed CC BY by

  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).
- Biodiversity, Species Loss, and Ecosystem Function and The Industrialization of Nature: A Modern History (1500 to the present) from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors. Download for free at CNX. (licensed under CC-BY)

---

This page titled 16.2: Measures of Biodiversity Loss is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by Melissa Ha and Rachel Schleiger (ASCCC Open Educational Resources Initiative) .

- **47.1: The Biodiversity Crisis** by OpenStax is licensed CC BY 4.0.

# 16.3: Threats to Biodiversity

The core threat to biodiversity on the planet, and therefore a threat to human welfare, is the combination of human population growth and resource exploitation. The human population requires resources to survive and grow, and those resources are being removed unsustainably from the environment. The three greatest proximate threats to biodiversity are **habitat loss**, **overharvesting**, and introduction of **exotic species**. The first two of these are a direct result of **human population growth** and resource use. The third results from increased mobility and trade. A fourth major cause of extinction, anthropogenic **climate change**, has not yet had a large impact, but it is predicted to become significant during this century. Global climate change is also a consequence of human population needs for energy and the use of fossil fuels to meet those needs (Figure **16.3.1**). Environmental issues, such as toxic **pollution**, have specific targeted effects on species, but they are not generally seen as threats at the magnitude of the others.

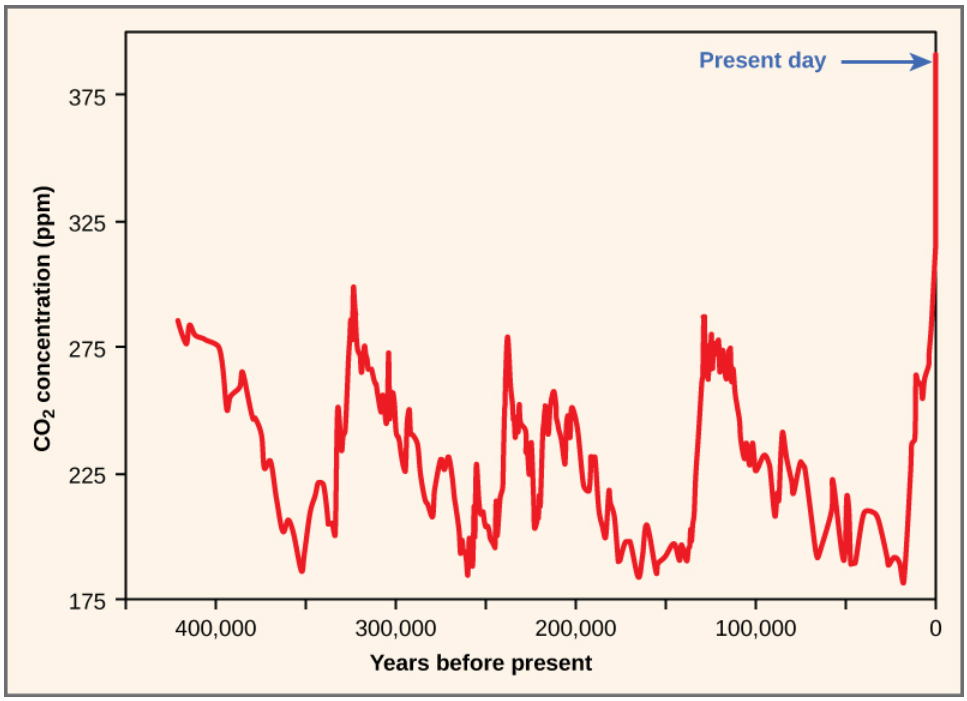



Figure **16.3.1**: Atmospheric carbon dioxide levels fluctuate in a cyclical manner. However, the burning of fossil fuels in recent history has caused a dramatic increase in the levels of carbon dioxide in the Earth's atmosphere, which have now reached levels never before seen in human history. Scientists predict that the addition of this "greenhouse gas" to the atmosphere is resulting in climate change that will significantly impact biodiversity in the coming century.

## Habitat Loss

Humans rely on technology to modify their environment and replace certain functions that were once performed by the natural ecosystem. Other species cannot do this. Elimination of their ecosystem—whether it is a forest, a desert, a grassland, a freshwater estuarine, or a marine environment—will kill the individuals in the species. Remove the entire habitat within the range of a species and, unless they are one of the few species that do well in human-built environments, the species will become extinct. Human destruction of habitats accelerated in the latter half of the twentieth century. Consider the exceptional biodiversity of Sumatra: it is home to one species of orangutan, a species of critically endangered elephant, and the Sumatran tiger, but half of Sumatra's forest is now gone. The neighboring island of Borneo, home to the other species of orangutan, has lost a similar area of forest. Forest loss continues in protected areas of Borneo. The orangutan in Borneo is listed as endangered by the International Union for Conservation of Nature (IUCN), but it is simply the most visible of thousands of species that will not survive the disappearance of the forests of Borneo. The forests are removed for timber and to plant palm oil plantations (Figure **16.3.2**). Palm oil is used in many products including food products, cosmetics, and biodiesel in Europe.



(a) (b)




(c)



(d) (e)

Figure **16.3.2**: (a) One species of orangutan, *Pongo pygmaeus*, is found only in the rainforests of Borneo, and the other species of orangutan (*Pongo abelii*) is found only in the rainforests of Sumatra. These animals are examples of the exceptional biodiversity of (c) the islands of Sumatra and Borneo. Other species include the (b) Sumatran tiger (*Panthera tigris sumatrae*) and the (d) Sumatran elephant (*Elephas maximus sumatranus*), both critically endangered species. Rainforest habitat is being removed to make way for (e) oil palm plantations such as this one in Borneo's Sabah Province. (credit a: modification of work by Thorsten Bachner; credit b: modification of work by Dick Mudde; credit c: modification of work by U.S. CIA World Factbook; credit d: modification of work by "Nonprofit Organizations"/Flickr; credit e: modification of work by Dr. Lian Pin Koh)

According to Global Forest Watch, 9.7% of tree cover was lost globally from 2002 to 2019, and 9% of that occurred in Indonesia and Malaysia (where Sumatra and Borneo are located). Figure **16.3.*b*** shows the average annual change in forest area around the

world from 1990 to 2015. In the tropics, these losses certainly also represent the extinction of species because of high levels of endemism (species unique to those areas).

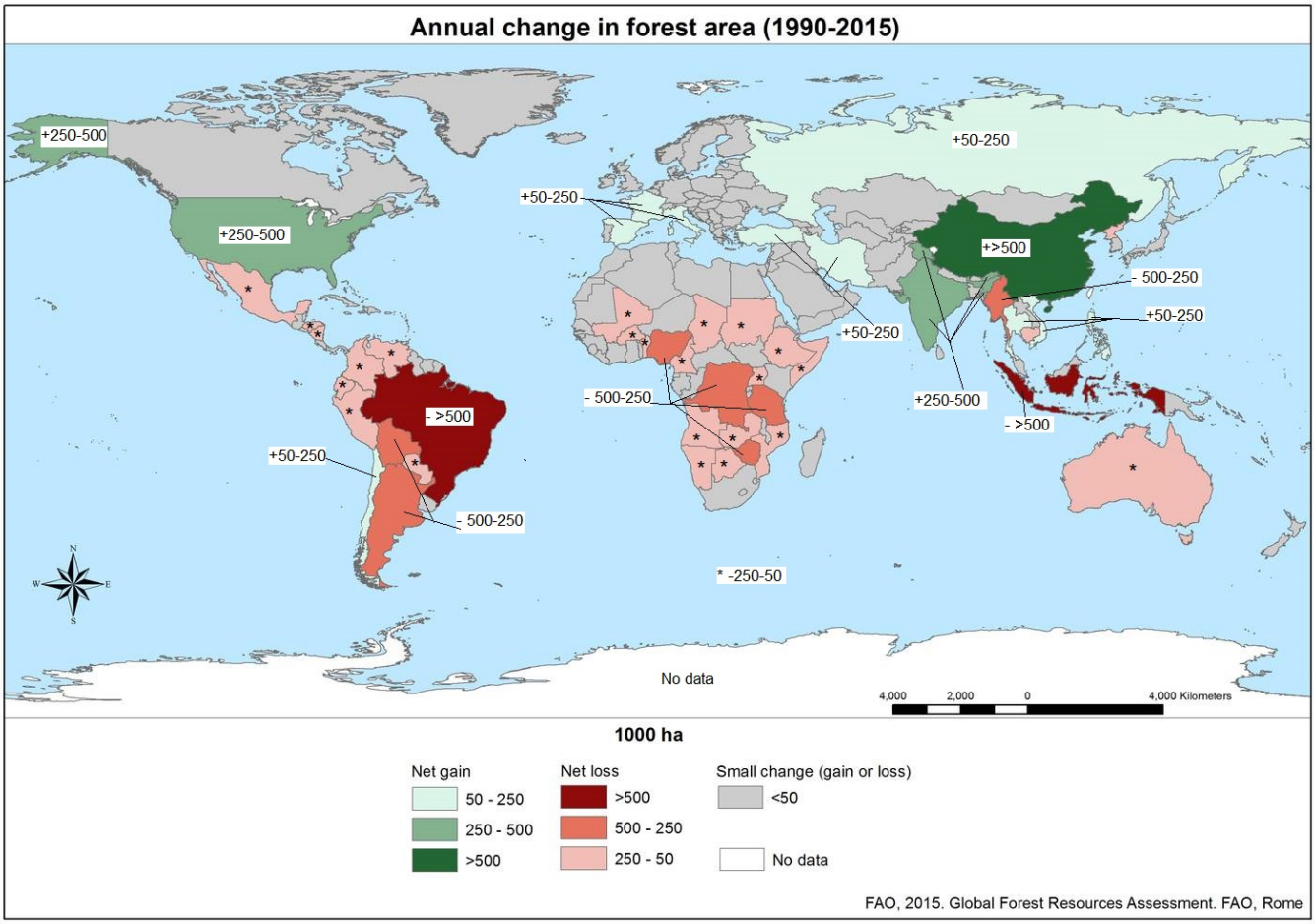



**Figure 16.3.*b***: Average annual change in forest area globally from 2005 to 2015. China is dark green, indicating that have gained more than 500 "kilohectares" (kha) of forest. Medium green indicates countries that have gained 250-500 kha, including the the United States, India, Ghutan, and Bangladesh. Light green indicates countries that have gained 50-250 khA, including Chile, Spain, France, Italy, Turkey, Iran, Russia, Vietnam, and Thailand. Indonesia and Brazil lost more than 500 kha of forest (marked with dark red). Countries that lost 500-250 kha are marked with medium red, including Bolivia, Argentina, Nigeria, Democratic Republic of the Congo, Tanzania, Zimbabwe, and Myanmar. Countries that lost 250-50 kha are marked with light red and an asterisk (*). These include Mexico, Honduras, Nicaragua, Venezuela, Colombia, Ecuador, Peru, Paraguay, Mali, Burkina Faso, Benin, Cameroon, Chad, Sudan, Ethiopia, Somalia, Uganda, Angola, Namibia, Botswana, Zimbabwe, Mozambique, Cambodia, North Korea, and Australia. No data has been collected for Antarctica. All other regions have had lost or gained less than 50 kha in forest area. Image by FAO, 2015. Global Forest Resources Assessment. FAO. Rome (Open Access Policy).

## Everyday Connection: Preventing Habitat Destruction with Wise Wood Choices

Most consumers do not imagine that the home improvement products they buy might be contributing to habitat loss and species extinctions. Yet the market for illegally harvested tropical timber is huge, and the wood products often find themselves in building supply stores in the United States. One estimate is that up to 10% of the imported timber stream in the United States, which is the world's largest consumer of wood products, is illegally logged. A 2012 United Nations and Interpol report estimated that the illegal timber trade is worth $30-100 billion each year. Most of the illegal products are imported from countries that act as intermediaries and are not the originators of the wood.



**Figure 16.3.*c*:** The sawmill in Uaxactun, Guatamala is Forest Stewardship Council (FSC) certified and provides good income from a sustainable resource for not only the saw operators but also many others who help keep the operation running. Image by Jason Houston for USAID (public domain).

How is it possible to determine if a wood product, such as flooring, was harvested sustainably or even legally? The Forest Stewardship Council (FSC) certifies sustainably harvested forest products; therefore, looking for their certification on flooring and other hardwood products is one way to ensure that the wood has not been taken illegally from a tropical forest (figure **16.3.*c***). Certification applies to specific products, not to a producer; some producers' products may not have certification while other products are certified. There are certifications other than the FSC, but these are run by timber companies creating a conflict of interest. Another approach is to buy domestic wood species. While it would be great if there was a list of legal versus illegal woods, it is not that simple. Logging and forest management laws vary from country to country; what is illegal in one country may be legal in another. Where and how a product is harvested and whether the forest from which it comes is being sustainably maintained all factor into whether a wood product will be certified by the FSC. It is always a good idea to ask questions about where a wood product came from and how the supplier knows that it was harvested legally.

Habitat destruction can affect ecosystems other than forests. Worldwide, for example, the conversion of land to agriculture and cultivation have led to significant losses in grassland ecosystems. In North America, nearly 70% of the tallgrass prairie ecosystem (which once covered 142 million acres) has been converted to agriculture, and losses from other causes, such as urban development, have brought the total to about 90%. Current estimates indicate that agricultural activity and cultivation systems now cover nearly 25% of the Earth's surface.

Rivers and streams are important ecosystems and are frequently the target of habitat modification through building and from damming or water removal. Damming of rivers affects flows and access to all parts of a river. Altering a flow regime can reduce or eliminate populations that are adapted to seasonal changes in flow. For example, an estimated 91 percent of river lengths in the United States have been have modifications like dams, to create energy or store water; levees, to prevent flooding; or dredging or rerouting, to create land that is more suitable for human development. Many fish species in the United States, especially rare species or species with restricted distributions, have seen declines caused by river damming and habitat loss. The category "wetlands" includes many types of ecosystems, but current estimates indicate that about 50% of the world's wetland habitat has been lost. The former extent of wetland habitats worldwide (fresh, brackish and salt) is difficult to determine but certainly exceeded a billion acres.

Research has confirmed that species of amphibians that must carry out parts of their life cycles in both aquatic and terrestrial habitats are at greater risk of population declines and extinction because of the increased likelihood that one of their habitats or access between them will be lost. This is of particular concern because amphibians have been declining in numbers and going extinct more rapidly than many other groups for a variety of possible reasons.

Link to Learning

Explore a U.S. Fish & Wildlife Service interactive map of critical habitat for endangered and threatened species in the United States. To begin, select “Visit the online mapper.”

**Habitat fragmentation** occurs when the living space of a species is divided into discontinuous patches. For example, a mountain highway could divide a forest habitat into separate patches. This is especially problematic for consumers at the top of the food chain, which require large ranges to find adequate prey. Additionally, habitat fragmentation separates individuals from potential mates. **Wildlife corridors** mitigate the damage of habitat fragmentation by connecting patches with suitable habitat. For example, the bridge over a highway could allow animals to move between habitat patches (figure **16.3.*d***). Riparian areas, areas of land adjacent to bodies of water, such as streams, can serve as natural wildlife corridors when left intact.



**Figure 16.3.*d***: This overpass on the Trans-Canada Highway between Banff and Lake Louise, Alberta, serves as a wildlife corridor. Image by WikiPedant at Wikimedia Commons (CC-BY-SA).

## Overharvesting

**Overexploitation** (**overharvesting**) involves hunting, fishing, or otherwise collecting organisms at a faster rate than they can be replenished. Aquatic species are particularly vulnerable to overexploitation, which is more specifically called **overfishing** in this case. For about one billion people, aquatic resources provide the main source of animal protein (figure **16.3.*b***), but since 1990, production from global **fisheries** (areas for catching wild or farmed fish or other aquatic animals) has declined. This devastates the ability of coastal and open-ocean ecosystems to provide a wide range of ecosystem services such as food provisioning, water filtration, and detoxification (Worm et al., 2006). The collapse of fisheries also has dramatic and long-lasting effects on local human populations that work in the fishery.




**Figure 16.3.*b*:** Fishing boats at marine fishers. Alaskan waters have been fished by people for thousands of years, but they are under pressure from modern fishing technologies and large-scale extraction. Source: National Oceanic and Atmospheric Administration

There are many examples of regulated fisheries (including hunting of marine mammals and harvesting of crustaceans and other species) monitored by fisheries scientists that have nevertheless collapsed. The western Atlantic cod fishery is the most spectacular recent collapse. While it was a hugely productive fishery for 400 years, the introduction of modern factory trawlers in the 1980s and the pressure on the fishery led to it becoming unsustainable. Bluefin tuna are in danger of extinction. The once-abundant Mediterranean swordfish fishery have been depleted to commercial and biological exhaustion. Figure **16.3.*c*** illustrates the extent of overfishing in the U.S. Despite considerable effort, few fisheries are managed sustainability.

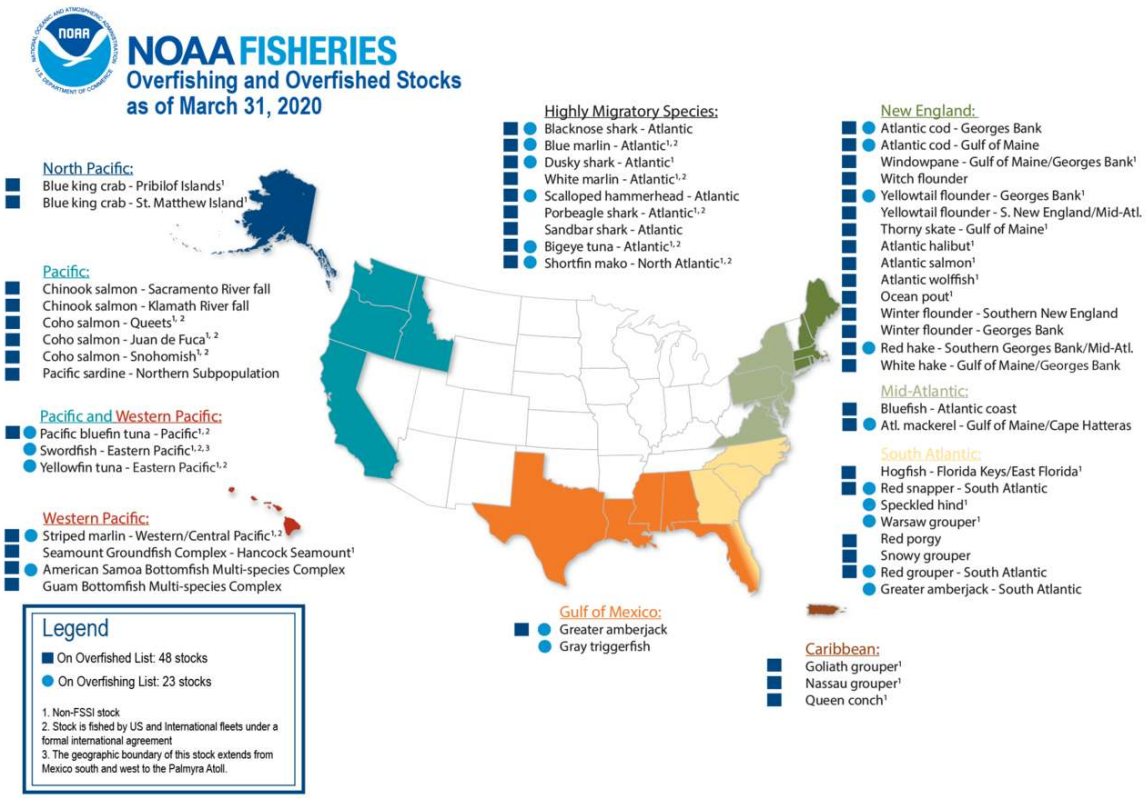



**Figure 16.3.*c*:** Map of overfishing and overfished stocks in the U.S. by region. Stocks on the overfishing list are being harvested too quickly, and those on the overfished list have population sizes that are too low. For example, stocks of Chinook salmon, Coho salmon, and Pacific sardines are overfished in the Pacific. Some species, including stocks of Pacific bluefin tuna and Atlantic cod, are on both the overfishing and overfished lists. Image by NOAA (public domain).

The causes of fishery collapse are both economic and political in nature. Most fisheries are managed as a common resource, available to anyone willing to fish, even when the fishing territory lies within a country's territorial waters. Common resources are subject to an economic pressure known as the **tragedy of the commons**, in which fishers have little motivation to exercise restraint in harvesting a fishery when they do not own the fishery. The general outcome of harvests of resources held in common is their overexploitation. While large fisheries are regulated to attempt to avoid this pressure, it still exists in the background. This overexploitation is exacerbated when access to the fishery is open and unregulated and when technology gives fishers the ability to overfish. In a few fisheries, the biological growth of the resource is less than the potential growth of the profits made from fishing if that time and money were invested elsewhere. In these cases—whales are an example—economic forces will drive toward fishing the population to extinction.

Early efforts to control overfishing used several kinds of regulations on quotas, fishing effort, and gear. Some forms of fishery management limit the number of fish that can be caught in an entire fishery. Under a total allowable catch (TAC) system, fishers can fish when and how they want, but once the quota for the fishery has been met, fishing must stop until the next season. Unfortunately, TAC policies do not solve the underlying problem that fishermen compete for the fish, and often yield perverse incentives and undesirable outcomes such as overcapitalization of the industry (Beddington, Agnew, & Clark, 2007) and races

between fishing boat crews to catch fish before the quota is reached. In the well-known case of the Alaskan halibut fishery, the race became so extreme that the fishing season was reduced to a single 24-hour mad dash; given that fish are perishable, this temporal clumping of the catch is not a desirable outcome. Resource economists developed the idea of a tradable permit scheme to help manage fisheries. Individual tradable quota (ITQ) schemes are cap-and-trade policies for fish, where total catch is limited but fishers in the fishery are given permits that guarantee them a right to a share of that catch. Players in the fishery can sell their quota shares to each other (helping the catch to flow voluntarily to the most efficient boats in the industry) and there is no incentive for captains to buy excessively large boats or fish too rapidly to beat the other boats to the catch. ITQ policies have rationalized the Alaskan halibut fishery completely: the fish stock is thriving, overcapitalization is gone, and the fish catch is spread out over time (Levy, 2010). ITQs have also been implemented in the fisheries of New Zealand, yielding large improvements in the biological status of the stocks (Annala, 1996).

For the most part, fishery extinction is not equivalent to biological extinction—the last fish of a species is rarely fished out of the ocean. But there are some instances in which true extinction is a possibility. Whales have slow-growing populations and are at risk of complete extinction through hunting. Also, there are some species of sharks with restricted distributions that are at risk of extinction. The groupers are another population of generally slow-growing fishes that, in the Caribbean, includes a number of species that are at risk of extinction from overfishing.

A related consequence of fishing practices is "**bycatch**," animals that fishers sometimes catch and discard because they do not want them, cannot sell them, or are not allowed to keep them. Bycatch can be fish, but also includes other animals such as dolphins, whales, sea turtles, and seabirds that become hooked or entangled in fishing gear. Fishing boats are forbidden in some places from using conventional longlines because that gear yields high levels of bycatch and kills endangered leatherback turtles.

Overfishing can result in a radical restructuring of the marine ecosystem in which a dominant species is so overexploited that it no longer serves its ecological function. For example, overfishing a tertiary consumer could causes populations of secondary consumers to increase. Secondary consumers would then feed on primary consumes (like zooplankton), decreasing their population size. With fewer zooplankton, populations of primary producers (**phytoplankton**, or photosynthetic microorganisms) would be unregulated.

Sustainable seafood is a movement that has gained momentum as more people become aware of overfishing and environmentally-destructive fishing methods. Sustainable seafood is seafood from either fished or farmed sources that can maintain or increase production in the future without jeopardizing the ecosystems from which it was acquired. In general, slow-growing fish that reproduce late in life, such as orange roughy, are vulnerable to overfishing and are considered unsustainable seafood. Seafood species that grow quickly and breed young, such as anchovies and sardines, are much more resistant to overfishing and are therefore labeled “sustainable” and promoted as good alternatives.

Coral reefs are extremely diverse marine ecosystems that face peril from several processes. Reefs are home to 1/3 of the world’s marine fish species—about 4,000 species—despite making up only 1 percent of marine habitat. Most home marine aquaria are stocked with wild-caught organisms, not cultured organisms. Although no species is known to have been driven extinct by the pet trade in marine species, there are studies showing that populations of some species have declined in response to harvesting, indicating that the harvest is not sustainable at those levels.

Terrestrial animals may be overexploited as sources of food, garments, jewelry, medicine, or pets. For example, the poaching of elephants for their valuable ivory and rhinos for their horns, which are used in traditional medicine, is a major threat to these species. There are also concerns about the effect of the pet trade on some terrestrial species such as turtles, amphibians, birds, plants, and even the orangutans. Harvesting of pangolins for their scales and meat, and as curiosities, has led to a drastic decline in population size (figure **16.3.*a***).



**Figure 16.3.*a*:** Pangolins are threatened by overexploitation. This work by David Brossard is licensed under CC-BY.

**Bush meat** is the generic term used for wild animals killed for food. Hunting is practiced throughout the world, but hunting practices, particularly in equatorial Africa and parts of Asia, are believed to threaten several species with extinction. Traditionally, bush meat in Africa was hunted to feed families directly. However, recent commercialization of the practice now has bush meat available in grocery stores, which has increased harvest rates to the level of unsustainability. Additionally, human population growth has increased the need for protein foods that are not being met from agriculture. Species threatened by the bush meat trade are mostly mammals including many monkeys and the great apes living in the Congo basin.

Some plant and fungal species are also overexploited, particularly if they are slow-growing. For example, stocks of wild ginseng, which is valued for its health benefits, are dwindling. Peyote cactus, which causes hallucinations and is used in sacred ceremonies, is also declining. Yarsagumba, dead moth larvae that were infected by fungal parasites (caterpillar fungus, *Ophiocordyceps sinensis*), is overexploited because it is highly valued in traditional medicine and used as an aphrodisiac (figure 16.3.*d*).



**Figure 16.3.*d*:** Yarsagumba is a combination of moth larvae and the fungus that infected and killed it. Image by Punya (CC-BY-SA).

## Exotic Species

**Non-native** (exotic) refers to species occurring outside of their historic distribution. Exotic species are species that have been intentionally or unintentionally introduced by humans into an ecosystem in which they did not evolve. If an introduced species is able to survive in its new habitat, that introduction is now reflected in the observed range of the species. Human transportation of

people and goods, including the intentional transport of organisms for trade, has dramatically increased the introduction of species into new ecosystems, sometimes at distances that are well beyond the capacity of the species to ever travel itself and outside the range of the species' natural predators.

Most exotic species introductions probably fail because of the low number of individuals introduced or poor adaptation to the ecosystem they enter. Some species, however, have characteristics that can make them especially successful in a new ecosystem. These exotic species often undergo dramatic population increases in their new habitat and reset the ecological conditions in the new environment, threatening the species that exist there. When this happens, the exotic species also becomes an invasive species. I Invasive species can cause ecological and economic damage. They threaten other species through competition for resources, predation, or disease. For example, Kudzu (*Pueraria lobata*), which is native to Japan, was introduced in the United States in 1876. It was later planted for soil conservation. Problematically, it grows too well in the southeastern United States—up to a foot a day. It is now a pest species and covers over 7 million acres in the southeastern United States. In the United States, invasive species like the purple loosestrife (*Lythrum salicaria*) and the zebra mussel (*Dreissena polymorpha*) have drastically altered the ecosystems they invaded. Some well-known invasive animals include the emerald ash borer (*Agrilus planipennis*) and the European starling (*Sturnus vulgaris*; figure **16.3.*a***). Whether enjoying a forest hike, taking a summer boat trip, or simply walking down an urban street, you have likely encountered an invasive species.

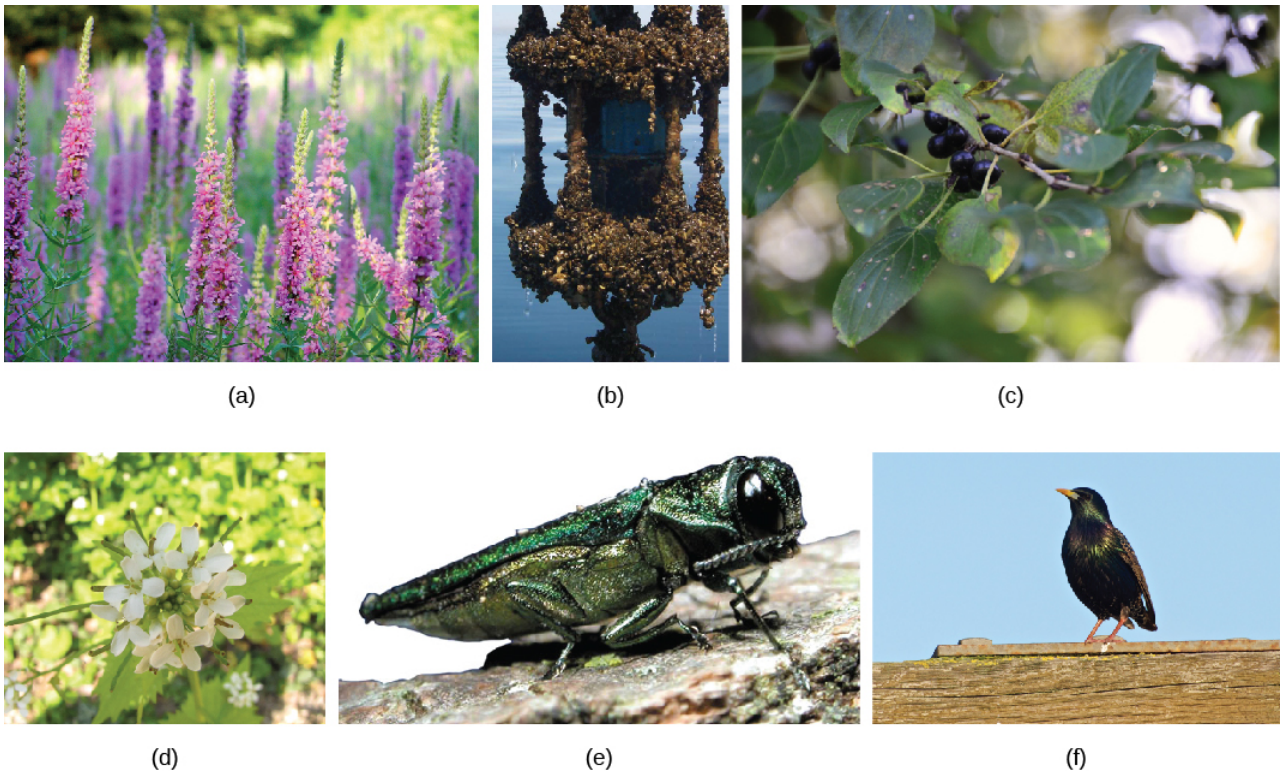



**Figure 16.3.*a*:** In the United States, invasive species like (a) purple loosestrife (*Lythrum salicaria*) and the (b) zebra mussel (*Dreissena polymorpha*) threaten certain aquatic ecosystems. Some forests are threatened by the spread of (c) common buckthorn (*Rhamnus cathartica*), (d) garlic mustard (*Alliaria petiolata)*, and (e) the emerald ash borer (*Agrilus planipennis*). The (f) European starling (*Sturnus vulgaris*) may compete with native bird species for nest holes. (credit a: modification of work by Liz West; credit b: modification of work by M. McCormick, NOAA; credit c: modification of work by E. Dronkert; credit d: modification of work by Dan Davison; credit e: modification of work by USDA; credit f: modification of work by Don DeBold)

Link to Learning

Explore an interactive global database of exotic or invasive species.

Lakes and islands are particularly vulnerable to extinction threats from introduced species. In Lake Victoria, the intentional introduction of the Nile perch was largely responsible for the extinction of about 200 species of cichlids. The accidental introduction of the brown tree snake via aircraft (Figure **16.3.3**) from the Solomon Islands to Guam in 1950 has led to the extinction of three species of birds and three to five species of reptiles endemic to the island. Several other species are still threatened. The brown tree snake is adept at exploiting human transportation as a means to migrate; one was even found on an aircraft arriving in Corpus Christi, Texas. Constant vigilance on the part of airport, military, and commercial aircraft personnel is required to prevent the snake from moving from Guam to other islands in the Pacific, especially Hawaii. Islands do not make up a large area of land on the globe, but they do contain a disproportionate number of endemic species because of their isolation from mainland ancestors.



**Figure 16.3.3:** The brown tree snake, *Boiga irregularis*, is an exotic species that has caused numerous extinctions on the island of Guam since its accidental introduction in 1950. (credit: NPS)

Many introductions of aquatic species, both marine and freshwater, have occurred when ships have dumped ballast water taken on at a port of origin into waters at a destination port. Water from the port of origin is pumped into tanks on a ship empty of cargo to increase stability. The water is drawn from the ocean or estuary of the port and typically contains living organisms such as plant parts, microorganisms, eggs, larvae, or aquatic animals. The water is then pumped out before the ship takes on cargo at the destination port, which may be on a different continent. The zebra mussel was introduced to the Great Lakes from Europe prior to 1988 in ship ballast. The zebra mussels in the Great Lakes have cost the industry millions of dollars in clean up costs to maintain water intakes and other facilities. The mussels have also altered the ecology of the lakes dramatically. They threaten native mollusk populations, but have also benefited some species, such as smallmouth bass. The mussels are filter feeders and have dramatically improved water clarity, which in turn has allowed aquatic plants to grow along shorelines, providing shelter for young fish where it did not exist before. The European green crab, *Carcinus maenas*, was introduced to San Francisco Bay in the late 1990s, likely in ship ballast water, and has spread north along the coast to Washington. The crabs have been found to dramatically reduce the abundance of native clams and crabs with resulting increases in the prey of native crabs.

One of the many recent proliferations of an invasive species concerns the Asian carp in the United States. Asian carp were introduced to the United States in the 1970s by fisheries (commercial catfish ponds) and by sewage treatment facilities that used the fish's excellent filter feeding abilities to clean their ponds of excess plankton. Some of the fish escaped, and by the 1980s they had

colonized many waterways of the Mississippi River basin, including the Illinois and Missouri Rivers. Voracious feeders and rapid reproducers, Asian carp may outcompete native species for food and could lead to their extinction. One species, the grass carp, feeds on phytoplankton and aquatic plants. It competes with **native** species (those that historically occurred in the area and are adapted to the local ecosystem) for these resources and alters habitats for other fish by removing aquatic plants. In some parts of the Illinois River, Asian carp constitute 95 percent of the community's biomass. Although edible, the fish is bony and not desired in the United States. The Great Lakes and their prized salmon and lake trout fisheries are being threatened by Asian carp. The carp are not yet present in the Great Lakes, and attempts are being made to prevent its access to the lakes through the Chicago Ship and Sanitary Canal, which is the only connection between the Mississippi River and Great Lakes basins. To prevent the Asian carp from leaving the canal, a series of electric barriers have been used to discourage their migration; however, the threat is significant enough that several states and Canada have sued to have the Chicago channel permanently cut off from Lake Michigan. Local and national politicians have weighed in on how to solve the problem. In general, governments have been ineffective in preventing or slowing the introduction of invasive species.

Invading exotic species can also be disease organisms. It now appears that the global decline in amphibian species recognized in the 1990s is, in some part, caused by the fungus *Batrachochytrium dendrobatidis*, which causes the disease chytridiomycosis (Figure **16.3.4**). There is evidence that the fungus is native to Africa and may have been spread throughout the world by transport of a commonly used laboratory and pet species: the African clawed toad (*Xenopus laevis*). It may well be that biologists themselves are responsible for spreading this disease worldwide. The North American bullfrog, *Rana catesbeiana*, which has also been widely introduced as a food animal but which easily escapes captivity, survives most infections of *Batrachochytrium dendrobatidis* and can act as a reservoir for the disease.



**Figure 16.3.4:** This Limosa Harlequin Frog (*Atelopus limosus*), an endangered species from Panama, died from a fungal disease called chytridiomycosis. The red lesions are symptomatic of the disease. (credit: Brian Gratwicke)

Early evidence suggests that another fungal pathogen, *Geomyces destructans*, introduced from Europe is responsible for white-nose syndrome, which infects cave-hibernating bats in eastern North America and has spread from a point of origin in western New York State (Figure **16.3.5**). The disease has decimated bat populations and threatens extinction of species already listed as endangered: the Indiana bat, *Myotis sodalis*, and potentially the Virginia big-eared bat, *Corynorhinus townsendii virginianus*. How the fungus was introduced is unclear, but one logical presumption would be that recreational cavers unintentionally brought the fungus on clothes or equipment from Europe.



**Figure 16.3.5:** This little brown bat in Greeley Mine, Vermont, March 26, 2009, was found to have white-nose syndrome. (credit: Marvin Moriarty, USFWS)

## Biological Control of Invasive Species

One reason why invasive species proliferate dramatically outside of their native range is due to **release from predators**. This means that parasites, predators, or herbivores that usually regulate their populations are not present, allowing them to outcompete or overpredate native species, which are still regulated. Based on this principle, organisms that regulate the invasive species populations have been introduced to the newly colonized areas in some cases. The release of organisms (or viruses) to limit population size is called **biological control**. As described in the examples below, biological control of invasive species has had varying success, exacerbating the problem in some cases and solving it in others.

Introduced into Australia, this Prickly-pear Cactus (*Opuntia*) soon spread over millions of hectares of range land driving out forage plants. In 1924, the cactus moth, *Cactoblastis cactorum*, was introduced (from Argentina) into Australia. The caterpillars of the moth are voracious feeders on prickly-pear cactus, and within a few years, the caterpillars had reclaimed the range land without harming a single native species. However, its introduction into the Caribbean in 1957 did not produce such happy results. By 1989, the cactus moth had reached Florida, and now threatens five species of native cacti there.

The leaf beetle (*Galerucella calmariensis*) has been introduced to suppress purple loosestrife, a noxious weed (figure 16.3.*e*). A combination of four biological controls, including the leaf beetle were released in Minnesota since 1992. While it has not eradicated populations of this invasive species, biological control largely removed leaves from 20% of the purple loosestrive populations where it was released, which could reduce competition for native species. The biological controls established populations in most locations where they were released and even spread to new patches of purple loosestrife.



**Figure 16.3.*e*:** Young larvae of the leaf-beetle feed in and on the developing buds of plants, often destroying them. This may stunt plant growth and delay or prevent flowering. Adults (shown) and older larvae feed on leaves and cause severe defoliation. Leaf-beetles can be used to as a biocontrol for invasive plants such as purple loosestrife.

In 1946 two species of *Chrysolina* beetles were introduced into California to control the Klamath weed (St. Johnswort) that was ruining millions of acres of range land in California and the Pacific Northwest. Before their release, the beetles were carefully tested to make certain that they would not turn to valuable plants once they had eaten all the Klamath weed they could find. The beetles succeeded beautifully, restoring about 99% of the endangered range land and earning them a commemorative plaque at the Agricultural Center Building in Eureka, California.

In 1859, the European rabbit was introduced into Australia for sport. With no important predator there, it multiplied explosively (figure **16.3.*f***). The raising of sheep (another imported species) suffered badly as the rabbits competed with them for forage. In 1950, the **myxoma virus** was brought from Brazil and released. The epidemic that followed killed off millions of rabbits (more than 99% of the population). Green grass returned, and sheep raising once again became profitable. Rabbit populations gradually increased, however, because the rabbits evolved to be more resistant to the virus, and the myxoma virus evolved to cause less damage. (Parasites, like viruses, benefit from multiplying inside the host and spreading to other individuals. If they kill their hosts too soon, they typically limit opportunities to multiply and spread.) More recently, the rabbit hemorrhagic disease virus has been used as biological control.

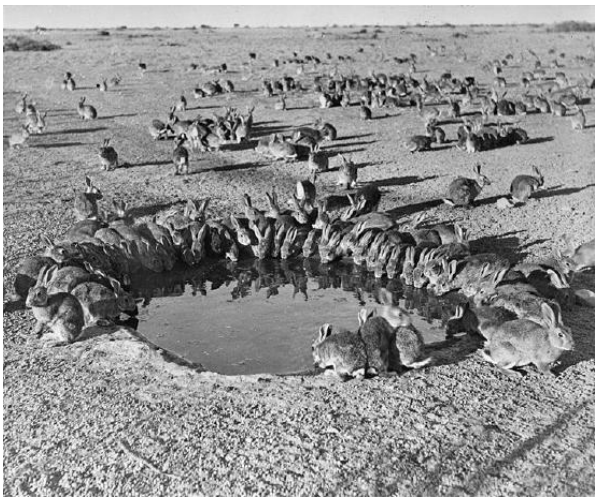


**Figure 16.3.*f*:** These rabbits in Australia removed all forage plants, which ordinarily supply them with water as well as food. They thus had to drink from a pool. Image by National Archives of Australia (public domain).

To summarize the lessons learned from biological control successes and failures, only candidates that have a very narrow target preference (eat only a sharply-limited range of hosts) should be chosen. Each candidate should be carefully tested to be sure that

once it has cleaned up the intended target, it does not turn to desirable species. Biological controls must not be used against native species. Finally, introduction of non-native species into the environment should be avoided because they could themselves be invasive.

## Climate Change

Climate change, and specifically the anthropogenic (meaning, caused by humans) warming trend presently underway, is recognized as a major extinction threat, particularly when combined with other threats such as habitat loss. Anthropogenic warming of the planet has been observed and is hypothesized to continue due to past and continuing emission of greenhouse gases, primarily carbon dioxide and methane, into the atmosphere caused by the burning of fossil fuels and deforestation. These gases decrease the degree to which Earth is able to radiate heat energy created by the sunlight that enters the atmosphere. The changes in climate and energy balance caused by increasing greenhouse gases are complex and our understanding of them depends on predictions generated from detailed computer models. Scientists generally agree the present warming trend is caused by humans and some of the likely effects include dramatic and dangerous climate changes in the coming decades.

However, there is still a lack of understanding about outcomes for specific species related to climate change. Scientists disagree about the likely magnitude of the effects, with extinction rate estimates ranging from 15 percent to 40 percent of species committed to extinction by 2050. Scientists do agree, however, that climate change will alter regional climates, including rainfall and snowfall patterns, making habitats less hospitable to the species living in them. The warming trend will shift colder climates toward the north and south poles, forcing species to move with their adapted climate norms while facing habitat gaps along the way. Range shifts are already being observed: for example, on average, European bird species ranges have moved 91 km (56.5 mi) northward. The same study suggested that the optimal shift based on warming trends was double that distance, suggesting that the populations are not moving quickly enough. Range shifts have also been observed in plants, butterflies, other insects, freshwater fishes, reptiles, amphibians, and mammals.

The shifting ranges will impose new competitive regimes on species as they find themselves in contact with other species not present in their historic range. One such unexpected species contact is between polar bears and grizzly bears (figure **16.3.*f***). Previously, these two species had separate ranges. Now, their ranges are overlapping and there are documented cases of these two species mating and producing viable offspring.




Figure **16.3.6**: Since 2008, grizzly bears (*Ursus arctos horribilis*) have been spotted farther north than their historic range, a possible consequence of climate change. As a result, grizzly bear habitat now overlaps polar bear (*Ursus maritimus*) habitat. The two kinds of bears, which are capable of mating and producing viable offspring, are considered separate species as historically they lived in different habitats and never met. However, in 2006 a hunter shot a wild grizzly-polar bear hybrid known as a grolar bear, the first wild hybrid ever found.

Changing climates also throw off species' delicate timing adaptations to seasonal food resources and breeding times. Many contemporary mismatches to shifts in resource availability and timing have already been documented.

Climate gradients will also move up mountains, eventually crowding species higher in altitude and eliminating the habitat for those species adapted to the highest elevations. Some climates will completely disappear. The rate of warming appears to be accelerated in the arctic, which is recognized as a serious threat to polar bear populations that require sea ice to hunt seals during the winter months: seals are the only source of protein available to polar bears. A trend to decreasing sea ice and glacier coverage has occurred since observations began in the mid-twentieth century. The rate of decline observed in recent years is far greater than previously predicted by climate models (Figure **16.3.6**).




**Figure 16.3.6:** The effect of global warming can be seen in the continuing retreat of Grinnell Glacier. The mean annual temperature in Glacier National Park has increased 1.33°C since 1900. The loss of a glacier results in the loss of summer meltwaters, sharply reducing seasonal water supplies and severely affecting local ecosystems. (credit: USGS, GNP Archives)

Finally, global warming will raise ocean levels due to meltwater from glaciers and the greater volume occupied by warmer water. Shorelines will be inundated, reducing island size, which will have an effect on some species, and a number of islands will disappear entirely. Additionally, the gradual melting and subsequent refreezing of the poles, glaciers, and higher elevation mountains—a cycle that has provided freshwater to environments for centuries—will be altered. This could result in an overabundance of salt water and a shortage of fresh water.

## References

Global Forest Watch. 2020. World Resources Institute. Accessed 2020-07-29.

Illegal logging nets organized crime up to 100 billion dollars a year, INTERPOL–UNEP report reveals. 2012. Interpol. Accessed 2020-07-29.

Acheson, J. M. (1988). The Lobster Gangs of Maine. Lebanon, NH: University of New England Press.

Annala, J. H. (1996). New Zealand's ITQ system: have the first eight years been a success or a failure? Reviews in Fish Biology and Fisheries. 6(1), 43–62. doi: 10.1007/BF00058519

Beddington, J. R., Agnew, D.J., & Clark, C. W. (2007). Current problems in the management of marine fisheries. Science, 316(5832), 1713-1716. doi:10.1126/science.1137362

Costello, C., Gaines, S. D., & Lynham, J. (2008). Can catch shares prevent fisheries collapse? Science, 321(5896), 1678 – 1681. doi10.1126/science.1159478

Levy, S. (2010). Catch shares management. BioScience, 60(10), 780–785. doi:10.1525/bio.2010.60.10.3

Montaigne, F. (2007, April). Still waters, the global fish crisis. National Geographic Magazine. Retrieved from http://ngm.nationalgeographic.com/print/2007/04/global-fisheries-crisis/montaigne-text.

Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., Hackson, J. B. C., Lotze, H. K., Micheli, F., Palumbi, S. R., Sala, E., Selkoe, K. A., Stachowicz, J. J., & Watson, R. (2006). Impacts of biodiversity loss on ocean ecosystem services. Science, 314(5800), 787 – 790. doi:10.1126/science.1132294

# Contributors and Attributions

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

- Threats to Biodiversity, Importance of Biodiversity, Community Ecology, and Sustainable Agriculture from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)

- Threats to Biodiversity from *General Biology* by OpenStax (licensed under CC-BY) by

  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).
- Case Study: Marine Fisheries, and Biodiversity, Species Loss, and Ecosystem Function from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors. Download for free at CNX. (licensed under CC-BY)
- Understanding Bycatch. NOAA Fisheries. Accessed 28 March 2021 (public domain)
- Biological Control and Symbiosis from *Biology* by John W. Kimball (licensed under CC-BY)
- *Environmental Science* by Kamala Dorsner
- **47.2C: Managing Fisheries** by Boundless has no license indicated.

---

This page titled 16.3: Threats to Biodiversity is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **47.3: Threats to Biodiversity** by OpenStax is licensed CC BY 4.0.
- **21.2: Threats to Biodiversity** by OpenStax is licensed CC BY 4.0.

# 16.4: Importance of Biodiversity

Healthy ecosystems contain a diversity of species, and each species plays a role in ecosystem function; therefore, species diversity as well as ecosystem diversity are essential to maintaining ecosystem services. Loss of biodiversity eventually threatens other species we do not impact directly because of their interconnectedness; as species disappear from an ecosystem other species are threatened by the changes in available resources. Biologists recognize that human populations are embedded in ecosystems and are dependent on them, just as is every other species on the planet. Agriculture began after early hunter-gatherer societies first settled in one place and heavily modified their immediate environment: the ecosystem in which they existed. This cultural transition has made it difficult for humans to recognize their dependence on living things other than crops and domesticated animals on the planet. Today our technology smooths out the extremes of existence and allows many of us to live longer, more comfortable lives, but ultimately the human species cannot exist without its surrounding ecosystems. Our ecosystems provide our food. This includes living plants that grow in soil ecosystems and the animals that eat these plants (or other animals) as well as photosynthetic organisms in the oceans and the other organisms that eat them. Our ecosystems have provided and will provide many of the medications that maintain our health, which are commonly made from compounds found in living organisms. Ecosystems provide our clean water, which is held in lake and river ecosystems or passes through terrestrial ecosystems on its way into groundwater.



**Figure 16.4.1:** This tropical lowland rainforest in Madagascar is an example of a high biodiversity habitat. This particular location is protected within a national forest, yet only 10 percent of the original coastal lowland forest remains, and research suggests half the original biodiversity has been lost. (credit: Frank Vassen)

Often called **ecological services** or **ecosystem services**, the products and processes associated with biological systems are of immense value to the well being of people. An incomplete list of these services and products includes the formation of soil and cycling of nutrients; provisioning of food, fresh water, fuel, fiber, and recreation opportunities; the regulation of climate, flooding, and disease. The value of these services is often overlooked or simply taken for granted, but one global estimate puts it somewhere between $16-64 trillion annually. From global food security, to a source of medicines, to even the oxygen in our air, we are dependent on biodiversity and the sustained integrity of ecological systems. Nature is also the basis for a significant part of aesthetic and spiritual values held by many cultures.

## Human Health

Many medications are derived from natural chemicals made by a diverse group of organisms. For example, many plants produce secondary plant compounds which are toxins used to protect the plant from insects and other animals that eat them. Some of these compounds also work as human medicines. Contemporary societies that live close to the land often have a broad knowledge of the medicinal uses of plants growing in their area. For centuries in Europe, older knowledge about the medical uses of plants was compiled in herbals—books that identified the plants and their uses. Humans are not the only animals to use plants for medicinal reasons. The other great apes, orangutans, chimpanzees, bonobos, and gorillas have all been observed self-medicating with plants.

Modern pharmaceutical science also recognizes the importance of these plant compounds. Examples of significant medicines derived from plant compounds include aspirin, codeine, digoxin, atropine, and vincristine (Figure **16.4.2**). Many medications were once derived from plant extracts but are now synthesized. It is estimated that, at one time, 25 percent of modern drugs contained at

least one plant extract. That number has probably decreased to about 10 percent as natural plant ingredients are replaced by synthetic versions of the plant compounds. Antibiotics, which are responsible for extraordinary improvements in health and lifespans in developed countries, are compounds largely derived from fungi and bacteria.



**Figure 16.4.2:** Catharanthus roseus, the Madagascar periwinkle, has various medicinal properties. Among other uses, it is a source of vincristine, a drug used in the treatment of lymphomas. (credit: Forest and Kim Starr)

In recent years, animal venoms and poisons have excited intense research for their medicinal potential. By 2007, the FDA had approved five drugs based on animal toxins to treat diseases such as hypertension, chronic pain, and diabetes. Another five drugs are undergoing clinical trials and at least six drugs are being used in other countries. Other toxins under investigation come from mammals, snakes, lizards, various amphibians, fish, snails, octopuses, and scorpions.

Aside from representing billions of dollars in profits, these medications improve people's lives. Pharmaceutical companies are actively looking for new natural compounds that can function as medicines. It is estimated that one third of pharmaceutical research and development is spent on natural compounds and that about 35 percent of new drugs brought to market between 1981 and 2002 were from natural compounds.

Finally, it has been argued that humans benefit psychologically from living in a biodiverse world. The chief proponent of this idea is entomologist E. O. Wilson. He argues that human evolutionary history has adapted us to living in a natural environment and that built environments generate stresses that affect human health and well-being. There is considerable research into the psychologically regenerative benefits of natural landscapes that suggest the hypothesis may hold some truth.

### Agricultural

Since the beginning of human agriculture more than 10,000 years ago, human groups have been breeding and selecting crop varieties. This crop diversity matched the cultural diversity of highly subdivided populations of humans. For example, potatoes were domesticated beginning around 7,000 years ago in the central Andes of Peru and Bolivia. The people in this region traditionally lived in relatively isolated settlements separated by mountains. The potatoes grown in that region belong to seven species and the number of varieties likely is in the thousands. Each variety has been bred to thrive at particular elevations and soil and climate conditions. The diversity is driven by the diverse demands of the dramatic elevation changes, the limited movement of people, and the demands created by crop rotation for different varieties that will do well in different fields.

Potatoes are only one example of agricultural diversity. Every plant, animal, and fungus that has been cultivated by humans has been bred from original wild ancestor species into diverse varieties arising from the demands for food value, adaptation to growing conditions, and resistance to pests. The potato demonstrates a well-known example of the risks of low crop diversity: during the tragic Irish potato famine (1845–1852 AD), the single potato variety grown in Ireland became susceptible to a potato blight—wiping out the crop. The loss of the crop led to famine, death, and mass emigration. Resistance to disease is a chief benefit to maintaining crop biodiversity and lack of diversity in contemporary crop species carries similar risks. Seed companies, which are the source of most crop varieties in developed countries, must continually breed new varieties to keep up with evolving pest organisms. These same seed companies, however, have participated in the decline of the number of varieties available as they focus on selling fewer varieties in more areas of the world replacing traditional local varieties.

The ability to create new crop varieties relies on the diversity of varieties available and the availability of wild forms related to the crop plant. These wild forms are often the source of new gene variants that can be bred with existing varieties to create varieties with new attributes. Loss of wild species related to a crop will mean the loss of potential in crop improvement. Maintaining the genetic diversity of wild species related to domesticated species ensures our continued supply of food.

Since the 1920s, government agriculture departments have maintained seed banks of crop varieties as a way to maintain crop diversity. This system has flaws because over time seed varieties are lost through accidents and there is no way to replace them. In 2008, the Svalbard Global seed Vault, located on Spitsbergen island, Norway, (Figure **16.4.3**) began storing seeds from around the world as a backup system to the regional seed banks. If a regional seed bank stores varieties in Svalbard, losses can be replaced from Svalbard should something happen to the regional seeds. The Svalbard seed vault is deep into the rock of the arctic island. Conditions within the vault are maintained at ideal temperature and humidity for seed survival, but the deep underground location of the vault in the arctic means that failure of the vault's systems will not compromise the climatic conditions inside the vault.

ART CONNECTION



**Figure 16.4.3:** The Svalbard Global Seed Vault is a storage facility for seeds of Earth's diverse crops. (credit: Mari Tefre, Svalbard Global Seed Vault)

The Svalbard seed vault is located on Spitsbergen island in Norway, which has an arctic climate. Why might an arctic climate be good for seed storage?

Although crops are largely under our control, our ability to grow them is dependent on the biodiversity of the ecosystems in which they are grown. That biodiversity creates the conditions under which crops are able to grow through what are known as ecosystem services—valuable conditions or processes that are carried out by an ecosystem. Crops are not grown, for the most part, in built environments. They are grown in soil. Although some agricultural soils are rendered sterile using controversial pesticide treatments, most contain a huge diversity of organisms that maintain nutrient cycles—breaking down organic matter into nutrient compounds that crops need for growth. These organisms also maintain soil texture that affects water and oxygen dynamics in the soil that are necessary for plant growth. Replacing the work of these organisms in forming arable soil is not practically possible. These kinds of processes are called ecosystem services. They occur within ecosystems, such as soil ecosystems, as a result of the diverse metabolic activities of the organisms living there, but they provide benefits to human food production, drinking water availability, and breathable air.

Other key ecosystem services related to food production are plant pollination and crop pest control. It is estimated that honeybee pollination within the United States brings in $1.6 billion per year; other pollinators contribute up to $6.7 billion. Over 150 crops in the United States require pollination to produce. Many honeybee populations are managed by beekeepers who rent out their hives' services to farmers. Honeybee populations in North America have been suffering large losses caused by a syndrome known as colony collapse disorder, a new phenomenon with an unclear cause. Other pollinators include a diverse array of other bee species and various insects and birds. Loss of these species would make growing crops requiring pollination impossible, increasing dependence on other crops.

Finally, humans compete for their food with crop pests, most of which are insects. Pesticides control these competitors, but these are costly and lose their effectiveness over time as pest populations adapt. They also lead to collateral damage by killing non-pest

species as well as beneficial insects like honeybees, and risking the health of agricultural workers and consumers. Moreover, these pesticides may migrate from the fields where they are applied and do damage to other ecosystems like streams, lakes, and even the ocean. Ecologists believe that the bulk of the work in removing pests is actually done by predators and parasites of those pests, but the impact has not been well studied. A review found that in 74 percent of studies that looked for an effect of landscape complexity (forests and fallow fields near to crop fields) on natural enemies of pests, the greater the complexity, the greater the effect of pest-suppressing organisms. Another experimental study found that introducing multiple enemies of pea aphids (an important alfalfa pest) increased the yield of alfalfa significantly. This study shows that a diversity of pests is more effective at control than one single pest. Loss of diversity in pest enemies will inevitably make it more difficult and costly to grow food. The world's growing human population faces significant challenges in the increasing costs and other difficulties associated with producing food.

### Wild Food Sources

In addition to growing crops and raising food animals, humans obtain food resources from wild populations, primarily wild fish populations. For about one billion people, aquatic resources provide the main source of animal protein. But since 1990, production from global fisheries has declined. Despite considerable effort, few fisheries on Earth are managed sustainability.

Fishery extinctions rarely lead to complete extinction of the harvested species, but rather to a radical restructuring of the marine ecosystem in which a dominant species is so over-harvested that it becomes a minor player, ecologically. In addition to humans losing the food source, these alterations affect many other species in ways that are difficult or impossible to predict. The collapse of fisheries has dramatic and long-lasting effects on local human populations that work in the fishery. In addition, the loss of an inexpensive protein source to populations that cannot afford to replace it will increase the cost of living and limit societies in other ways. In general, the fish taken from fisheries have shifted to smaller species and the larger species are overfished. The ultimate outcome could clearly be the loss of aquatic systems as food sources.

## Psychological and Moral Value

Finally, it has been clearly shown that humans benefit psychologically from living in a biodiverse world. A chief proponent of this idea is Harvard entomologist E. O. Wilson. He argues that human evolutionary history has adapted us to live in a natural environment and that city environments generate psychological stressors that affect human health and well-being. There is considerable research into the psychological regenerative benefits of natural landscapes that suggests the hypothesis may hold some truth. In addition, there is a moral argument that humans have a responsibility to inflict as little harm as possible on other species.

## Footnotes

1. 1 International Institute for Species Exploration (IISE), *2011 State of Observed Species (SOS)*. Tempe, AZ: IISE, 2011. Accessed May, 20, 2012. species.asu.edu/SOS.
2. Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., et al. (2011, March). Has the Earth's sixth mass extinction already arrived? *Nature*, *471*, 51-57. doi:10.1038/nature09678

   Vie, J-C, Hilton-Taylor, C. & Stuart S.N. (Eds.). (2009). *Wildlife in a Changing World: An Analysis of the 2008 IUCN Red List of Threatened Species*™. Gland, Switzerland: IUCN. Retrieved from http://data.iucn.org/dbtw-wpd/edocs/RL-2009-001.pdf.

### Contributors and Attributions

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

- Community Ecology and Importance of Biodiversity from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- *Environmental Science* by *Kamala Dorsner*
- Biodiversity, Species Loss, and Ecosystem Function from *Sustainability: A Comprehensive Foundation* by Tom Theis and Jonathan Tomkin, Editors. Download for free at CNX. (licensed under CC-BY)
- Preserving Biodiversity, **21.1: Importance of Biodiversity**, 47.2 The Importance of Biodiversity to Human Life, and **47.1: The Biodiversity Crisis** by OpenStax, is licensed CC BY by

  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).

Access for free at https://openstax.org/books/biology-2e/pages/1-introduction

---

This page titled 16.4: Importance of Biodiversity is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **21.1: Importance of Biodiversity** by OpenStax is licensed CC BY 4.0.

# 16.5: Preserving Biodiversity

Preserving biodiversity is an extraordinary challenge that must be met by greater understanding of biodiversity itself, changes in human behavior and beliefs, and various preservation strategies. Today, the main efforts to preserve biodiversity involve legislative approaches to regulate human and corporate behavior, setting aside protected areas, and habitat restoration.

## Changing Human Behavior

Legislation has been enacted to protect species throughout the world. The legislation includes international treaties as well as national and state laws. The **Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)** treaty came into force in 1975. The treaty, and the national legislation that supports it, provides a legal framework for preventing approximately 35,000 "listed" species from being transported across nations' borders, thus protecting them from being caught or killed when the purpose involves international trade. Species can be listed in one of three CITES appendices. Trade is banned for Appendix I species, which are threatened with extinction. For example the trade of ivory is banned by CITES. Trade is regulated for Appendix II species, such as the meat and shells of the queen conch or big-leaf mahogany (figure **16.5.*b***). Appendix III species are protected in at least one country, and the local government needs a coordinated response through CITES. For example, several species of red and pink corals have been added to Appendix III at the request of China. The treaty is limited in its reach because it only deals with international movement of organisms or their parts. It is also limited by various countries' ability or willingness to enforce the treaty and supporting legislation. The illegal trade in organisms and their parts is probably a market in the hundreds of millions of dollars. Illegal wildlife trade is monitored by another non-profit: Trade Records Analysis of Flora and Fauna in Commerce (TRAFFIC).



**Figure 16.5.1:** Big-leaf mahogany (*Swietenia macrophylla*) timber is used to make furniture, musical instruments, boats, and other products. It listed in CITES Appendix II, meaning that it is regulated. Specifically, a permit is required to ship big-leaf mahogany timber. Image by Forest and Kim Starr (CC-BY).

Within many countries there are laws that protect endangered species and that regulate hunting and fishing. In the United States, the **Endangered Species Act** (ESA) was enacted in 1973. When an at-risk species is listed by the Act, the U.S. Fish & Wildlife Service is required by law to develop a management plan to protect the species and bring it back to sustainable numbers. The ESA, and others like it in other countries, is a useful tool, but it suffers because it is often difficult to get a species listed or to get an effective management plan in place once a species is listed. The ESA does not automatically protect species categorized as

threatened on the Red List. Instead, the U.S. Fish and Wildlife Service (FWS), which enforces the ESA, assesses candidates for protected status as threatened or endangered (figure **16.5.2**). Consideration of candidate species can be initiated by the FWS itself or at the request of the public. Through the Species Status Assessment Framework, the FWS compiles biological data, such as habitat and population information and current threats to the species. This biological data is used to inform decisions. Additionally, species may be controversially taken off the list without necessarily having had a change in their situation. More fundamentally, the approach to protecting individual species rather than entire ecosystems is both inefficient and focuses efforts on a few highly visible and often charismatic species, perhaps at the expense of other species that go unprotected. At the same time, the ESA has a critical habitat provision outlined in the recovery mechanism that may benefit species other than the one targeted for management.



**Figure 16.5.2:** Public Affairs Specialist for the U.S. Fish and Wildlife Service, Joanna Gilkeson, takes photos of federally endangered San Diego fairy shrimp in a vernal pool in Otay Mesa, California. Vernal pools are shallow seasonal ponds, and vernal pool species are uniquely adapted to changes in water availability. Image by Maideline Sanchez/USFWS (public domain).

The 1972 **Marine Mammals Protection Act** prohibits the "take" of marine mammals—including harassment, hunting, capturing, collecting, or killing—in U.S. waters and by U.S. citizens on the high seas. The act also makes it illegal to import marine mammals and marine mammal products into the United States without a permit. State laws can also aid in conservation. Through **California Endangered Species Act** (CESA), originally passed in 1970 and subsequently amended, the California Fish and Game Commission assesses species to be listed as threatened or endangered by the state. A listed species, or any part or product of the plant or animal, may not be imported into the state, exported out of the state, "taken" (killed), possessed, purchased, or sold without proper authorization. Note that endangered is a subcategory of threatened on the Red List, but for ESA and CESA, threatened and endangered are separate categories, with the latter representing at the greater risk of extinction.

The **Migratory Bird Treaty Act** (MBTA) is an agreement between the United States and Canada that was signed into law in 1918 in response to declines in North American bird species caused by hunting. The Act now lists over 800 protected species. It makes it illegal to disturb or kill the protected species or distribute their parts (much of the hunting of birds in the past was for their feathers). Examples of protected species include northern cardinals, the red-tailed hawk, and the American black vulture.

Global warming is expected to be a major driver of biodiversity loss. Many governments are concerned about the effects of anthropogenic global warming, primarily on their economies and food resources. Because greenhouse gas emissions do not respect national boundaries, the effort to curb them is international. The international response to global warming has been mixed. The **Kyoto Protocol**, an international agreement that came out of the United Nations Framework Convention on Climate Change that committed countries to reducing greenhouse gas emissions by 2012, was ratified by some countries, but spurned by others. Two countries that were especially important in terms of their potential impact that did not ratify the Kyoto protocol were the United States and China. Some goals for reduction in greenhouse gasses were met and exceeded by individual countries, but, worldwide,

the effort to limit greenhouse gas production is not succeeding. A renegotiated 2016 treaty, called the **Paris Agreement**, once again brought nations together to take meaningful action on climate change. But like before, some nations are reluctant to participate.

As already mentioned, the private non-profit sector plays a large role in the conservation effort both in North America and around the world. The approaches range from species-specific organizations to the broadly focused IUCN and TRAFFIC. The Nature Conservancy takes a novel approach. It purchases land and protects it in an attempt to set up preserves for ecosystems. Ultimately, human behavior will change when human values change. At present, the growing urbanization of the human population is a force that poses challenges to the valuing of biodiversity.

## Conservation in Preserves

Establishment of wildlife and ecosystem preserves is one of the key tools in conservation efforts (Figure **16.5.3**). A **preserve** is an area of land set aside with varying degrees of protection for the organisms that exist within the boundaries of the preserve. Governments or private organizations establish nature preserves. Preserves can be effective in the short term for protecting both species and ecosystems, but they face challenges that scientists are still exploring to strengthen their viability as long-term solutions.

Due to the way protected lands are allocated (they tend to contain less economically valuable resources rather than being set aside specifically for the species or ecosystems at risk) and the way biodiversity is distributed, determining a target percentage of land or marine habitat that should be protected to maintain biodiversity levels is challenging. The IUCN World Parks Congress estimated that 11.5 percent of Earth's land surface was covered by preserves of various kinds in 2003. This area is greater than previous goals; however, it only represents 9 out of 14 recognized major biomes. Research has shown that 12 percent of all species live only outside preserves; these percentages are much higher when only threatened species and high quality preserves are considered. For example, high quality preserves include only about 50 percent of threatened amphibian species. The conclusion must be that either the percentage of area protected must increase, or the percentage of high quality preserves must increase, or preserves must be targeted with greater attention to biodiversity protection. Researchers argue that more attention to the latter solution is required.

It is important to protect natural areas for several reasons. Some people feel a cultural or spiritual connection to the wilderness. Every year, millions of people visit recreational lands such as parks and wilderness areas to experience attractions of the great outdoors: hiking among the giant sequoias in California, traveling on a photo safari in Kenya or just picnicking at a local county park. Besides providing people with obvious health benefits and aesthetic pleasures, recreational lands also generate considerable tourist money for government and local economies. Outdoor recreation activities such as hiking and camping benefit tourist industries and manufacturers of outdoor clothes and equipment.




**Figure 16.5.3:** Mequon Nature Preserve in Wisconsin. Image by Jennifer Tomaloff (CC-BY-NC-SA).

## Preserve Design

There has been extensive research into optimal preserve designs for maintaining biodiversity. The fundamental principle behind much of the research has been the seminal theoretical work of Robert H. MacArthur and Edward O. Wilson published in 1967 on island biogeography.[1] This work sought to understand the factors affecting biodiversity on islands. The fundamental conclusion was that biodiversity on an island was a function of the origin of species through migration, speciation, and extinction on that island. Islands farther from a mainland are harder to get to, so migration is lower and the equilibrium number of species is lower. Within island populations, evidence suggests that the number of species gradually increases to a level similar to the numbers on the mainland from which the species is suspected to have migrated. In addition, smaller islands are harder to find, so their immigration rates for new species are lower. Smaller islands are also less geographically diverse so there are fewer niches to promote speciation. And finally, smaller islands support smaller populations, so the probability of extinction is higher. As islands get larger, the number of species accelerates, although the effect of island area on species numbers is not a direct correlation.

Conservation preserves can be seen as "islands" of habitat within "an ocean" of non-habitat. In general, large preserves are better because they support more species, including species with large home ranges; they have more core area of optimal habitat for individual species; they have more niches to support more species; and they attract more species because they can be found and reached more easily. One large preserve is better than the same area of several smaller preserves because there is more core habitat unaffected by less hospitable ecosystems outside the preserve boundary. For this same reason, preserves in the shape of a square or circle will be better than a preserve with many thin "arms." If preserves must be smaller, then providing **wildlife corridors** (narrow strips of protected land) between two preserves is important so that species and their genes can move between them. All of these factors are taken into consideration when planning the nature of a preserve before the land is set aside.

For a species to persist in a preserve, the preserve must be large enough. The critical size depends, in part, on the home range that is characteristic of the species. A preserve for wolves, which range hundreds of kilometers, must be much larger than a preserve for butterflies, which might range within ten kilometers during its lifetime. But larger preserves have more core area of optimal habitat for individual species, they have more niches to support more species, and they attract more species because they can be found and reached more easily.

Preserves perform better when there are buffer zones around them of suboptimal habitat. The buffer allows organisms to exit the boundaries of the preserve without immediate negative consequences from predation or lack of resources. One large preserve is better than the same area of several smaller preserves because there is more core habitat unaffected by edges. For this same reason, preserves in the shape of a square or circle will be better than a preserve with many thin "arms." If preserves must be smaller, then providing wildlife corridors between them so that individuals and their genes can move between the preserves, for example along rivers and streams, will make the smaller preserves behave more like a large one. All of these factors are taken into consideration when planning the nature of a preserve before the land is set aside.

In addition to the physical, biological, and ecological specifications of a preserve, there are a variety of policy, legislative, and enforcement specifications related to uses of the preserve for functions other than protection of species. These can include anything from timber extraction, mineral extraction, regulated hunting, human habitation, and nondestructive human recreation. Many of these policy decisions are made based on political pressures rather than conservation considerations. In some cases, wildlife protection policies have been so strict that subsistence-living indigenous populations have been forced from ancestral lands that fell within a preserve. In other cases, even if a preserve is designed to protect wildlife, if the protections are not or cannot be enforced, the preserve status will have little meaning in the face of illegal poaching and timber extraction. This is a widespread problem with preserves in areas of the tropics.

## Limitations on Preserves

Some of the limitations on preserves as conservation tools are evident from the discussion of preserve design. Political and economic pressures typically make preserves smaller, never larger, so setting aside areas that are large enough is difficult. If the area set aside is sufficiently large, there may not be sufficient area to create a buffer around the preserve. In this case, an area on the outer edges of the preserve inevitably becomes a riskier suboptimal habitat for the species in the preserve. Enforcement of protections is also a significant issue in countries without the resources or political will to prevent poaching and illegal resource extraction.

Climate change will create inevitable problems with the location of preserves. The species within them will migrate to higher latitudes as the habitat of the preserve becomes less favorable. Scientists are planning for the effects of global warming on future preserves and striving to predict the need for new preserves to accommodate anticipated changes to habitats; however, the end effectiveness is tenuous since these efforts are prediction based.

Finally, an argument can be made that conservation preserves reinforce the cultural perception that humans are separate from nature, can exist outside of it, and can only operate in ways that do damage to biodiversity. Creating preserves reduces the pressure on human activities outside the preserves to be sustainable and non-damaging to biodiversity. Ultimately, the political, economic, and human demographic pressures will degrade and reduce the size of conservation preserves if the activities outside them are not altered to be less damaging to biodiversity.

Link to Learning



An interactive global data system of protected areas can be found at website. Review data about individual protected areas by location or study statistics on protected areas by country or region.

On an international level, important wilderness lands have been designated by the United Nations through its "**Man and the Biosphere Program**." This program was established in 1973 to protect examples of major natural regions throughout the world, and provide opportunities for ecological research and education. **Biosphere reserves** are organized into three interrelated zones: the **core area**, the **buffer zone** and the **transition area**. The core area contains the landscape and ecosystems to be preserved. The buffer zone is an area where activities are controlled to protect the core area. The outer transition area contains a variety of agricultural activities, human settlements and other uses. Local communities, conservation agencies, scientists and private enterprises that have a stake in the management of the region work together to make the reserves work. Mt Kenya in Africa and the Galapagos Islands are examples of wilderness areas protected under this provision.

## Types of Protected Areas in the United States

The public lands described below differ in their level of protection. For example, national parks and forests allow camping whereas wildlife refuges place more limitations on human activities. National parks and forests and wildlife refuges can contain wilderness areas.

**Wilderness areas**, comprise ecosystems in which human activity has not significantly affected the plant and animal populations or their environment. Natural processes predominate. According to the "Wilderness Act of 1964," wilderness areas are defined as being those areas where the nearest road is at least five miles away and where no permanent buildings stand. Activities that could disrupt native species, such as the use of motorized vehicles is prohibited. More than 100 million acres of land are now preserved as wilderness under this act. Sparsely populated Alaska contains the largest chunk of wilderness areas, over half of it. Although wilderness areas are scattered among most of the lower 48 states, the largest percentage is found in the western states. Few undesignated areas in the contiguous states remain that would qualify as wilderness. California contains significant wilderness areas, with over 4 million acres of National Forest Wilderness areas, and 1.5 million acres of mostly desert wilderness in the Mojave Desert National Preserve (figure **16.5.4**). Wilderness areas provide an essential habitat for a wide array of fish, wildlife, and plants, and are particularly important in protecting endangered species. For scientists, wilderness areas serve as natural laboratories, where studies can be performed that would not be possible in developed areas.



**Figure 16.5.4:** A variety of unique plant life grows at Mojave Desert National Preserve. Image by John Fowler (CC-BY).

The United States has set aside more land for public recreational use than any other country. The **National Park System** manages more than 380 parks, recreation areas, seashores, trails, monuments, memorials, battlefields, and other historic sites. It consists of more than 80 million acres nationwide (figure 16.5.5). The largest national park is Wrangell–St. Elias National Park and Preserve in Alaska with over 13 million acres. California has eight national parks: Channel Islands, Death Valley, Joshua Tree, Lassen, Redwood, Sequoia, Kings Canyon, and Yosemite. Many national parks such as Yosemite, Yellowstone and the Grand Canyon are such popular recreation destinations that the ecosystems of those parks are being severely tested by human activities.



**Figure 16.5.5:** National parks, such as Grand Teton National Park in Wyoming, help conserve biodiversity. (credit: Don DeBold)

Every state has also set aside significant amounts of land for recreational use. The **California State Park System** manages more than one million acres of parklands including: coastal wetlands, estuaries, scenic coastlines, lakes, mountains and desert areas. California's largest state park is Anza-Borrego Desert State Park, which is the largest state park in the United States with 600,000 acres. The stated mission of the California State Park System is "to provide for the health, inspiration and education of the people of California by helping to preserve the state's extraordinary biological diversity, protecting its most valued natural and cultural resources and creating opportunities for high-quality outdoor recreation". This is the basic goal of all recreational lands: to manage and conserve natural ecosystems, while supporting a sustainable and balanced level of human use of those areas. Unfortunately, it is a goal which is sometimes difficult to achieve due to the increasing popularity and use of recreational lands.

The **National Forest System**, managed by the U.S. Forest Service (part of the United States Department of Agriculture), consists of more than 170 forestlands and grasslands, which are available for activities such as camping, fishing, hiking and hunting. These are managed as multiple use lands, which balance the needs for recreation, grazing, timber, watershed protection, wildlife and fish, and wilderness. Some examples of national forests are the Sierra National Forest in California and the White Mountain National Forest in New Hampshire. The Coronado National Forest in Arizona is famous for "sky islands", or steep mountain ranges surrounded by low-lying areas. The dramatic increase in elevation is associated with changes in the flora and fauna (figure 16.5.6). Explore national forests using this interactive map.



**Figure 16.5.6:** The Santa Teresa Mountains of the Coronado National Forest form "sky islands". Image by Jstuby (public domain).

The U.S. Fish and Wildlife Service manage more than 500 national **wildlife refuges** (figure 16.5.7), which not only protect animal habitats and breeding areas but also provide recreational facilities. Find a Refuge is an interactive map for locating wildlife refuges.



**Figure 16.5.7:** Bobcat at Sonny Bono National Wildlife Refuge in southern California. Image by Mark Stewart/USFWS (public domain).

Several other types of public lands complement the designated wilderness land system. These include: national forest roadless areas, national trails system, natural research areas and state and private wilderness lands. The national forest roadless areas consist of millions of acres of wild, undeveloped land without roads that exist on National Forest land outside of designated wilderness lands. The "**National Trail System**," established by Congress in 1968, includes trails in wilderness areas and other public lands. **Research Natural Areas** located throughout the country on public lands serve as outdoor laboratories to study natural systems. They are intended in part to serve as gene pools for rare and endangered species and as examples of significant natural ecosystems. Some wilderness lands are maintained by states or private organizations. For example, the state of New York has long preserved a region of the Adirondacks as wilderness.

## Habitat Restoration, Remediation, and Reclamation

Habitat restoration holds considerable promise as a mechanism for restoring and maintaining biodiversity. Of course once a species has become extinct, its restoration is impossible. However, restoration can improve the biodiversity of degraded ecosystems. Reintroducing wolves, a top predator, to Yellowstone National Park in 1995 led to dramatic changes in the ecosystem that increased biodiversity. The wolves (Figure 16.5.8) function to suppress elk and coyote populations and provide more abundant resources to the guild of carrion eaters. Reducing elk populations has allowed revegetation of riparian areas, which has increased the diversity

of species in that habitat. Decreasing the coyote population has increased the populations of species that were previously suppressed by this predator. The number of species of carrion eaters has increased because of the predatory activities of the wolves. In this habitat, the wolf is a keystone species, meaning a species that is instrumental in maintaining diversity in an ecosystem. Removing a keystone species from an ecological community may cause a collapse in diversity. The results from the Yellowstone experiment suggest that restoring a keystone species can have the effect of restoring biodiversity in the community. Ecologists have argued for the identification of keystone species where possible and for focusing protection efforts on those species; likewise, it also makes sense to attempt to return them to their ecosystem if they have been removed.

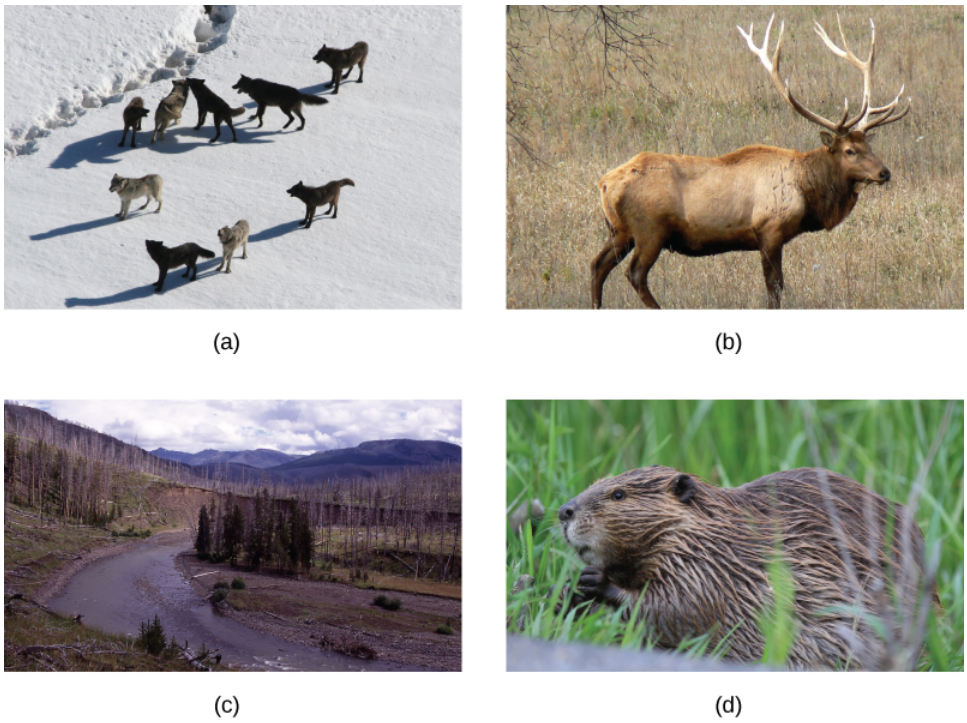



**Figure 16.5.8:** (a) The Gibbon wolf pack in Yellowstone National Park, March 1, 2007, represents a keystone species. The reintroduction of wolves into Yellowstone National Park in 1995 led to a change in the grazing behavior of (b) elk. To avoid predation, the elk no longer grazed exposed stream and riverbeds, such as (c) the Lamar Riverbed in Yellowstone. This allowed willow and cottonwood seedlings to grow. The seedlings decreased erosion and provided shading to the creek, which improved fish habitat. A new colony of (d) beaver may also have benefited from the habitat change. (credit a: modification of work by Doug Smith, NPS; credit c: modification of work by Jim Peaco, NPS; credit d: modification of work by "Shiny Things"/Flickr)

Other large-scale restoration experiments underway involve dam removal. In the United States, since the mid-1980s, many aging dams are being considered for removal rather than replacement because of shifting beliefs about the ecological value of free-flowing rivers and because many dams no longer provide the benefit and functions that they did when they were first built. The measured benefits of dam removal include restoration of naturally fluctuating water levels (the purpose of dams is frequently to reduce variation in river flows), which leads to increased fish diversity and improved water quality. In the Pacific Northwest, dam removal projects are expected to increase populations of salmon, which is considered a keystone species because it transports key nutrients to inland ecosystems during its annual spawning migrations. In other regions such as the Atlantic coast, dam removal has allowed the return of spawning anadromous fish species (species that are born in fresh water, live most of their lives in salt water, and return to fresh water to spawn). Some of the largest dam removal projects have yet to occur or have happened too recently for the consequences to be measured. The large-scale ecological experiments that these removal projects constitute will provide valuable data for other dam projects slated either for removal or construction.

Besides physical processes, socioeconomic factors must also be considered in a restoration project. Actions of humans have historically been important in shaping ecosystems, and are important in determining the success of restoration efforts. Because the cost to restore an individual site can involve millions of dollars, government support is a necessity.

Danger to human health from both historic and modern pollution requires that cleanup measures be implemented. **Remediation** is aimed at neutralization, containment, and/or removal of the polluting chemicals. The goal is to prevent the spread of the pollution,

or to reduce it to levels that will not appreciably risk human health. Many times, it is physically impossible or financially unfeasible to completely clear all contamination. Often, experts and the public disagree on how clean is clean enough.

Many communities are struggling to find the funds and technological expertise needed to clean up polluted areas. Some settings, such as brownfields, can be remediated fairly easily. **Brownfields** are abandoned industrial or commercial facilities or blighted urban areas that need to be cleansed of contamination before they can be redeveloped. Other areas, because of their size or the extreme toxicity of their contaminants, require very expensive, complex, and long-term remediation. Many of these have been designated as Superfund sites.

**Superfund sites** are areas with the most toxic contamination in the United States. The contamination may not only make the site itself too dangerous to inhabit, but often leaks toxic levels of pollutants into the surrounding soil, water, or air. An example of a Superfund site is Love Canal in Niagara Falls, New York (figure **16.5.9**). The canal was a chemical waste dump for many years, then in the 1950's was covered with soil and sold to the city. Over time, many homes and a school were built over the former dump. In the 1970's, heavy rains raised the water table and carried contaminants back to the surface. Residents noticed foul smells, and gardens and trees turned black and died. Soon after, rates of birth defects, cancer, and other illnesses began to rise sharply. In 1977, the State of New York and the federal government began remediation work. Buildings were removed, and all residents were bought out and relocated, contaminated deposits and soils were excavated, and remaining soils and groundwater were treated and sealed off to prevent further spread of the contamination. Remediation activities have now been completed at this site.



**Figure 16.5.9:** Love Canal. Source: US Environmental Protection Agency

The type of pollution and the medium affected (air, water, or soil) determine remediation methods. Methods include incineration, absorption onto carbon, chemical methods, or bioremediation. **Bioremediation** is the use of plants, bacteria, or fungus to "digest" the contaminant to a non-toxic or less toxic form. All of these methods tend to be expensive and time-consuming.

**Reclamation** involves salvaging some features of a degraded habitat, but it may not restore the ecosystem fully (figure **16.5.10**). For example, instead of abandoning a mined area once resources have been collected, it can be reclaimed by planting vegetation, reshaping the landscape, and redirecting water flow. However the reclaimed land still lacks many features of the original ecosystem, such as complex topography, vegetation that took tens or hundreds of years to grow, soil quality, and an intricate network of streams.

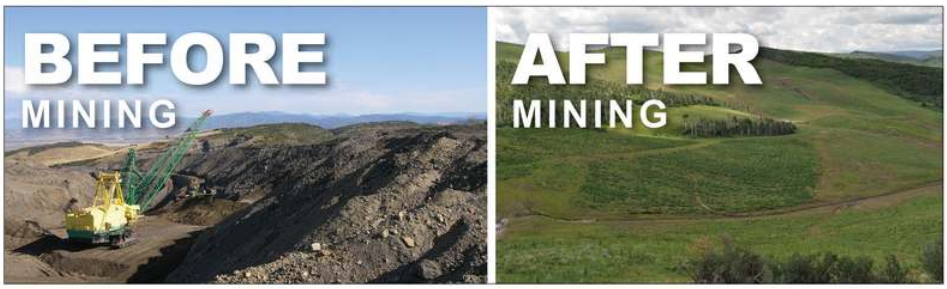



**Figure 16.5.10:** The Seneca Yoast coal once land has been cleared in preparation for mining (left) and after reclamation (right). Image by Peabody Energy, Inc. (CC-BY).

Sometimes, actions can be taken to avoid, reduce or compensate for the effects of environmental damage. Such **mitigation** efforts have been taken by the Army Corps of Engineers during construction projects. The native plants are removed from a site before construction begins and transplanted at a special holding site. After the construction project is completed, the native plants are replanted using those from the holding site. Another example of mitigation might involve the creation or enhancement of wetlands in one area, in order to compensate for permitted wetland losses in another area. Mitigation often goes hand-in-hand with restoration. Texaco, in conjunction with environmental groups and the United States Fish and Wildlife Service, restored 500 acres of agricultural lands in the lower Mississippi Delta to bottomland hardwoods. Texaco received environmental credits for the mitigating effects of the new woodlands on air quality.

## The Role of Captive Breeding

Zoos have sought to play a role in conservation efforts both through captive breeding programs and education (Figure 16.5.11). The transformation of the missions of zoos from collection and exhibition facilities to organizations that are dedicated to conservation is ongoing and gaining strength. In general, it has been recognized that, except in some specific targeted cases, captive breeding programs for endangered species are inefficient and often prone to failure when the species are reintroduced to the wild. However, captive breeding programs have yielded some success stories, such as the California condor reintroduction to the Grand Canyon and the reestablishment of the Whooping Crane along the Midwest flyway. Unfortunately, zoo facilities are far too limited to contemplate captive breeding programs for the numbers of species that are now at risk. Education is another potential positive impact of zoos on conservation efforts, particularly given the global trend to urbanization and the consequent reduction in contacts between people and wildlife. A number of studies have been performed to look at the effectiveness of zoos on people's attitudes and actions regarding conservation; at present, the results tend to be mixed.



**Figure 16.5.11:** Zoos and captive breeding programs help preserve many endangered species, such as this golden lion tamarin. (credit: Garrett Ziegler)

## Economic Influences on Conservation

Economics greatly impacts conservation success. Short-term profits can incentivize individuals, companies, or governments to harvesting resources at an unsustainable rate and at the expense of ecosystem health. In impoverished regions, compromising habitat to grow high-value crops, such as coffee or oil palms, or poaching endangered species may seem like the only source of income. One solution is **debt-for-nature swaps** through which one country forgives the debt of another if the latter agrees to protect natural areas. These conservation efforts can provide a new source of income for residents near the protected areas. For example, the United States forgave $20 million in debt from Costa Rica. In exchange, Costa Rica invested in expanding its protected areas and developing the ecotourism industry, which provides jobs to many of its residents (figure 16.5.12). **Ecotourism** involves visiting and enjoying natural areas while minimizing ecological damage. Ecotourism can benefit local economies and alleviate poverty especially if the earnings from it are reinvested into the communities living near tourist destinations. It generates jobs such as park operators, sellers of local crafts, and tour guides.



**Figure 16.5.12:** A tourist ziplines over the Costa Rican rainforest. Image by Khaufle at the English language Wikipedia (CC-BY-SA).

## Supplemental Reading

America's Public Lands Explained. 2016. U.S. Department of the Interior.

## Footnotes

1. 1 Robert H. MacArthur and Edward O. Wilson, E. O., *The Theory of Island Biogeography* (Princeton, N.J.: Princeton University Press, 1967).

## Contributors and Attributions

Modified by Kyle Whittinghill and Melissa Ha from the following sources

- Preserving Biodiversity by OpenStax is licensed under CC BY 4.0 by

  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).

- Threatened and Endangered Species by California Department of Fish and Wildlife (public domain)
- Marine Mammal Protection by NOAA Fisheries (public domain)
- Ecotourism from *Life Sciences Grade 10* by Siyavula (licensed under CC-BY)
- Preserving Biodiversity and Case Study - The Love Canal Disaster from *Environmental Biology* by Matthew R. Fisher (licensed under CC-BY)
- Issues and Opinions, Biological and Land from *AP Environmental Science* by University of California College Prep, University of California (licensed under CC-BY). Download for free at CNX.
- Multiple Use Lands. 12 May 2015. The United States Department of Justice. Accessed 28 March 2021 (public domain)

---

This page titled 16.5: Preserving Biodiversity is shared under a CC BY-NC-SA license and was authored, remixed, and/or curated by OpenStax.

- **47.4: Preserving Biodiversity** by OpenStax is licensed CC BY 4.0.
- **5.5: Preserving Biodiversity** by Matthew R. Fisher is licensed CC BY 4.0. Original source: https://openoregon.pressbooks.pub/envirobiology.

# Index


## A

acid
3.4: Water
acid rain
8.5: Acid Deposition
adaptation
4.1: Understanding Evolution
adaptive radiation
4.2: Formation of New Species
16.1: Introduction to Biodiversity
adhesion
3.4: Water
age structure
14.3: Human Population Growth
allopatric speciation
4.2: Formation of New Species
allopolyploid
4.2: Formation of New Species
analytical model
15.4: Ecology of Ecosystems
aneuploidy
4.2: Formation of New Species
anion
3.2: The Building Blocks of Molecules
apex consumer
15.4: Ecology of Ecosystems
Aposematic Coloration
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
applied science
2: The Process of Science
atom
3.5: Themes and Concepts of Biology
atomic number
3.2: The Building Blocks of Molecules
autopolyploid
4.2: Formation of New Species

## B

base
3.4: Water
basic science
2: The Process of Science
Batesian mimicry
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
behavioral isolation
4.2: Formation of New Species
biochemistry
3.5: Themes and Concepts of Biology
biodiversity
16.1: Introduction to Biodiversity
biodiversity hotspot
16.1: Introduction to Biodiversity
bioenergetics
3.1: Energy
biosphere
3.5: Themes and Concepts of Biology
botany
3.5: Themes and Concepts of Biology
buffer
3.4: Water
bush meat
16.3: Threats to Biodiversity

## C

Calvin cycle
5.4: Photosynthesis
camouflage
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
capillary action
3.4: Water
cation
3.2: The Building Blocks of Molecules
cell
3.5: Themes and Concepts of Biology
cellular respiration
5: Cycling of Matter in the Earth System
chemical bond
3.2: The Building Blocks of Molecules
chemical diversity
16.1: Introduction to Biodiversity
chemoautotroph
5.4: Photosynthesis
chloroplast
5.4: Photosynthesis
chytridiomycosis
16.3: Threats to Biodiversity
climax community
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
cohesion
3.4: Water
commensalism
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
community
3.5: Themes and Concepts of Biology
competitive exclusion principle
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
conceptual model
15.4: Ecology of Ecosystems
control
2: The Process of Science
convergent evolution
4.1: Understanding Evolution
covalent bond
3.2: The Building Blocks of Molecules
Crop rotations
13.7: Sustainable Agriculture

## D

deductive reasoning
2: The Process of Science
demography
14.1: Population Demography
descriptive science
2: The Process of Science
detrital food web
15.4: Ecology of Ecosystems
dispersal
4.2: Formation of New Species
dissociation
3.4: Water
divergent evolution
4.1: Understanding Evolution
DNA barcoding
16.5: Preserving Biodiversity

## E

ecosystem
3.5: Themes and Concepts of Biology
15.4: Ecology of Ecosystems
ecosystem diversity
16.1: Introduction to Biodiversity
ecosystem dynamics
15.4: Ecology of Ecosystems
electron
3.2: The Building Blocks of Molecules
electron transfer
3.2: The Building Blocks of Molecules
element
3.2: The Building Blocks of Molecules
Emsleyan/Mertensian mimicry
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
Endemic species
16.1: Introduction to Biodiversity
environmental disturbance
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
equilibrium
15.4: Ecology of Ecosystems
eukaryote
3.5: Themes and Concepts of Biology
evaporation
3.4: Water
evolution
3.5: Themes and Concepts of Biology
4.1: Understanding Evolution
exotic species
16.3: Threats to Biodiversity
extinction
16.1: Introduction to Biodiversity
extinction rate
16.1: Introduction to Biodiversity

## F

falsifiable
2: The Process of Science
floods
6.4.4: Floods

food chain
15.4: Ecology of Ecosystems
food security
13.1: Food Security
food web
15.4: Ecology of Ecosystems
foundation species
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
frontier ethic
1.3: Environmental Ethics and Environmental Justice

## G

gametic barrier
4.2: Formation of New Species
genetic diversity
16.1: Introduction to Biodiversity
genetic engineering
13.5: Selective Breeding and Genetic Engineering
granum
5.4: Photosynthesis
grazing food web
15.4: Ecology of Ecosystems

## H

habitat isolation
4.2: Formation of New Species
heat of vaporization of water
3.4: Water
heterogeneity
16.1: Introduction to Biodiversity
heterotroph
5.4: Photosynthesis
holistic ecosystem model
15.4: Ecology of Ecosystems
homeostasis
3.5: Themes and Concepts of Biology
homologous structures
4.1: Understanding Evolution
host
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
hybrid
4.2: Formation of New Species
hydrogen bond
3.2: The Building Blocks of Molecules
hydrophilic
3.4: Water
hydrophobic
3.4: Water
hypothesis
2: The Process of Science

## I

inductive reasoning
2: The Process of Science
integrated pest management
13.7: Sustainable Agriculture
intercropping
13.7: Sustainable Agriculture
interglacials
9.2: Prehistoric Climate Change and Proxys for Studying Past Climates
interspecific competition
14.2: Population Growth and Regulation
ion
3.2: The Building Blocks of Molecules
ionic bond
3.2: The Building Blocks of Molecules
island biogeography
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
isotope
3.2: The Building Blocks of Molecules

## K

keystone species
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions

## L

life science
2: The Process of Science
life table
14.1: Population Demography
litmus paper
3.4: Water

## M

Müllerian Mimicry
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
macromolecule
3.5: Themes and Concepts of Biology
mark and recapture
14.1: Population Demography
mass number
3.2: The Building Blocks of Molecules
matter
3.2: The Building Blocks of Molecules
megafauna
16.1: Introduction to Biodiversity
Mertensian mimicry
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
mesocosm
15.4: Ecology of Ecosystems
mesophyll
5.4: Photosynthesis
microbiology
3.5: Themes and Concepts of Biology
microcosm
15.4: Ecology of Ecosystems
minimal tillage
13.7: Sustainable Agriculture
molecular biology
3.5: Themes and Concepts of Biology
molecule
3.5: Themes and Concepts of Biology
mortality rate
14.1: Population Demography
mutualism
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions

## N

natural science
2: The Process of Science
natural selection
4.1: Understanding Evolution
neurobiology
3.5: Themes and Concepts of Biology
neutron
3.2: The Building Blocks of Molecules
nonpolar covalent bond
3.2: The Building Blocks of Molecules
nucleus
3.2: The Building Blocks of Molecules

## O

octet rule
3.2: The Building Blocks of Molecules
organ
3.5: Themes and Concepts of Biology
organ system
3.5: Themes and Concepts of Biology
organelle
3.5: Themes and Concepts of Biology
organic farming
13.7: Sustainable Agriculture
organism
3.5: Themes and Concepts of Biology
ozone depletion
8.6: Stratospheric Ozone Depletion

## P

paleontology
3.5: Themes and Concepts of Biology
paleosols
6.6: Soils
parasite
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
periodic table of elements
3.2: The Building Blocks of Molecules
Persistent organic pollutants (POPs)
13.3: Pests and Pesticides
pest
13.3: Pests and Pesticides
pesticides
13.3: Pests and Pesticides
pH paper
3.4: Water
pH scale
3.4: Water
photoautotroph
5.4: Photosynthesis
phylogenetic tree
3.5: Themes and Concepts of Biology
physical science
2: The Process of Science

pigment
5.4: Photosynthesis
pioneer species
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
polar covalent bond
3.2: The Building Blocks of Molecules
pollination
13.6: Pollination and Seed Dispersal
population
3.5: Themes and Concepts of Biology
population density
14.1: Population Demography
population size
14.1: Population Demography
postzygotic barrier
4.2: Formation of New Species
preservation
16.5: Preserving Biodiversity
prezygotic barrier
4.2: Formation of New Species
primary consumer
15.4: Ecology of Ecosystems
primary producer
15.4: Ecology of Ecosystems
primary succession
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
prokaryote
3.5: Themes and Concepts of Biology
proton
3.2: The Building Blocks of Molecules

Q

quadrat
14.1: Population Demography

R

radioactive isotope
3.2: The Building Blocks of Molecules
relative species abundance
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
Remediation
12.3: Water Pollutants and Their Sources
reproductive isolation
4.2: Formation of New Species
resilience
15.4: Ecology of Ecosystems
resilience (ecological)
15.4: Ecology of Ecosystems
resistance
15.4: Ecology of Ecosystems
resistance (ecological)
15.4: Ecology of Ecosystems

S

science
2: The Process of Science
scientific law
2: The Process of Science
scientific method
2: The Process of Science
scientific theory
2: The Process of Science
secondary consumer
15.4: Ecology of Ecosystems
secondary succession
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
seed dispersal
13.6: Pollination and Seed Dispersal
simulation model
15.4: Ecology of Ecosystems
Snowball Earth hypothesis
9.2: Prehistoric Climate Change and Proxys for Studying Past Climates
Soil processes
6.6: Soils
Soil Profiles
6.6: Soils
solvent
3.4: Water
speciation
4.2: Formation of New Species
species
4.2: Formation of New Species
species dispersion pattern
14.1: Population Demography
species richness
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
specific heat capacity
3.4: Water
sphere of hydration
3.4: Water
stoma
5.4: Photosynthesis
stroma
5.4: Photosynthesis
surface tension
3.4: Water
survivorship curve
14.1: Population Demography
sustainable agriculture
13.7: Sustainable Agriculture
symbiosis
15: Community and Ecosystem Ecology
15.1: Community Ecology- Species Interactions
sympatric speciation
4.2: Formation of New Species

T

temporal isolation
4.2: Formation of New Species
tertiary consumer
15.4: Ecology of Ecosystems
thylakoid
5.4: Photosynthesis
thylakoid lumen
5.4: Photosynthesis
Tragedy of the Commons
16.3: Threats to Biodiversity
trophic level
15.4: Ecology of Ecosystems

V

van der Waals interaction
3.2: The Building Blocks of Molecules
variable
2: The Process of Science
variation
4.1: Understanding Evolution
vestigial structure
4.1: Understanding Evolution
vicariance
4.2: Formation of New Species

Z

zoology
3.5: Themes and Concepts of Biology

# Glossary

**Sample Word 1** | Sample Definition 1

# Detailed Licensing

## Overview

**Title:** Environmental Science (Whittinghill)

**Webpages:** 113

**Applicable Restrictions:** Noncommercial

**All licenses found:**

- CC BY-NC-SA 4.0: 91.2% (103 pages)
- CC BY 4.0: 6.2% (7 pages)
- Undeclared: 2.7% (3 pages)

## By Page


- Environmental Science (Whittinghill) - *CC BY-NC-SA 4.0*
  - Front Matter - *CC BY-NC-SA 4.0*
    - Front Matter - *CC BY-NC-SA 4.0*
    - TitlePage - *CC BY-NC-SA 4.0*
    - InfoPage - *CC BY-NC-SA 4.0*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Introduction to Environmental Science - *CC BY-NC-SA 4.0*
    - 1.1: What is Environmental Science? - *CC BY-NC-SA 4.0*
    - 1.2: Introduction to Sustainability - *CC BY-NC-SA 4.0*
    - 1.3: Environmental Ethics and Environmental Justice - *CC BY 4.0*
    - 1.4: Environmental Policy - *CC BY-NC-SA 4.0*
  - 2: The Process of Science - *CC BY-NC-SA 4.0*
  - 3: Matter, Energy, Life - *CC BY-NC-SA 4.0*
    - 3.1: Energy - *CC BY 4.0*
    - 3.2: The Building Blocks of Molecules - *CC BY-NC-SA 4.0*
    - 3.3: A Cell is the Smallest Unit of Life - *CC BY 4.0*
    - 3.4: Water - *CC BY-NC-SA 4.0*
    - 3.5: Themes and Concepts of Biology - *CC BY-NC-SA 4.0*
  - 4: Evolution and the Origin of Species - *CC BY-NC-SA 4.0*
    - 4.1: Understanding Evolution - *CC BY-NC-SA 4.0*
    - 4.2: Formation of New Species - *CC BY-NC-SA 4.0*
  - 5: Cycling of Matter in the Earth System - *CC BY-NC-SA 4.0*
    - 5.1: Introduction - Systems, Cycles, Reservoirs, and Fluxes - *CC BY-NC-SA 4.0*
    - 5.2: Water (Hydrologic) Cycle - *CC BY-NC-SA 4.0*
    - 5.3: Carbon Cycle - *CC BY-NC-SA 4.0*
    - 5.4: Photosynthesis - *CC BY-NC-SA 4.0*
    - 5.5: Nitrogen Cycle - *CC BY-NC-SA 4.0*
    - 5.6: Phosphorus Cycle - *CC BY-NC-SA 4.0*
    - 5.7: Eutrophication and Dead Zones - *CC BY-NC-SA 4.0*
    - 5.8: Sulfur Cycle - *CC BY-NC-SA 4.0*
  - 6: Geology - *CC BY-NC-SA 4.0*
    - 6.1: Introduction to Geology - *CC BY-NC-SA 4.0*
    - 6.2: Fundamentals of Plate Tectonics - *CC BY-NC-SA 4.0*
    - 6.3: The Rock Cycle - *CC BY-NC-SA 4.0*
    - 6.4: Geologic Hazards - *CC BY-NC-SA 4.0*
      - 6.4.1: Earthquakes - *CC BY-NC-SA 4.0*
      - 6.4.2: Volcanoes - *CC BY-NC-SA 4.0*
      - 6.4.3: Mass Wasting - *CC BY-NC-SA 4.0*
      - 6.4.4: Floods - *CC BY-NC-SA 4.0*
    - 6.5: Geologic Resources - *CC BY-NC-SA 4.0*
      - 6.5.1: Mineral Resources - *CC BY-NC-SA 4.0*
      - 6.5.2: Mining - *CC BY-NC-SA 4.0*
      - 6.5.3: Fossil Fuels- Formation and Mining - *CC BY-NC-SA 4.0*
    - 6.6: Soils - *CC BY-NC-SA 4.0*
  - 7: Energy - *CC BY-NC-SA 4.0*
    - 7.1: Energy from Fossil Fuels - *CC BY-NC-SA 4.0*
    - 7.2: Nuclear Energy - *CC BY-NC-SA 4.0*
    - 7.3: Solar Energy - *CC BY-NC-SA 4.0*
    - 7.4: Hydropower - *CC BY-NC-SA 4.0*
    - 7.5: Geothermal Energy - *CC BY-NC-SA 4.0*
    - 7.6: Wind Energy - *CC BY-NC-SA 4.0*
    - 7.7: Biofuels (Biomass Energy) - *CC BY-NC-SA 4.0*
    - 7.8: The Electrical Grid, Battery Challenges, and Energy Conservation - *CC BY-NC-SA 4.0*
  - 8: Atmosphere and Air Pollution - *CC BY 4.0*

- 8.1: The Atmosphere, Climate, and Weather - *CC BY-NC-SA 4.0*
- 8.2: Earth's Energy Balance - *CC BY-NC-SA 4.0*
- 8.3: Atmospheric and Ocean Circulation - *CC BY-NC-SA 4.0*
- 8.4: Air Pollution - *CC BY-NC-SA 4.0*
- 8.5: Acid Deposition - *CC BY-NC-SA 4.0*
- 8.6: Stratospheric Ozone Depletion - *CC BY-NC-SA 4.0*

9: Climate Change - *CC BY-NC-SA 4.0*

- 9.1: What Makes the Climate Change? - *CC BY-NC-SA 4.0*
- 9.2: Prehistoric Climate Change and Proxys for Studying Past Climates - *CC BY-NC-SA 4.0*
- 9.3: Anthropogenic Climate Change - *CC BY-NC-SA 4.0*
- 9.4: Implications of Climate Change - *CC BY-NC-SA 4.0*

10: Urbanization - *CC BY-NC-SA 4.0*

- 10.1: Urbanization - *CC BY-NC-SA 4.0*
- 10.2: Environmental Challenges of Urbanization - *CC BY-NC-SA 4.0*
- 10.3: Sustainable Cities - *CC BY-NC-SA 4.0*
- 10.4: Case Study- Sustainable Stormwater Management - *CC BY-NC-SA 4.0*

11: Solid Waste Management - *CC BY-NC-SA 4.0*

- 11.1: Waste Generation - *CC BY-NC-SA 4.0*
- 11.2: Waste Reduction - *CC BY-NC-SA 4.0*
- 11.3: Life Cycle Assessment - *CC BY-NC-SA 4.0*
- 11.4: Waste Disposal - *CC BY-NC-SA 4.0*
- 11.5: Waste Regulations - *CC BY-NC-SA 4.0*
- 11.6: Solid Waste and Marine Life - *CC BY-NC-SA 4.0*
- 11.7: Case Study- Electronic Waste and Extended Producer Responsibility - *CC BY-NC-SA 4.0*

12: Water Supply and Water Pollution - *CC BY-NC-SA 4.0*

- 12.1: Fresh Water Supply - *CC BY-NC-SA 4.0*
- 12.2: Case Study - The Aral Sea - Going, Going, Gone - *CC BY 4.0*
- 12.3: Water Pollutants and Their Sources - *CC BY-NC-SA 4.0*
- 12.4: Water and Wastewater Treatment - *CC BY 4.0*
- 12.5: Mitigating Water Pollution - *CC BY-NC-SA 4.0*

13: Agriculture - *CC BY-NC-SA 4.0*

- 13.1: Food Security - *CC BY-NC-SA 4.0*
- 13.2: Industrial Agriculture - *CC BY-NC-SA 4.0*
- 13.3: Pests and Pesticides - *CC BY 4.0*
- 13.4: Soil Degradation - *CC BY-NC-SA 4.0*
- 13.5: Selective Breeding and Genetic Engineering - *CC BY-NC-SA 4.0*
- 13.6: Pollination and Seed Dispersal - *CC BY-NC-SA 4.0*
- 13.7: Sustainable Agriculture - *CC BY-NC-SA 4.0*

14: Population Ecology - *CC BY-NC-SA 4.0*

- 14.1: Population Demography - *CC BY-NC-SA 4.0*
- 14.2: Population Growth and Regulation - *CC BY-NC-SA 4.0*
- 14.3: Human Population Growth - *CC BY-NC-SA 4.0*

15: Community and Ecosystem Ecology - *CC BY-NC-SA 4.0*

- 15.1: Community Ecology- Species Interactions - *CC BY-NC-SA 4.0*
- 15.2: Community Structure - *CC BY-NC-SA 4.0*
- 15.3: Community Dynamics - *CC BY-NC-SA 4.0*
- 15.4: Ecology of Ecosystems - *CC BY-NC-SA 4.0*
- 15.5: Energy Flow Through Ecosystems - *CC BY-NC-SA 4.0*

16: Conservation Biology - *CC BY-NC-SA 4.0*

- 16.1: Introduction to Biodiversity - *CC BY-NC-SA 4.0*
- 16.2: Measures of Biodiversity Loss - *CC BY-NC-SA 4.0*
- 16.3: Threats to Biodiversity - *CC BY-NC-SA 4.0*
- 16.4: Importance of Biodiversity - *CC BY-NC-SA 4.0*
- 16.5: Preserving Biodiversity - *CC BY-NC-SA 4.0*

Back Matter - *CC BY-NC-SA 4.0*

- Index - *CC BY-NC-SA 4.0*
- Glossary - *CC BY-NC-SA 4.0*
- Detailed Licensing - *Undeclared*