

# The Habitable Planet

## Unit 1 : Many Planets, One Earth



Surfaces of Mars, Moon, Venus, Earth. Source: NASA

### Overview

Astronomers have discovered dozens of planets orbiting other stars, and space probes have explored many parts of our solar system, but so far scientists have only discovered one place in the universe where conditions are suitable for complex life forms: Earth. In this unit, examine the unique characteristics that make our planet habitable and learn how these conditions were created.

### Sections:

1. Introduction
2. Many Planets, One Earth
3. Reading Geologic Records
4. Carbon Cycling and Earth's Climate
5. Testing the Thermostat: Snowball Earth
6. Atmospheric Oxygen
7. Early Life: Single-Celled Organisms
8. The Cambrian Explosion and the Diversification of Animals
9. The Age of Mammals
10. Further Reading

## 1. Introduction

Earth's long history tells a story of constant environmental change and of close connections between physical and biological environments. It also demonstrates the robustness of life. Simple organisms first appeared on Earth some 3.8 billion years ago, and complex life forms emerged approximately 2 billion years ago. Life on Earth has endured through many intense stresses, including ice ages, warm episodes, high and low oxygen levels, mass extinctions, huge volcanic eruptions, and meteorite impacts. Untold numbers of species have come and gone, but life has survived even the most extreme fluxes.

To understand why Earth has been so conducive to life, we need to identify key conditions that make it habitable and ask why they exist here but not on neighboring planets. This unit describes how Earth's carbon cycle regulates its climate and keeps surface temperatures within a habitable range. It also examines another central factor: the rise of free oxygen in the atmosphere starting more than 2 billion years ago. Next we briefly survey the evolution of life on Earth from simple life forms through the Cambrian explosion and the diversification of multicellular organisms—including, most recently, humans. This unit also describes how scientists find evidence in today's geologic records for events that took place millions or even billions of years ago (Fig. 1).



Figure 1. Evidence of glaciation in seaside rocks, Massachusetts

Humans are latecomers in geologic time: when Earth's history is mapped onto a 24-hour time scale, we appear less than half a minute before the clock strikes midnight (footnote 1). But even though

# The Habitable Planet

humans have been present for a relatively short time, our actions are changing the environment in many ways, which are addressed in units 5 through 13 of this course. Life on Earth will persist in spite of these human impacts. But it remains to be seen how our species will manage broad-scale challenges to our habitable planet, especially those that we create. As history shows, Earth has maintained conditions over billions of years that are uniquely suitable for life on Earth, but those conditions can fluctuate widely. Human impacts add to a natural level of ongoing environmental change.

## 2. Many Planets, One Earth

Our solar system formed from a solar nebula, or cloud of gas and dust, that collapsed and condensed about 4.56 billion years ago. Most of this matter compacted together to form the sun, while the remainder formed planets, asteroids, and smaller bodies. The outer planets, Jupiter, Saturn, Uranus, and Neptune, condensed at cold temperatures far from the sun. Like the sun, they are made mostly of hydrogen and helium. In contrast, the terrestrial planets, Mercury, Venus, Earth and Mars, formed closer to the sun where temperatures were too high to allow hydrogen and helium to condense. Instead they contain large amounts of iron, silicates (silicon and oxygen), magnesium, and other heavier elements that condense at high temperatures.

The young Earth was anything but habitable. Radioactive elements decaying within its mass and impacts from debris raining down from space generated intense heat—so strong that the first eon of Earth's history, from about 4.5 to 3.8 billion years ago, is named the Hadean after **hades**, the Greek word for hell. Most original rock from this period was melted and recycled into Earth's crust, so very few samples remain from our planet's formative phase. But by studying meteorites—stony or metallic fragments up to 4.5 billion years old that fall to Earth from space—scientists can see what materials were present when the solar system was formed and how similar materials may have been melted, crystallized, and transformed as Earth took shape (Fig. 2).

# The Habitable Planet



Figure 2. The Willamette Meteorite, the largest ever found in the United States (15 tons)

© Denis Finnin, American Museum of Natural History.

About 4 billion years ago, conditions on Earth gradually began to moderate. The planet's surface cooled, allowing water vapor to condense in the atmosphere and fall back as rain. This early hydrologic cycle promoted rock weathering, a key part of the carbon-silicate cycle that regulates Earth's climate (discussed in section 4). Evidence from ancient sediments indicates that oceans existed on Earth as long ago as 3.5 billion years.

Conditions evolved very differently on adjoining planets. Venus, which has nearly the same size and density as Earth and is only about 30 percent closer to the sun, is sometimes referred to as our "sister planet." Scientists once thought that conditions on Venus were much like those on Earth, just a little bit warmer. But in reality Venus is a stifling inferno with an average surface temperature greater than 460°C (860°F). This superheated climate is produced by Venus's dense atmosphere, which is about 100 times thicker than Earth's atmosphere and is made up almost entirely of carbon dioxide (CO<sub>2</sub>) (Fig. 3). As we will see in Unit 2, "Atmosphere," CO<sub>2</sub> is a greenhouse gas that traps heat reflected back from planetary surfaces, warming the planet. To make conditions even more toxic, clouds on Venus consists mainly of sulfuric acid droplets.

Paradoxically, if Venus had an atmosphere with the same composition as Earth's, Venus would be colder even though it is closer to the sun and receives approximately twice as much solar radiation as Earth does. This is because Venus has a higher **albedo** (its surface is brighter than Earth's surface), so it reflects a larger fraction of incoming sunlight back to space. Venus is hot because its dense atmosphere functions like a thick blanket and traps this outgoing radiation. An atmosphere with the same makeup as Earth's would function like a thinner blanket, allowing more radiation to escape back to space (Fig. 3).



Figure 3. Comparison of Venus and Earth

Courtesy NASA/JPL-Caltech.

Mars is not much farther from the sun than Earth, but it is much colder. Clouds of ice and frozen  $\text{CO}_2$  (dry ice) drift over its surface. Frozen ice caps at the poles, which can be seen from Earth with a telescope, reflect sunlight. Although Mars's atmosphere consists mainly of  $\text{CO}_2$ , it is 100 times thinner than Earth's atmosphere, so it provides only a small warming effect. Early in its history, the "Red Planet" had an atmosphere dense and warm enough to sustain liquid water, and it may even have had an ocean throughout its northern hemisphere. Today, however, all water on Mars is frozen.

Why is Venus so hot? Why is Mars so cold? And why has the Earth remained habitable instead of phasing into a more extreme state like Mars or Venus? The key difference is that an active carbon cycle has kept Earth's temperature within a habitable range for the past 4 billion years, despite changes in the brightness of the sun during that time. This process is described in detail in section 4, "Carbon Cycling and Earth's Climate." Moderate surface temperatures on Earth have created other important conditions for life, such as a hydrologic cycle that provides liquid water.

How unique are the conditions that allowed life to develop and diversify on Earth? Some scientists contend that circumstances on Earth were extremely unusual and that complex life is very unlikely to find such favorable conditions elsewhere in our universe, although simple life forms like microbes may be very common (footnote 2). Other scientists believe that Earth's history may not be the only environment in which life could develop, and that other planets with very different sets of conditions could foster complex life. What is generally agreed, however, is that no other planet in our solar

system has developed along the same geologic and biologic path as Earth. Life as we know it is a direct result of specific conditions that appear thus far to be unique to our planet.

### 3. Reading Geologic Records

Scientists have divided Earth's history into a series of time segments that are collectively referred to as the geologic time scale (Figure 4). Each of these units is defined based on geologic and fossil records, with divisions between the units marking some major change such as the appearance of a new class of living creatures or a mass extinction. Geologic time phases become shorter as we move forward from Earth's formation toward the present day because records grow increasingly rich. Newer rocks and fossils are better preserved than ancient deposits, so more information is available to categorize recent phases in detail and pinpoint when they began and ended.

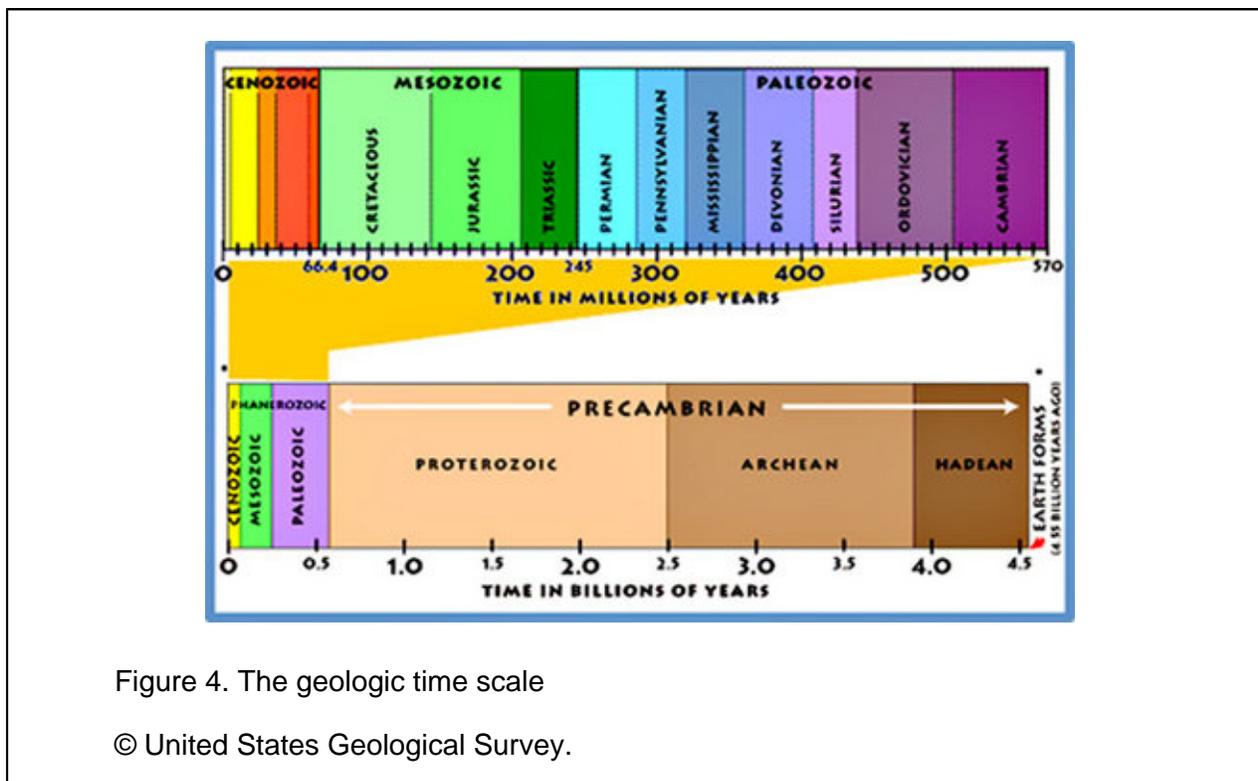


Figure 4. The geologic time scale

© United States Geological Survey.

Most of what we know about our planet's history is based on studies of the [stratigraphic record](#)—rock layers and fossil remains embedded in them. These rock records can provide insights into questions such as how geological formations were created and exposed, what role was played by living organisms, and how the compositions of oceans and the atmosphere have changed through geologic time.

Scientists use stratigraphic records to determine two kinds of time scales. Relative time refers to sequences—whether one incident occurred before, after, or at the same time as another. The geologic time scale shown in Figure 4 reads upwards because it is based on observations from sedimentary rocks, which accrete from the bottom up (wind and water lay down sediments, which are then compacted and buried). However, the sedimentary record is discontinuous and incomplete because **plate tectonics** are constantly reshaping Earth's crust. As the large plates on our planet's surface move about, they split apart at some points and collide or grind horizontally past each other at others. These movements leave physical marks: volcanic rocks intrude upward into sediment beds, plate collisions cause folding and faulting, and erosion cuts the tops off of formations thrust up to the surface.

Geologists have some basic rules for determining relative ages of rock layers. For example, older beds lie below younger beds in undisturbed formations, an intruding rock is younger than the layers it intrudes into, and faults are younger than the beds they cut across. In the geologic cross-section shown in Figure 5, layers E, F, G, H, I, and J were deposited through sedimentation, then cut by faults L and K, then covered by layers D, C, and B. A is a volcanic intrusion younger than the layers it penetrates.

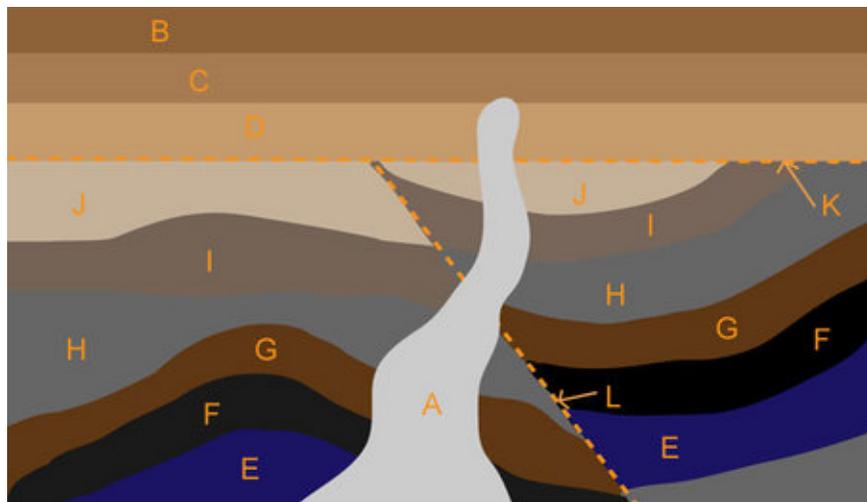


Figure 5. Sample geologic cross-section

Scientists also use fossil records to determine relative age. For example, since fish evolved before mammals, a rock formation at site A that contains fish fossils is older than a formation at site B that contains mammalian fossils. And environmental changes can leave telltale geologic imprints in rock records. For example, when free oxygen began to accumulate in the atmosphere, certain types of

rocks appeared for the first time in sedimentary beds and others stopped forming (for more details, see section 6, "Atmospheric Oxygen"). Researchers study mineral and fossil records together to trace interactions between environmental changes and the evolution of living organisms.

Until the early twentieth century, researchers could only assign relative ages to geologic records. More recently, the expanding field of nuclear physics has enabled scientists to calculate the absolute age of rocks and fossils using **radiometric dating**, which measures the decay of radioactive isotopes in rock samples. This approach has been used to determine the ages of rocks more than 3.5 billion years old (footnote 3). Once they establish the age of multiple formations in a region, researchers can correlate strata among those formations to develop a fuller record of the entire area's geologic history (Fig. 6).

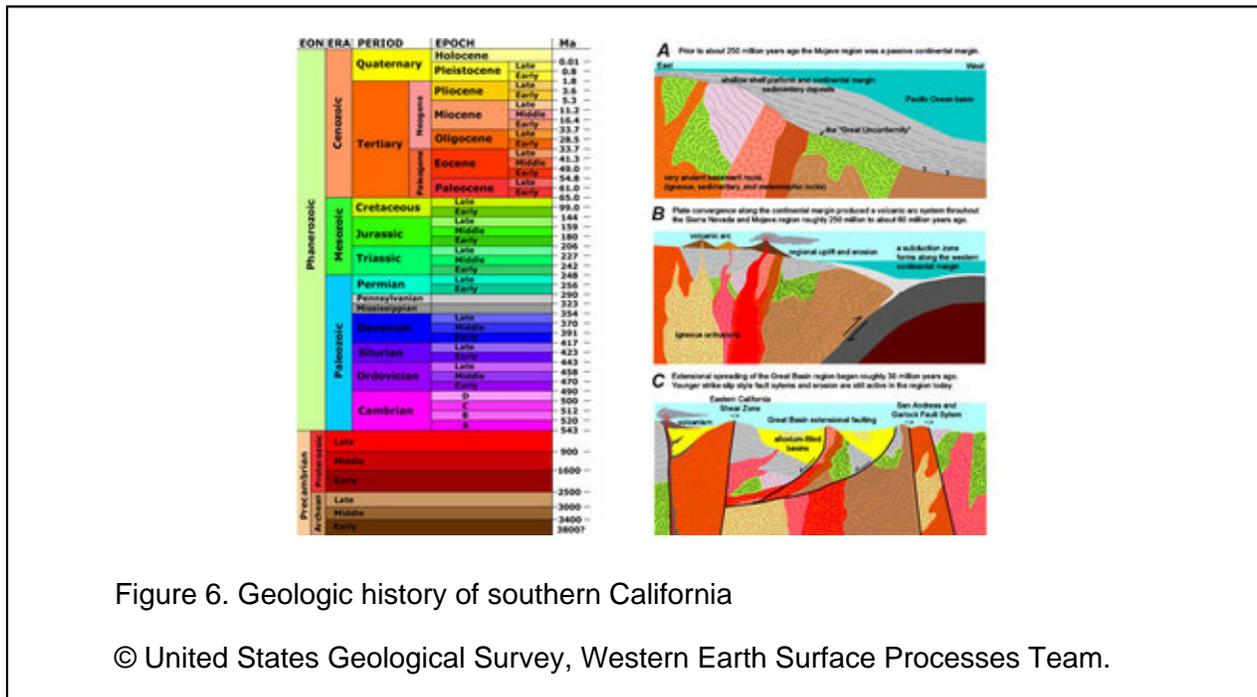


Figure 6. Geologic history of southern California

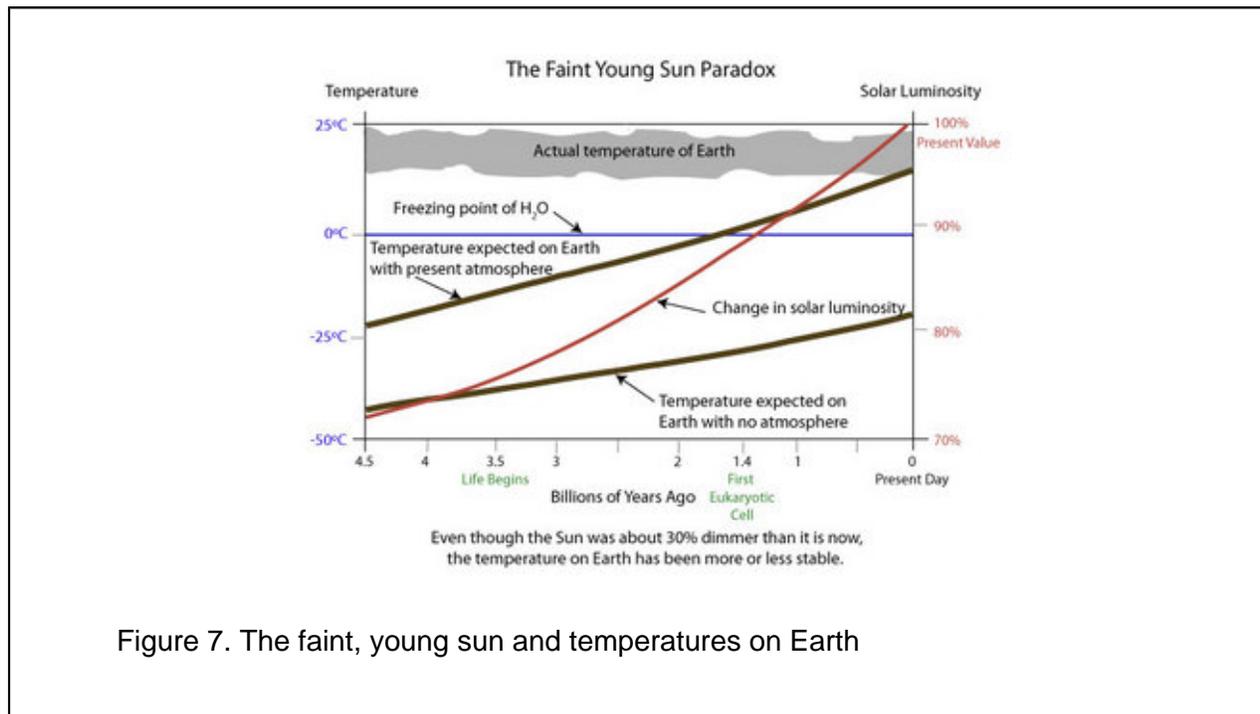
© United States Geological Survey, Western Earth Surface Processes Team.

Our understanding of Earth's history and the emergence of life draws on other scientific fields along with geology and paleontology. Biologists trace genealogical relationships among organisms and the expansion of biological diversity. And climate scientists analyze changes in Earth's atmosphere, temperature patterns, and geochemical cycles to determine why events such as ice ages and rapid warming events occurred. All of these perspectives are relevant because, as we will see in the following sections, organisms and the physical environment on Earth have developed together and influenced each other's evolution in many ways.

## 4. Carbon Cycling and Earth's Climate

How did early Earth transition from a hell-like environment to temperatures more hospitable to life? Early in the Archean (ancient) eon, about 3.8 billion years ago, the rain of meteors and rock bodies from space ended, allowing our planet's surface to cool and solidify. Water vapor in the atmosphere condensed and fell as rain, creating oceans. These changes created the conditions for **geochemical cycling**—flows of chemical substances between reservoirs in Earth's atmosphere, hydrosphere (water bodies), and lithosphere (the solid part of Earth's crust).

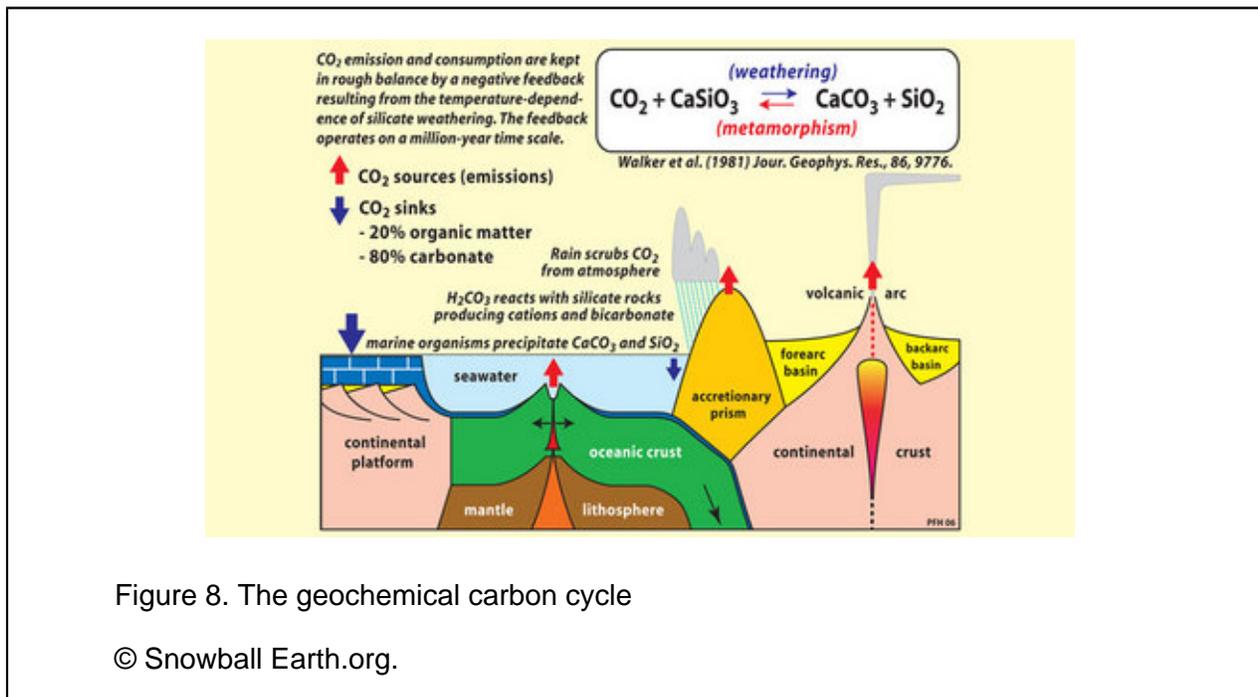
At this time the sun was about 30 percent dimmer than it is today, so our planet received less solar radiation. Earth's surface should have been well below the freezing point of water, too cold for life to exist, but evidence shows that liquid water was present and that simple life forms appeared as far back as 3.5 billion years ago. This contradiction is known as the "faint young sun" paradox (Fig. 7). The unexpected warmth came from greenhouse gases in Earth's atmosphere, which retained enough heat to keep the planet from freezing over.



The Archean atmosphere was a mix of gases including nitrogen, water vapor, methane ( $\text{CH}_4$ ), and  $\text{CO}_2$ . (As discussed in section 6, "Atmospheric Oxygen," free oxygen did not accumulate in the atmosphere until more than two billion years after Earth was formed.) Volcanoes emitted  $\text{CO}_2$  as a byproduct of heating within the Earth's crust. But instead of developing a runaway greenhouse effect

like that on Venus, Earth's temperatures remained within a moderate range because the carbon cycle includes a natural **sink**—a process that removes excess carbon from the atmosphere. This sink involves the weathering of silicate rocks, such as granites and basalts, that make up much of Earth's crust.

As illustrated in Figure 8, this process has four basic stages. First, rainfall scrubs CO<sub>2</sub> out of the air, producing carbonic acid (H<sub>2</sub>CO<sub>3</sub>), a weak acid. Next, this solution reacts on contact with silicate rocks to release calcium and other **cations** and leave behind carbonate and bicarbonate ions dissolved in the water. This solution is washed into the oceans by rivers, and then calcium carbonate (CaCO<sub>3</sub>), also known as limestone, is precipitated in sediments. (Today most calcium carbonate precipitation is caused by marine organisms, which use calcium carbonate to make their shells.) Over long time scales, oceanic crust containing limestone sediments is forced downward into Earth's mantle at points where plates collide, a process called **subduction**. Eventually, the limestone heats up and turns the limestone back into CO<sub>2</sub>, which travels back up to the surface with magma. Volcanic activity then returns CO<sub>2</sub> to the atmosphere.



Many climatic factors influence how quickly this process takes place. Warmer temperatures speed up the chemical reactions that take place as rocks weather, and increased precipitation may flush water more rapidly through soil and sedimentary rocks. This creates a **negative feedback** relationship between rock weathering and climatic changes: when Earth's climate warms or cools, the

system responds in ways that moderate the temperature change and push conditions back toward equilibrium, essentially creating a natural thermostat.

For example, when the climate warms, weathering rates accelerate and convert an increasing fraction of atmospheric  $\text{CO}_2$  to calcium carbonate, which is buried on the ocean floor. Atmospheric concentrations of  $\text{CO}_2$  decline, modifying the greenhouse effect and cooling Earth's surface. In the opposite instance, when the climate cools weathering slows down but volcanic outgassing of  $\text{CO}_2$  continues, so atmospheric  $\text{CO}_2$  levels rise and warm the climate.

This balance between  $\text{CO}_2$  outgassing from volcanoes and  $\text{CO}_2$  conversion to calcium carbonate through silicate weathering has kept the Earth's climate stable through most of its history. Because this feedback takes a very long time, typically hundreds of thousands of years, it cannot smooth out all the fluctuations like a thermostat in one's home. As a result, our planet's climate has fluctuated dramatically, but it has never gone to permanent extremes like those seen on Mars and Venus.

Why is Venus a runaway greenhouse? Venus has no water on its surface, so it has no medium to dissolve  $\text{CO}_2$ , form carbonic acid, and react with silicate rocks. As a result volcanism on Venus continues to emit  $\text{CO}_2$  without any carbon sink, so it accumulates in the atmosphere. Mars may have had such a cycle early in its history, but major volcanism stopped on Mars more than 3 billion years ago, so the planet eventually cooled as  $\text{CO}_2$  escaped from the atmosphere. On Earth, plate tectonics provide continuing supplies of the key ingredients for the carbon-silicate cycle:  $\text{CO}_2$ , liquid water, and plenty of rock.

## 5. Testing the Thermostat: Snowball Earth

We can see how durable Earth's silicate weathering "thermostat" is by looking at some of the most extreme climate episodes on our planet's history: severe glaciations that occurred during the Proterozoic era. The first "Snowball Earth" phase is estimated to have occurred about 2.3 billion years ago, followed by several more between about 750 and 580 million years ago. Proponents of the Snowball Earth theory believe that Earth became so cold during several glacial cycles in this period that it essentially froze over from the equator to the poles for spans of ten million years or more. But ultimately, they contend, the carbon-silicate cycle freed Earth from this deep-freeze state.

If Earth has a natural thermostat, how could it become cold enough for the entire planet to freeze over? One possible cause is continental drift. Researchers believe that around 750 million years ago, most of the continents may have been clustered in the tropics following the breakup of the supercontinent Rodinia (Fig. 9), and that such a configuration would have had pronounced effects on Earth's climate (footnote 4).

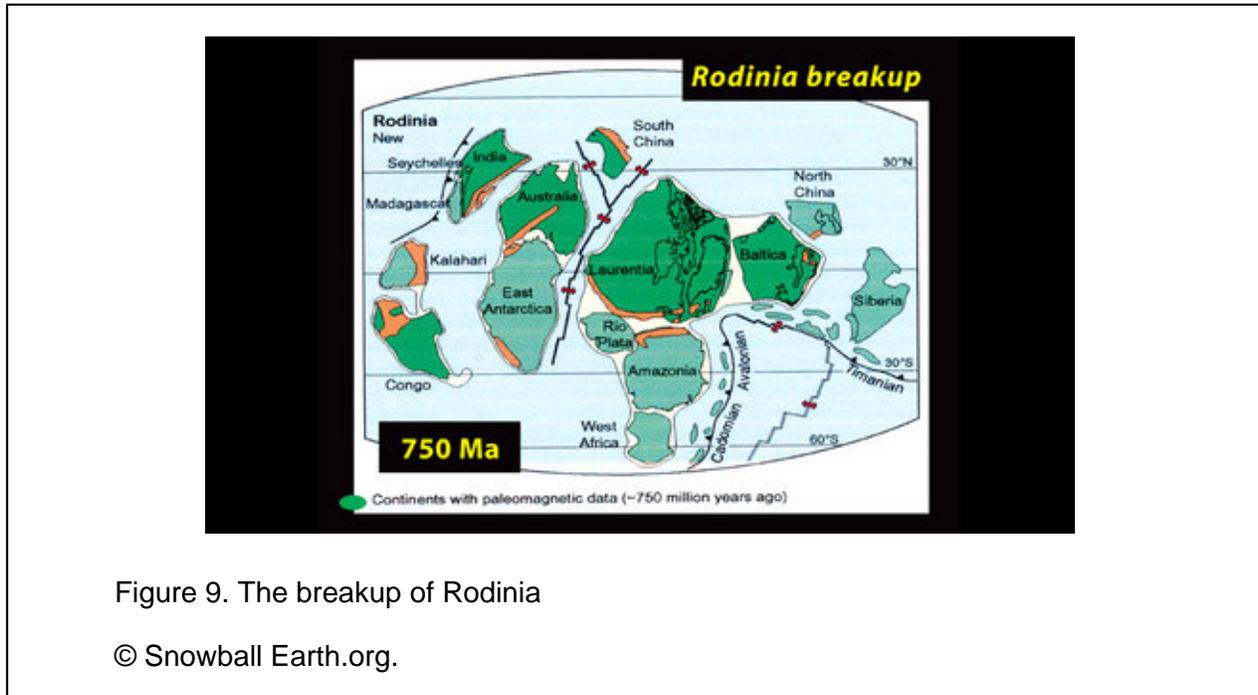


Figure 9. The breakup of Rodinia

© Snowball Earth.org.

Had the continents been located closer to the poles as they are today, ice sheets would have developed at high latitudes as the planet cooled. Ice cover would prevent the rocks beneath from weathering, thus slowing the rate at which carbon was removed from the atmosphere and allowing CO<sub>2</sub> from volcanic eruptions to build up in the atmosphere. As a result, Earth's surface temperature would warm.

But if continents were clustered at low latitudes, Earth's land masses would have remained ice-free for a long time even as ice sheets built up in the polar oceans and reflected a growing fraction of solar energy back to space. Because most continental area was in the tropics, the weathering reactions would have continued even as the Earth became colder and colder. Once sea ice reached past about 30 degrees latitude, Snowball Earth scholars believe that a runaway ice-albedo effect occurred: ice reflected so much incoming solar energy back to space, cooling Earth's surface and causing still more ice to form, that the effect became unstoppable. Ice quickly engulfed the planet and oceans froze to an average depth of more than one kilometer.

The first scientists who imagined a Snowball Earth believed that such a sequence must have been impossible, because it would have cooled Earth so much that the planet would never have warmed up. Now, however, scientists believe that Earth's carbon cycle saved the planet from permanent deep-freeze. How would a Snowball Earth thaw? The answer stems from the carbon-cycle thermostat discussed earlier.



formations were produced when iron concentrations in the deep ocean were very high but some oxygen existed in the surface ocean and atmosphere. Mixing of iron up from the deep ocean into the more oxidized ocean would cause the chemical precipitation of iron, producing iron formations.

Geologists do not find iron formations after about 1.8 billion years ago, once oxygen levels in the atmosphere and ocean were high enough to remove almost all of the dissolved iron. However, iron formations are found once again around 700 million years ago within snowball glacial deposits. The explanation appears to be that during a Snowball Earth episode sea ice formed over most of the ocean's surface, making it difficult for oxygen to mix into the water. Over millions of years iron then built up in seawater until the ice started to melt. Then atmospheric oxygen could mix with the ocean once again, and all the iron was deposited in these unusual iron formations (Fig. 11).

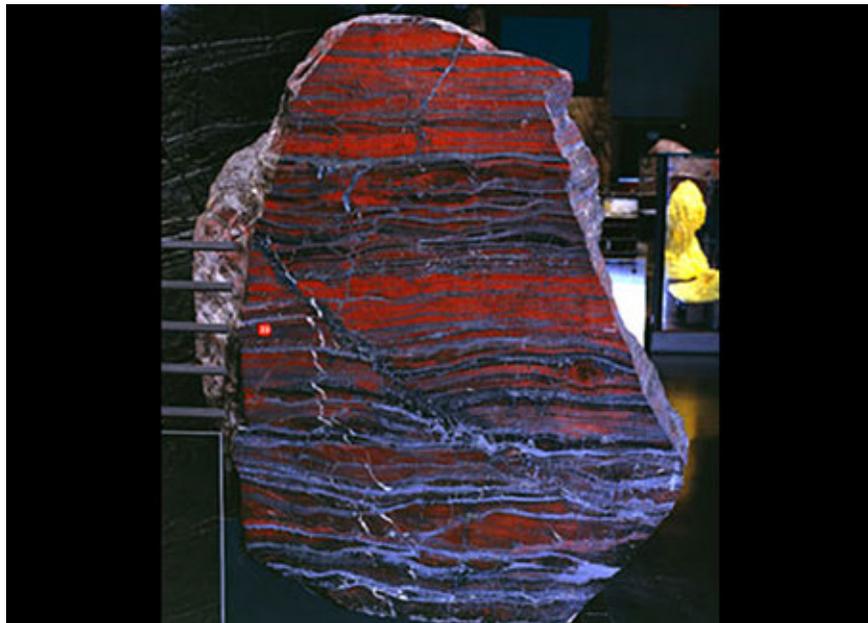


Figure 11. Banded iron formation from Ontario, Canada

© Denis Finnin, American Museum of Natural History.

The Snowball Earth is still a controversial hypothesis. Some scientists argue that the evidence is not sufficient to prove that Earth really did freeze over down to the equator. But the hypothesis is supported by more and more unusual geological observations from this time, and also carries some interesting implications for the evolution of life.

How could life survive a snowball episode? Paradoxically, scientists theorize that these deep freezes may have indirectly spurred the development of complex life forms. The most complex life forms on Earth at the time of the Neoproterozoic glaciations were primitive algae and protozoa. Most of these existing organisms were undoubtedly wiped out by glacial episodes. But recent findings have shown that some microscopic organisms can flourish in extremely challenging conditions—for example, within the channels inside floating sea ice and around vents on the ocean floor where superheated water fountains up from Earth's mantle. These environments may have been the last reservoirs of life during Snowball Earth phases. Even a small amount of geothermal heat near any of the tens of thousands of natural hot springs that exist on Earth would have been sufficient to create small holes in the ice. And those holes would have been wonderful refuges where life could persist.

Organisms adaptable enough to survive in isolated environments would have been capable of rapid genetic evolution in a short time. The last hypothesized Snowball Earth episode ended just a few million years before the Cambrian explosion, an extraordinary diversification of life that took place from 575 to 525 million years ago (discussed in section 8, "Multi-Celled Organisms and the Cambrian Explosion"). It is possible, although not proven, that the intense selective pressures of snowball glaciations may have fostered life forms that were highly adaptable and ready to expand quickly once conditions on Earth's surface moderated.

## 6. Atmospheric Oxygen

A stable climate is only one key requirement for the complex life forms that populate Earth today. Multi-cellular organisms also need a ready supply of oxygen for respiration. Today oxygen makes up about 20 percent of Earth's atmosphere, but for the first two billion years after Earth formed, its atmosphere was anoxic (oxygen-free). About 2.3 billion years ago, oxygen increased from a trace gas to perhaps one percent of Earth's atmosphere. Another jump took place about 600 million years ago, paving the way for multi-cellular life forms to expand during the Cambrian Explosion.

Oxygen is a highly reactive gas that combines readily with other elements like hydrogen, carbon, and iron. Many metals react directly with oxygen in the air to form metal oxides. For example, rust is an oxide that forms when iron reacts with oxygen in the presence of water. This process is called [oxidation](#), a term for reactions in which a substance loses electrons and become more positively charged. In this case, iron loses electrons to oxygen (Fig. 12). What little free oxygen was produced in Earth's atmosphere during the Archean eon would have quickly reacted with other gases or with minerals in surface rock formations, leaving none available for respiration.

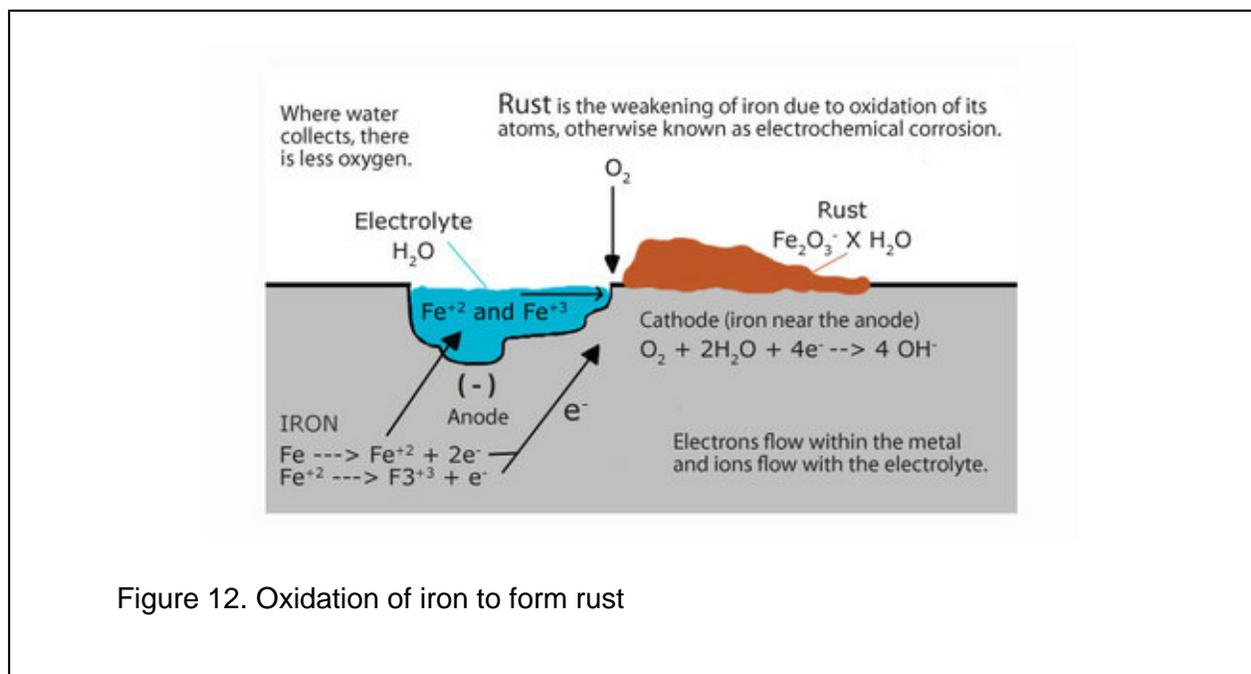


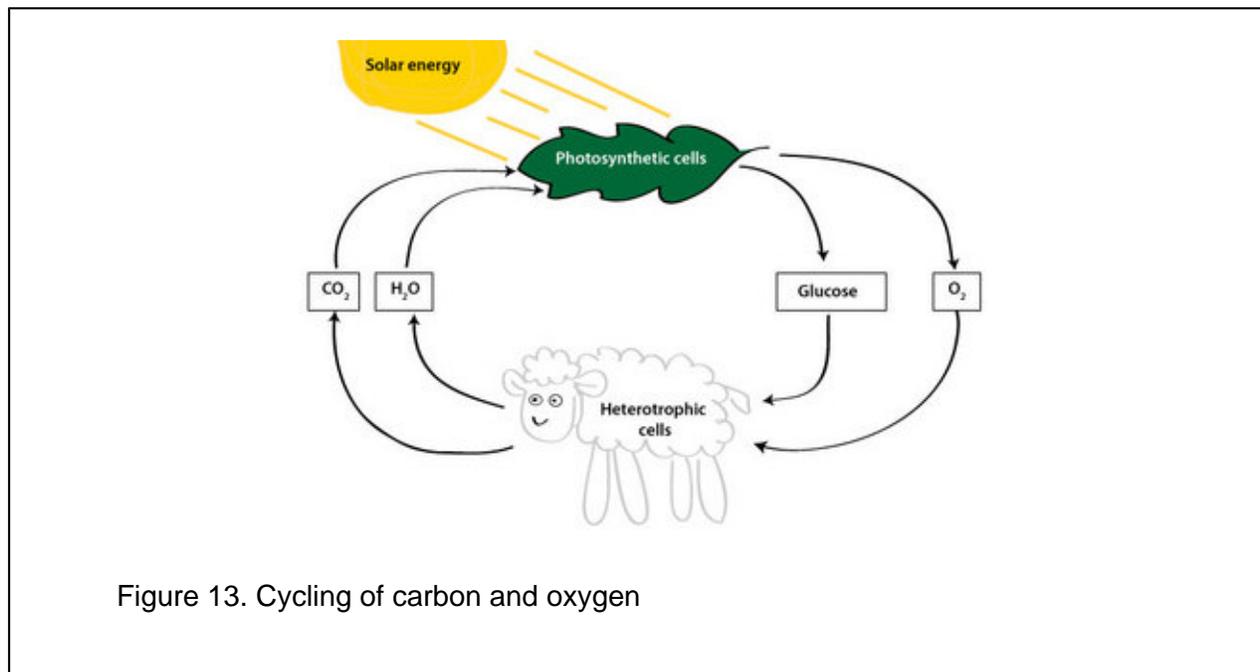
Figure 12. Oxidation of iron to form rust

Geologists trace the rise of atmospheric oxygen by looking for oxidation products in ancient rock formations. We know that very little oxygen was present during the Archean eon because sulfide minerals like pyrite (fool's gold), which normally oxidize and are destroyed in today's surface environment, are found in river deposits dating from that time. Other Archean rocks contain banded iron formations (BIFs)—the sedimentary beds described in section 5 that record periods when waters contained high concentrations of iron. These formations tell us that ancient oceans were rich in iron, creating a large sink that consumed any available free oxygen.

Scientists agree that atmospheric oxygen levels increased about 2.3 billion years ago to a level that may have constituted about 1 percent of the atmosphere. One indicator is the presence of rock deposits called red beds, which started to form about 2.2 billion years ago and are familiar to travelers who have visited canyons in Arizona or Utah. These strata of reddish sedimentary rock, which formed from soils rich in iron oxides, are basically the opposite of BIFs: they indicate that enough oxygen had accumulated in the atmosphere to oxidize iron present in soil. If the atmosphere had still been anoxic, iron in these soils would have remained in solution and would have been washed away by rainfall and river flows. Other evidence comes from changes in sulfur isotope ratios in rocks, which indicate that about 2.4 billion years ago sulfur chemistry changed in ways consistent with increasing atmospheric oxygen.

Why did oxygen levels rise? [Cyanobacteria](#), the first organisms capable of producing oxygen through photosynthesis, emerged well before the first step up in atmospheric oxygen concentrations, perhaps as early as 2.7 billion years ago. Their oxygen output helped to fill up the chemical sinks, such as iron

in soils, that removed oxygen from the air. But plant photosynthesis alone would not have provided enough oxygen to account for this increase, because **heterotrophs** (organisms that are not able to make their own food) respire oxygen and use it to metabolize organic material. If all new plant growth is consumed by animals that feed on living plants and decomposers that break down dead plant material, carbon and oxygen cycle in what is essentially a closed loop and net atmospheric oxygen levels remain unchanged (Fig. 13).



However, material can leak out of this loop and alter carbon-oxygen balances. If organic matter produced by photosynthesis is buried in sediments before it decomposes (for example, dead trees may fall into a lake and sink into the lake bottom), it is no longer available for respiration. The oxygen that decomposers would have consumed as they broke it down goes unused, increasing atmospheric oxygen concentrations. Many researchers theorize that this process caused the initial rise in atmospheric oxygen.

Some scientists suspect that atmospheric oxygen increased again about 600 million years ago to levels closer to the composition of our modern atmosphere. The main evidence is simply that many different groups of organisms suddenly became much larger at this time. Biologists argue that it is difficult for large, multicellular animals to exist if oxygen levels are extremely low, as such animals cannot survive without a fairly high amount of oxygen. However, scientists are still not sure what caused a jump in oxygen at this time.

One clue may be the strange association of jumps in atmospheric oxygen with snowball glaciations. Indeed, the jumps in atmospheric oxygen at 2.3 billion years ago and 600 million years ago do seem to be associated with Snowball Earth episodes (Fig. 14). However, scientists are still unsure exactly what the connection might be between the extreme ice ages and changes in the oxygen content of the atmosphere.

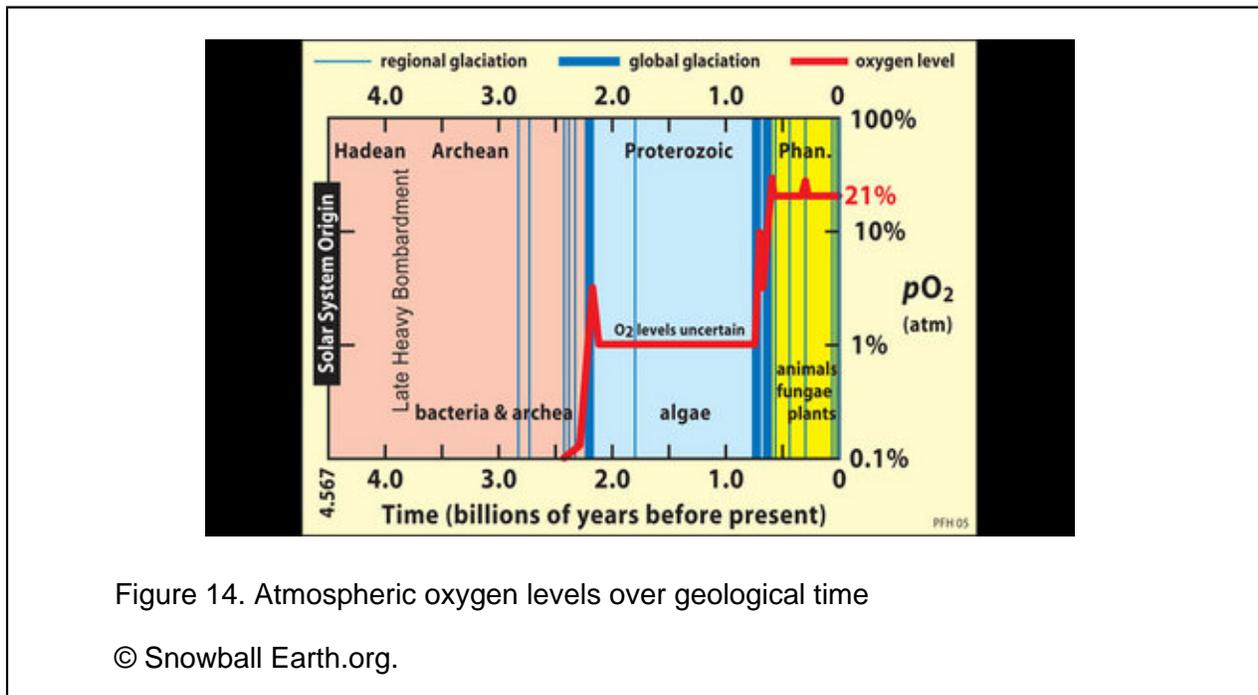


Figure 14. Atmospheric oxygen levels over geological time

© Snowball Earth.org.

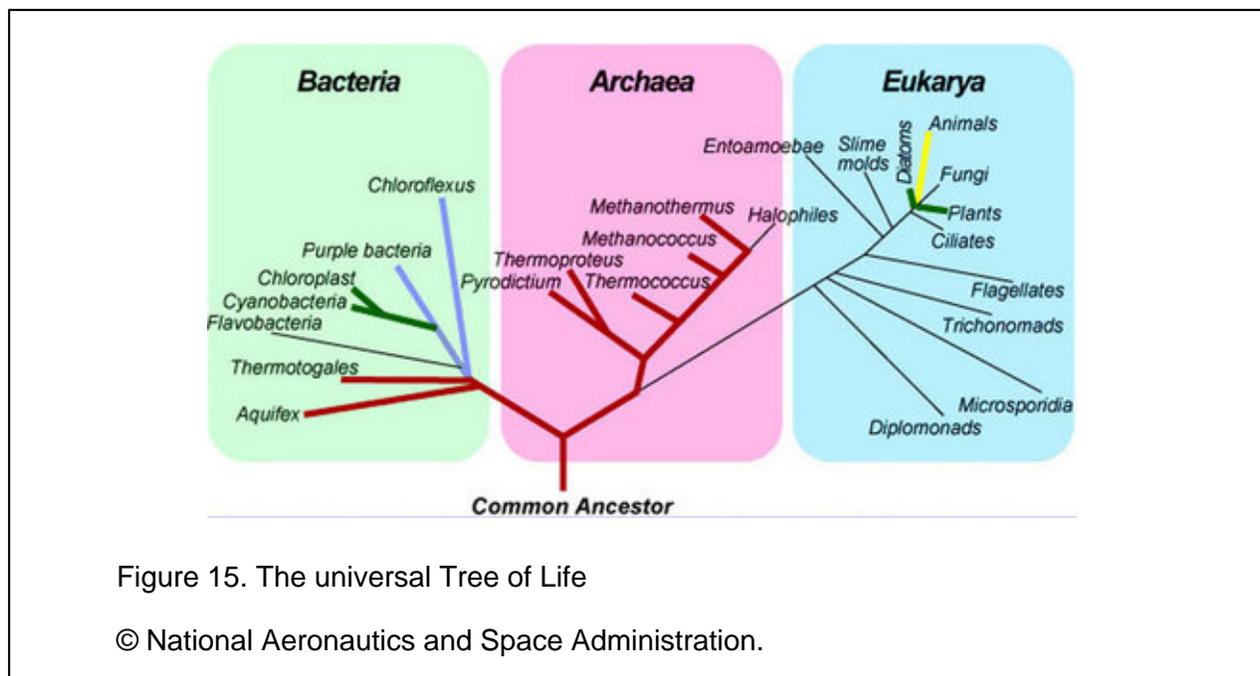
Why have atmospheric oxygen levels stayed relatively stable since this second jump? As discussed above, the carbon-oxygen cycle is a closed system that keeps levels of both elements fairly constant. The system contains a powerful negative feedback mechanism, based on the fact that most animals need oxygen for respiration. If atmospheric oxygen levels rose substantially today, marine zooplankton would eat and respire organic matter produced by algae in the ocean at an increased rate, so a lower fraction of organic matter would be buried, canceling the effect. Falling oxygen levels would reduce feeding and respiration by zooplankton, so more of the organic matter produced by algae would end up in sediments and oxygen would rise again. Fluctuations in either direction thus generate changes that push oxygen levels back toward a steady state.

Forest fires also help to keep oxygen levels steady through a negative feedback. Combustion is a rapid oxidation reaction, so increasing the amount of available oxygen will promote a bigger reaction. Rising atmospheric oxygen levels would make forest fires more common, but these fires would consume large amounts of oxygen, driving concentrations back downward.

## 7. Early Life: Single-Celled Organisms

We do not know exactly when life first appeared on Earth. There is clear evidence that life existed at least 3 billion years ago, and the oldest sediments that have been discovered—rocks formed up to 3.8 billion years ago—bear marks that some scientists believe could have been left by primitive microorganisms. If this is true it tells us that life originated soon after the early period of bombardment by meteorites, on an Earth that was probably much warmer than today and had only traces of oxygen in its atmosphere.

For the first billion years or so, life on Earth consisted of **bacteria** and **archaea**, microscopic organisms that represent two of the three genealogical branches on the Tree of Life (Fig. 15). Both groups are **prokaryotes** (single-celled organisms without nuclei). Archaea were recognized as a unique domain of life in the 1970s, based on some distinctive chemical and genetic features. The order in which branches radiate from the Tree of Life shows the sequence in which organisms evolved. Reading from the bottom up in the same way in which a tree grows and branches outward, we can see that eucarya (multi-celled animals) were the last major group to diverge and that animals are among the newest subgroups within this domain.



Life on Earth existed for many millions of years without atmospheric oxygen. The lowest groups on the Tree of Life, including thermatogales and nearly all of the archaea, are anaerobic organisms that cannot tolerate oxygen. Instead they use hydrogen, sulfur, or other chemicals to harvest energy

through chemical reactions. These reactions are key elements of many chemical cycles on Earth, including the carbon, sulfur, and nitrogen cycles. "Prokaryotic metabolisms form the fundamental ecological circuitry of life," writes paleontologist Andrew Knoll. "Bacteria, not mammals, underpin the efficient and long-term functioning of the biosphere" (footnote 5).

Some bacteria and archaea are **extremophiles**—organisms that thrive in highly saline, acidic, or alkaline conditions or other extreme environments, such as the hot water around hydrothermal vents in the ocean floor. Early life forms' tolerances and anaerobic metabolisms indicate that they evolved in very different conditions from today's environment.

Microorganisms are still part of Earth's chemical cycles, but most of the energy that flows through our biosphere today comes from photosynthetic plants that use light to produce organic material. When did photosynthesis begin? Archaean rocks from western Australia that have been dated at 3.5 billion years old contain organic material and fossils of early cyanobacteria, the first photosynthetic bacteria (footnote 6). These simple organisms jump-started the oxygen revolution by producing the first traces of free oxygen through photosynthesis: Knoll calls them "the working-class heroes of the Precambrian Earth" (footnote 7).

Cyanobacteria are widely found in tidal flats, where the organic carbon that they produced was buried, increasing atmospheric oxygen concentrations. Mats of cyanobacteria and other microbes trapped and bound sediments, forming wavy structures called stromatolites (layered rocks) that mark the presence of microbial colonies (Fig. 16).



Figure 16. Stromatolites at Hamelin Pool, Shark Bay, Australia

© National Aeronautics and Space Administration, JSC Astrobiology Institute.

The third domain of life, **eukaryotes**, are organisms with one or more complex cells. A eukaryotic cell contains a nucleus surrounded by a membrane that holds the cell's genetic material. Eukaryotic cells also contain organelles—sub-components that carry out specialized functions such as assembling proteins or digesting food. In plant and eukaryotic algae cells, chloroplasts carry out photosynthesis. These organelles developed through a process called endosymbiosis in which cyanobacteria took up residence inside host cells and carried out photosynthesis there. Mitochondria, the organelles that conduct cellular respiration (converting energy into usable forms) in eukaryotic cells, are also descended from cyanobacteria.

The first eukaryotic cells evolved sometime between 1.7 and 2.5 billion years ago, perhaps coincident with the rise in atmospheric oxygen around 2.3 billion years ago. As the atmosphere and the oceans became increasingly oxygenated, organisms that used oxygen spread and eventually came to dominate Earth's biosphere. Chemosynthetic organisms remained common but retreated into sediments, swamps, and other anaerobic environments.

Throughout the Proterozoic era, from about 2.3 billion years ago until around 575 million years ago, life on Earth was mostly single-celled and small. Earth's biota consisted of bacteria, archaea, and eukaryotic algae. Food webs began to develop, with amoebas feeding on bacteria and algae. Earth's land surfaces remained harsh and largely barren because the planet had not yet developed a protective ozone layer (this screen formed later as free oxygen increased in the atmosphere), so it was bombarded by intense ultraviolet radiation. However, even shallow ocean waters shielded microorganisms from damaging solar rays, so most life at this time was aquatic.

As discussed in sections 5 and 6, global glaciations occurred around 2.3 billion years ago and again around 600 million years ago. Many scientists have sought to determine whether there is a connection between these episodes and the emergence of new life forms around the same times. For example, one Snowball Earth episode about 635 million years ago is closely associated with the emergence of multicellularity in microscopic animals (footnote 8). However, no causal relationship has been proved.

## 8. The Cambrian Explosion and the Diversification of Animals

The first evidence of multicellular animals appears in fossils from the late Proterozoic era, about 575 million years ago, after the last snowball glaciation. These impressions were made by soft-bodied organisms such as worms, jellyfish, sea pens, and polyps similar to modern sea anemones (Fig. 17). In contrast to the microorganisms that dominated the Proterozoic era, many of these fossils are at least several centimeters long, and some measure up to a meter across.



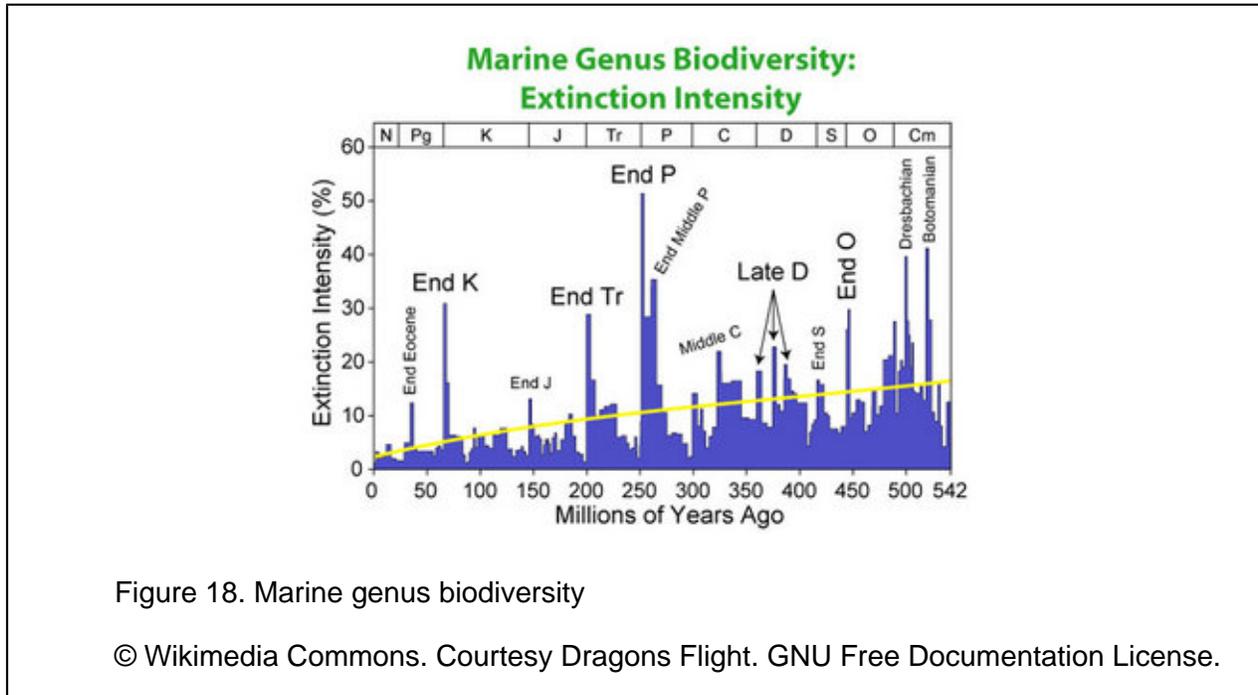
Figure 17. Fossils of Kimberella (thought to be a jellyfish)

Courtesy Wikimedia Commons. GNU Free Documentation License.

Shortly after this time, starting about 540 million years ago, something extraordinary happened: the incredible diversification of complex life known as the **Cambrian Explosion**. Within 50 million years every major animal **phylum** known in fossil records quickly appeared. The Cambrian Explosion can be thought of as multicellular animals' "big bang"—an incredible radiation of complexity.

What triggered the Cambrian Explosion? Scientists have pointed to many factors. For example, the development of predation probably spurred the evolution of shells and armor, while the growing complexity of ecological relationships created distinct roles for many sizes and types of organisms. Rising atmospheric and oceanic oxygen levels promoted the development of larger animals, which need more oxygen than small ones in order to move blood throughout their bodies. And some scientists believe that a mass extinction at the end of the Proterozoic era created a favorable environment for new life forms to evolve and spread.

Following the Cambrian Explosion, life diversified in several large jumps that took place over three eras: Paleozoic, Mesozoic, and Cenozoic (referring back to Fig. 4, the Cambrian period was the first slice of the Paleozoic era). Together these eras make up the Phanerozoic eon, a name derived from the Greek for "visible life." The Phanerozoic, which runs from 540 million years ago to the present, has also been a tumultuous phase in the evolution of life on Earth, with mass extinctions at the boundaries between each of its three geologic eras. Figure 18 shows the scale of historic mass extinctions as reflected in marine fossil records.



Early in the Paleozoic most of Earth's fauna lived in the sea. Many Cambrian organisms developed hard body parts like shells and bones, so fossil records became much more abundant and diverse. The Burgess Shale, rock beds in British Columbia made famous in paleontologist Stephen Jay Gould's book **Wonderful Life**, are ancient reef beds in the Canadian Rockies of British Columbia that are filled with fossil deposits from the mid-Cambrian period (footnote 9).

Land plants emerged between about 500 and 400 million years ago. Once established, they stabilized soil against erosion and accelerated the weathering of rock by releasing chemicals from their roots. Since faster weathering pulls increased amounts of carbon out of the atmosphere, plants reduced the greenhouse effect and cooled Earth's surface so dramatically that they are thought to have helped cause several ice ages and mass extinctions during the late Devonian period, about 375 million years ago. By creating shade, they also provided habitat for the first amphibians to move from water to land.

The most severe of all mass extinctions took place at the end of the Paleozoic era at the Permian/Triassic boundary, wiping out an estimated 80 to 85 percent of all living species. Scientists still do not understand what caused this crisis. Geologic records indicate that deep seas became anoxic, which suggest that something interfered with normal ocean mixing, and that Earth's climate suddenly became much warmer and drier. Possible causes for these developments include massive volcanic eruptions or a melting of methane hydrate deposits (huge reservoirs of solidified methane), both of which could have sharply increased the greenhouse effect.

The Mesozoic era, spanning the Triassic, Jurassic, and Cretaceous periods, was the era of reptiles, which colonized land and air more thoroughly than the amphibians that preceded them out of the water. Dinosaurs evolved in the Triassic, about 215 million years ago, and became the largest and most dominant animals on Earth for the next 150 million years. This period also saw the emergence of modern land plants, including the first angiosperms (flowering plants); small mammals; and the first birds, which evolved from dinosaurs. Figure 19 shows a model of a fossilized Archaeopteryx, a transitional species from the Jurassic period with both avian and dinosaur features.



Figure 19. Model of Archaeopteryx fossil

© Wikimedia Commons. CeCILL license.

Another mass extinction at the end of the Mesozoic, 65 million years ago, killed all of the dinosaurs except for birds, along with many other animals. For many years scientists thought that climate change caused this extinction, but in 1980 physicist Louis Alvarez, his son, Walter, a geologist, and other colleagues published a theory that a huge meteorite had hit Earth, causing impacts like shock waves, severe atmospheric disturbances, and a global cloud of dust that would have drastically cooled the planet. Their most important evidence was widespread deposits of iridium—a metal that is

extremely rare in Earth's crust but that falls to Earth in meteorites—in sediments from the so-called K-T (Cretaceous-Tertiary) boundary layer.

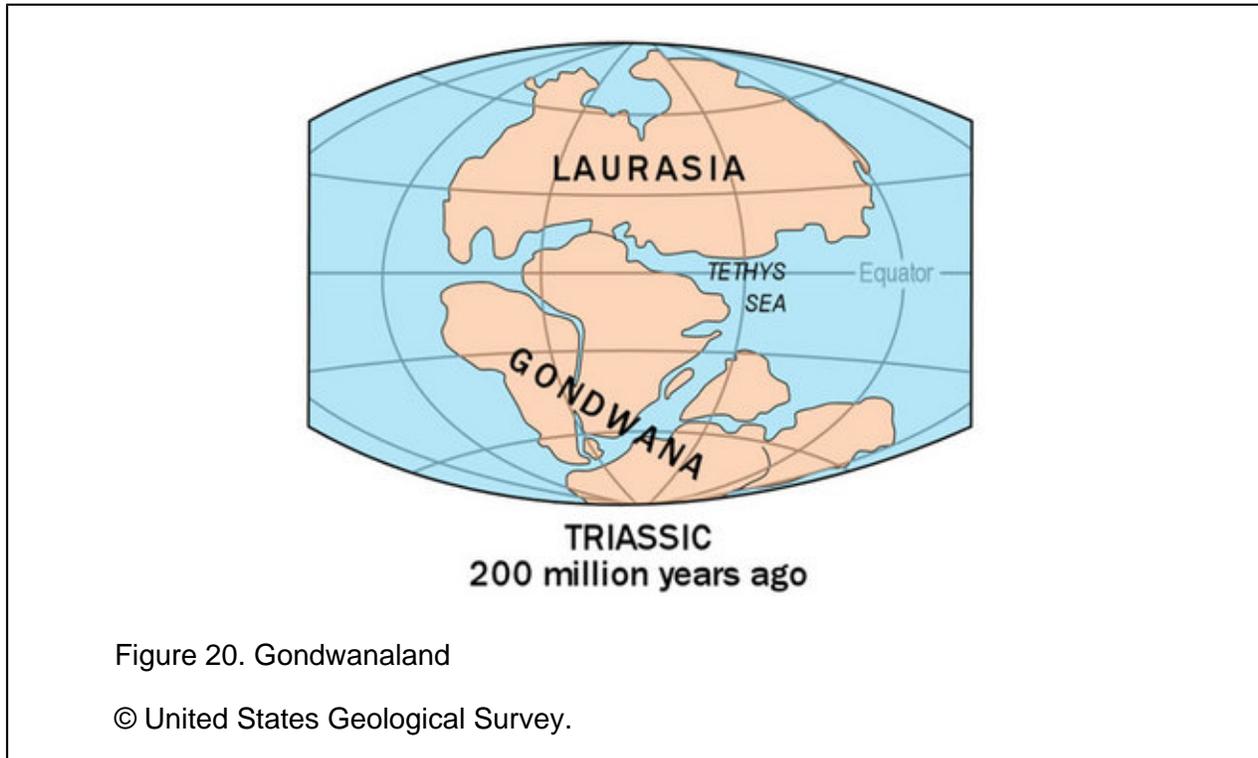
Further evidence discovered since 1980 supports the meteor theory, which is now widely accepted. A crater has been identified at Chicxulub, in Mexico's Yucatan peninsula, that could have been caused by a meteorite big enough to supply the excess iridium, and grains of shocked quartz from the Chicxulub region have been found in sediments thousands of kilometers from the site that date to the K-T boundary era.

## 9. The Age of Mammals

The first mammals on Earth were rodent-sized animals that evolved in the shadow of dinosaurs during the Jurassic and Triassic periods. After the K-T boundary extinction eliminated dinosaurs as predators and competitors, mammals radiated widely. Most of the modern mammal orders, from bats to large types like primates and whales, appeared within about 10 million years after dinosaurs died out. Because mammals could maintain a relatively constant internal temperature in hot or cold environments, they were able to adapt to temperature changes more readily than cold-blooded animals like reptiles, amphibians, and fish. This characteristic helped them to populate a wide range of environments.

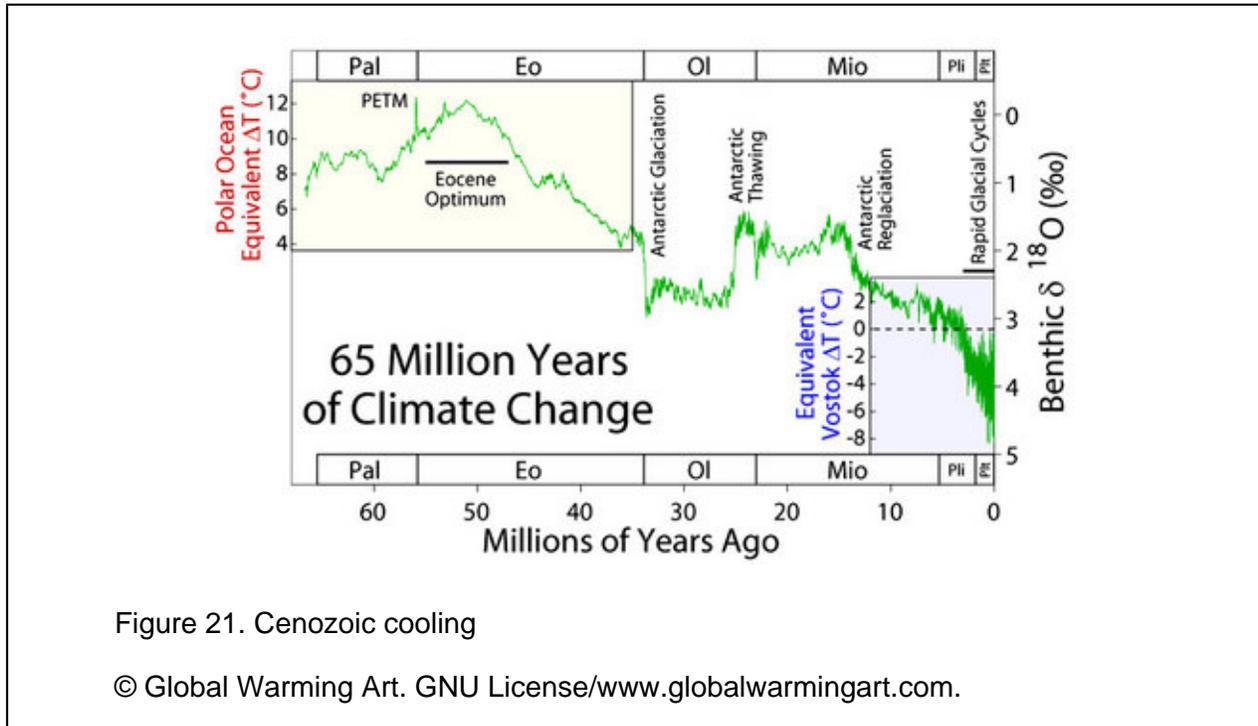
Another important ecological shift was the spread of angiosperms (flowering plants), which diversified and became the dominant form of land plants. Unlike earlier plants like ferns and conifers, angiosperms' seeds were enclosed within a structure (the flower) that protected developing embryos. Their flower petals and fruits, which grew from plants' fertilized ovaries, attracted animals, birds, and insects that helped plants spread by redistributing pollen and seeds. These advantages enabled angiosperms to spread into more diverse habitats than earlier types of plants.

Earth's climate continued to fluctuate during the Cenozoic, posing challenges for these new life forms. After an abrupt warming about 55 million years ago, the planet entered a pronounced cooling phase that continued up to the modern era. One major cause was the ongoing breakup of Gondwanaland, a supercontinent that contained most of the land masses in today's southern hemisphere, including Africa, South America, Australia, and India (Fig. 20).



Once these fragments started to separate about 160 million years ago, ocean currents formed around Antarctica. Water trapped in these currents circulated around the pole and became colder and colder. As a result, Antarctica cooled and developed a permanent ice cover, which in turn cooled global atmospheric and ocean temperatures. Climates became dryer, with grasslands and arid habitat spreading into many regions that previously had been forested.

Continued cooling through the Oligocene and Miocene eras, from about 35 million to 5 million years ago, culminated in our planet's most recent ice age: a series of glacial advances and retreats during the Pleistocene era, starting about 3.2 million years ago (Fig. 21). During the last glacial maximum, about 20,000 years ago, ice sheets covered most of Canada and extended into what is now New England and the upper Midwestern states.



Human evolution occurred roughly in parallel with the modern ice age and was markedly influenced by geologic and climate factors. Early hominids (members of the biological family of the great apes) radiated from earlier apes in Africa between 5 and 8 million years ago. Humans' closest ancestor, **Australopithecus**, was shorter than modern man and is thought to have spent much of its time living in trees. The human genus, **Homo**, which evolved about 2.5 million years ago, had a larger brain, used hand tools, and ate a diet heavier in meat than **Australopithecus**. In sum, **Homo** was better adapted for life on the ground in a cooler, drier climate where forests were contracting and grasslands were expanding.

By 1.9 million years ago, **Homo erectus** had migrated from Africa to China and Eurasia, perhaps driven partly by climate shifts and resulting changes to local environments. **Homo sapiens**, the modern human species, is believed to have evolved in Africa about 200,000 years ago. **Homo sapiens** gradually migrated outward from Africa, following dry land migration routes that were exposed as sea levels fell during glacial expansions. By about 40,000 years ago **Homo sapiens** had settled Europe, and around 10,000 years ago man reached North America. Today, archaeologists, anthropologists, and geneticists are working to develop more precise maps and histories of the human migration out of Africa, using mitochondrial DNA (maternally inherited genetic material) to assess when various areas were settled.



Early in their history, humans found ways to manipulate and affect their environment. Mass extinctions of large mammals, such as mammoths and saber-toothed cats, occurred in North and South America, Europe, and Australia roughly when humans arrived in these areas. Some researchers believe that over-hunting, alone or in combination with climate change, may have been the cause. After humans depleted wildlife, they went on to domesticate animals, clear forests, and develop agriculture, with steadily expanding impacts on their surroundings that are addressed in units 5 through 12 of this text.

## 10. Further Reading

University of California Museum of Paleontology, Web Geological Time Machine, <http://www.ucmp.berkeley.edu/help/timeform.html>. An era-by-era guide through geologic time using stratigraphic and fossil records.

Science Education Resource Center, Carleton College, "Microbial Life in Extreme Environments," <http://serc.carleton.edu/microbelife/extreme/index.html>. An online compendium of information about extreme environments and the microbes that live in them.

James Shreve, "Human Journey, Human Origins," National Geographic, March 2006, <http://www7.nationalgeographic.com/ngm/0603/feature2/index.html?fs=www3.nationalgeographic.com&fs=plasma.nationalgeographic.com>. An overview of what DNA evidence tells us about human migration out of Africa, with additional online resources.

## Footnotes

1. American Museum of Natural History, "Our Dynamic Planet: Rock Around the Clock," <http://www.amnh.org/education/resources/rfl/web/earthmag/peek/pages/clock.htm>.
2. Peter D. Ward and Donald Brownlee, **Rare Earth: Why Complex Life Is Uncommon in the Universe** (New York: Springer-Verlag, 2000).
3. U.S. Geological Survey, "Radiometric Time Scale," <http://pubs.usgs.gov/gip/geotime/radiometric.html>, and "The Age of the Earth," <http://geology.wr.usgs.gov/parks/gtime/ageofearth.html>.
4. Paul F. Hoffman and Daniel P. Schrag, "Snowball Earth," **Scientific American**, January 2000, pp. 68–75.
5. Andrew Knoll, **Life on a Young Planet: The First Three Billion Years of Evolution on Earth** (Princeton University Press, 2003), p. 23.
6. University of California Museum of Paleontology, "Cyanobacteria: Fossil Record," <http://www.ucmp.berkeley.edu/bacteria/cyanofr.html>.
7. Knoll, **Life on a Young Planet**, p. 42.

8. "Did the Snowball Earth Kick-Start Complex Life?", <http://www.snowballearth.org/kick-start.html>.
9. Stephen Jay Gould, **Wonderful Life: The Burgess Shale and the Nature of History** (New York: Norton, 1990).

## Glossary

**albedo** : The fraction of electromagnetic radiation reflected after striking a surface.

**archaea** : A major division of microorganisms. Like bacteria, Archaea are single-celled organisms lacking nuclei and are therefore prokaryotes, classified as belonging to kingdom Monera in the traditional five-kingdom taxonomy.

**bacteria** : Microscopic organisms whose single cells have neither a membrane-bounded nucleus nor other membrane-bounded organelles like mitochondria and chloroplasts.

**Cambrian explosion** : Between about 570 and 530 million years ago, when a burst of diversification occurred, with the eventual appearance of the lineages of almost all animals living today.

**cation** : An ion with a positive charge.

**cyanobacteria** : A phylum of Bacteria that obtain their energy through photosynthesis. They are often referred to as blue-green algae, although they are in fact prokaryotes, not algae.

**eukaryotes** : A single-celled or multicellular organism whose cells contain a distinct membrane-bound nucleus.

**extremophiles** : Microorganisms belonging to the domains Bacteria and Archaea that can live and thrive in environments with extreme conditions such as high or low temperatures and pH levels, high salt concentrations, and high pressure.

**geochemical cycling** : Flows of chemical substances between reservoirs in Earth's atmosphere, hydrosphere (water bodies), and lithosphere (the solid part of Earth's crust).

**heterotrophs** : An organism that requires organic substrates to get its carbon for growth and development.

**negative feedback** : When part of a system's output, inverted, feeds into the system's input; generally with the result that fluctuations are weakened.

**oxidation** : An array of reactions involving several different types of chemical conversions: (1) loss of electrons by a chemical, (2) combination of oxygen and another chemical, (3) removal of hydrogen atoms from organic compounds during biological metabolism, (4) burning of some material, (5) biological metabolism that results in the decomposition of organic material, (6) metabolic conversions in toxic materials in biological organism, (7) stabilization of organic pollutants during wastewater treatment, (8) conversion of plant matter to compost, (9) decomposition of pollutants or toxins that contaminate the environment.



**phylum** : The largest generally accepted groupings of animals and other living things with certain evolutionary traits.

**plate tectonics** : A concept stating that the crust of the Earth is composed of crustal plates moving on the molten material below.

**prokaryotes** : Organisms without a cell nucleus, or any other membrane-bound organelles. Most are unicellular, but some prokaryotes are multicellular. The prokaryotes are divided into two domains: the bacteria and the archaea.

**radiometric dating** : A technique used to date materials based on a knowledge of the decay rates of naturally occurring isotopes, and the current abundances. It is the principal source of information about the age of the Earth and a significant source of information about rates of evolutionary change.

**Snowball Earth** : Hypothesis that proposes that the Earth was entirely covered by ice in part of the Cryogenian period of the Proterozoic eon, and perhaps at other times in the history of Earth

**stratigraphic record** : Sequences of rock layers. Correlating the sequences of rock layers in different areas enables scientists to trace a particular geologic event to a particular period.

**subduction** : The process in which one plate is pushed downward beneath another plate into the underlying mantle when plates move towards each other.

# The Habitable Planet

## Unit 2 : Atmosphere



Utah sky.

### Overview

The atmosphere is a critical system that helps to regulate Earth's climate and distribute heat around the globe. In this unit, discover the fundamental processes that cause atmospheric circulation and create climate zones and weather patterns, and learn how carbon cycling between atmosphere, land, and ocean reservoirs helps to regulate Earth's climate.

### Sections:

1. Introduction
2. The Structure of the Atmosphere
3. Radiative Balance and the Natural Greenhouse Effect
4. Major Greenhouse Gases
5. Vertical Motion in the Atmosphere
6. Atmospheric Circulation Patterns
7. Climate, Weather, and Storms
8. The Global Carbon Cycle
9. Feedbacks in the Atmosphere
10. Further Reading



## 1. Introduction

Earth's atmosphere is a critical system for life on our planet. Together with the oceans, the atmosphere shapes Earth's climate and weather patterns and makes some regions more habitable than others. But Earth's climate is not static. How variable is it, and how quickly does it change? What physical factors control climate, and how do they interact with one another?

To see how and why climate fluctuates, we need to learn about the basic characteristics of the atmosphere and some physical concepts that help us understand weather and climate. This unit describes the structure of the atmosphere and examines some of its key functions, including screening out harmful solar radiation, warming Earth through the natural greenhouse effect, and cycling carbon. It then summarizes how physical processes shape the distributions of pressures and temperatures on Earth to create climate zones, weather patterns, and storms, creating conditions suitable for life around the planet.

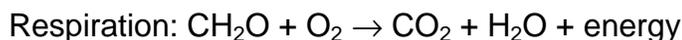
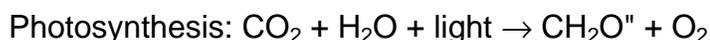
The atmosphere is a complex system in which physical and chemical reactions are constantly taking place. Many atmospheric processes take place in a state of **dynamic balance**—for example, there is an average balance between the heat input to, and output from, the atmosphere. This condition is akin to a leaky bucket sitting under a faucet: when the tap is turned on and water flows into the bucket, the water level will rise toward a steady state where inflow from the tap equals outflow through the leaks. Once this condition is attained, the water level will remain steady even though water is constantly flowing in and out of the bucket.

Similarly, Earth's climate system maintains a dynamic balance between solar energy entering and radiant energy leaving the atmosphere. Levels of oxygen in the atmosphere are regulated by a dynamic balance in the natural carbon cycle between processes that emit oxygen through photosynthesis and others that consume oxygen, such as respiration. The strength of atmospheric circulation is also controlled by a dynamic balance. Some parts of the planet receive more energy from the sun than others, and this uneven heating creates wind motions that act to move heat from warm to cold regions. (The process by which differential heating triggers atmospheric motion is discussed below in Section 5, "Vertical Motion in the Atmosphere.")

Today human actions are altering key dynamic balances in the atmosphere. Most importantly, humans are increasing greenhouse gas levels in the troposphere, which raises Earth's surface temperature by increasing the amount of heat radiated from the atmosphere back to the ground. The broad impacts of global warming are discussed in Unit 12, "Earth's Changing Climate," but it should be noted here that climate change will alter factors that are key determinants of environmental conditions upon which ecosystems depend. As the following sections will show, changing global surface temperatures and precipitation patterns will have major impacts on Earth's climate and weather.

## 2. The Structure of the Atmosphere

The atmosphere is composed of nitrogen, oxygen, argon, water vapor, and a number of trace gases (Table 1). This composition has remained relatively constant throughout much of Earth's history. Chemical reactions maintain the ratios of major constituents of the atmosphere to each other. For example, oxygen is released into the atmosphere by photosynthesis and consumed by respiration. The concentration of oxygen in the atmosphere is maintained by a balance between these two processes:



"CH<sub>2</sub>O" denotes the average composition of organic matter.

Many gases play critical roles in the atmosphere even though they are present in relatively low concentrations. Important substances that will receive further attention throughout this text include tropospheric ozone, which is addressed in Unit 11, "Atmospheric Pollution," and greenhouse gases (CO<sub>2</sub>, methane, N<sub>2</sub>O, and chlorofluorocarbons), which are discussed in Section 4 of this unit and in Unit 12, "Earth's Changing Climate."

**Table 1.** Atmospheric gas composition (average). Concentrations of gases shown in color are rising due to human activities.

Gas	Mole fraction
Nitrogen (N)	0.78
Oxygen (O)	0.21
Water (H <sub>2</sub> O)	0.04 to $< 5 \times 10^{-3}$ ; $4 \times 10^{-6}$ — strat
Argon (Ar)	0.0093
Carbon Dioxide (CO <sub>2</sub> )	$370 \times 10^{-6}$ (date: 2000)
Neon (Ne)	$18.2 \times 10^{-6}$
Ozone (O <sub>3</sub> )	$0.02 \times 10^{-6}$ to $10 \times 10^{-6}$
Helium (He)	$5.2 \times 10^{-6}$
Methane (CH <sub>4</sub> )	$1.7 \times 10^{-6}$
Krypton (Kr)	$1.1 \times 10^{-6}$
Hydrogen (H)	$0.55 \times 10^{-6}$

Gas	Mole fraction
Nitrous Oxide (N <sub>2</sub> O)	$0.32 \times 10^{-6}$
Carbon Monoxide (CO)	$0.03 \times 10^{-6}$ to $0.3 \times 10^{-6}$
Chlorofluorocarbons	$3.0 \times 10^{-9}$
Carbonyl Sulfide (COS)	$0.1 \times 10^{-9}$

Earth's atmosphere extends more than 560 kilometers (348 miles) above the planet's surface and is divided into four layers, each of which has distinct thermal, chemical, and physical properties (Fig. 1).

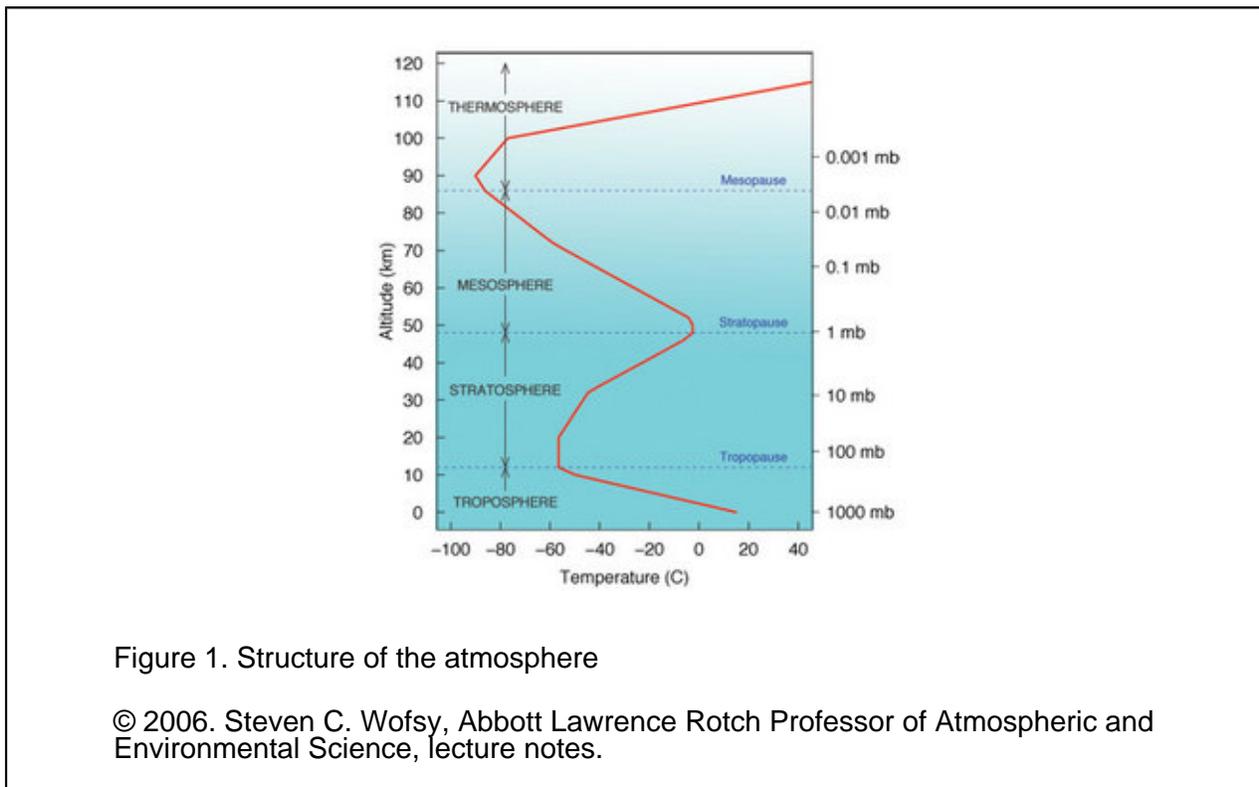


Figure 1. Structure of the atmosphere

© 2006. Steven C. Wofsy, Abbott Lawrence Rotch Professor of Atmospheric and Environmental Science, lecture notes.

Almost all weather occurs in the troposphere, the lowest layer of the atmosphere, which extends from the surface up to 8 to 16 kilometers above Earth's surface (lowest toward the poles, highest in the tropics). Earth's surface captures solar radiation and warms the troposphere from below, creating rising air currents that generate vertical mixing patterns and weather systems, as detailed further below. Temperatures decrease by about 6.5°C with each kilometer of altitude. At the top

of the troposphere is the tropopause, a layer of cold air (about  $-60^{\circ}\text{C}$ ), which forms the top of the troposphere and creates a "cold trap" that causes atmospheric water vapor to condense.

The next atmospheric layer, the stratosphere, extends upward from the tropopause to 50 kilometers. In the stratosphere temperatures increase with altitude because of absorption of sunlight by stratospheric ozone. (About 90 percent of the ozone in the atmosphere is found in the stratosphere.) The stratosphere contains only a small amount of water vapor (only about one percent of total atmospheric water vapor) due to the "cold trap" and the tropopause, and vertical air motion in this layer is very slow. The stratopause, where temperatures peak at about  $-3^{\circ}\text{C}$ , marks the top of the stratosphere.

In the third atmospheric layer, the mesosphere, temperatures once again fall with increasing altitude, to a low of about  $-93^{\circ}\text{C}$  at an altitude of 85 kilometers. Above this level, in the thermosphere, temperatures again warm with altitude, rising higher than  $1700^{\circ}\text{C}$ .

The atmosphere exerts pressure at the surface equal to the weight of the overlying air. Figure 1 also shows that atmospheric pressure declines exponentially with altitude—a fact familiar to everyone who has felt pressure changes in their ears while flying in an airplane or climbed a mountain and struggled to breathe at high levels. At sea level, average atmospheric pressure is 1013 millibars, corresponding to a mass of 10,000 kg (10 tons) per square meter or a weight of 100,000 Newtons per square meter (14.7 pounds per square inch) for a column of air from the surface to the top of the atmosphere.

Pressure falls with increasing altitude because the weight of the overlying air decreases. It falls exponentially because air is compressible, so most of the mass of the atmosphere is compressed into its lowest layers. About half of the mass of the atmosphere lies in the lowest 5.5 kilometers (the summit of Mt. Everest at 8850 m extends above about roughly two-thirds of the atmosphere), and 99 percent is within the lowest 30 kilometers.

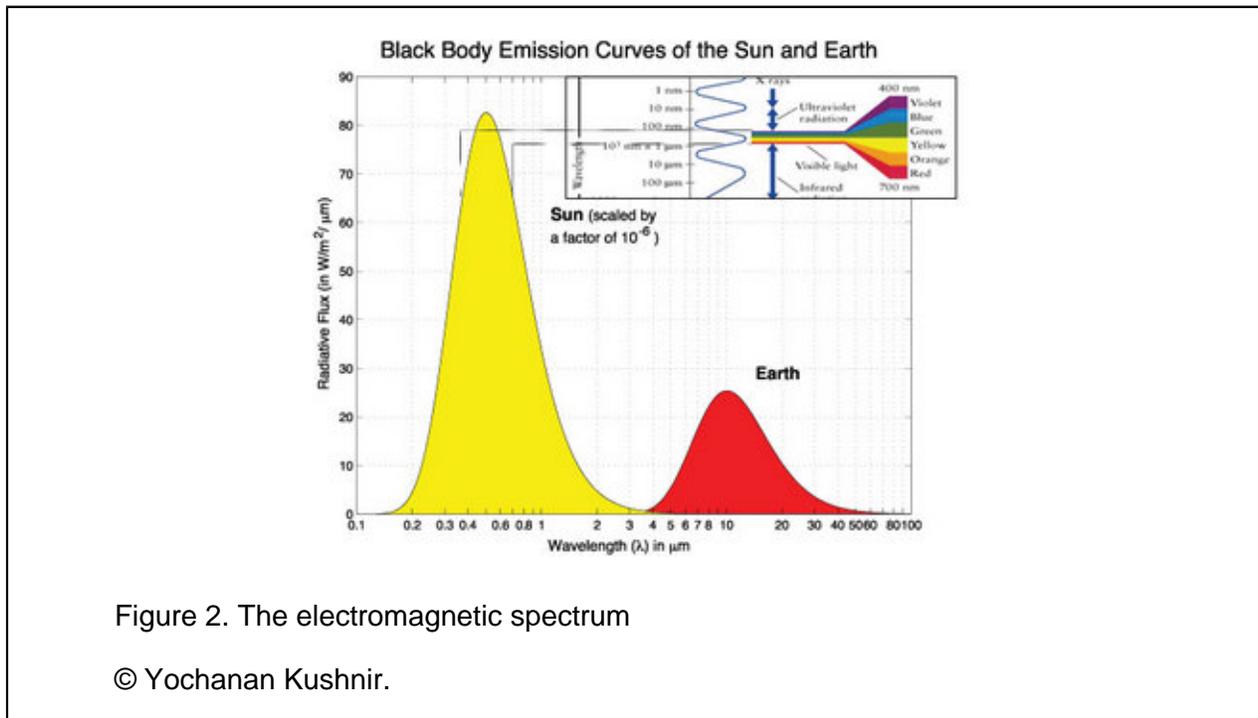
### 3. Radiative Balance and the Natural Greenhouse Effect

Earth's surface temperature has been remarkably constant over geologic time. Even the dramatic cooling that occurred during the most recent ice age represented a change of only  $3^{\circ}\text{C}$  in the global average surface temperature, occurring over thousands of years. Seasonal changes in temperature, although large in a particular place, correspond to very tiny changes in global mean temperature. Why have temperatures held so steady?

Earth exchanges energy with its environment primarily through transfers of electromagnetic radiation. At any time our planet is simultaneously absorbing energy from the sun and radiating energy back into space. The temperature remains stable over long periods of time because the planet radiates energy back to space at a rate that closely balances the energy input it receives from the sun (i.e., the planet is close to being in radiative energy balance).

Earth receives energy from the sun in the form of solar radiation—radiation with varying wavelengths along the electromagnetic spectrum. The sun emits strongly in the visible light range, but it also

produces ultraviolet and infrared radiation. The earth radiates heat back to space mostly at much longer wavelengths than solar radiation (Fig. 2).



When visible solar radiation reaches Earth, it may be absorbed by clouds, the atmosphere, or the planet's surface. Once absorbed it is transformed into heat energy, which raises Earth's surface temperature. However, not all solar radiation intercepted by the Earth is absorbed. The fraction of incoming solar radiation that is reflected back to space constitutes Earth's **albedo**, as shown below in Figure 3.

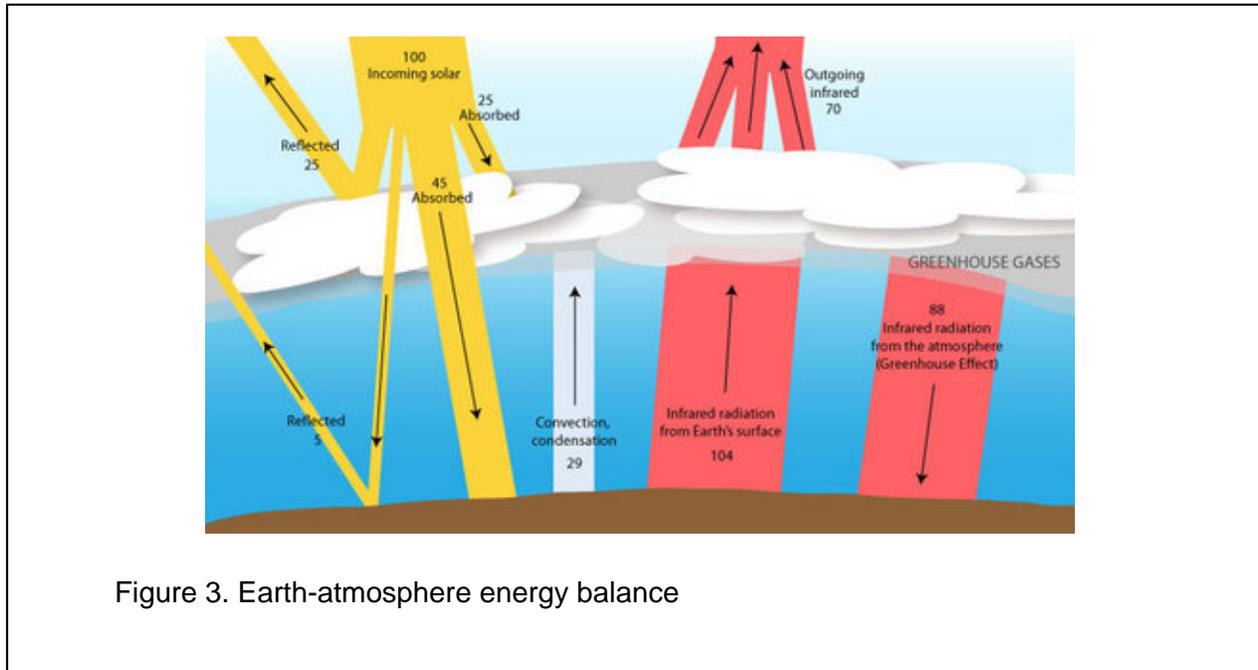


Figure 3. Earth-atmosphere energy balance

Any form of matter emits radiation if its temperature is above absolute zero (zero degrees Kelvin). Incoming solar radiation warms Earth, and the planet emits infrared radiation back to outer space. Note that Earth emits radiation at a longer wavelength—i.e., a lower energy level—than the sun (Fig. 2). This difference occurs because the total energy flux from an object varies with the fourth power of the object's absolute temperature, and the sun is much hotter than the Earth.

Some outgoing infrared energy emitted from the Earth is trapped in the atmosphere and prevented from escaping to space, through a natural process called the "greenhouse effect." The most abundant gases in the atmosphere—nitrogen, oxygen, and argon—neither absorb nor emit terrestrial or solar radiation. But clouds, water vapor, and some relatively rare [greenhouse gases](#) (GHGs) such as carbon dioxide, methane, and nitrous oxide in the atmosphere can absorb long-wave radiation (terrestrial radiation, see Figure 2). Molecules that can absorb radiation of a particular wavelength can also emit that radiation, so GHGs in the atmosphere therefore will radiate energy both to space and back towards Earth. This back-radiation warms the planet's surface.

In Figure 3, 100 units of solar radiation are intercepted by the Earth each second. On average 30 units are reflected, 5 by the surface and 25 by clouds. Energy balance is achieved by Earth's emission of 70 units of infrared ("terrestrial") radiation to space. The earth's surface is warmed directly by only 45 units of solar energy, with almost twice as much energy (88 units) received from thermal radiation due to greenhouse gases and clouds in the atmosphere. Energy is removed from the



surface by radiation of infrared energy back to the atmosphere and space (88 units) and by other processes such as evaporation of water and direct heat transfer (29 units).

Note that the amount of heat received by the surface is actually much larger (3x) than the amount the surface receives in solar radiation, due to the natural greenhouse effect. The result is a surface temperature on average around 15°C (60°F), as compared to temperatures colder than -18°C (0°F) if there were no greenhouse effect.

#### 4. Major Greenhouse Gases

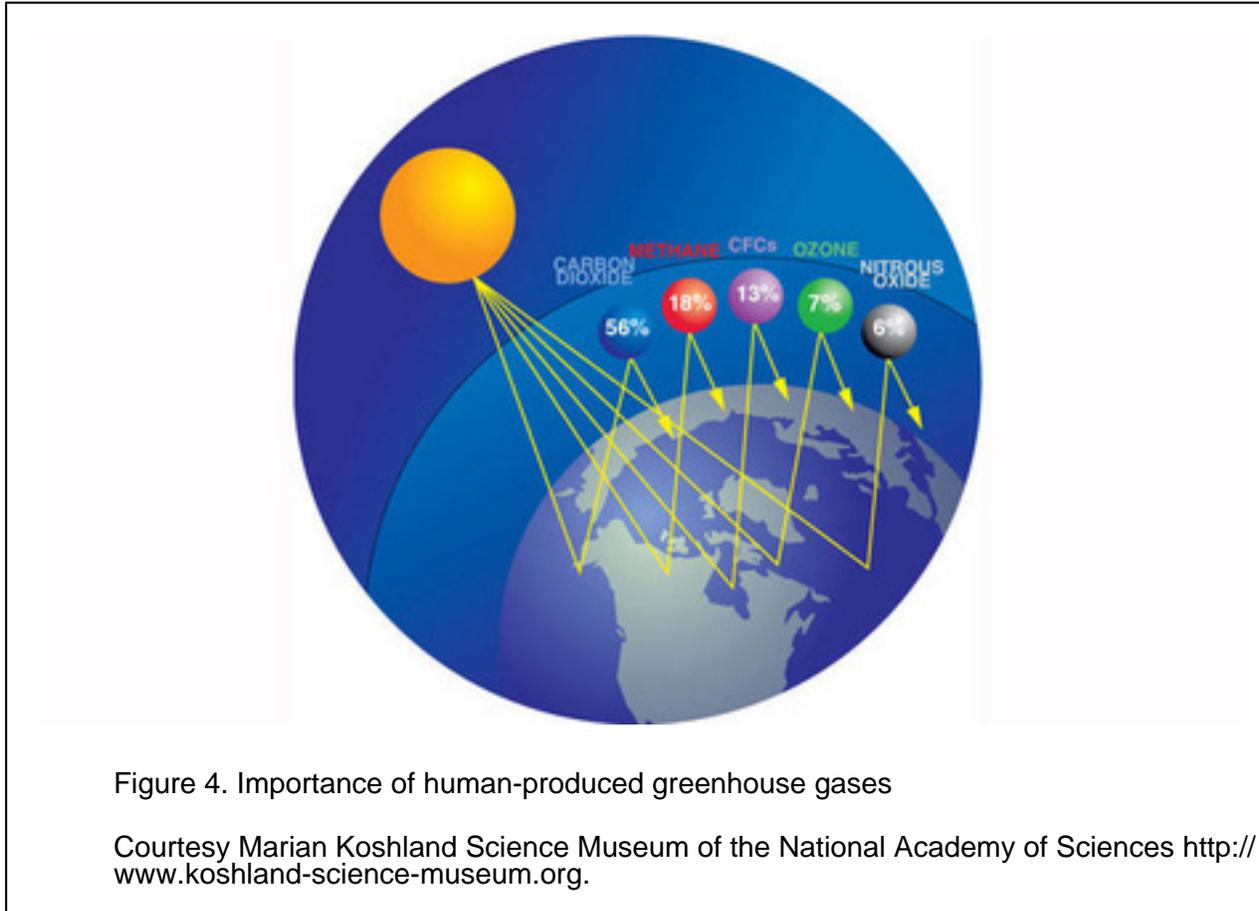
Many GHGs, including water vapor (the most important), ozone, carbon dioxide, methane, and nitrous oxide, are naturally present in the atmosphere. Other GHGs are synthetic chemicals that are emitted only as a result of human activity. Anthropogenic (human) activities are significantly increasing atmospheric concentrations of many GHGs.

- **Carbon dioxide (CO<sub>2</sub>)**, the most significant GHG directly affected by anthropogenic activity, is the product of the oxidation of carbon in organic matter, either through combustion of carbon-based fuels or the decay of biomass. Natural CO sources include volcanic eruptions, **respiration** of organic matter in natural ecosystems, **natural fires**, and exchange of dissolved CO with the oceans. The main anthropogenic sources are (a) fossil fuel combustion and (b) **deforestation** and land use changes (such as converting agricultural land or forests to urban development), which release stored organic matter and reduce the ability of natural ecosystems to store carbon.
- **Methane (CH<sub>4</sub>)** is produced by **anaerobic** decay of organic material in landfills, wetlands, and rice fields; enteric fermentation in the digestive tracts of **ruminant** animals such as cattle, goats, and sheep; manure management; wastewater treatment; fossil fuel combustion; and leaks from natural gas transportation and distribution systems and abandoned coal mines.
- **Nitrous oxide (N<sub>2</sub>O)** is produced by fertilizer use, animal waste management, fossil fuel combustion, and industrial activities.

# The Habitable Planet

- **Hydrofluorocarbons (HFCs) and perfluorocarbons (PFCs)** are synthetic chemicals that are used in a variety of industrial production processes such as semiconductor manufacturing. PFCs are also produced as a by-product of aluminum smelting. Both groups of chemicals are finding increasing use as substitutes for ozone-depleting chlorofluorocarbons (CFCs), which are being phased out under the 1987 [Montreal Protocol on Substances that Deplete the Ozone Layer](#). HFCs and PFCs are replacing CFCs in applications such as refrigeration and foam-blowing for insulation.

When atmospheric GHG concentrations increase, Earth temporarily traps infrared radiation more efficiently, so the natural radiative balance is disturbed until its surface temperature rises to restore equilibrium between incoming and outgoing radiation. It takes many decades for the full effect of greenhouse gases to be realized in higher surface temperatures, because the oceans have a huge capacity to store heat. They must be gradually warmed by excess infrared radiation from the atmosphere. Figure 4 illustrates the relative contributions from man-made emissions of various GHGs to climate change.



As we will see in section 8, "The Global Carbon Cycle," CO<sub>2</sub> emitted from combustion of fossil fuel cycles between the atmosphere and land and ocean "sinks" (carbon storage reservoirs), which are absorbing a large fraction of anthropogenic carbon emissions. Ultimately, though, there are limits to the amount of carbon that these sinks can absorb. These sinks are more likely to delay than to prevent human actions from altering Earth's radiative balance.

Higher surface temperatures on Earth will have profound impacts on our planet's weather and climate. Before we consider those impacts, however, we need to understand how variables such as pressure, temperature, and moisture combine to create air currents, drive normal atmospheric circulation patterns, and create the overall climate.

## 5. Vertical Motion in the Atmosphere

To see how atmospheric motion generates climate and weather systems, it is useful to start by picturing an air parcel—a block of air with uniform temperature and pressure throughout. The air parcel can change over time by rising, falling, or emitting or absorbing heat. If the pressure of the surrounding environment changes, the pressure of the parcel changes, but it does not exchange heat or chemicals with the surroundings and therefore may behave differently from its surroundings.

Many weather patterns start with rising air. According to the Perfect Gas Law, if an air parcel stays at a constant pressure, increasing its temperature will cause it to expand and become less dense. Contrary to popular belief, however, air does not rise simply because it is warm. Rather, air that becomes warmer than its surroundings—for example, if it is warmed by heat radiating from the Earth's surface—becomes **buoyant** and floats on top of cooler, denser air in the same way that oil floats on top of water. This buoyant motion is called **convection**.

Water is also a key factor in weather and climate. As shown in Table 1, two to three percent of the atmosphere typically consists of water vapor. Weather forecasts often report the **relative humidity**, which compares the amount of water vapor in the air to the maximum amount that air can hold at that temperature. When relative humidity reaches 100 percent, the air has reached saturation pressure and cannot absorb any more water vapor. (High humidity levels make people feel uncomfortable because little or none of their sweat can evaporate and draw heat from their skin.) As air warms, the amount of water vapor that it can hold rises exponentially. Consequently, atmospheric water vapor concentrations are highest in warm regions and decrease toward the poles (Fig. 5).

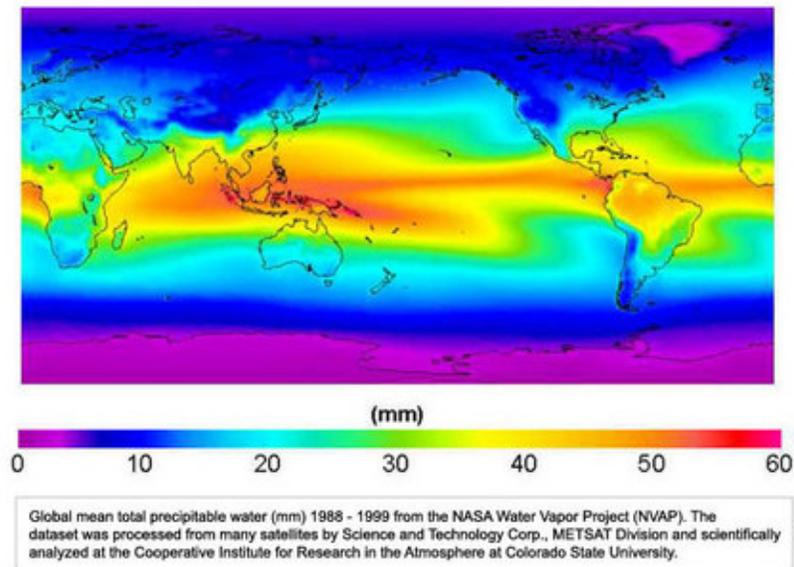


Figure 5. Mean distribution of atmospheric water vapor above Earth's surface, 1988–1999

Courtesy Cooperative Institute for Research in the Atmosphere, Colorado State University.

Atmospheric water vapor contributes to weather patterns in several ways. First, adding water vapor to the air reduces its density, so adding moisture to dry air may make it become buoyant and rise. Secondly, moist air carries **latent energy**, the potential for condensation of water vapor to heat the air. Liquid water absorbs energy when it evaporates, so when this water vapor condenses, energy is released and warms the surrounding environment. (As we will see in the next section, thunderstorms and hurricanes draw energy from the release of latent heat.) The **dew point**, another key weather variable, denotes the temperature to which air would have to cool to reach 100 percent relative humidity. When an air parcel cools to its dew point, water vapor begins condensing and forming cloud droplets or ice crystals, which may ultimately grow large enough to fall as rain or snow.

When a rising air parcel expands it pushes away the surrounding atmosphere, and in doing this work it expends energy. If heat is not added or removed as this hypothetical parcel moves—a scenario called an **adiabatic process**—the only source of energy is the motion of molecules in the air parcel, and therefore the parcel will cool as it rises. (Recall from Figure 1 that that in the troposphere, temperature falls 6.5°C on average with each kilometer of altitude. The actual decrease under real-world conditions, which may vary from region to region, is called the **atmospheric lapse rate**.)

A dry air parcel (one whose relative humidity is less than 100 percent) cools by  $9.8^{\circ}\text{C}$  for each thousand meters that it rises, a constant decrease called the **dry adiabatic lapse rate**. However, if the parcel cools enough that its relative humidity reaches 100 percent, water starts to condense and form cloud droplets. This condensation process releases latent heat into the parcel, so the parcel cools at a lower rate as it moves upward, called the **moist adiabatic lapse rate**.

Atmospheric conditions can be stable or unstable, depending on how quickly the temperature of the environment declines with altitude. An unstable atmosphere is more likely to produce clouds and storms than a stable atmosphere. If atmospheric temperature decreases with altitude faster than the dry adiabatic lapse rate (i.e., by more than  $9.8^{\circ}\text{C}$  per kilometer), the atmosphere is unstable: rising air masses will be warmer and less dense than the surrounding air, so they experience buoyancy and will continue to rise and form clouds that can generate storms. If temperature falls more gradually with altitude than the dry adiabatic lapse rate but more steeply than the wet adiabatic lapse rate, the atmosphere is conditionally unstable. In this case, air masses may rise and form clouds if they contain enough water vapor to warm them as they expand (Fig. 6), but they have to get a fairly strong push upwards to start the condensation process (up to 4000 meters in the figure). If temperature falls with altitude more slowly than the moist adiabatic lapse rate, the atmosphere is stable: rising air masses will become cooler and denser than the surrounding atmosphere and sink back down to where they started.

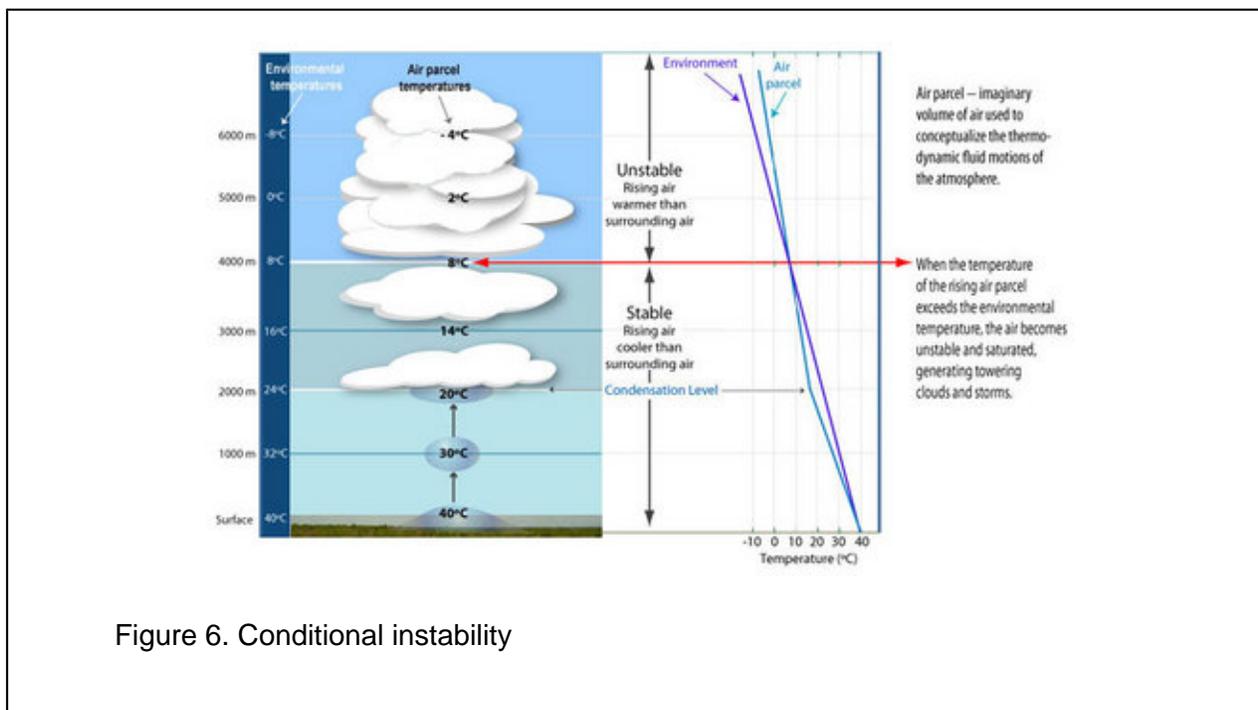


Figure 6. Conditional instability

Convection is not the only process that lifts air from lower to higher altitudes. When winds run into mountains and are forced upward the air cools, often forming clouds over windward slopes and the crests of hills. **Convergence** occurs when air masses run together, pushing air upward, as happens often in the tropics and in warm summer conditions in midlatitudes, generating thunderstorms. And when warm and cold air fronts collide, the denser cold air slides underneath the warm air layer and lifts it. In each case, if warm air is lifted high enough to reach its dew point, clouds will form. If lifting forces are strong, the system will produce tall, towering clouds that can generate intense rain or snow storms.

As discussed in section 9, "Feedbacks in the Atmosphere," clouds are important factors in Earth's energy balance. Their net impact is hard to measure and model because different types of clouds have different impacts on climate. Low-altitude clouds emit and absorb infrared radiation much as the ground does, so they are roughly the same temperature as Earth's surface and thus do not increase atmospheric temperatures. However, they have a cooling effect because they reflect a portion of incoming solar radiation back into space, increasing Earth's albedo and reducing the total input of solar energy to the planet's surface.

In contrast, high-altitude clouds tend to be thinner, so they do not reflect significant levels of incoming solar radiation. However, since they reside in a higher, cooler area of the atmosphere, they efficiently absorb outgoing thermal radiation and warm the atmosphere, and they radiate heat back to the surface from a part of the atmosphere that would otherwise not contribute to the greenhouse effect.

## 6. Atmospheric Circulation Patterns

Atmospheric circulation is set up when mass moves in the atmosphere. This motion may be vertical, as when warm air rises and becomes buoyant. It can also be horizontal: wind is created by air moving from high pressure areas, where air is densely compressed, to low-pressure areas, where air is less dense, although horizontal winds follow curved trajectories due to the rotation of the earth (see below). Atmospheric forces cause the air to move, modifying the difference in pressure. On a weather map, pressure differences are demarcated by parallel lines called **isobars** that show changes in pressure, usually in increments of 2 to 4 millibars.

Sea breezes show how vertical and horizontal movements combine to modify temperature and pressure at a local level. During the day coastal land regions heat up more than the sea because land warms more quickly than water. Air over the land is thus warmed and rises, increasing pressure in the atmosphere above the surface, where it starts to cool and form clouds. It then flows at altitude from the area of high pressure over land to lower pressure over the sea. Because there is then less mass over the land and more over the sea, pressure at the surface is higher at sea, so air flows in from the sea to the land. At night, when land cools more quickly than the ocean, the cycle is reversed (Fig. 7).

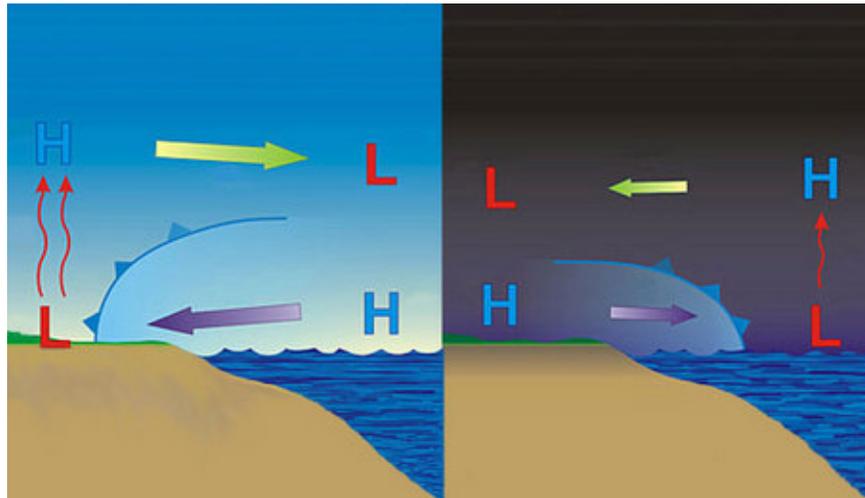


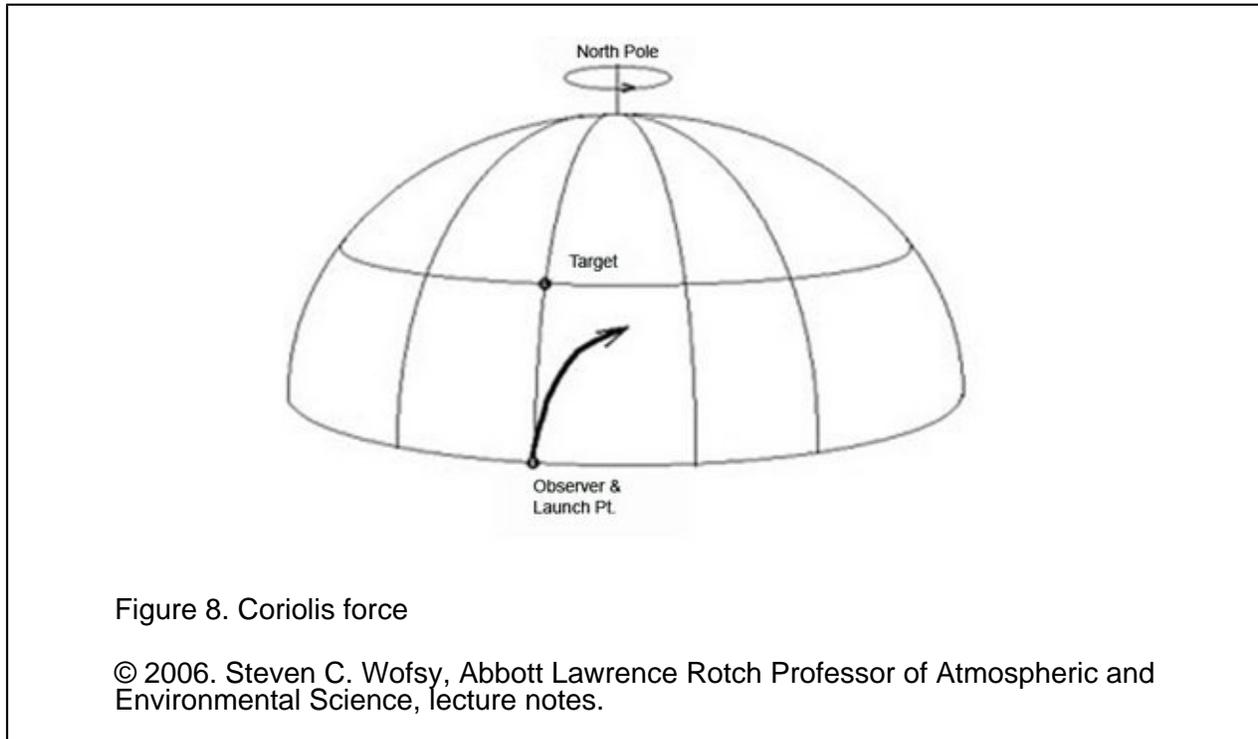
Figure 7. Sea breeze

Adapted from graphic by National Oceanic Atmospheric Administration, Jet Stream.

The sea breezes in this example flow directly between two points, but many larger weather systems follow less-direct courses. Their paths are not random, however. Winds that move over very long distances appear to curve because of the **Coriolis force**, an apparent force caused by Earth's rotation. This phenomenon occurs because all points on the planet's surface rotate once around Earth's axis every 24 hours, but different points move at different speeds: air at a point on the equator rotates at 1,700 kilometers per hour, compared to 850 kilometers per hour for a point that lies at 60 degrees latitude, closer to Earth's spin axis.

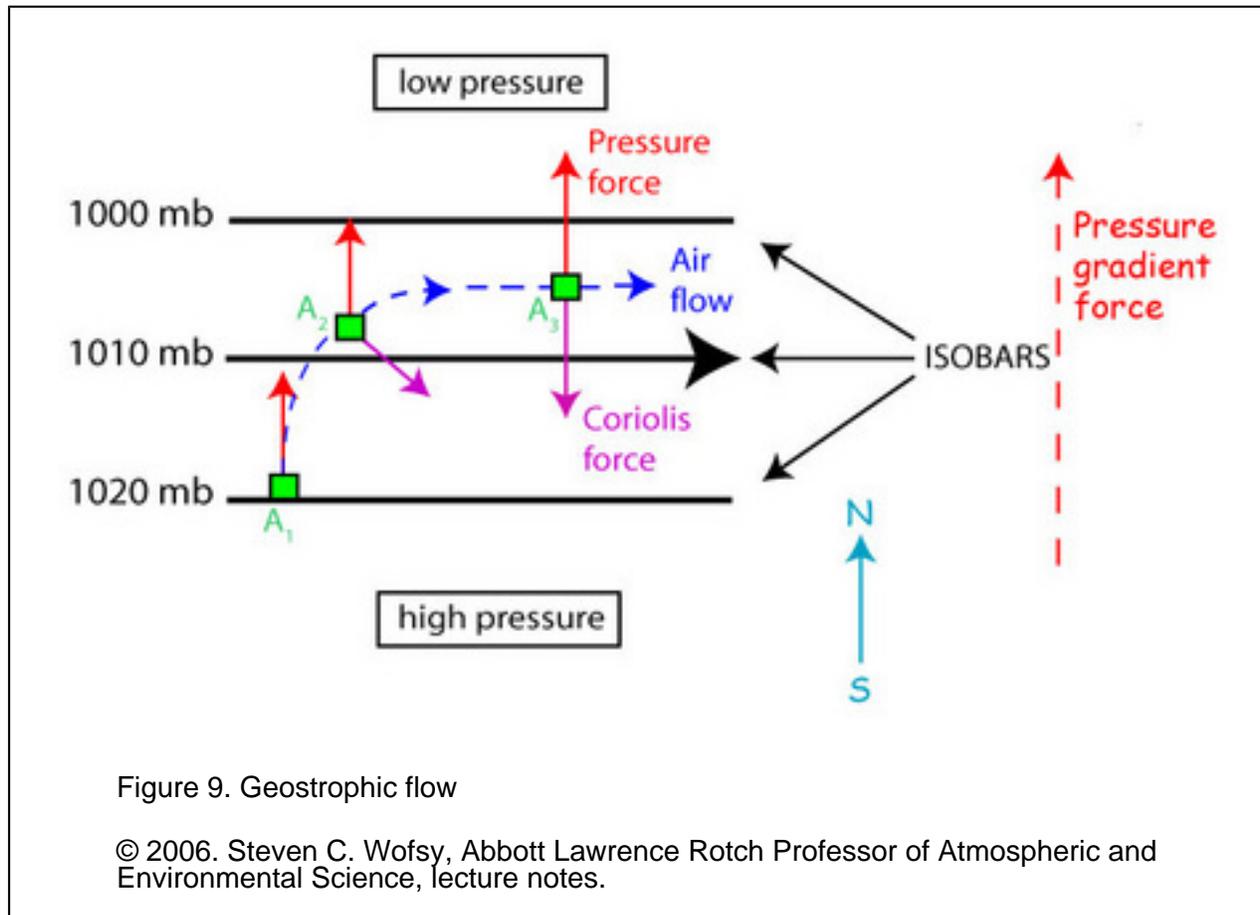
Because Earth spins, objects on its surface have **angular momentum**, or energy of motion, which defines how a rotating object moves around a reference point. An object's angular momentum is the product of its mass, its velocity, and its distance from the reference point (its radius). Angular momentum is conserved as an object moves on the Earth, so if its radius of spin decreases (as it moves from low latitude to high latitude), its velocity must increase. This relationship is what makes figure skaters rotate faster when they pull their arms in close to their bodies during spins.

The same process affects a parcel of air moving north from the Equator toward the pole: its radius of spin around Earth decreases as it moves closer to Earth's axis of rotation, so its rate of spin increases. The parcel's angular velocity is greater than the angular velocity of Earth's surface at the higher latitude, so it deflects to the right of its original trajectory relative to the planet's surface (Fig. 8). In the Southern hemisphere, the parcel would appear to deflect to the left.



This effect was discovered by French scientist Gustave-Gaspard Coriolis, who sought to explain why shots fired from long-range cannons were falling wide to the right of their targets. The Coriolis force only affects masses that travel over long distances, so it is not apparent in local weather patterns such as sea breezes. Nor, contrary to an oft-repeated misbelief, does it make water draining from a sink or toilet rotate in one direction in the Northern Hemisphere and the other direction in the Southern Hemisphere. But the Coriolis force makes winds appear to blow almost parallel to isobars, rather than directly across them from high to low pressure.

The Coriolis force makes the winds in low-pressure weather systems such as hurricanes rotate (counterclockwise in the Northern Hemisphere and clockwise in the Southern Hemisphere), curving into spirals. Air initially starts to move through the atmosphere under the influence of pressure gradients that push it from high pressure to low pressure areas. As it travels, the Coriolis force starts to bend its course. The motion tends toward a state called [geostrophic flow](#), where the pressure gradient force and the Coriolis force exactly balance each other. At this point the air parcel is no longer moving from a high-pressure to a low-pressure zone. Instead, it follows a course parallel to the isobars. In Figure 9, the air parcel is in geostrophic flow at point  $A_3$ .



When a low-pressure region develops in the Northern Hemisphere, pressure forces direct air from the outside toward the low. Air that moves in as a response to this force is deflected to the right and rotates counter-clockwise around the system. In contrast, a region of high pressure produces a pressure force directed away from the high. Air starting to move in response to this force is deflected to the right (in the Northern Hemisphere), producing a clockwise circulation pattern around a region of high pressure (Fig. 10).

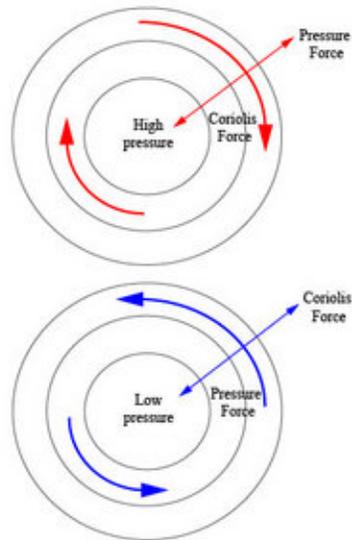


Figure 10. Circulation of air around regions of high and low pressure in the Northern Hemisphere

© 2006. Steven C. Wofsy, Abbott Lawrence Rotch Professor of Atmospheric and Environmental Science, lecture notes.

This pattern is modified at altitudes below 1 kilometer as friction with objects on the ground slows winds down. As wind speed declines, so does the Coriolis force, but pressure gradient forces stay constant. As a result, winds near the ground are deflected toward low pressure areas. Air parcels will spiral into low pressure areas near the surface, then rise once they reach the center. As the air rises, it cools, producing condensation, clouds, and rain. In contrast, air parcels will spiral away from high pressure areas near the surface toward low pressure areas. To maintain barometric balance, air will descend from above. In the process, the descending air will warm and its relative humidity will decrease, usually producing sunny weather (Fig. 11).

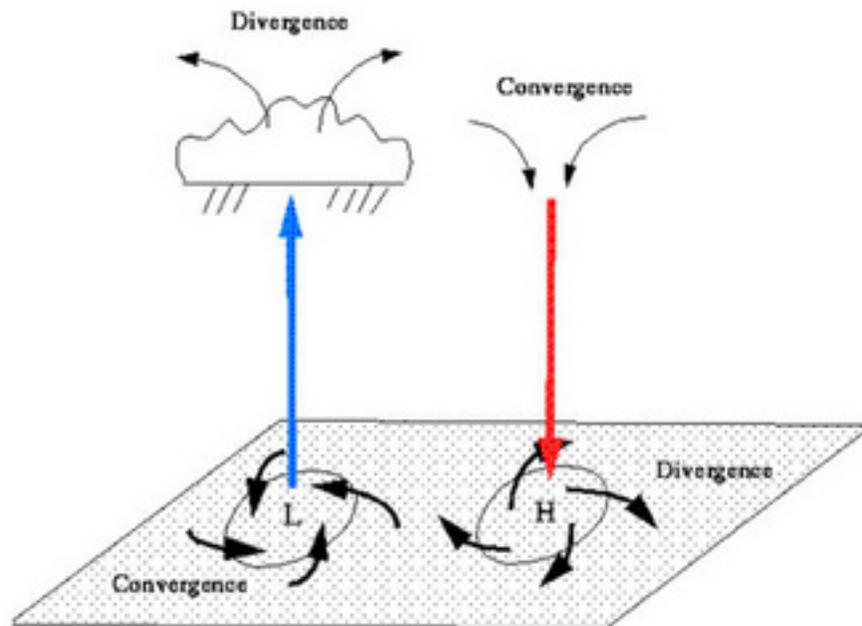


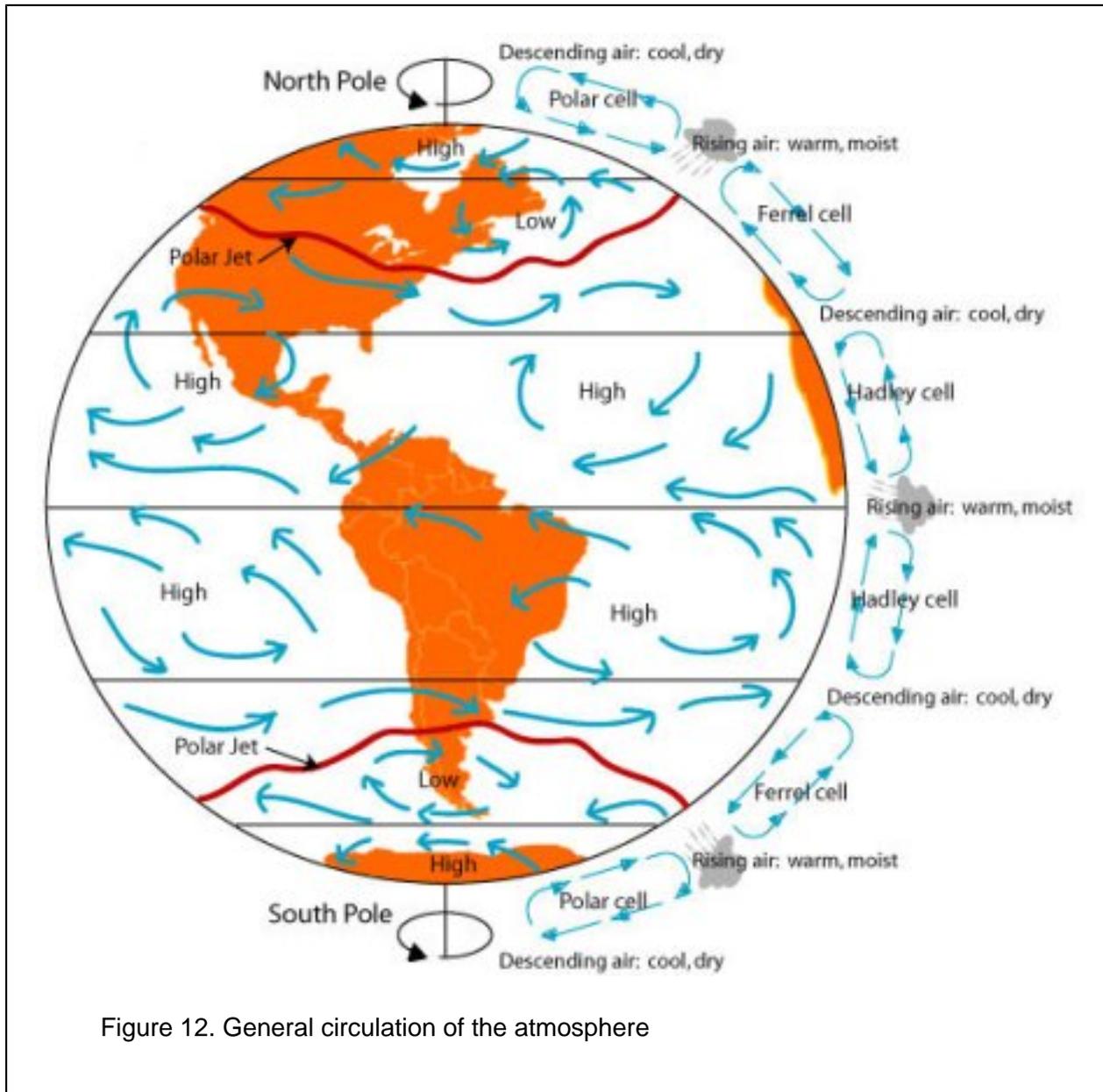
Figure 11. Winds around highs and lows

© 2006. Steven C. Wofsy, Abbott Lawrence Rotch Professor of Atmospheric and Environmental Science, lecture notes.

The first attempt to show how weather patterns combined to produce a general circulation of the atmosphere was offered in 1735 by English meteorologist George Hadley. Hadley pictured global-scale circulation as a large-scale version of the local system pictured above in Figure 7: a vast sea breeze with warm air rising over the equator and sinking over the poles. Hadley wanted to explain why sailors encountered westerly winds at midlatitudes and easterly "trade winds" near the equator. He deduced that this trend was caused by the Earth's rotation.

Hadley's model was accurate in many respects. Due to differential heating of the Earth with more warmth near the equator and cooling by radiation to space, buoyancy develops at low latitudes and mass is moved upward and poleward in the atmosphere, creating pressure gradients. The atmosphere tries to set up a simple circulation, upwelling near the equator and descending in polar regions, but in reality the [Hadley circulation](#) terminates at a latitude of about 30°. At this point, air sinks to the ground and flows back to the tropics, deflected by the Coriolis force, which produces easterly winds near the surface at low latitudes ("trade winds") and westerly winds at high latitudes. Farther north and south, this pattern repeats in two more sets of circulation zones, or "cells," between

the tropics and the poles (Fig. 12). The strength of the atmospheric circulation is controlled by a dynamic balance between motions caused by differential heating and friction that slows down the winds.



## 7. Climate, Weather, and Storms

Climate and weather are tightly connected, but it is important to distinguish between them. Climate refers to long-term weather trends and the range of variations that can be detected over decades in a specific region. Specific weather trends like annual snowfall may vary widely from one year to another (as a famous saying puts it, "Climate is what you expect, weather is what you get"), but forecasters can predict that these trends will fall within certain ranges over time based on long-term climate records. For example, southern Arizona has a hot, dry climate but its weather patterns include heavy rainstorms in July and August.

The global circulation patterns described in the preceding sections create predictable regional climate zones (Fig. 13). Low-pressure belts at the equator and at 50° to 60° north and south latitudes generate abundant precipitation. At latitudes around 30° north and south, dry descending air from high pressure belts—like the descending flow in a sea breeze circulation—produces arid zones that include Earth's major deserts.

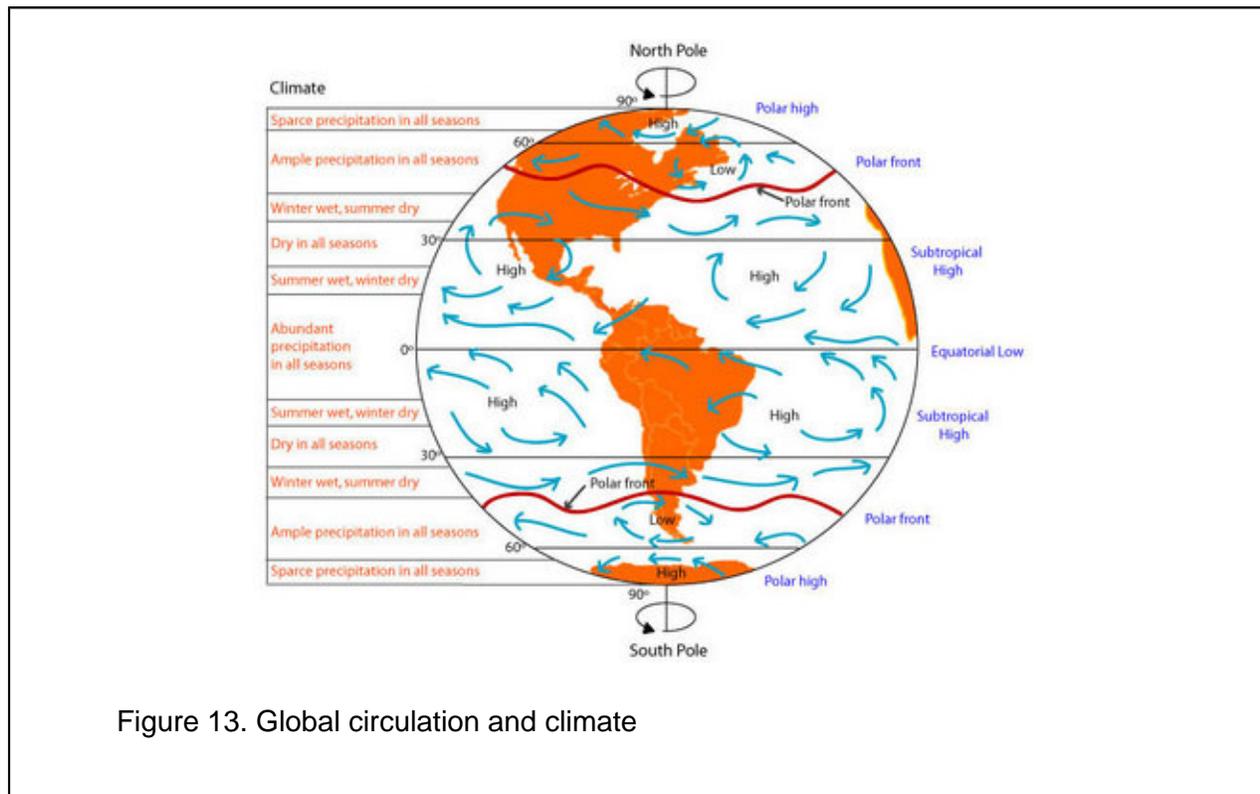


Figure 13. Global circulation and climate

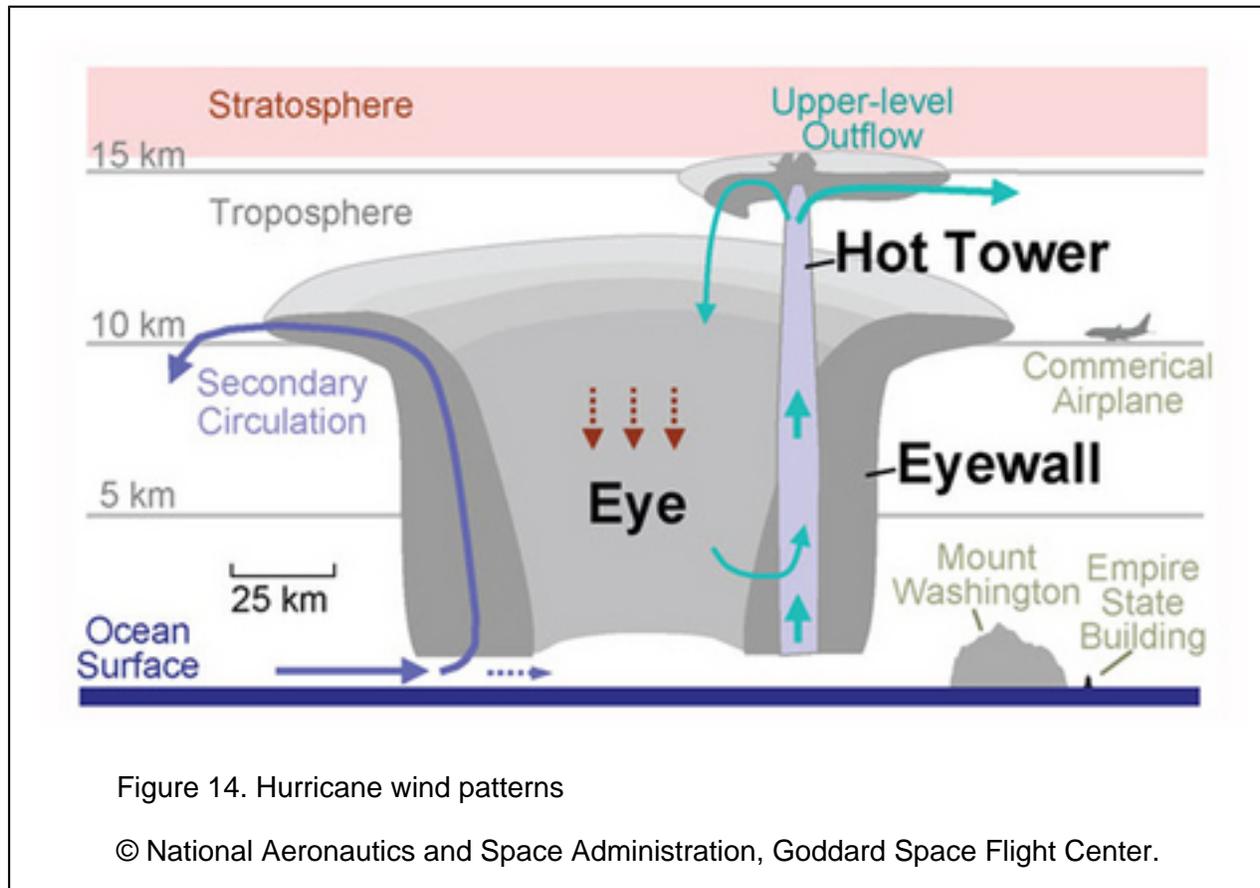
Because the Coriolis effect prevents mass and heat from moving readily to polar latitudes, temperatures decline and pressures increase sharply between middle latitudes and the polar regions.

This sharp pressure gradient creates powerful **jet stream** winds flowing from west to east at the boundary area. Jet stream winds meander and transport heat as they shift northward and southward. In the process they bring much of the weather system activity in the middle latitudes. When the mid-latitude jet stream dips down from Canada into the United States during the winter, it can carry arctic air and winter storms into the southeastern states.

To see how local climatic conditions create specific weather patterns, consider two types of storms: hurricanes and mid-latitude cyclones. Hurricanes form over tropical waters (between 8° and 20° latitude) in areas with high humidity, light winds, and warm sea surface temperatures, typically above 26.5°C (80°F). The most active area is the western Pacific, which contains a wide expanse of very warm ocean water. More hurricanes occur annually in the Pacific than in the Atlantic, which is a smaller area and therefore provides a smaller expanse of warm ocean water.

The first sign of a potential hurricane is the appearance of a tropical disturbance (cluster of thunderstorms). At the ocean's surface a feedback loop sometimes develops: falling pressure pulls in more air at the surface, which makes more warm air rise and release latent heat, which further reduces surface pressure. The Coriolis force will lead the converging winds into a counterclockwise circulation around the storm's lowest-pressure area.

Meanwhile, air pressure near the top of the storm starts to rise in response to latent heat warming. This high-pressure zone makes air diverge (flow outward) around the top of the center of the system. It then drops to the ground, forming powerful winds (Fig. 14). This upper-level area of high pressure acts like a chimney to vent the tropical system and keeps the air converging at the surface from piling up around the center. If air were to pile up at the center, surface pressure would rise inside the storm and ultimately weaken or destroy it.



At the surface hurricanes can diminish quickly if they move over cooler water or land and lose their supplies of warm, moist tropical air, or if they move into an area where the large-scale flow aloft is not favorable for continued development or maintenance of the circulation.

Mid-latitude cyclones cause most of the stormy weather in the United States, especially during the winter season. They occur when warm tropical and cold polar air masses meet at the polar front (coincident with the jet stream). Typically, warm air is lifted over the colder air and the system starts to move into a spiral. Because mid-latitude systems create buoyancy through lifting, their strongest wind velocities are at high altitudes (Fig. 15). In contrast, hurricanes generate buoyancy from rising warm air, so their highest velocities are at the surface where pressure differences are greatest.

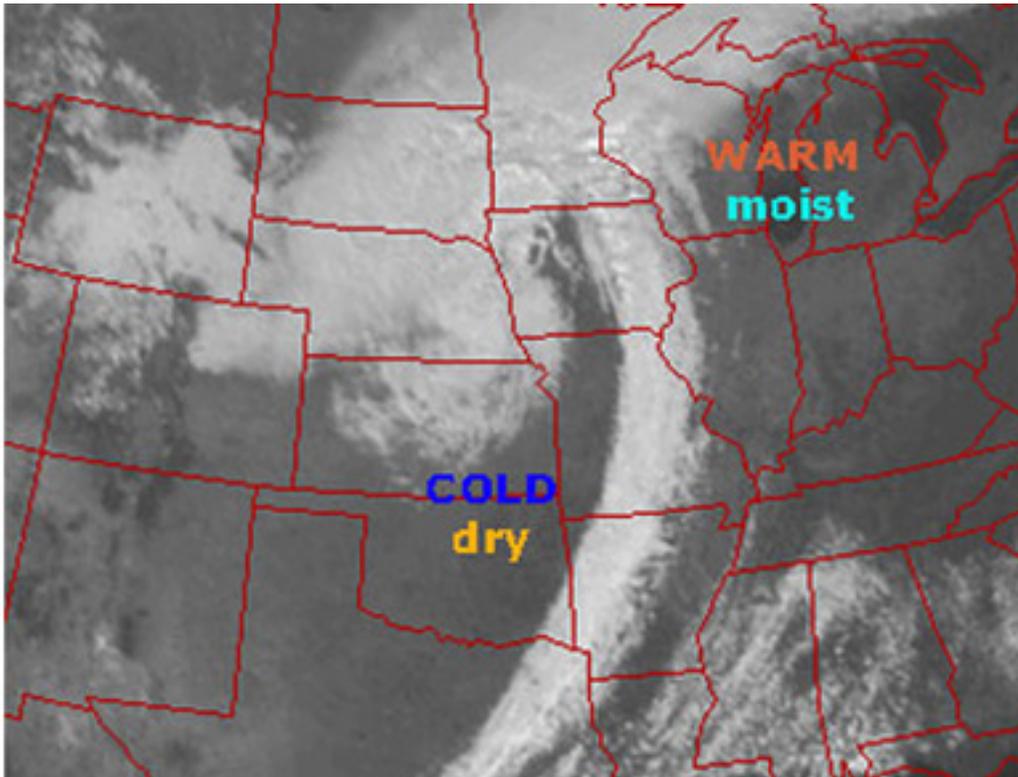


Figure 15. Mid-latitude cyclones along the polar front

© Dr. Michael Pidwimy, University of British Columbia Okanagan.

In many parts of the globe, atmospheric dynamics and ocean circulation patterns interact to create other distinct climate cycles that occur over longer periods than a single storm. Examples include seasonal monsoon rainstorms in Asia and the American southwest and multi-year patterns such as the [El Niño Southern Oscillation \(ENSO\)](#). These climate cycles are discussed in detail in Unit 3, "Oceans."

## 8. The Global Carbon Cycle

As this unit shows, temperature and moisture levels are major variables shaping Earth's weather patterns. By increasing atmospheric concentrations of greenhouse gases through activities such as burning fossil fuels, humans are changing the planet's radiative balance. This process is altering global temperatures and moisture levels, so we can expect that it will change Earth's weather

patterns. One of the key issues in current atmospheric science research is understanding how GHG emissions affect natural cycling of carbon between the atmosphere, oceans, and land. The rate at which land and ocean sinks take up carbon will determine what fraction of man-made CO<sub>2</sub> emissions remain in the atmosphere and alter Earth's radiative balance.

Atmospheric levels of CO<sub>2</sub>, the most important anthropogenic greenhouse gas, are controlled by a dynamic balance among biological and inorganic processes that make up the carbon cycle. These processes operate on very diverse time scales ranging from months to geological epochs. Today, human intervention in the carbon cycle is disturbing this natural balance. As a result, atmospheric CO<sub>2</sub> concentrations are rising rapidly and are already significantly higher than any levels that have existed for at least the past 650,000 years.

In recent decades, only about half of the CO<sub>2</sub> added to the atmosphere by human activities has stayed in the atmosphere. The rest has been taken up and stored in the oceans and in terrestrial ecosystems. The basic processes through which land and ocean sinks (storage reservoirs) take up carbon are well understood, but there are many questions about how much anthropogenic carbon these sinks can absorb, which sinks are taking up the largest shares, and how sensitive these sinks are to various changes in the environment. These issues are concerns for atmospheric scientists because carbon that cannot be taken up by land and ocean sinks will ultimately end up in the atmosphere. By monitoring atmospheric concentrations of CO<sub>2</sub> and other greenhouse gases, scientists are working to understand the operation of natural carbon sinks more accurately (Fig. 16).

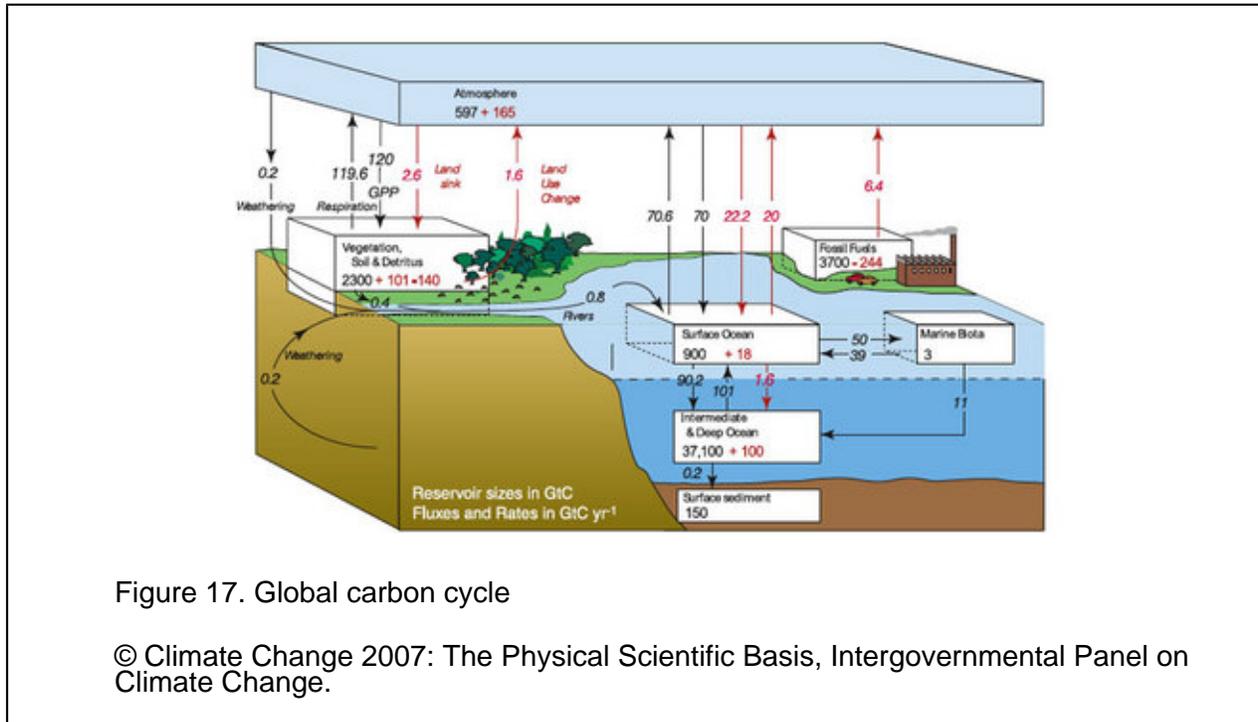
*"We use the atmosphere as a diagnostic to get a handle on these processes—to quantify where they take place and how long they are. If we can get an understanding of what the Earth itself is doing with these excess gases, we can make better prognoses of what future climate change might be like."*

Dr. Pieter Tans, National Oceanographic and Atmospheric Administration



Figure 16. Tall tower monitoring station for atmospheric sampling

The carbon cycle can be viewed as a set of reservoirs or compartments, each of which holds a form of carbon (such as calcium carbonate in rocks or  $\text{CO}_2$  and methane in the atmosphere), with carbon moving at various natural rates of transfer between these reservoirs (Fig. 17). The total amount of carbon in the system is fixed by very long-term geophysical processes such as the weathering of rock. Human actions that affect the carbon cycle, such as fossil fuel combustion and deforestation, change the rate at which carbon moves between important reservoirs. Burning fossil fuels speeds up the "weathering" of buried hydrocarbons, and deforestation accelerates the natural pace at which forests die and decompose, releasing carbon back to the atmosphere.



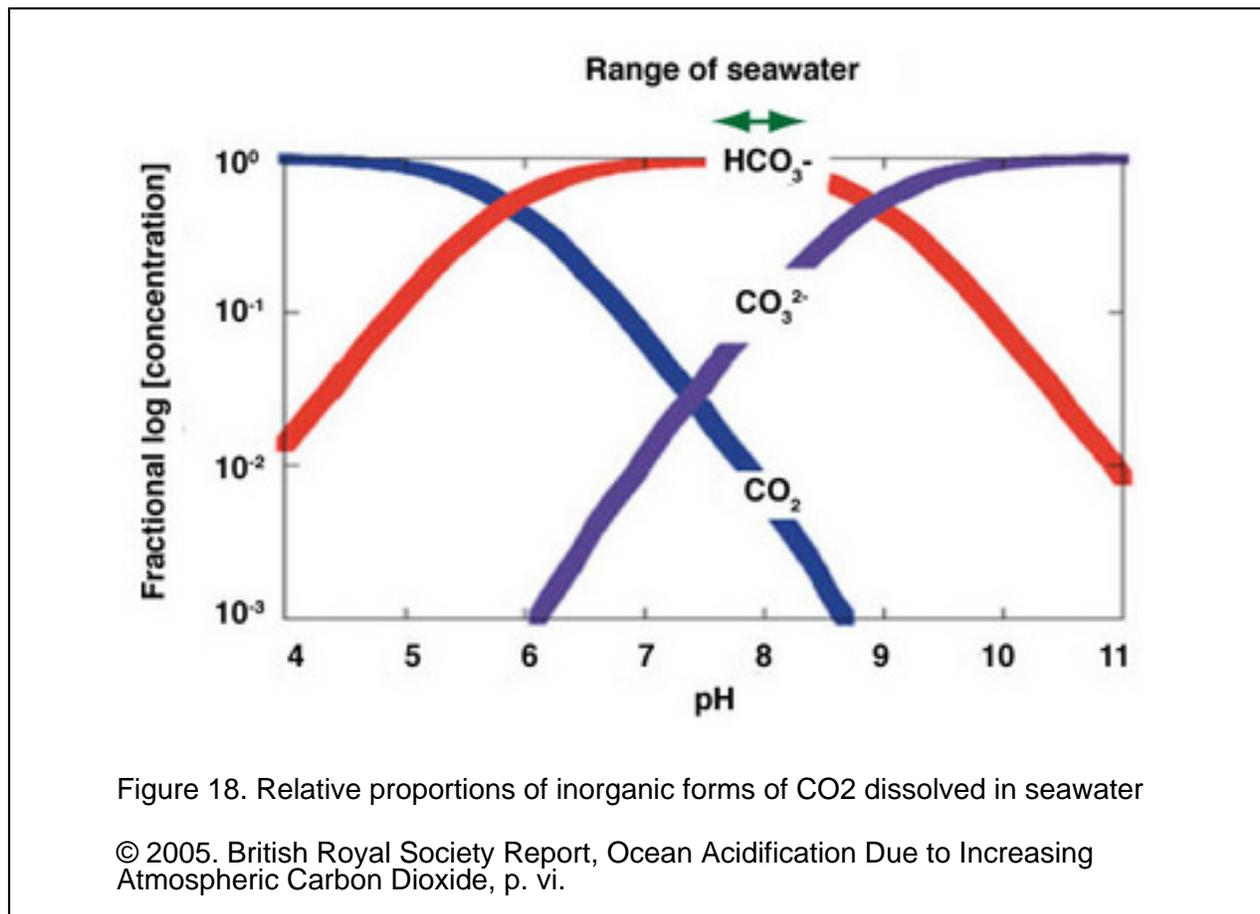
The residence time of carbon varies widely among different reservoirs. On average a carbon atom spends about 5 years in the atmosphere, 10 years in terrestrial vegetation, and 380 years in intermediate and deep ocean waters. Carbon can remain locked up in ocean sediments or fossil fuel deposits for millions of years. Fast cycling processes that take place in months or a few years have rapid effects but only influence small CO<sub>2</sub> reservoirs, so they do not change long-term CO<sub>2</sub> levels significantly. Slow processes that take place over centuries, millennia, or geologic epochs have greater influence on CO<sub>2</sub> concentrations over the long term.

Two processes remove CO<sub>2</sub> from the atmosphere: photosynthesis by land plants and marine organisms, and dissolution in the oceans. There is an important distinction between these processes in terms of permanence. CO<sub>2</sub> taken up through photosynthesis is converted into organic plant material, whereas CO<sub>2</sub> dissolved in the oceans is transferred to a new carbon reservoir but remains in inorganic form. Organic carbon in plant tissues can remain sequestered for thousands or millions of years if it is buried in soils or deep ocean sediments, but it returns to the atmosphere quickly from material such as leaf litter. Similarly, CO<sub>2</sub> dissolved in the oceans will stay a long time if sequestered in deep water, but will escape more readily back into the atmosphere if ocean mixing brings it to the surface.

Oceans and land ecosystems thus serve as both sources and sinks for carbon. Until recently these processes were in rough equilibrium, but the balance is being disrupted today as human activities add

more carbon to the atmosphere and a large fraction of that anthropogenic carbon is transferred to the oceans. Therefore, it is important to understand the chemical and biological processes through which the oceans take up  $\text{CO}_2$ .

Atmospheric  $\text{CO}_2$  dissolves into surface waters, where it reacts with liquid water to form carbonic acid, carbonate, and bicarbonate. This process makes the oceans an important buffer against global climate change, but there are limits to how much  $\text{CO}_2$  the oceans can absorb. Seawater is slightly basic, with a pH value of 8.2. Adding  $\text{CO}_2$  acidifies the water. Dissolved  $\text{CO}_2$  gas reacts with carbonate ( $\text{CO}_3^{2-}$ ) ions in the water, increasing concentrations of  $\text{H}^+$  and other hydrogen ions, which drives pH values lower (Fig. 18).



Over the long term, reducing the concentration of carbonate ions will slow the rate at which oceans take up  $\text{CO}_2$ . However, this process could significantly alter ocean chemistry. The British Royal Society estimated in a 2005 report that uptake of anthropogenic  $\text{CO}_2$  emissions had already reduced

The logo features the text "The Habitable Planet" in a sans-serif font. "The" is in orange, "Habitable" is in green, and "Planet" is in blue. A red circular arc is positioned behind the text, partially overlapping the letters "H" and "P".

# The Habitable Planet

the pH of the oceans by 0.1 units, and that the average pH of the oceans could fall by 0.5 units by 2100 if CO<sub>2</sub> emissions from human activities continued to rise at their current pace (footnote 1).

Theoretically the oceans could absorb nearly all of the CO<sub>2</sub> that human activities are adding to the atmosphere. However, only a very small portion of the ocean (the mixed layer, discussed further in Unit 3, "Oceans") comes into close contact with the atmosphere in a year. It would take about 500 years for all ocean water to come into contact with the atmosphere. As we will see in Unit 12, "Earth's Changing Climate," solutions to climate change are needed on a faster scale.

As noted above, biological uptake in the oceans occurs when phytoplankton in surface waters use CO<sub>2</sub> during photosynthesis to make organic matter. The organic carbon stored in these organisms is then transferred up the food chain, where most is turned back into CO<sub>2</sub>. However, some ultimately falls to lower depths and is stored in deep ocean waters or in ocean sediments, a mechanism called the "biological pump" (for more details, see Unit 3, "Oceans").

Forests take up CO<sub>2</sub> through photosynthesis and store carbon in plant tissue, forest litter, and soils. Forests took up a rising share of CO<sub>2</sub> from fossil fuel combustion in the 1980s and 1990s. Scientists believe that this occurred mainly because forests in the Northeastern United States and similar areas in Europe, many of which were clear-cut or used for agriculture in the 1700s and 1800s, have been growing back with the decline of agriculture in the region (Fig. 19).



Figure 19. Farm abandonment (1850) and hardwood forest regrowth (1930) in central New England

© Harvard Forest Dioramas, Fisher Museum, Harvard Forest, Petersham, MA.  
Photography is by John Green and David Foster.

Can forests solve the problem of rising atmospheric CO<sub>2</sub> levels? If lands are managed to optimize CO<sub>2</sub> uptake through sustainable forestry practices, forests can continue to sequester a significant fraction of the carbon that human activities are adding to the atmosphere. However, this share is unlikely to grow much beyond its current level (about 10% of anthropogenic emissions) because the rate of carbon uptake levels off as forests mature. Forests can help, but are not a total solution. (For more details, see Unit 13, "Looking Forward: Our Global Experiment.")

## 9. Feedbacks in the Atmosphere

Understanding how rising GHG levels may affect Earth's energy balance is complicated because of **feedbacks** in Earth's climate system. Feedbacks are interactions between climate variables such as temperature, precipitation, and vegetation and elements that control the greenhouse effect, such as clouds and albedo. Positive feedbacks amplify temperature change by making the greenhouse effect stronger or by reducing albedo, so they make the climate system more sensitive to the properties that trigger them. Negative feedbacks have a dampening effect on temperature change, making the climate system less sensitive to the factors that trigger them.

Feedbacks can be very complex processes and may take place over short or long time spans. Important feedbacks in Earth's atmosphere include:

- **Water vapor feedback (positive).** The atmosphere can hold increasing amounts of water vapor as the temperature rises, because the pressure of water vapor in equilibrium with liquid water increases exponentially with temperature. The presence of more water vapor as temperature rises increases the greenhouse effect, as well as the absorption of solar radiation, which further raises temperature. This is the strongest and best-understood feedback mechanism in the atmosphere, because it is based on the straightforward fact that warm air can hold more water vapor than cool air.
- **Cloud feedback on terrestrial radiation (positive).** Because warmer temperatures increase water vapor amounts, they can increase cloudiness and further raise temperature. This is a very strong feedback that is not well understood. It is hard to know whether or how much cloudiness will increase as temperature does, because cloudiness depends more on upward air motion than on temperature or water vapor levels directly. (For details on how clouds form, see Section 5, "Vertical Motion in the Atmosphere.")
- **Cloud feedback on solar radiation (negative).** As temperature increases and atmospheric water vapor levels rise, cloudiness may increase. Greater cloudiness raises Earth's albedo, reflecting an increasing fraction of solar radiation back into space and decreasing temperature, although some cloud types are more reflective than others (Fig. 20). This is another very strong feedback that is not well understood because it is hard to know whether or how much cloudiness will increase with temperature. Also, as noted in the previous example, clouds can also absorb infrared radiation, raising temperatures.

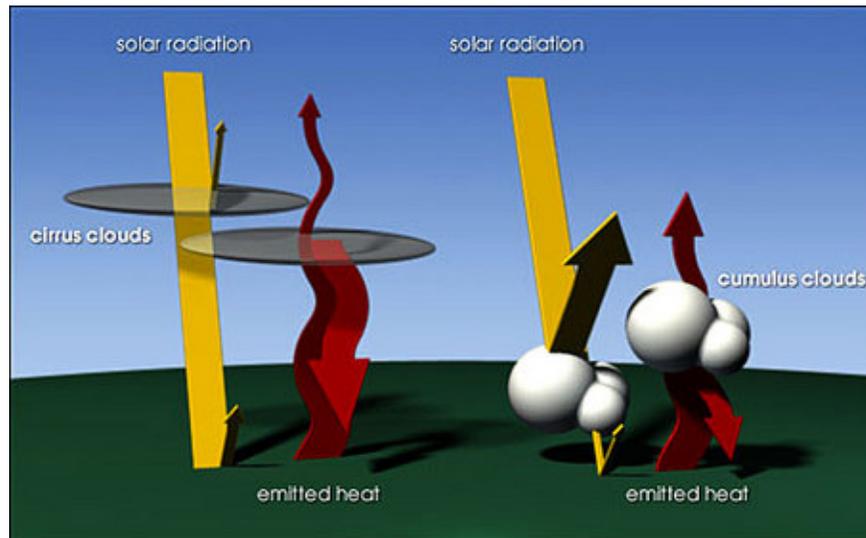


Figure 20. Effects of cirrus and cumulus clouds on Earth's energy balance

© National Aeronautics and Space Administration. Earth Observatory.

- **Vegetation feedback on solar radiation (negative).** As temperatures rise, deserts may expand, increasing Earth's albedo and decreasing temperature. This is a very complex feedback. It is uncertain whether deserts will expand, or conversely, whether higher CO<sub>2</sub> levels might stimulate higher plant growth levels and increase vegetation instead of reducing it.
- **Ice-albedo feedback on solar radiation (positive).** Rising temperatures cause polar glaciers and floating ice sheets to recede, decreasing Earth's albedo and raising temperatures. This feedback is very strong at times when polar ice has expanded widely, such as at the peak of ice ages. It can work in both directions, helping ice sheets to advance as Earth cools and accelerating the retreat of ice sheets during warming periods. There is relatively little polar ice on land today, so this feedback is not likely to play a major role in near-term climate change. However, temperature increases large enough to melt most or all of the floating ice in the Arctic could sharply accelerate global climate change, because ocean water absorbs almost all of the incident solar radiation whereas ice reflects most sunlight.



Feedbacks cause much of the uncertainty in today's climate change models, and more research is needed to understand how these relationships work. A 2003 National Research Council study called for better measurement of many factors that affect climate feedbacks, including temperature, humidity, the distribution and properties of clouds, the extent of snow cover and sea ice, and atmospheric GHG concentrations (footnote 2).

## 10. Further Reading

National Aeronautics and Space Administration, "The Importance of Understanding Clouds," NASA Facts, FS-2005-0-073-GSFC, [http://www.nasa.gov/pdf/135641main\\_clouds\\_trifold21.pdf](http://www.nasa.gov/pdf/135641main_clouds_trifold21.pdf).

## Footnotes

1. The Royal Society, **Ocean Acidification Due to Increasing Atmospheric Carbon Dioxide**, June 2005, p. vi, <http://www.royalsoc.ac.uk/displaypagedoc.asp?id=13539>.
2. National Research Council, **Understanding Climate Change Feedbacks** (Washington, DC: National Academy Press, 2003), p. 3.

## Glossary

**albedo** : The fraction of electromagnetic radiation reflected after striking a surface.

**anaerobic** : Describes an organism that is able to live without oxygen. Also used to describe environments that are devoid of gaseous or dissolved molecular oxygen.

**angular momentum** : The measure of the extent to which an object will continue to rotate about a point unless acted upon by an external torque.

**atmospheric (adiabatic) lapse rate** : The constant decline in temperature of an air parcel as it rises in the atmosphere due to pressure drop and gas expansion.

**buoyant** : Capable of floating.

**convection** : The transfer of heat by a moving fluid, such as air or water.

**convergence** : The flowing together of air masses.

**Coriolis force** : The apparent force, resulting from the rotation of the Earth, that deflects air or water movement.

**deforestation** : Removal of trees and other vegetation on a large scale, usually to expand agricultural or grazing lands.

**dew point** : The temperature at which air becomes saturated with water vapor and condenses into water called dew.



**dry adiabatic lapse rate** : The rate at which the temperature of a parcel of dry air decreases as the parcel is lifted in the atmosphere.

**dynamic balance** : Condition of a system in which inflow of materials or energy equals flow.

**El Niño Southern Oscillation (ENSO)** : A global event arising from large-scale interactions between the ocean and the atmosphere, usually an oscillation in the surface pressure (atmospheric mass) between the southeastern tropical Pacific and the Australian-Indonesian regions.

**feedback** : Corrective information or a signal generated within a self-regulating system or process that is intended to induce a change in that system or process.

**geostrophic flow** : A current in the atmosphere in which the Coriolis force and the pressure gradient are in balance.

**greenhouse gases** : Atmospheric gases or vapors that absorb outgoing infrared energy emitted from the Earth naturally or as a result of human activities. Greenhouse gases are components of the atmosphere that contribute to the Greenhouse effect.

**Hadley circulation** : A general circulation pattern in which air rises near the equator, flows north and south away from the equator at high altitudes, sinks near the poles, and flows back along the surface from both poles to the equator.

**isobars** : Lines on a map connecting points having the same barometric pressure.

**jet stream** : Fast flowing, relatively narrow air currents found in the atmosphere at around 11 kilometres (36,000 ft) above the surface of the Earth, just under the tropopause.

**latent energy** : Energy supplied externally, normally as heat, that does not bring about a change in temperature.

**moist adiabatic lapse rate** : The rate at which the temperature of a parcel of saturated air decreases as the parcel is lifted in the atmosphere. The moist adiabatic lapse rate is not a constant like the dry adiabatic lapse rate but is dependent on parcel temperature and pressure.

**Montreal Protocol on Substances That Deplete the Ozone Layer** : A 1987 international agreement, subsequently amended in 1990, 1992, 1995, and 1997, that establishes in participating countries a schedule for the phaseout of chloroflourocarbons and other substances with an excessive ozone-depleting potential.

**natural fires** : A rapid, persistent chemical reaction that releases heat and light; especially the combination of any substance that's easy to burn with oxygen that releases heat. Most natural fires start when a lightning bolt strikes a tree trunk and knocks the tree down.

**relative humidity** : The ratio of the amount of water vapor present in a specified volume of air to the maximal amount that can be held by the same volume of air at a specified temperature and pressure.

# The Habitable Planet

**respiration** : Metabolism of an individual cell, tissue, or organism that results in the release of chemical energy derived from organic nutrients.

**ruminant** : Any hooved animal that digests its food in two steps, first by eating the raw material and regurgitating a semi-digested form known as cud, then eating (chewing) the cud, a process called ruminating.

**sinks** : Habitats that serve to trap or otherwise remove chemicals such as plant nutrients, organic pollutants, or metal ions through natural processes.

## Unit 3 : Oceans

### Overview

Oceans cover three-quarters of the Earth's surface, but many parts of the deep oceans have yet to be explored. Learn about the large-scale ocean circulation patterns that help to regulate temperatures and weather patterns on land, and the microscopic marine organisms that form the base of marine food webs.



Night time Prochlorococcus sampling in the Pacific Ocean.

### Sections:

1. Introduction
2. Ocean Structure and Composition
3. Ocean Currents
4. Thermohaline Circulation
5. Ocean Circulation and Climate Cycles
6. Biological Activity in the Upper Ocean
7. The "Biological Pump"
8. Further Reading

## 1. Introduction

Although it may seem counterintuitive, the oceans play a major part in creating conditions for life on land. Together with the atmosphere, oceans regulate global temperatures, shape weather and climate patterns, and cycle elements through the biosphere. They also contain nearly all of the water on Earth's surface and are an important food source. Life on Earth originated in the oceans, and they are home to many unique ecosystems that are important sources of biodiversity, from coral reefs to polar sea ice communities (Fig. 1).



Figure 1. Coral reef in the Hawaiian islands

© United States Geological Survey.

Scientific understanding of the oceans made great advances during the 20th century, but we still know relatively little about such central issues as the abundance and diversity of marine species, the declining health of coral reefs, and how future changes in global climate might affect ocean circulation. Technologies such as remote sensing from satellites are making it possible to collect and analyze more of the physical, chemical, and biological data that researchers need to address these questions.

However, the sheer size of the oceans and the difficulties involved in exploring them—especially at depths where no light penetrates and overlying water generates crushing pressures—make it an ongoing challenge to understand how the seas work. In an effort to fill this gap, several U.S. expert



panels recently have recommended investing in a national ocean exploration program and a broad-scale system like the National Weather Service to observe and forecast ocean conditions (footnote 1).

*"[T]he ocean remains one of the least explored and understood environments on the planet—a frontier for discoveries that could provide important benefits . . . . Ocean science and technology will play an increasingly central role in the multidisciplinary study and management of the whole-Earth system."*

U.S. Commission on Ocean Policy, **An Ocean Blueprint for the 21st Century** (2004)

This unit explores the working of ocean currents and circulation patterns and their influence on global climate cycles. It then turns to biological activity in the oceans, focusing on microscopic plankton that form the base of ocean food webs, and the influence of physical conditions like temperatures and currents on ocean food production.

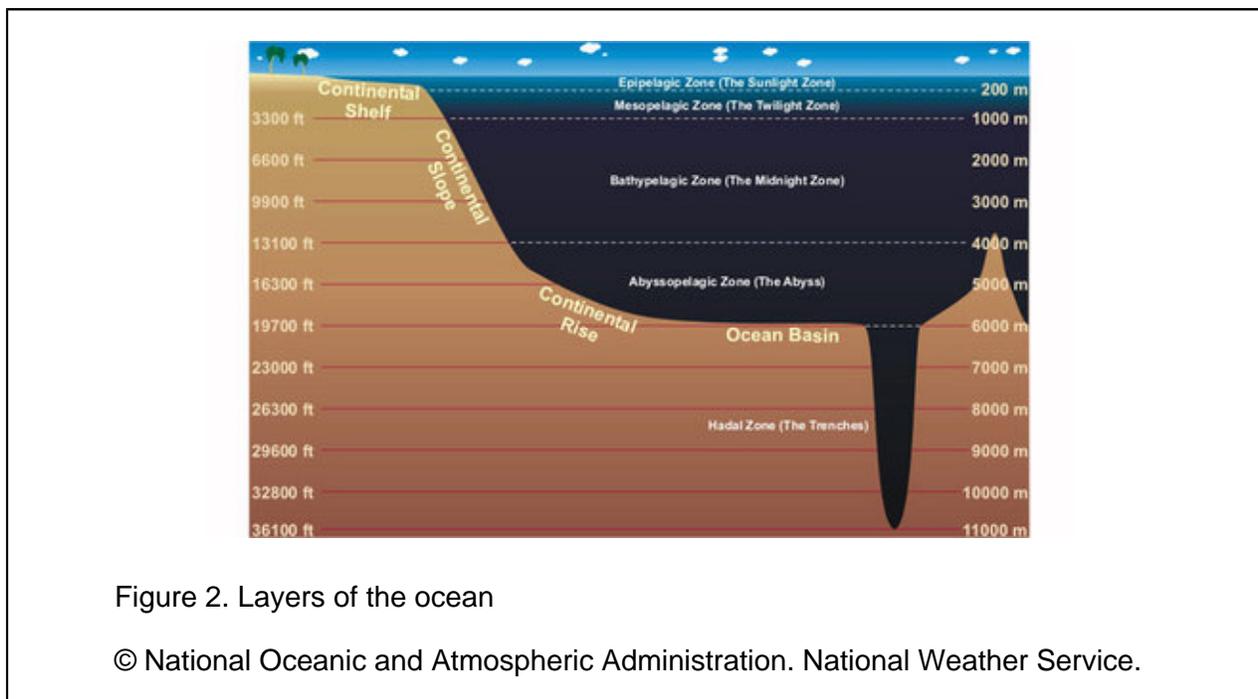
Finally, this unit looks at one of the most important global regulating mechanisms: the so-called "biological pump" in which plankton take up carbon from the atmosphere, then carry it to the deep ocean where it can remain for thousands of years. This process is centrally important to life on Earth for several reasons. First, as discussed in Unit 2, "Atmosphere," the process of photosynthesis releases one molecule of oxygen to the atmosphere for every carbon atom that is packaged in organic carbon. Phytoplankton photosynthesis produces about half of the world's oxygen supply. And as we will see, by exporting carbon to deep waters, the biological pump helps to regulate Earth's energy balance and partially offsets rising CO<sub>2</sub> concentrations in the atmosphere.

Despite their global scope, the oceans are highly vulnerable to human impacts, including marine pollution (discussed in Unit 8, "Water Resources") and over-fishing (addressed in Unit 9, "Biodiversity Decline"). Most profoundly, scientists increasingly agree that human-induced climate change could affect ocean temperatures, circulation, and carbon cycling over the next century. Scientists widely agree today that global climate change is causing ocean temperatures to rise and is changing fresh and salt water balances in many areas of the seas. Before we consider how ocean systems may be affected by global climate change, however, it is essential to understand how biological and physical processes interact normally in the oceans and how the oceans help to shape conditions for life on Earth. (For more details on how global climate change affects the oceans, see Unit 12, "Earth's Changing Climate," and Unit 13, "Looking Forward: Our Global Experiment.")

## 2. Ocean Structure and Composition

Like the atmosphere, the oceans are not uniformly mixed but are structured in layers with distinct properties (Fig. 2). Pressure increases with depth as the weight of the overlying air and

water increase. Unlike the atmosphere, however, pressure changes at a linear rate rather than exponentially because water is almost impossible to compress, so its mass is equally distributed throughout a vertical water column. Atmospheric pressure at sea level is 14.7 pounds per square inch (also referred to as "one atmosphere"), and pressure increases by an additional atmosphere for every 10 meters of descent under water. This gradient is well known to scuba divers who have experienced painful "ear squeeze" from pressure differences between the air in their ears and the seawater around them.



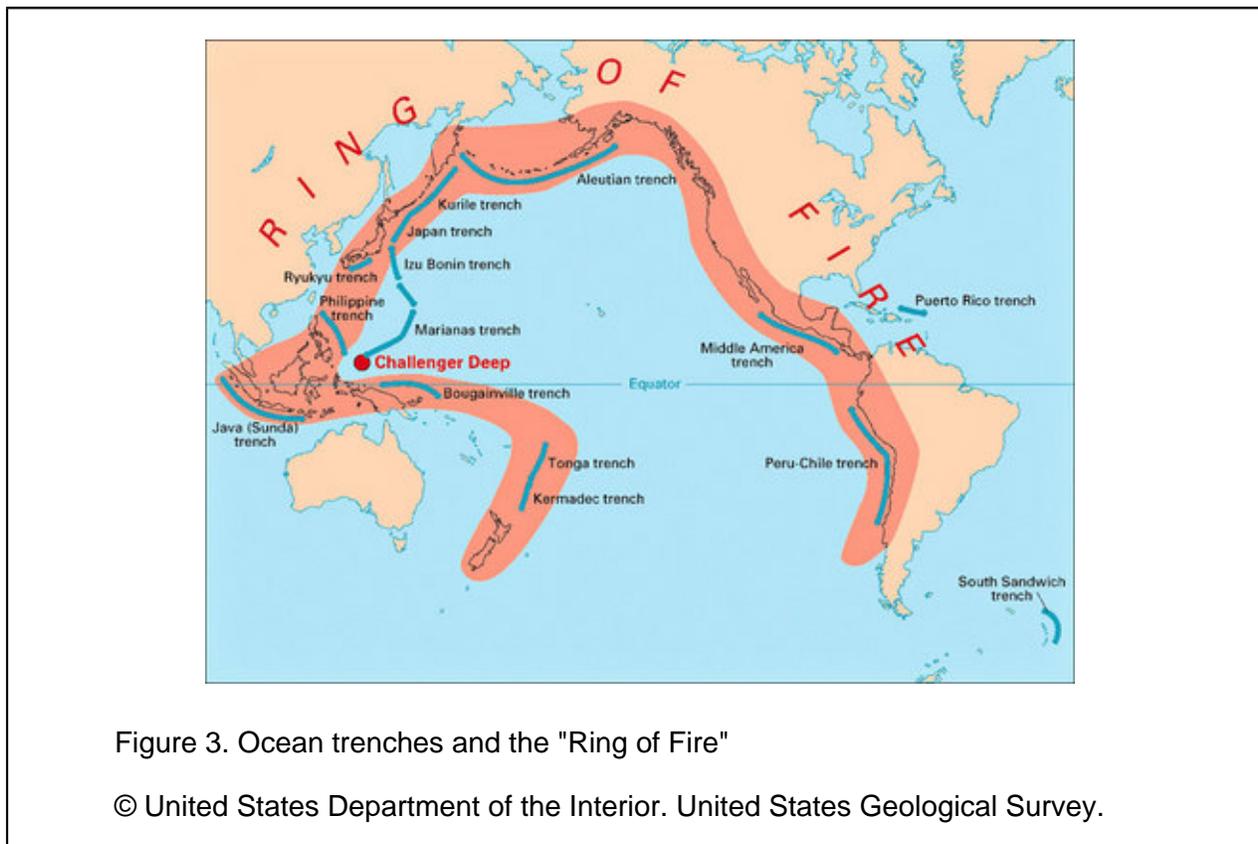
The Epipelagic, or sunlight, zone (so called because most visible light in the oceans is found here) comprises the first 200 meters below the surface, and is warm and mixed by winds and wave action. Surface waters account for about 2 percent of total worldwide ocean volume. At a depth of about 200 meters, the Continental Shelf (the submerged border of the continents) begins to slope more sharply downward, marking the start of the Mesopelagic, or twilight, zone. Here water temperature falls rapidly with depth to less than 5°C at 1,000 meters. This sharp transition, which is called the **thermocline**, inhibits vertical mixing between denser, colder water at depths and warmer water nearer the surface. About 18 percent of the total volume of the oceans is within this zone.

Below 1,000 meters, in the Bathypelagic, or midnight, zone, water is almost uniformly cold, approximately 4°C. No sunlight penetrates to this level, and pressure at the bottom of the zone (around 4,000 meters depth) is about 5,880 pounds per square inch. Little life exists at the

Abyssopelagic (abyssal) zone, which reaches to the ocean floor at a depth of about 6,000 meters. Together, these cold, deep layers contain about 80 percent of the total volume of the ocean.

The deepest points in the ocean lie in long, narrow trenches that occur at convergence zones—points where two oceanic plates collide and one is driven beneath the other. This region is called the Hadal zone. The deepest oceanic trench measured to date is the Marianas Trench near the Philippines, which reaches more than 10,000 meters below sea level. Highly specialized life forms, including fish, shrimps, sea cucumbers, and microbes, survive even at these depths.

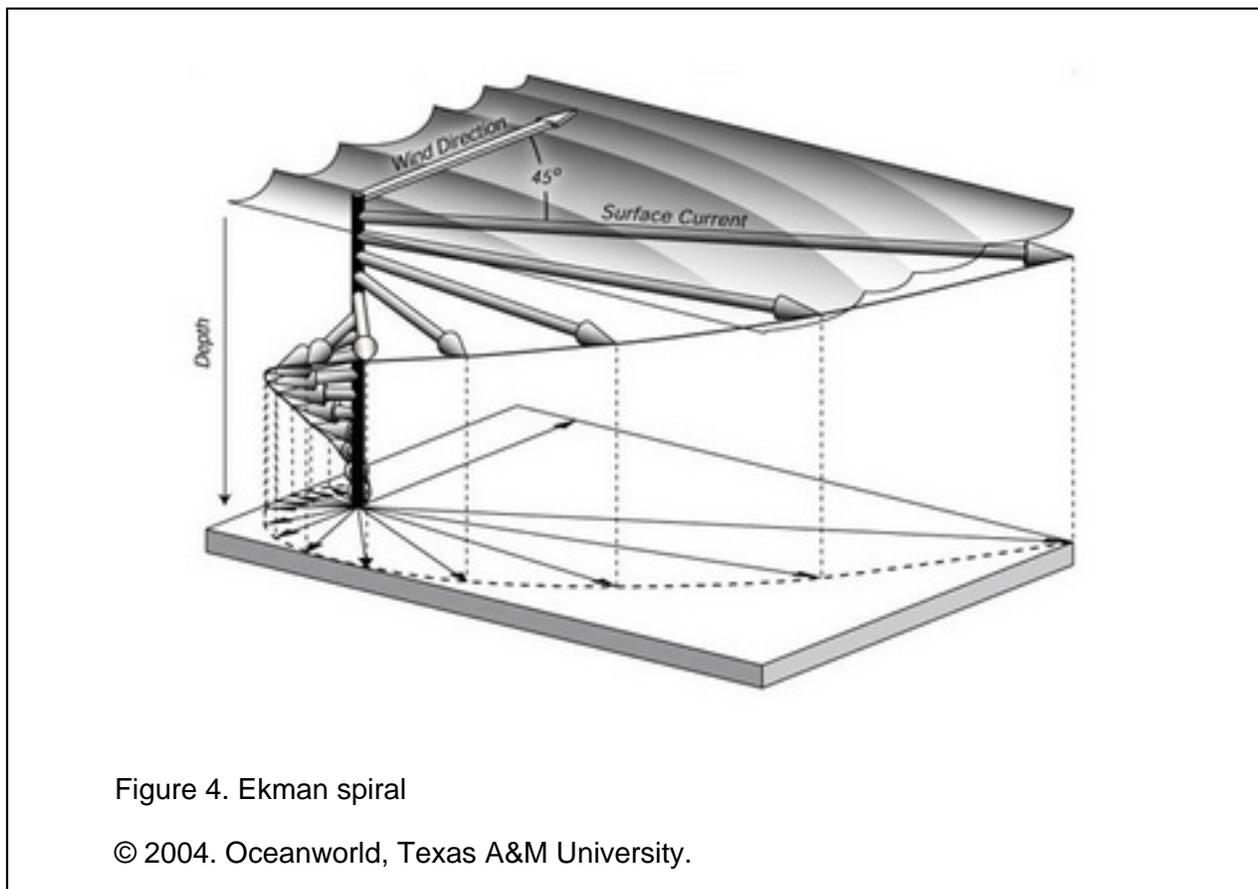
Movement along colliding plates in convergence zones frequently generates earthquakes and tsunamis; an earthquake measuring 9.15 on the Richter scale off the coast of Sumatra triggered the Indian Ocean tsunami that killed more than 230,000 people on December 26, 2004. Volcanoes often erupt near convergence zones when hot magma escapes through rock fractures. Many of the world's largest ocean trenches, therefore, are located along the "Ring of Fire," an arc of volcanoes around the Pacific Ocean that denotes convergent plates margins (Fig. 3).



## 3. Ocean Currents

Mixing is a key dynamic in the oceans, creating currents and exchanges between cold, deep waters and warmer surface waters. These processes redistribute heat from low to high latitudes, carry nutrients from deep waters to the surface, and shape the climates of coastal regions.

Several types of forces cause ocean mixing. Waves and surface currents are caused mainly by winds. When winds "pile up" water in the upper ocean, they create an area of high pressure and water flows from high to low pressure zones. Ocean currents tend to follow Earth's major wind patterns, but with a difference: the **Coriolis force** deflects surface currents at an angle of about 45 degrees to the wind—to the right in the Northern Hemisphere, left in the Southern Hemisphere. (For more about the Coriolis force, see Unit 2, "Atmosphere.") This pattern is called **Ekman transport**, after Swedish oceanographer Vagn Ekman. Each layer of the ocean transfers momentum to the water beneath it, which moves further to the right (left), producing a spiral effect (Fig. 4).



# The Habitable Planet

At deeper levels, ocean mixing is caused by differences in density between colder, saltier water and warmer, fresher water. Because the density of water increases as it becomes colder and saltier, it sinks at high latitudes and is replaced by warm water flowing northward from the tropics. (This pattern, called the [Thermohaline Circulation](#), is a key mechanism that helps to regulate Earth's climate and is discussed further below and in Section 4.) Cold water typically flows below warmer water, but when winds blowing along coastlines deflect warm surface currents away from shore through Ekman transport, they allow cold, nutrient-rich water to rise to the surface. This coastal [upwelling](#) process occurs against western coastlines in the Atlantic, Pacific, and Indian oceans.

The combined effects of these forces create circular currents called [gyres](#) in the world's largest oceans, centered at about 25° to 30° north and south latitudes. Gyres rotate clockwise in the Northern Hemisphere and counter-clockwise in the Southern Hemisphere, driven by easterly winds at low latitudes and westerly winds at high latitudes. Due to a combination of friction and planetary rotation, currents on the western boundaries of ocean gyres are narrower and flow faster than eastern boundary currents. Warm surface currents flow out of ocean gyres from the tropics to higher latitudes, and cold surface currents flow from colder latitudes toward the equator (Fig. 5).

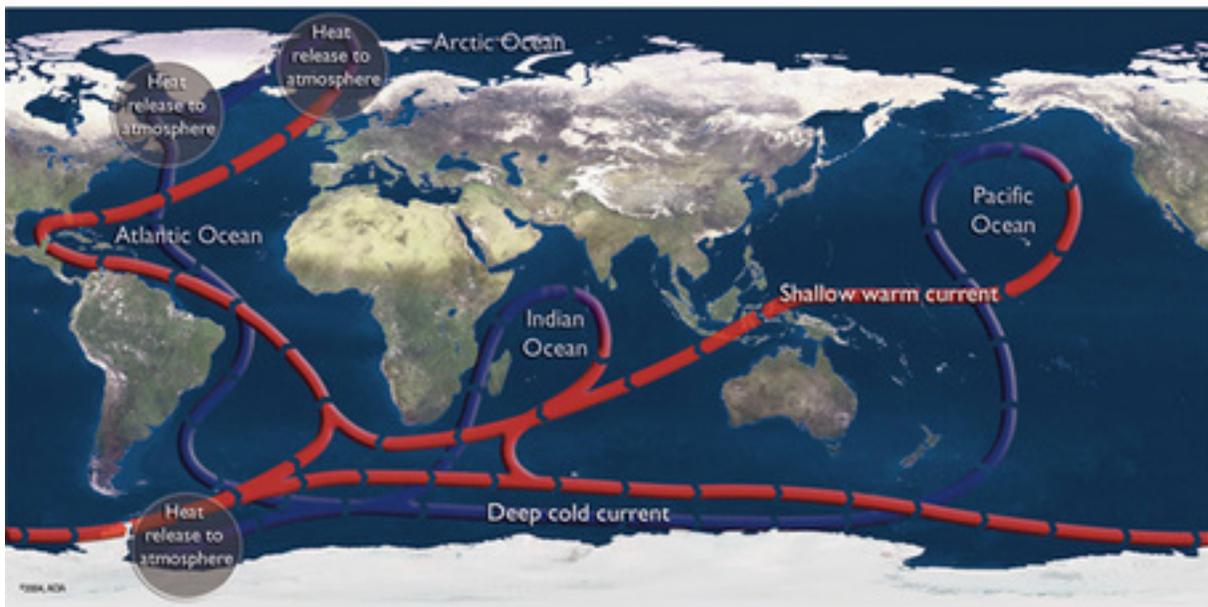


Figure 5. Ocean currents

© 2004. Arctic Climate Impact Assessment, Graphics Set 1, p. 21.

How does ocean circulation affect Earth's climate? The oceans redistribute heat from high to low latitudes by moving warm water from the equator toward the poles. They also cause a net transfer of heat from the Southern to the Northern Hemisphere. As currents flow, they warm or cool the overlying atmosphere. The most famous example is the Gulf Stream, the fast-moving western boundary current that flows north through the Atlantic Ocean and makes northern Europe much warmer than Canadian provinces lying at the same latitudes.

In areas where coastal upwelling brings cold water up from the depths, cold currents have the opposite effect. As one illustration, San Diego, California, and Columbia, South Carolina, lie at the same latitude, but coastal upwelling in the eastern Pacific brings cold water into the California Current system, which runs south along the California coast. This process helps to keep peak summer temperatures in San Diego at about 78°F, compared to 95°F in Columbia.

Ocean waters are warmest in the tropics and coldest at the poles because the sun heats the equator more strongly than high latitudes. Surface water temperatures can be 30°C warmer at the equator than in polar regions (Fig. 6). Short-term variations in ocean surface temperatures, from day to night and summer to winter, are mainly influenced by the sun's energy. However, as we will see below in Section 5, some longer-term temperature changes are not driven directly by the sun but rather by complex atmosphere/ocean interactions that occur on seasonal, annual, or multi-year cycles.

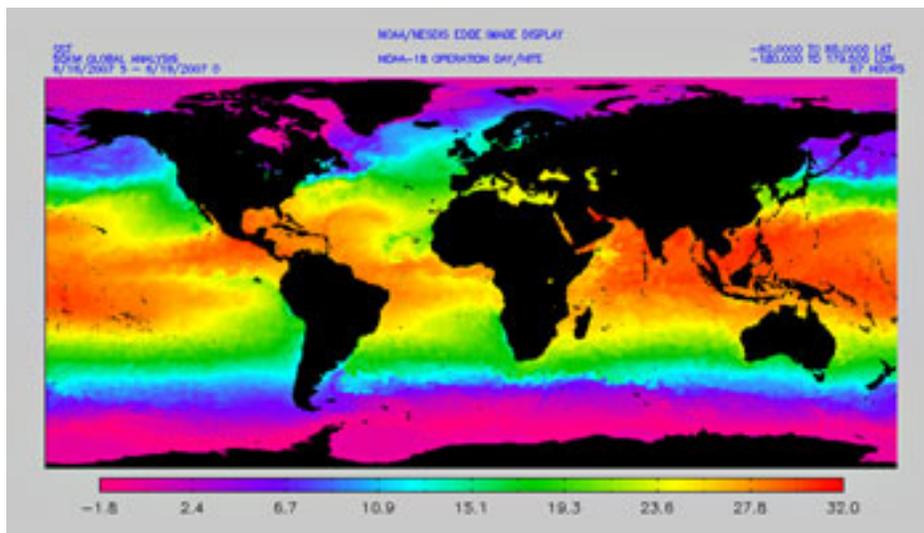


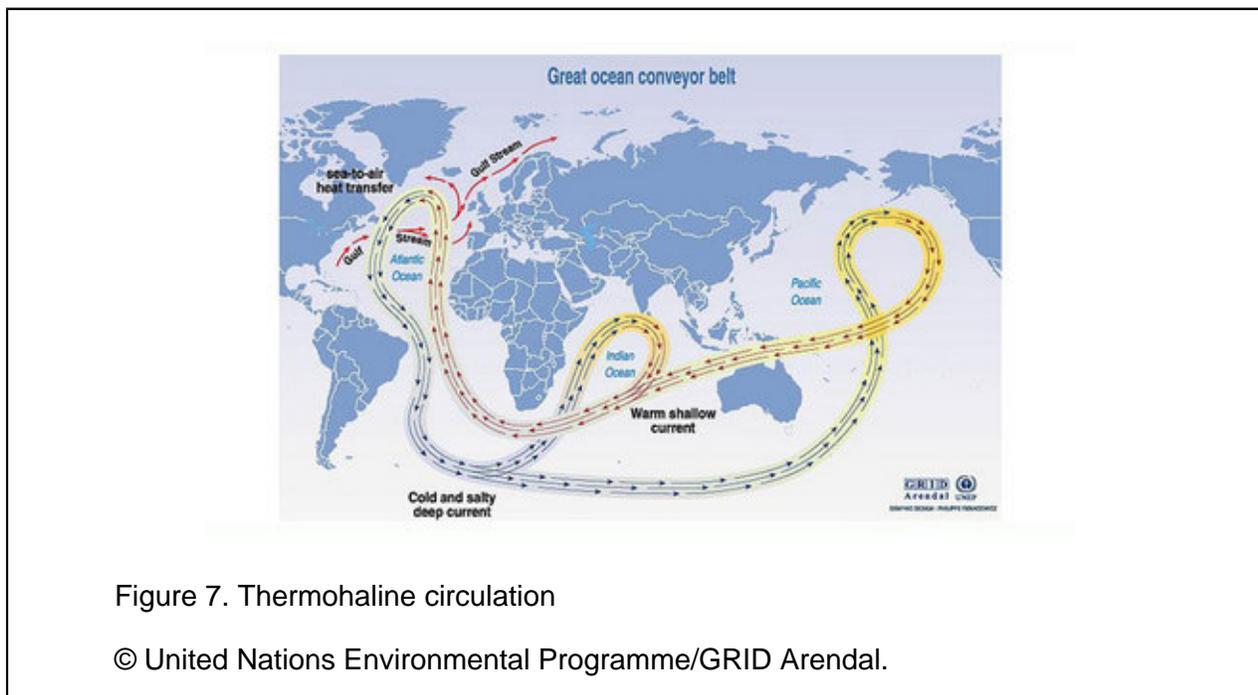
Figure 6. Global sea surface temperatures, July 1–4, 2005

© NOAA Satellite and Information Service. National Environment Satellite, Data, and Information Services.

The oceans respond to temperature changes more slowly than the atmosphere because water has a higher **specific heat capacity (SHC)** than air, which means that it takes more energy input to increase its temperature (about four times more for liquid water). Because water warms and cools more slowly than land, oceans tend to moderate climates in many coastal areas: seaside regions are typically warmer in winter and cooler in summer than inland locations. On a larger scale, by absorbing heat the oceans are delaying the full impact of rising temperatures due to global climate change by decades to centuries. (For more details, see Unit 13, "Looking Forward: Our Global Experiment.")

## 4. Thermohaline Circulation

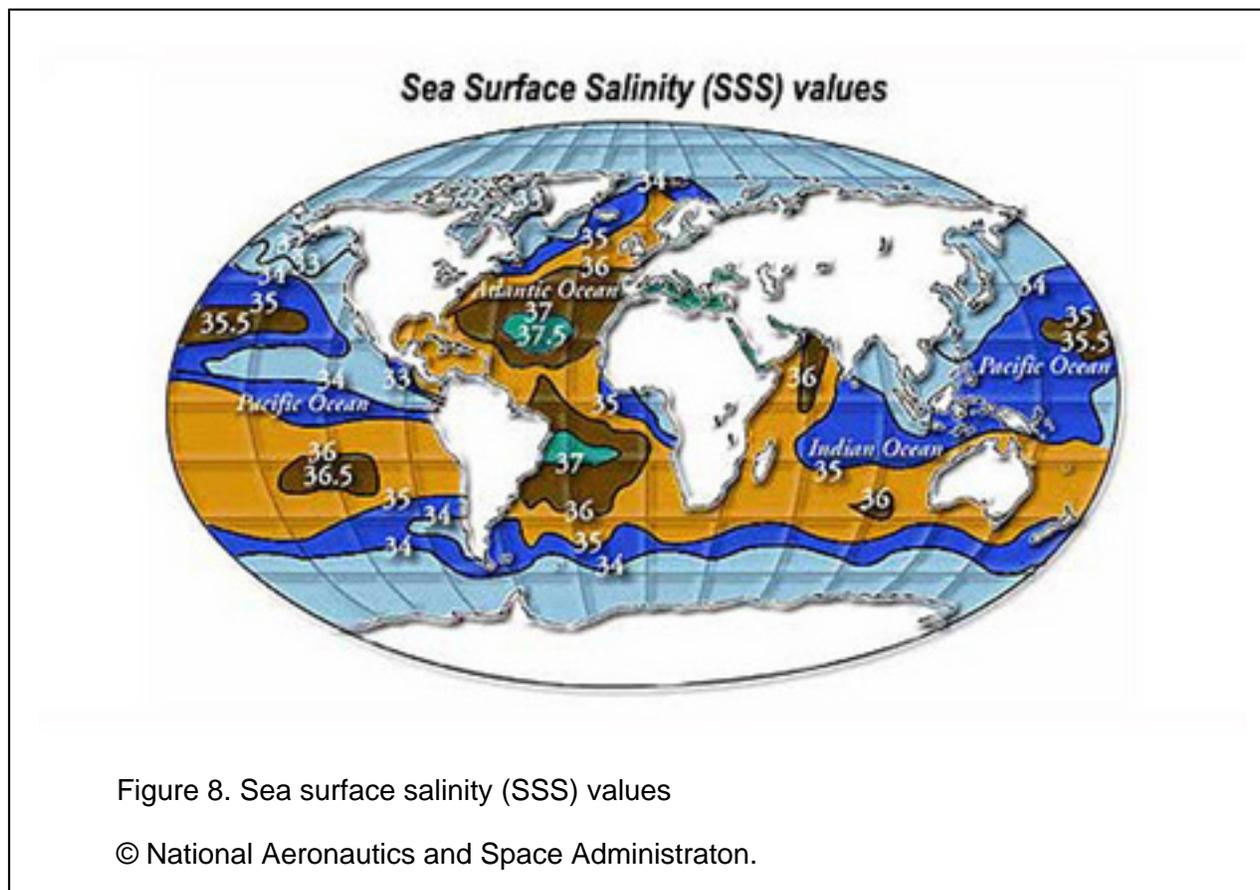
The thermohaline circulation is often referred to as the "global conveyor belt" because it moves large volumes of water along a course through the Atlantic, Pacific, and Indian oceans. Cold, salty water sinks in the Norwegian Sea and travels south to the Antarctic, then east to the Pacific. Here water warms and rises, then reverses course and follows an upper-ocean course back through the Indian Ocean and around Cape Horn to the Atlantic (Fig. 7). The current has a flow equal to that of 100 Amazon rivers.



The thermohaline circulation is driven by **buoyancy** differences in the upper ocean that arise from temperature differences (thermal forcing) and salinity differences (haline forcing). As noted in the previous section, ocean temperatures are lower in polar regions and higher at the equator because

the latter receives more radiant energy from the sun. Cold water sinks in polar areas and warmer, less dense water floats on top.

In contrast, salinity differences are caused by evaporation, precipitation, freshwater runoff, and sea ice formation. Evaporation rates are highest in subtropical regions where Hadley circulation loops produce descending currents of warm, dry air (for more details on [Hadley circulation](#), see Unit 2, "Atmosphere"). When sea water evaporates or freezes, most of its salt content is left behind in the ocean, so high rates of evaporation in the subtropics raise salinity levels. Sea water is relatively less saline at higher latitudes because these regions have more precipitation than evaporation. In addition, melting sea ice returns fresh water to the oceans (Fig. 8).



Putting these two factors together, water cools in the North Atlantic and becomes dense enough to sink. Surface currents carry warm, salty water poleward to replace it. In the Indian and Pacific oceans, deep water returns to the surface through upwellings.

As sea ice forms and melts at the poles, it influences ocean circulation by altering the salinity of surface waters. When sea water freezes into ice, it ejects its salt content into the surrounding water,

so waters near the surface become saltier and dense enough to sink. This process propels cold waters to depth and draws warmer waters northward in their place. Two factors can make polar waters less salty and reduce this flow. First, warmer temperatures on land increase glacial melting, which sends higher flows of fresh water into the oceans. Since fresh water is less dense than salt water, it floats on the ocean's surface like a film of oil, reducing vertical mixing and slowing the formation of deepwater. Warmer air temperatures also reduce the formation of sea ice, so less salt is ejected into northern waters.

Within the ocean, distinct water masses with physical properties that are different from the surrounding water form and circulate, much like air masses in the atmosphere. Several important water masses help to drive the thermohaline circulation. North Atlantic Deep Water (NADW), the biggest water mass in the oceans, forms in the North Atlantic and runs down the coast of Canada, eastward into the Atlantic, and south past the tip of South America. NADW forms in the area where the North Atlantic Drift (the northern extension of the Gulf Stream) ends, so it helps to pull the Gulf Stream northward. If the NADW were to slow down or stop forming, as has happened in the past, this could weaken the Gulf Stream and the North Atlantic Drift and cool the climate of northwest Europe. (For more details, see Unit 13, "Looking Forward: Our Global Experiment.")

Another cold water mass, Antarctic Bottom Water (AABW), is the densest water mass in the oceans. It forms when cold, salty water sinks in the seas surrounding Antarctica, carrying oxygen and nutrients with it, and flows northward along the sea floor underneath the North Atlantic Deep Water, displacing the waters above it and helping to propel the Thermohaline Circulation (Fig. 9).

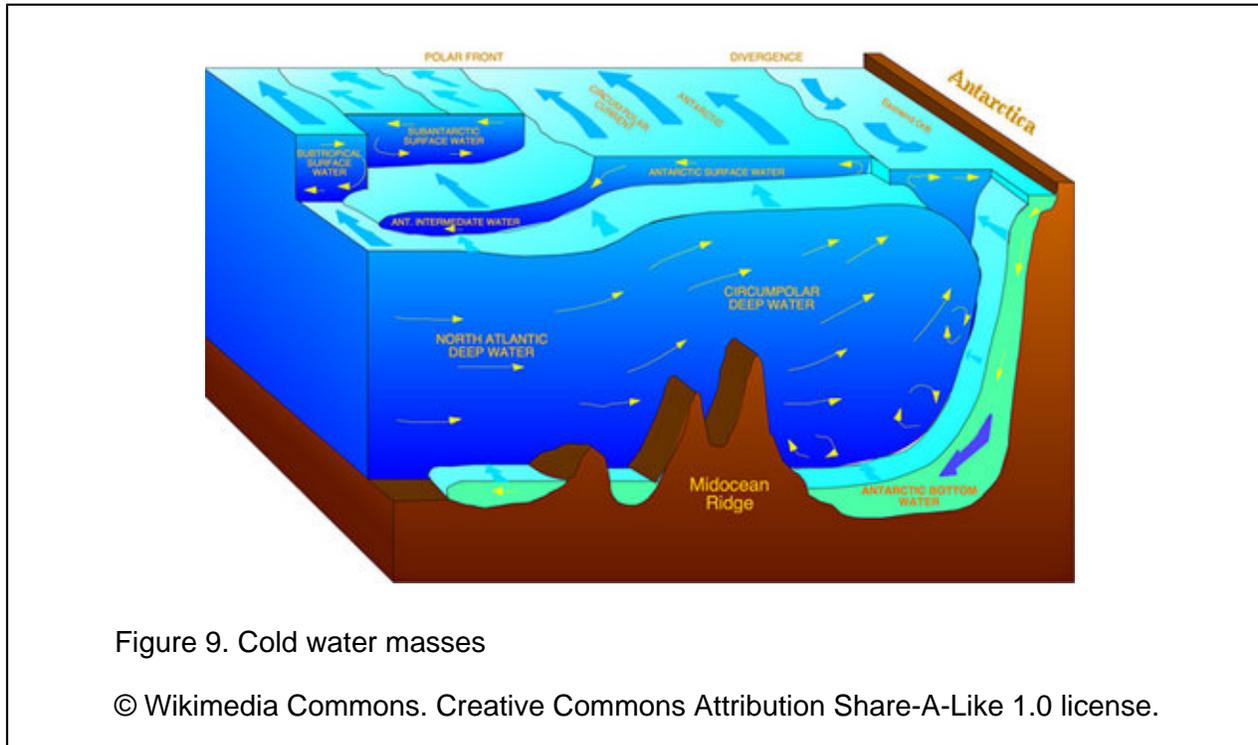


Figure 9. Cold water masses

© Wikimedia Commons. Creative Commons Attribution Share-A-Like 1.0 license.

This circulation pattern is not constant or permanent. Studies of Earth's climate history have linked changes in the strength of the thermohaline circulation to broader climate changes. At times when the conveyor belt slowed down, temperatures in the Northern Hemisphere fell; when the circulation intensified, temperatures in the region rose. Current analyses suggest that as the oceans warm in response to global climate change, the Thermohaline circulation could weaken again, possibly shutting down completely in extreme scenarios. (For more details, see Unit 13, "Looking Forward: Our Global Experiment.")

## 5. Ocean Circulation and Climate Cycles

Ocean circulation patterns are important contributing factors in many well-known and predictable climate variations. Weather patterns in a given region of the Earth can be affected at any given time by one or more of these climate cycles, depending on the areas that they cover and whether they overlap in time. Measuring the variables that signal switches in climate cycles, such as changes in ocean temperature and atmospheric pressure, is an important research focus for ocean and atmospheric scientists who are working to make better predictions of climate and weather cycles.

Monsoons are a well-known example of a seasonal climate cycle. These storms arise when land and water heat up at sharply different rates. As land temperatures increase during summer months,

hot air masses rise over the land and create low-pressure zones. At the surface, ocean winds blow toward land carrying moist ocean air. When these winds flow over land and are lifted up by mountains, their moisture condenses and produces torrential rainfalls.

In winter this pattern reverses: land cools more quickly than water, so air rises over the warm oceans and draws winds from the continents, producing high pressure and clear weather over land. In Asia, winter monsoons occur from December through March and summer monsoons take place from June through September (Fig. 10). Summer monsoons also occur from July through September in Mexico and the southwestern United States.



Figure 10. Monsoon rain clouds near Nagercoil, India, August 2006

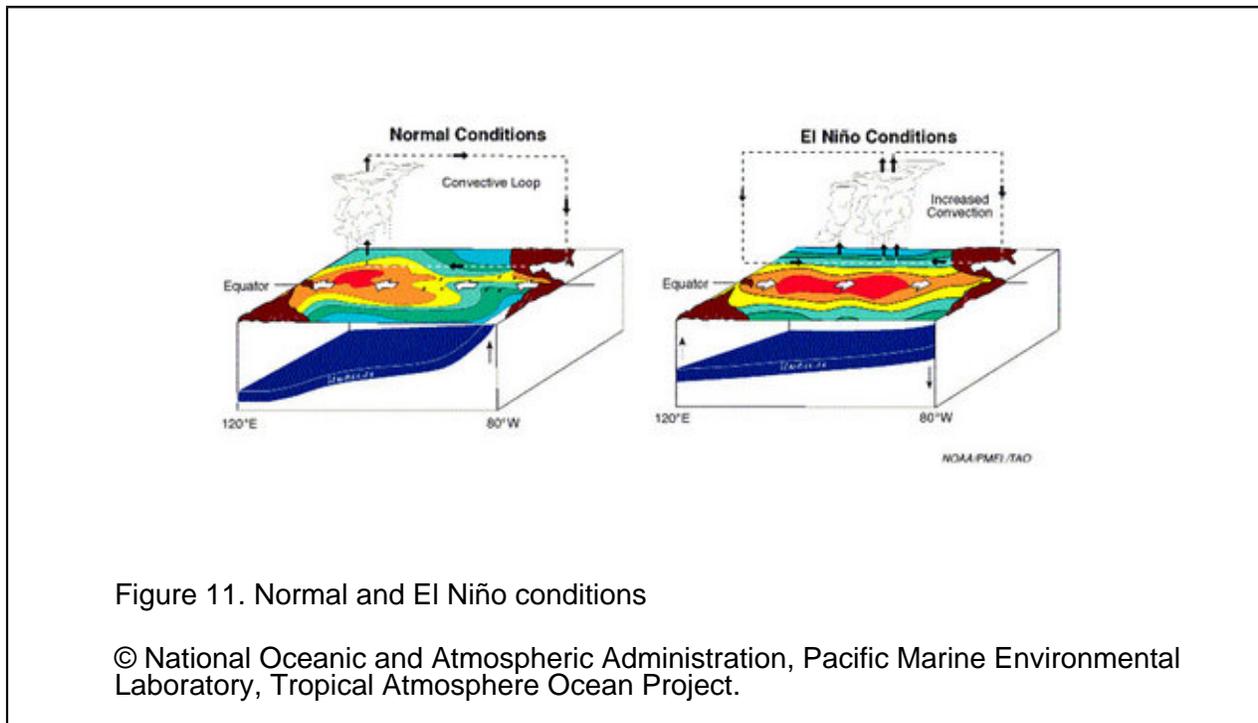
© Wikimedia Commons. Creative Commons Attribution Share-A-Like 1.0 license.

Hurricanes develop on an annual cycle generated by atmospheric and ocean conditions that occur from June through November in the Atlantic and from May through November in the eastern Pacific. The main requirements for hurricanes to develop are warm ocean waters (at least 26.5°C/80°F), plenty of atmospheric moisture, and weak easterly trade winds (for more details, see Unit 2, "Atmosphere"). A typical North Atlantic hurricane season produces eleven named storms, of which six become hurricanes, including two major hurricanes (Category 3 strength or higher).

The best-known multi-annual climate cycle is the [El Niño Southern Oscillation \(ENSO\)](#), which is caused every three to seven years by changes in atmospheric and ocean conditions over the Pacific Ocean. Normally, as shown in Fig. 5 above, a large gyre sits over the southern Pacific. When this gyre is strong, winds along the coast of South America and trade winds that blow west along the Equator are strong. These winds pile warm surface waters up in the western Pacific and produce

coastal upwelling of cold, nutrient-rich water that supports major fisheries near the coast of South America (Fig. 11, left image). This pattern is called the **Walker Circulation**, named for its discoverer, British scientist Sir Gilbert Walker.

In an El Niño year, the gyre weakens and this pattern flips into reverse. Atmospheric pressure rises over Asia and falls over South America, equatorial trade winds weaken, and warm water moves eastward toward South and Central America and California (Fig. 11, right image). Coastal upwelling in the eastern Pacific dwindles or stops. Warm, moist air rises over the west coasts of North and South America, causing heavy rains and landslides as droughts occur in Indonesia and other Asian countries. The event was originally named El Niño (meaning "the Christ Child") by South American fishermen because coastal waters often began warming around Christmas. Major El Niño events occurred in 1982–1983 and 1997–1998: each cycle killed several thousand people (mostly through flooding), triggered major forest fires in Asia, and caused millions of dollars' worth of property damage.



During a La Niña event (an unusually intensive reverse phase of El Niño), Pacific water temperatures become unusually cold and cold-water upwelling increases along western South America. La Niña episodes produce unusually wet weather in Asia and drier than normal conditions in much of the United States. A full El Niño/La Niña cycle lasts for about four years.

Other ocean/atmosphere climate cycles happen on even longer time frames. The **Pacific Decadal Oscillation (PDO)** is a 20- to 30-year cycle in the North Pacific Ocean. Positive PDO indices (warm phases) are characterized by warm Sea Surface Temperature (SST) anomalies along the Pacific coast and cool SST anomalies in the central North Pacific. Negative PDO indices (cold phases) correspond to the opposite anomalies along the coast and offshore. They intensify the Aleutian Low pressure cell, a semi-permanent atmospheric pressure cell that settles over the North Pacific from late fall to spring.

Cool PDO phases are well correlated with cooler and wetter than average weather in the western United States. During the warm phase of the PDO, the western Pacific cools and the eastern Pacific warms, producing weather that is slightly warmer and drier than normal in the western states. Since the North American climate anomalies associated with PDO extremes are similar to those associated with the ENSO cycle, the PDO can be seen as a long-lived El Niño-like pattern in Pacific climate variability.

The **North Atlantic Oscillation (NAO)**, another multi-decadal cycle, refers to a north-south oscillation in the intensities of a low-pressure region south of Iceland and a high-pressure region near the Azores. Positive NAO indices mark periods when the differences in Sea Level Pressures (SLP) are greatest between these two regions. Under these conditions, the westerly winds that pass from North America between the high and low pressure regions and on to Europe are unusually strong, and the strength of Northeast Trade Winds is also enhanced. This strong pressure differential produces warm, mild winters in the eastern United States and warm, wet winters in Europe as storms crossing the Atlantic are steered on a northerly path.

In the negative phase, pressure weakens in the subtropics, so winter storms cross the Atlantic on a more direct route from west to east. Both the eastern United States and Europe experience colder winters, but temperatures are milder in Greenland because less cold air reaches its latitude.

## 6. Biological Activity in the Upper Ocean

Most living organisms in the ocean, including most that live in the deepest waters, are nourished directly or indirectly by plankton that live at the surface. Marine biologists are still discovering new life forms of all shapes and sizes in the sea, but some of the most important findings in recent decades concern the ocean's tiniest life forms: microbes that produce organic matter (mostly through photosynthesis) and form the base of ocean food chains.

Measured in caloric output per year, the oceans account for nearly half of the world's gross primary production. Nearly all of it takes place in a thin upper layer of the oceans and is done by **phytoplankton**—microorganisms that typically measure one-twentieth of a millimeter across at the largest and live for one to five days (Fig. 12). But what phytoplankton lack in size, they make up in numbers. Two types of cyanobacteria (blue-green algae), **Synechococcus** and **Prochlorococcus**,

may be the two most abundant organisms on Earth and are found in concentrations of up to 500,000 cells per milliliter of sea water (footnote 2).



Figure 12. Marine phytoplankton

© United States Geological Survey.

Since average temperatures are much less variable in the ocean than on land, temperature is less of a limiting factor for primary production in the oceans. However, light is key. As sunlight penetrates down through the ocean, it is absorbed or scattered by water and by particles floating in the water. The **compensation depth**, where net energy produced from photosynthesis equals the energy that producers use for respiration, occurs at a depth where light is reduced to about 1 percent of its strength at the surface. In clear water this may be as deep as 110 meters, but in turbid coastal water it may be less than 20 meters. Activities that make water more murky, such as dredging or water pollution, thus are likely to reduce biological productivity in the areas where they occur.

Mixing also affects how much light phytoplankton receive. If winds stir the ocean so that the mixed layer extends far below the **compensation zone**, phytoplankton will be pushed down levels where there is not enough light for photosynthesis, so their net production will be lower than if they drift constantly in well-lit water. In areas of the ocean where wind speeds vary widely from season to season, primary productivity may rise and fall accordingly.

In addition to light, phytoplankton need nutrients for photosynthesis. Carbon is available in the form of  $\text{CO}_2$  from the atmosphere. Other important macronutrients including nitrogen (N), phosphorus (P), and silicon (Si), and micronutrients such as iron (Fe), are dissolved in seawater. On average, phytoplankton use nutrients in a ratio of 106 C: 16N: 1P: .001 Fe, so if one of these elements runs out

in the mixed layer, productivity goes down. Ocean mixing and coastal upwelling bring new supplies of nutrients up from deeper waters.

When optimal light, temperature, and nutrient conditions occur, and limited numbers of grazers (larger organisms that feed on phytoplankton) are present, plankton population explosions called **blooms** occur. Grazers reproduce more slowly than phytoplankton, so it takes time for them to catch up to their food supply. Major blooms can color large stretches of ocean water red, brown, or yellow-green depending on what species is present (hence the popular term "red tide," although tides do not cause blooms). Many of these events are not harmful in themselves, but they deplete oxygen in the water when the organisms die and decompose. Some types of phytoplankton algae produce neurotoxins, so blooms of these varieties are dangerous to swimmers and consumers of fish or shellfish from the affected area.

Blooms can be triggered by runoff that carries fertilizer or chemicals into ocean waters or by storms that mix ocean waters and bring nutrients to the surface. They often occur in spring, when rising water temperatures and longer daylight hours stimulate phytoplankton to increase their activity levels after slow or dormant periods during winter. Most plankton blooms are beneficial to ocean life because they increase the availability of organic material, much like the flowering that takes place on land in spring.

A massive spring bloom occurs each year in the North Atlantic from March through June, raising ocean productivity levels from North Carolina to Canada. This bloom extends all the way north to the southern edge of Arctic sea ice and moves northward through the spring and summer as ice melts. Figure 13 shows a portion of the bloom.

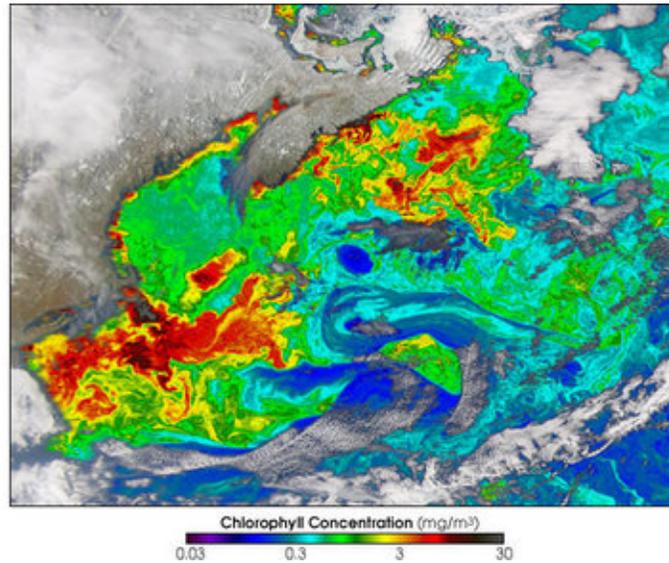


Figure 13. North Atlantic spring bloom, March 28, 2003 (Cape Cod to Newfoundland shown)

© National Aeronautics and Space Administration. Earth Observatory.

Phytoplankton are eaten by animal microorganisms such as [zooplankton](#), which range from single-celled creatures like [foraminifera](#) (small amoebae with calcium carbonate shells) to larger organisms such as krill and jellyfish. Copepods, the most abundant type of small crustaceans, are one to two millimeters long and serve as important food sources for fish, whales, and seabirds. Phytoplankton blooms attract zooplankton, which in turn are eaten by larger predators. "Phytoplankton are the plants of the ocean and the base of the food web," says MIT biologist Penny Chisholm. "If they weren't there, there would be nothing else living in the oceans."

Climate cycles can have major impacts on biological productivity in the oceans. When Pacific coastal upwelling off South America slows or stops during an El Niño event, plankton growth falls, reducing food supplies for anchovies and sardines that prey on the plankton. These fish die off or move to colder waters, which in turn reduces food for large predators like tuna, sea lions, and seabirds. Major El Niño events in the 20th century devastated the South American fishing industry and killed thousands of predator species as far north as the Bering Straits.

Other climate cycles have similar impacts. Generally, warm phases of the Pacific Decadal Oscillation decrease productivity off of the western United States and cold phases increase it. Decadal fluctuations in salmon, several ground fish, albacore, seabirds, and marine mammals in the North Pacific have been associated with the PDO. The mechanisms underlying these associations are

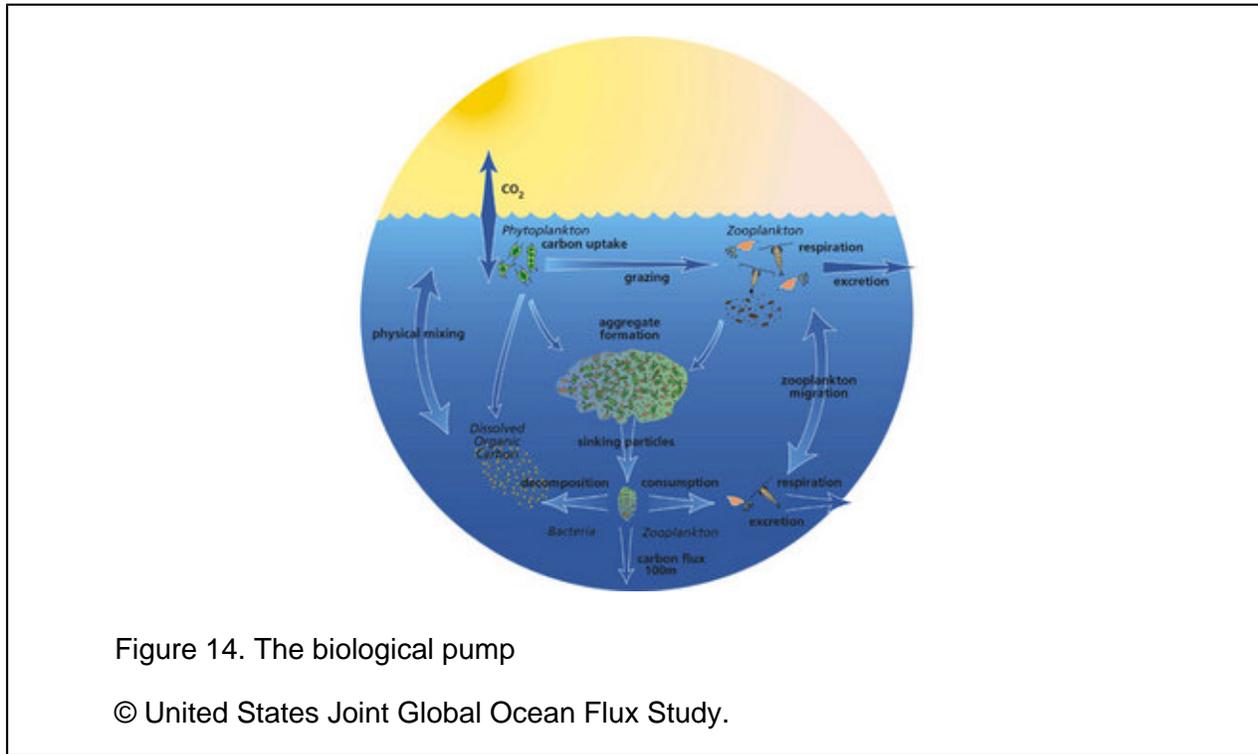
speculative, but probably represent a combined effect of atmospheric conditions such as wind strength; upper ocean physical conditions, such as current strength, depth of wind mixing, and nutrient availability; and biological responses to these conditions across many trophic levels. Similarly, successful [recruitment](#) of several species of fish, including sardines and cod, has been associated with different phases of the NAO in different regions of the North Atlantic.

## 7. The "Biological Pump"

Deep waters provide nutrients that plankton need for primary production in the upper ocean, but how do these nutrients get to the ocean depths? They are carried down from the surface in a rain of particles often referred to as [marine snow](#), which includes fecal pellets from zooplankton, shells from dead plankton, and other bits of organic material from dead or dying microorganisms. Many of these bits are almost invisible and too light to sink (a fecal pellet is about the size of a typeset comma), but when they stick together and form clumps called aggregates, the clumps eventually become heavier. Bacteria, plankton, and larger creatures feed on the pellets as they sink, then excrete wastes that add to the snowfall. Particles can travel to the deep ocean floor in spans of days or weeks.

When marine snow reaches deep waters, some is consumed by bottom-dwellers and microbes who depend on it as a food source since they live below the zones where light penetrates. Some is oxidized, releasing CO<sub>2</sub>, nitrate, and phosphate and recycling nutrients into deep waters. The remainder is buried in sediments; organic-rich carbon buried thousands of years ago in ancient inland seas and coastal margins is the source of today's offshore oil and gas deposits.

In addition to recycling nutrients and providing food for deep-dwelling species, this flow of particles to ocean depths is a critical link in the global carbon cycle, dubbed the [biological pump](#) by oceanographer and climate scientist Roger Revelle. Plankton take up carbon from the atmosphere in two ways: they fix CO<sub>2</sub> as organic carbon during photosynthesis and form shells from calcium carbonate. Marine snow carries both of these forms of carbon away from the atmosphere and surface waters to reservoirs in the deep oceans and ocean sediments, where it remains stored for centuries (Fig. 14).



Without this mechanism, concentrations of  $CO_2$  in the atmosphere would be substantially higher—and since atmospheric  $CO_2$  traps heat, Earth's surface temperature would be significantly higher. (For more details on the greenhouse effect, see Unit 2, "Atmosphere.") "If there were no phytoplankton—if the biological pump did not exist and the oceans all mixed from top to bottom and all that  $CO_2$  in the deep oceans equilibrated with the atmosphere—the concentration of  $CO_2$  in the atmosphere would more than double," says MIT's Chisholm. "Phytoplankton keep that pump pumping downward."

The overall efficiency of the biological pump depends on a combination of physical and biogeochemical factors. Both light and nutrients must be available in sufficient quantities for plankton to package more energy than they consume. Enough particles must sink to recycle nutrients into deep waters, and upwelling must occur to bring nutrients back to the surface. Factors that can impede this process include ocean warming (which makes the sea's layers more stratified, preventing mixing that brings up nutrients), and pollution and turbulence, which can reduce the penetration of sunlight at the surface.

## 8. Further Reading

National Academy of Sciences, Office on Public Understanding of Science, "El Niño and La Niña: Tracing the Dance of Ocean and Atmosphere," March 2000, <http://www7.nationalacademies.org/>



[opus/el\\_nino.html](#). A summary showing how atmospheric and oceanographic research have improved our capability to predict climate fluctuations.

National Weather Service Climate Prediction Center, <http://www.cpc.ncep.noaa.gov/>. Assessments, weather forecasts, graphics, and information on climate cycles including El Niño/La Niña, NAO, PDO, and others.

Woods Hole Oceanographic Institute, "Ocean Instruments: How they work, what they do, and why they do it," <http://www.whoi.edu/science/instruments/>. A guide to gravity corers, seismometers, and other ocean research tools.

## Footnotes

1. U.S. Commission on Ocean Policy, **An Ocean Blueprint for the 21st Century**, Final Report (Washington, DC, 2004), [http://www.oceancommission.gov/documents/full\\_color\\_rpt/welcome.html](http://www.oceancommission.gov/documents/full_color_rpt/welcome.html), pp. 312–29; National Research Council, **Exploration of the Seas: Voyage Into the Unknown** (Washington, DC: National Academies Press, 2003).
2. John Waterbury, "Little Things Matter A Lot," **Oceanus**, March 11, 2005.

## Glossary

**biological pump** : The sum of a suite of biologically-mediated processes that transport carbon from the surface euphotic zone (the depth of the water that is exposed to sufficient sunlight for photosynthesis to occur) to the ocean's interior.

**blooms** : A relatively rapid increase in the population of (usually) phytoplankton algae in an aquatic system. Algal blooms may occur in freshwater or marine environments.

**buoyancy** : The upward force on an object produced by the surrounding fluid (i.e., a liquid or a gas) in which it is fully or partially immersed, due to the pressure difference of the fluid between the top and bottom of the object.

**compensation depth** : Depth at which light intensity reaches a level at which oxygen evolved from a photosynthesizing organism equals that consumed by its respiration.

**compensation zone** : The point at which there is just enough light for a plant to survive. At this point all the food produced by photosynthesis is used up by respiration. For aquatic plants, the compensation point is the depth of water at which there is just enough light to sustain life (deeper water = less light = less photosynthesis).

**Coriolis force** : The apparent force, resulting from the rotation of the Earth, that deflects air or water movement.

**Ekman transport** : The change in wind direction with altitude caused by the varying effect of surface friction.

**El Niño Southern Oscillation (ENSO)** : A global event arising from large-scale interactions between the ocean and the atmosphere, usually an oscillation in the surface pressure (atmospheric mass) between the southeastern tropical Pacific and the Australian-Indonesian regions.

**foraminifera** : A large group of amoeboid protists with reticulating pseudopods—fine strands of cytoplasm that branch and merge to form a dynamic net. They typically produce a test, or shell, which can have either one or multiple chambers, some becoming quite elaborate in structure.

**gyre** : A circular or spiral motion, especially a circular ocean current.

**Hadley circulation** : A general circulation pattern in which air rises near the equator, flows north and south away from the equator at high altitudes, sinks near the poles, and flows back along the surface from both poles to the equator.

**marine snow** : The tiny leftovers of animals, plants, and non-living matter in the ocean's sun-suffused upper zones. Among these particles are chains of single-celled plants called diatoms, shreds of zooplankters' mucous food traps, soot, fecal pellets, dust motes, radioactive fallout, sand grains, pollen, and pollutants. Microorganisms also live inside and on top of these odd-shaped flakes.

**North Atlantic Oscillation (NAO)** : A major disturbance of the atmospheric circulation and climate of the North Atlantic-European region, linked to a waxing and waning of the dominant middle-latitude westerly wind flow during winter. The NAO Index is based on the pressure difference between various stations to the north (Iceland) and south (Azores) of the middle latitude westerly flow. It is, therefore, a measure of the strength of these winds.

**Pacific Decadal Oscillation (PDO)** : A pattern of Pacific climate variability that shifts phases on at least inter-decadal time scale, usually about 20 to 30 years. The PDO is detected as warm or cool surface waters in the Pacific Ocean, north of lat. 20° N. During a "warm" or "positive" phase, the west Pacific becomes cool and part of the eastern ocean warms; during a "cool" or "negative" phase, the opposite pattern occurs.

**phytoplankton** : Microscopic plants that live in the water column of oceans, seas, and bodies of fresh water and are the foundation of the marine food chain.

**specific heat capacity** : The amount of heat, measured in calories, required to raise the temperature of one gram of a substance by one Celsius degree.

**thermocline** : A layer within a body of water or air where the temperature changes rapidly with depth. The thermocline varies with latitude and season: it is permanent in the tropics, variable in the temperate climates (strongest during the summer), and weak to nonexistent in the polar regions, where the water column is cold from the surface to the bottom.

**thermohaline circulation** : The global density-driven circulation of the oceans.

# The Habitable Planet

**upwelling** : An oceanographic phenomenon that involves wind-driven motion of dense, cooler, and usually nutrient-rich water towards the ocean surface, replacing the warmer, usually nutrient-deplete surface water.

**Walker circulation** : An atmospheric circulation of air at the equatorial Pacific Ocean, responsible for creating ocean upwelling off the coasts of Peru and Ecuador. This brings nutrient-rich cold water to the surface, increasing fishing stocks.

**zooplankton** : Microscopic animals that live in the water column of oceans, seas, and bodies of fresh water. The smallest zooplankton can be characterized as recyclers of water-column nutrients and often are closely tied to measures of nutrient enrichment. Larger zooplankton are important food for forage fish species and larval stages of all fish.

## Unit 4 : Ecosystems



Elk in Yellowstone National Park.

### Overview

Why are there so many living organisms on Earth, and so many different species? How do the characteristics of the nonliving environment, such as soil quality and water salinity, help determine which organisms thrive in particular areas? These questions are central to the study of ecosystems—communities of living organisms in particular places and the chemical and physical factors that influence them. Learn how scientists study ecosystems to predict how they may change over time and respond to human impacts.

### Sections:

1. Introduction
2. Major Terrestrial and Aquatic Biomes
3. Energy Flow Through Ecosystems
4. Biogeochemical Cycling in Ecosystems
5. Population Dynamics
6. Regulation of Ecosystem Functions
7. Ecological Niches
8. Evolution and Natural Selection in Ecosystems
9. Natural Ecosystem Change
10. Further Reading

## 1. Introduction

Ecology is the scientific study of relationships in the natural world. It includes relationships between organisms and their physical environments (physiological ecology); between organisms of the same species (population ecology); between organisms of different species (community ecology); and between organisms and the fluxes of matter and energy through biological systems (ecosystem ecology).

Ecologists study these interactions in order to understand the abundance and diversity of life within Earth's ecosystems—in other words, why there are so many plants and animals, and why there are so many different types of plants and animals (Fig. 1). To answer these questions they may use field measurements, such as counting and observing the behavior of species in their habitats; laboratory experiments that analyze processes such as predation rates in controlled settings; or field experiments, such as testing how plants grow in their natural setting but with different levels of light, water, and other inputs. Applied ecology uses information about these relationships to address issues such as developing effective vaccination strategies, managing fisheries without over-harvesting, designing land and marine conservation reserves for threatened species, and modeling how natural ecosystems may respond to global climate change.



Figure 1. Tropical ecologist Stuart Davies in the field

Change is a constant process in ecosystems, driven by natural forces that include climate shifts, species movement, and ecological succession. By learning how ecosystems function, we can improve our ability to predict how they will respond to changes in the environment. But since living

organisms in ecosystems are connected in complex relationships, it is not always easy to anticipate how a step such as introducing a new species will affect the rest of an ecosystem.

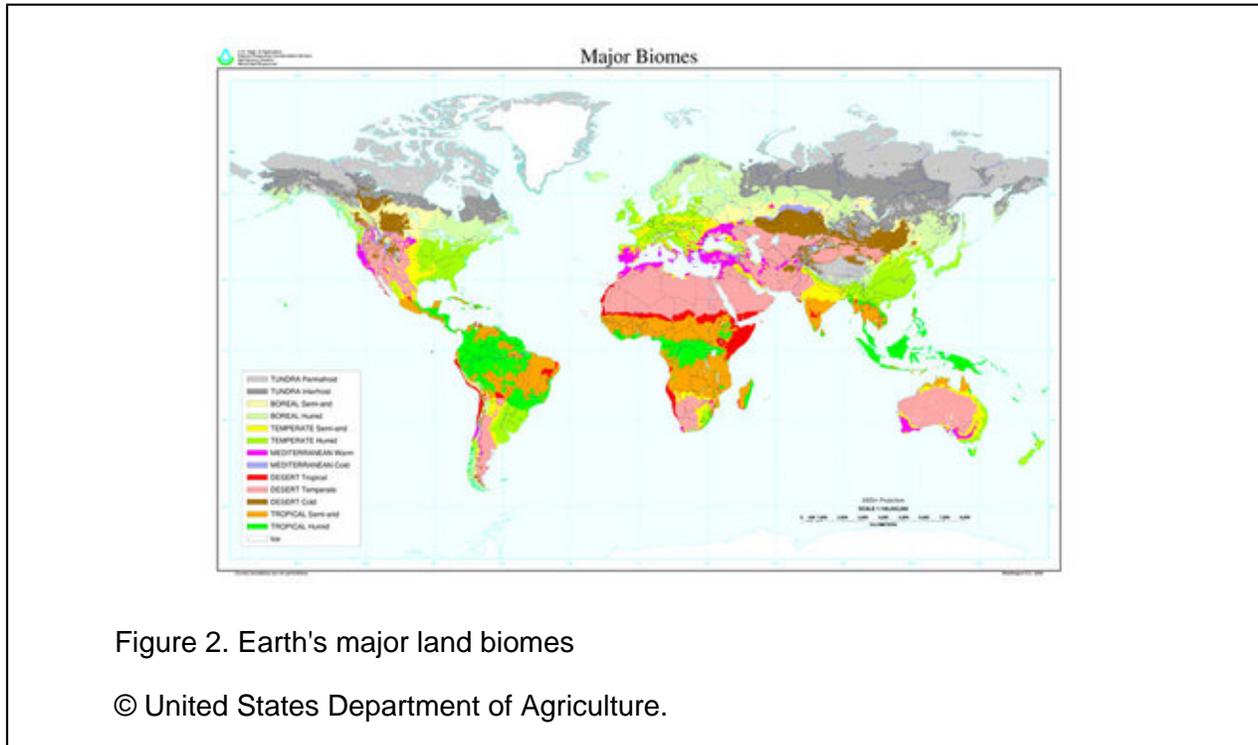
Human actions are also becoming major drivers of ecosystem change. Important human-induced stresses on ecosystems are treated in later units of this text. Specifically, Unit 7 ("Agriculture") examines how agriculture and forestry create artificial, simplified ecosystems; Unit 9 ("Biodiversity Decline") discusses the effects of habitat loss and the spread of invasive species; and Unit 12 ("Earth's Changing Climate") considers how climate change is affecting natural ecosystems.

## 2. Major Terrestrial and Aquatic Biomes

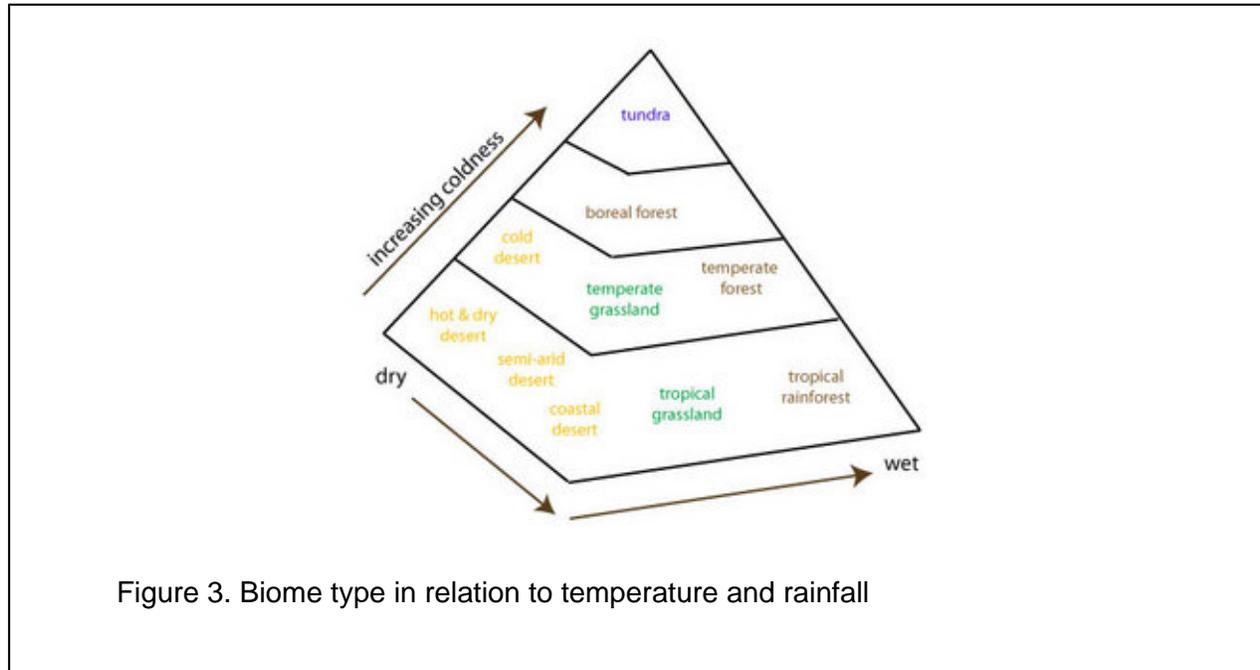
Geography has a profound impact on ecosystems because global circulation patterns and climate zones set basic physical conditions for the organisms that inhabit a given area. The most important factors are temperature ranges, moisture availability, light, and nutrient availability, which together determine what types of life are most likely to flourish in specific regions and what environmental challenges they will face.

As discussed in Unit 2, "Atmosphere," and Unit 3, "Oceans," Earth is divided into distinct climate zones that are created by global circulation patterns. The tropics are the warmest, wettest regions of the globe, while subtropical high-pressure zones create dry zones at about 30° latitude north and south. Temperatures and precipitation are lowest at the poles. These conditions create **biomes**—broad geographic zones whose plants and animals are adapted to different climate patterns. Since temperature and precipitation vary by latitude, Earth's major terrestrial biomes are broad zones that stretch around the globe (Fig. 2). Each biome contains many ecosystems (smaller communities) made up of organisms adapted for life in their specific settings.

# The Habitable Planet



Another way to visualize major land biomes is to compare them based on their average temperature ranges and rainfall levels, which shows how these variables combine to create a range of climates (Fig. 3).



Land biomes are typically named for their characteristic types of vegetation, which in turn influence what kinds of animals will live there. Soil characteristics also vary from one biome to another, depending on local climate and geology. compares some key characteristics of three of the forest biomes.

**Table 1.** Forest biomes.

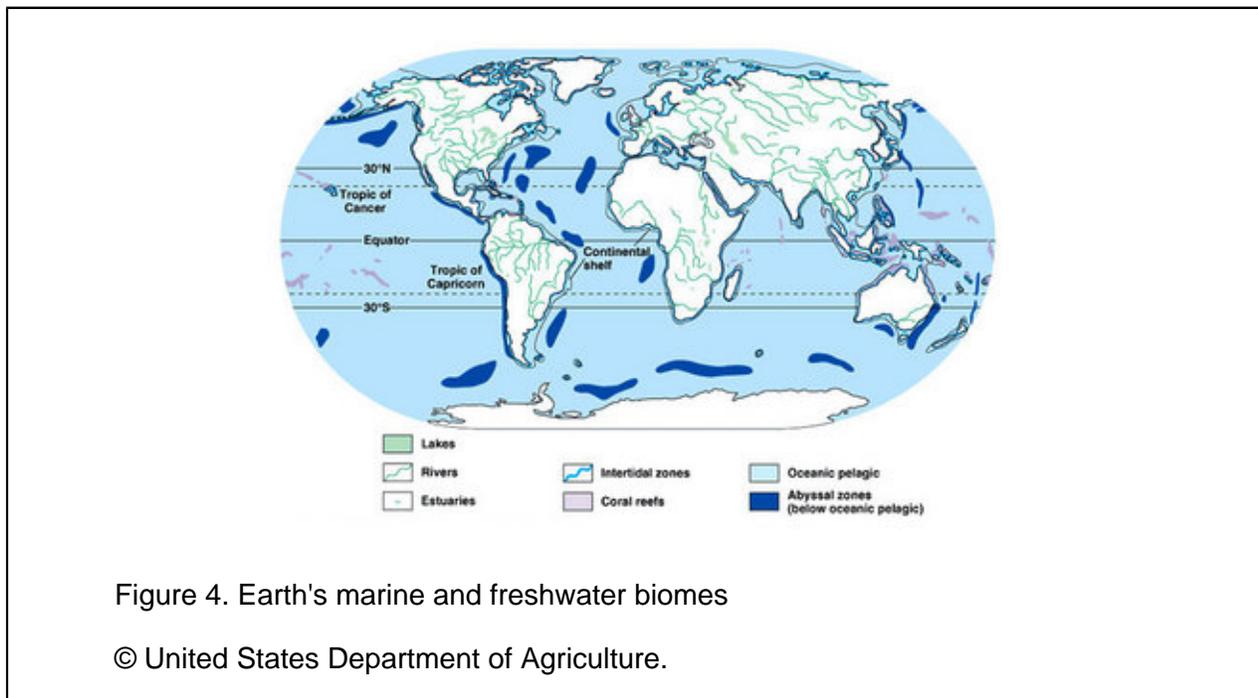
Forest type	Temperature	Precipitation	Soil	Flora
Tropical	20-25°C	>200 cm/yr	Acidic, low in nutrients	Diverse (up to 100 species/km <sup>2</sup> )
Temperate	-30 to 30°C	75-150 cm/yr	Fertile, high in nutrients	3-4 tree species/km <sup>2</sup>
Boreal (taiga)	Very low	40-100 cm/year, mostly snow	Thin, low in nutrients, acidic	Evergreens

Aquatic biomes (marine and freshwater) cover three-quarters of the Earth's surface and include rivers, lakes, coral reefs, estuaries, and open ocean (Fig. 4). Oceans account for almost all of this area. Large bodies of water (oceans and lakes) are stratified into layers: surface waters are warmest and contain most of the available light, but depend on mixing to bring up nutrients from deeper levels

(for more details, see Unit 3, "Oceans"). The distribution of temperature, light, and nutrients set broad conditions for life in aquatic biomes in much the same way that climate and soils do for land biomes.

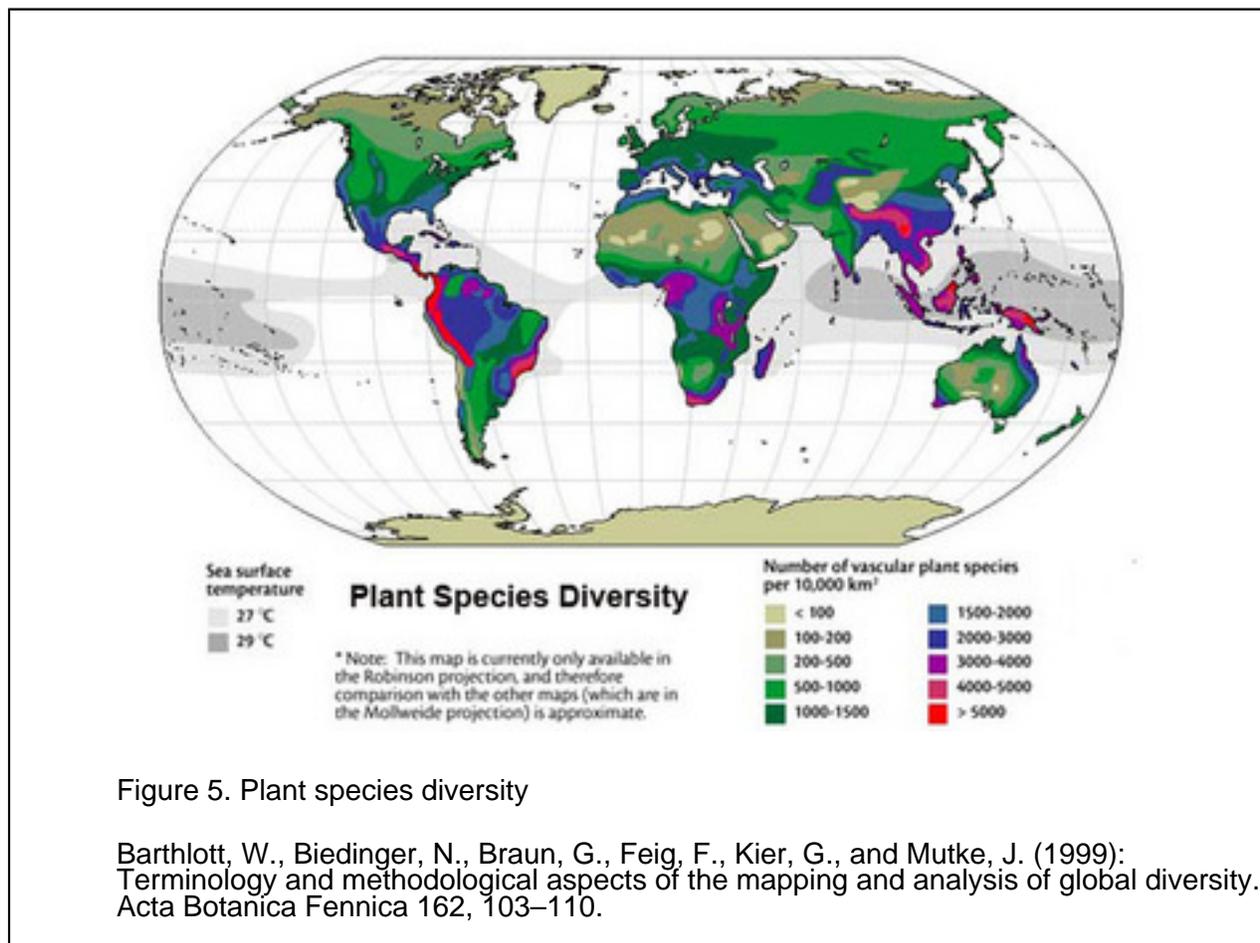
Marine and freshwater biomes change daily or seasonally. For example, in the intertidal zone where the oceans and land meet, areas are submerged and exposed as the tide moves in and out. During the winter months lakes and ponds can freeze over, and wetlands that are covered with water in late winter and spring can dry out during the summer months.

There are important differences between marine and freshwater biomes. The oceans occupy large continuous areas, while freshwater habitats vary in size from small ponds to lakes covering thousands of square kilometers. As a result, organisms that live in isolated and temporary freshwater environments must be adapted to a wide range of conditions and able to disperse between habitats when their conditions change or disappear.



Since biomes represent consistent sets of conditions for life, they will support similar kinds of organisms wherever they exist, although the species in the communities in different places may not be taxonomically related. For example, large areas of Africa, Australia, South America, and India are covered by savannas (grasslands with scattered trees). The various grasses, shrubs, and trees that grow on savannas all are generally adapted to hot climates with distinct rainy and dry seasons and periodic fires, although they may also have characteristics that make them well-suited to specific conditions in the areas where they appear.

Species are not uniformly spread among Earth's biomes. Tropical areas generally have more plant and animal biodiversity than high latitudes, measured in **species richness** (the total number of species present) (footnote 1). This pattern, known as the **latitudinal biodiversity gradient**, exists in marine, freshwater, and terrestrial ecosystems in both hemispheres. Figure 5 shows the gradient for plant species, but it also holds true for animals.



Why is biodiversity distributed in this way? Ecologists have proposed a number of explanations:

- Higher productivity in the tropics allows for more species;
- The tropics were not severely affected by glaciation and thus have had more time for species to develop and adapt;

- Environments are more stable and predictable in the tropics, with fairly constant temperatures and rainfall levels year-round;
- More predators and pathogens limit competition in the tropics, which allows more species to coexist; and
- Disturbances occur in the tropics at frequencies that promote high successional diversity.

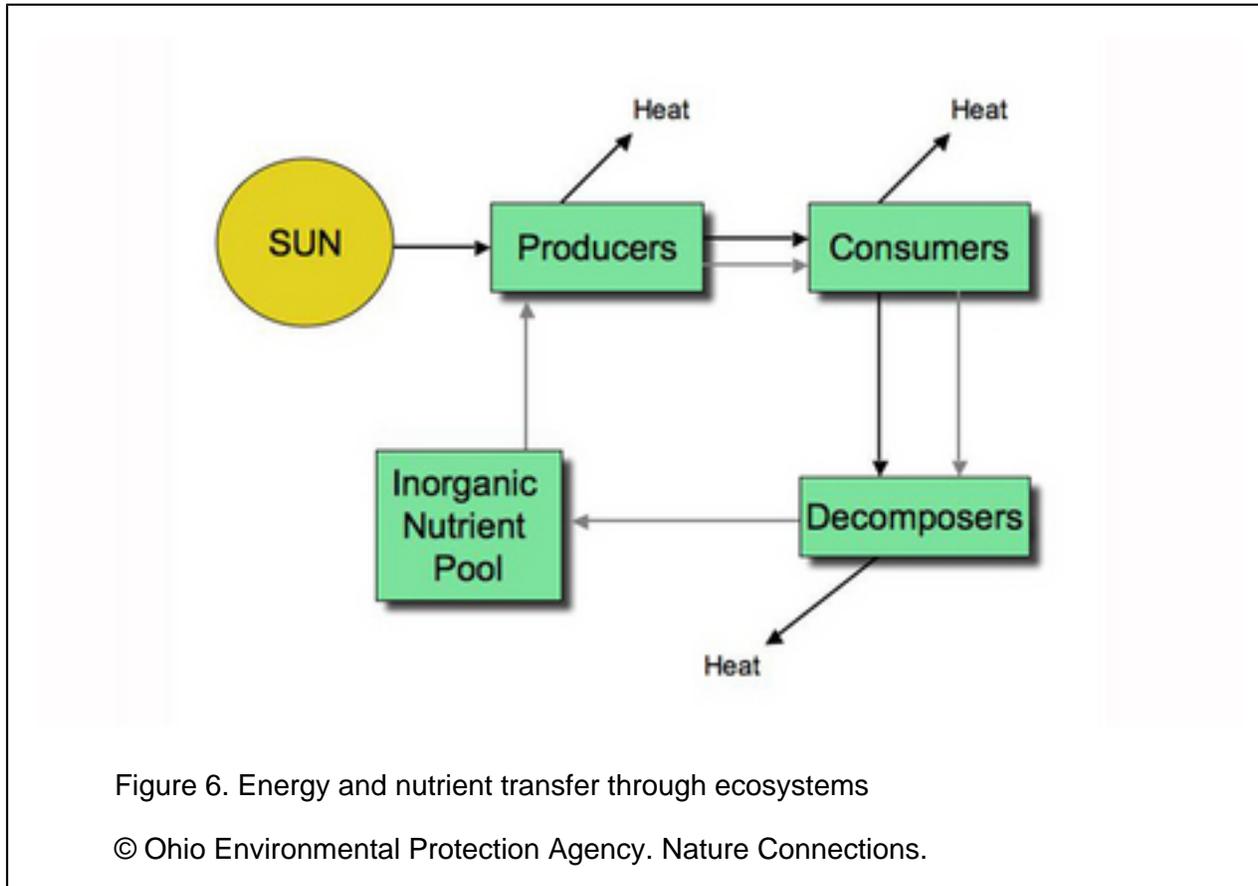
Of these hypotheses, evidence is strongest for the proposition that a stable, predictable environment over time tends to produce larger numbers of species. For example, both tropical ecosystems on land and deep sea marine ecosystems—which are subject to much less physical fluctuation than other marine ecosystems, such as estuaries—have high species diversity. Predators that seek out specific target species may also play a role in maintaining species richness in the tropics.

### 3. Energy Flow Through Ecosystems

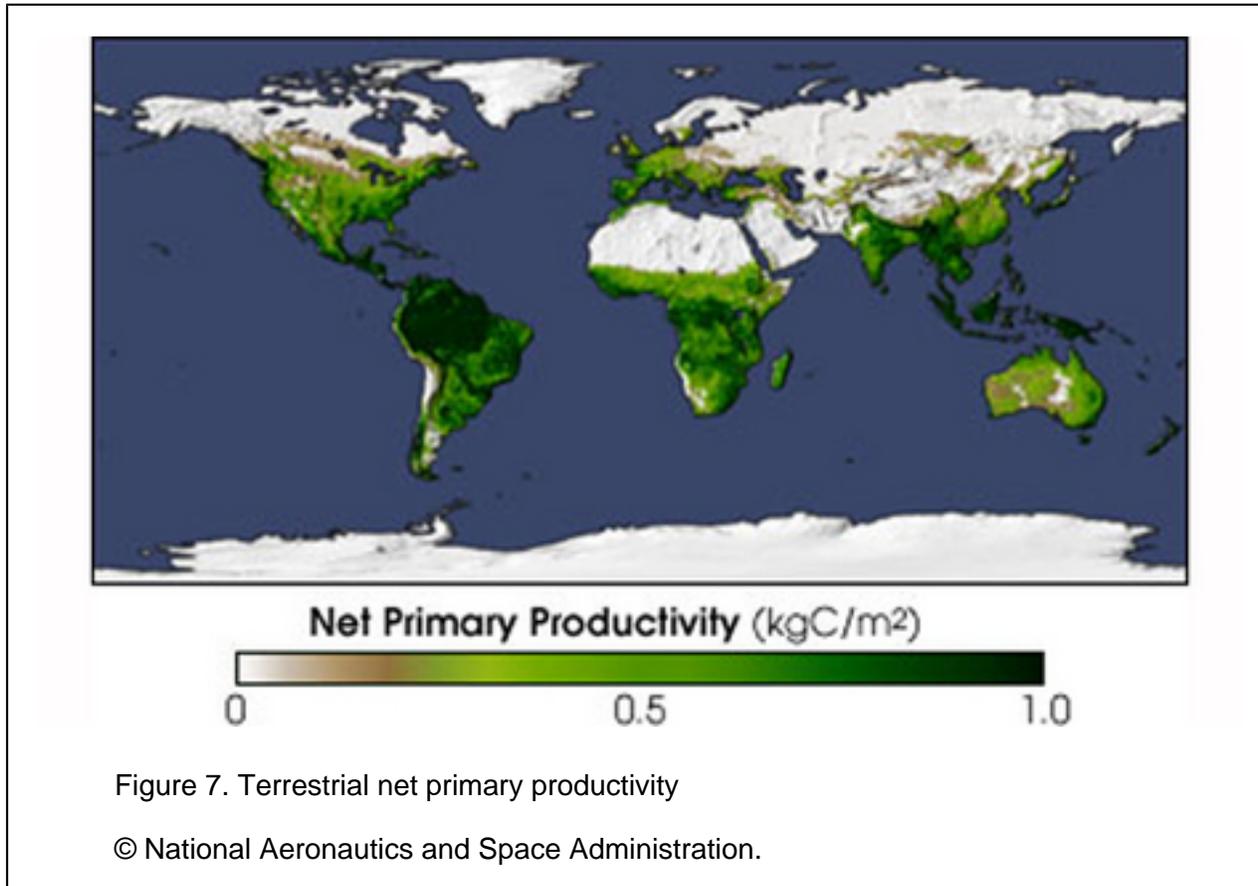
Ecosystems maintain themselves by cycling energy and nutrients obtained from external sources. At the first **trophic level**, **primary producers** (plants, algae, and some bacteria) use solar energy to produce organic plant material through photosynthesis. Herbivores—animals that feed solely on plants—make up the second trophic level. Predators that eat herbivores comprise the third trophic level; if larger predators are present, they represent still higher trophic levels. Organisms that feed at several trophic levels (for example, grizzly bears that eat berries and salmon) are classified at the highest of the trophic levels at which they feed. Decomposers, which include bacteria, fungi, molds, worms, and insects, break down wastes and dead organisms and return nutrients to the soil.

On average about 10 percent of net energy production at one trophic level is passed on to the next level. Processes that reduce the energy transferred between trophic levels include respiration, growth and reproduction, defecation, and nonpredatory death (organisms that die but are not eaten by consumers). The nutritional quality of material that is consumed also influences how efficiently energy is transferred, because consumers can convert high-quality food sources into new living tissue more efficiently than low-quality food sources.

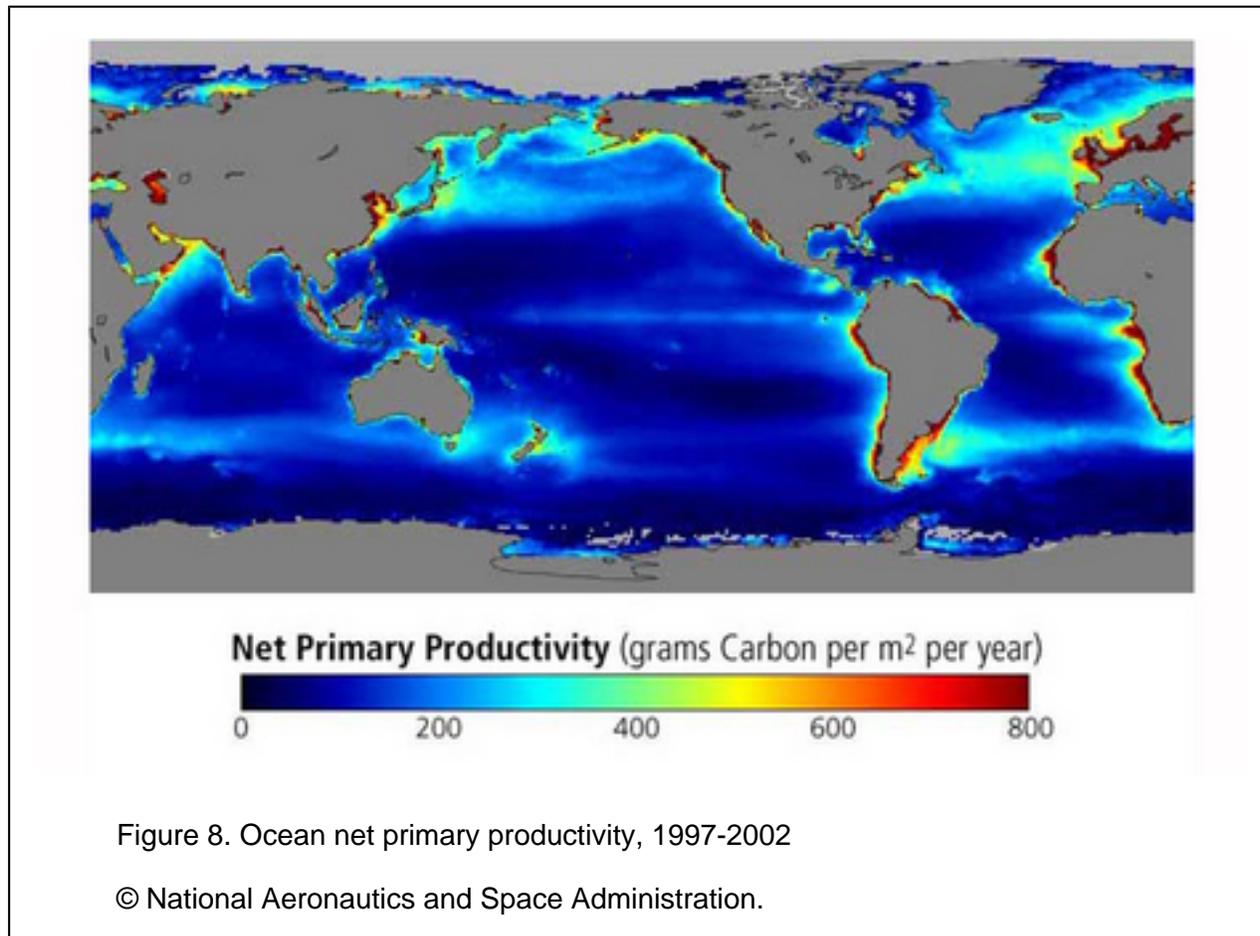
The low rate of energy transfer between trophic levels makes decomposers generally more important than producers in terms of energy flow. Decomposers process large amounts of organic material and return nutrients to the ecosystem in inorganic form, which are then taken up again by primary producers. Energy is not recycled during decomposition, but rather is released, mostly as heat (this is what makes compost piles and fresh garden mulch warm). Figure 6 shows the flow of energy (dark arrows) and nutrients (light arrows) through ecosystems.



An ecosystem's **gross primary productivity (GPP)** is the total amount of organic matter that it produces through photosynthesis. **Net primary productivity (NPP)** describes the amount of energy that remains available for plant growth after subtracting the fraction that plants use for respiration. Productivity in land ecosystems generally rises with temperature up to about 30°C, after which it declines, and is positively correlated with moisture. On land primary productivity thus is highest in warm, wet zones in the tropics where tropical forest biomes are located. In contrast, desert scrub ecosystems have the lowest productivity because their climates are extremely hot and dry (Fig. 7).



In the oceans, light and nutrients are important controlling factors for productivity. As noted in Unit 3, "Oceans," light penetrates only into the uppermost level of the oceans, so photosynthesis occurs in surface and near-surface waters. Marine primary productivity is high near coastlines and other areas where upwelling brings nutrients to the surface, promoting plankton blooms. Runoff from land is also a source of nutrients in estuaries and along the continental shelves. Among aquatic ecosystems, algal beds and coral reefs have the highest net primary production, while the lowest rates occur in the open due to a lack of nutrients in the illuminated surface layers (Fig. 8).



How many trophic levels can an ecosystem support? The answer depends on several factors, including the amount of energy entering the ecosystem, energy loss between trophic levels, and the form, structure, and physiology of organisms at each level. At higher trophic levels, predators generally are physically larger and are able to utilize a fraction of the energy that was produced at the level beneath them, so they have to forage over increasingly large areas to meet their caloric needs.

Because of these energy losses, most terrestrial ecosystems have no more than five trophic levels, and marine ecosystems generally have no more than seven. This difference between terrestrial and marine ecosystems is likely due to differences in the fundamental characteristics of land and marine primary organisms. In marine ecosystems, microscopic phytoplankton carry out most of the photosynthesis that occurs, while plants do most of this work on land. Phytoplankton are small organisms with extremely simple structures, so most of their primary production is consumed and used for energy by grazing organisms that feed on them. In contrast, a large fraction of the biomass

that land plants produce, such as roots, trunks, and branches, cannot be used by herbivores for food, so proportionately less of the energy fixed through primary production travels up the food chain.

Growth rates may also be a factor. Phytoplankton are extremely small but grow very rapidly, so they support large populations of herbivores even though there may be fewer algae than herbivores at any given moment. In contrast, land plants may take years to reach maturity, so an average carbon atom spends a longer residence time at the primary producer level on land than it does in a marine ecosystem. In addition, locomotion costs are generally higher for terrestrial organisms compared to those in aquatic environments.

The simplest way to describe the flux of energy through ecosystems is as a food chain in which energy passes from one trophic level to the next, without factoring in more complex relationships between individual species. Some very simple ecosystems may consist of a food chain with only a few trophic levels. For example, the ecosystem of the remote wind-swept Taylor Valley in Antarctica consists mainly of bacteria and algae that are eaten by nematode worms (footnote 2). More commonly, however, producers and consumers are connected in intricate food webs with some consumers feeding at several trophic levels (Fig. 9).

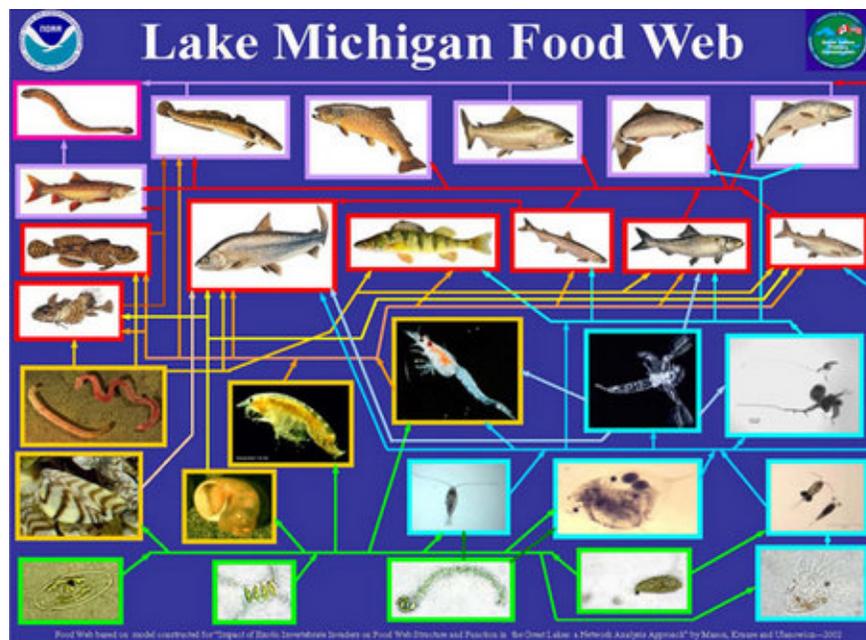


Figure 9. Lake Michigan food web

Courtesy of NOAA Great Lakes Environmental Research Laboratory and the Great Lakes Fishery Commission.

An important consequence of the loss of energy between trophic levels is that contaminants collect in animal tissues—a process called **bioaccumulation**. As contaminants bioaccumulate up the food web, organisms at higher trophic levels can be threatened even if the pollutant is introduced to the environment in very small quantities.

The insecticide DDT, which was widely used in the United States from the 1940s through the 1960s, is a famous case of bioaccumulation. DDT built up in eagles and other raptors to levels high enough to affect their reproduction, causing the birds to lay thin-shelled eggs that broke in their nests. Fortunately, populations have rebounded over several decades since the pesticide was banned in the United States. However, problems persist in some developing countries where toxic bioaccumulating pesticides are still used.

Bioaccumulation can threaten humans as well as animals. For example, in the United States many federal and state agencies currently warn consumers to avoid or limit their consumption of large predatory fish that contain high levels of mercury, such as shark, swordfish, tilefish, and king mackerel, to avoid risking neurological damage and birth defects.

#### 4. Biogeochemical Cycling in Ecosystems

Along with energy, water and several other chemical elements cycle through ecosystems and influence the rates at which organisms grow and reproduce. About 10 major nutrients and six trace nutrients are essential to all animals and plants, while others play important roles for selected species (footnote 3). The most important biogeochemical cycles affecting ecosystem health are the water, carbon, nitrogen, and phosphorus cycles.

As noted earlier, most of the Earth's area that is covered by water is ocean. In terms of volume, the oceans dominate further still: nearly all of Earth's water inventory is contained in the oceans (about 97 percent) or in ice caps and glaciers (about 2 percent), with the rest divided among groundwater, lakes, rivers, streams, soils, and the atmosphere. In addition, water moves very quickly through land ecosystems. These two factors mean that water's residence time in land ecosystems is generally short, on average one or two months as soil moisture, weeks or months in shallow groundwater, or up to six months as snow cover.

But land ecosystems process a lot of water: almost two-thirds of the water that falls on land as precipitation annually is transpired back into the atmosphere by plants, with the rest flowing into rivers and then to the oceans. Because cycling of water is central to the functioning of land ecosystems, changes that affect the hydrologic cycle are likely to have significant impacts on land ecosystems. (Global water cycling is discussed in more detail in Unit 8, "Water Resources.")

Both land and ocean ecosystems are important sinks for carbon, which is taken up by plants and algae during photosynthesis and fixed as plant tissue. Table 2 compares the quantities of carbon stored in Earth's major reservoirs.

**Table 2.** Global carbon storage.

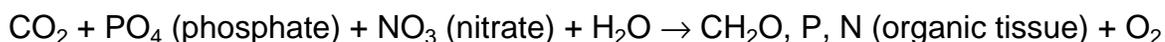
Location	Amount (gigatons carbon)
Atmosphere	750
Land plants	610
Soil and detritus	1,500
Surface ocean	1,020
Intermediate and deep ocean	37,890
Sediments	78,000,000

Carbon cycles relatively quickly through land and surface-ocean ecosystems, but may remain locked up in the deep oceans or in sediments for thousands of years. The average residence time that a molecule of carbon spends in a terrestrial ecosystem is about 17.5 years, although this varies widely depending on the type of ecosystem: carbon can be held in old-growth forests for hundreds of years, but its residence time in heavily grazed ecosystems where plants and soils are repeatedly turned over may be as short as a few months.

Human activities, particularly fossil fuel combustion, emit significant amounts of carbon each year over and above the natural carbon cycle. Currently, human activities generate about 7 billion tons of carbon per year, of which 3 billion tons remain in the atmosphere. The balance is taken up in roughly equal proportions by oceans and land ecosystems. Identifying which ecosystems are absorbing this extra carbon and why this uptake is occurring are pressing questions for ecologists.

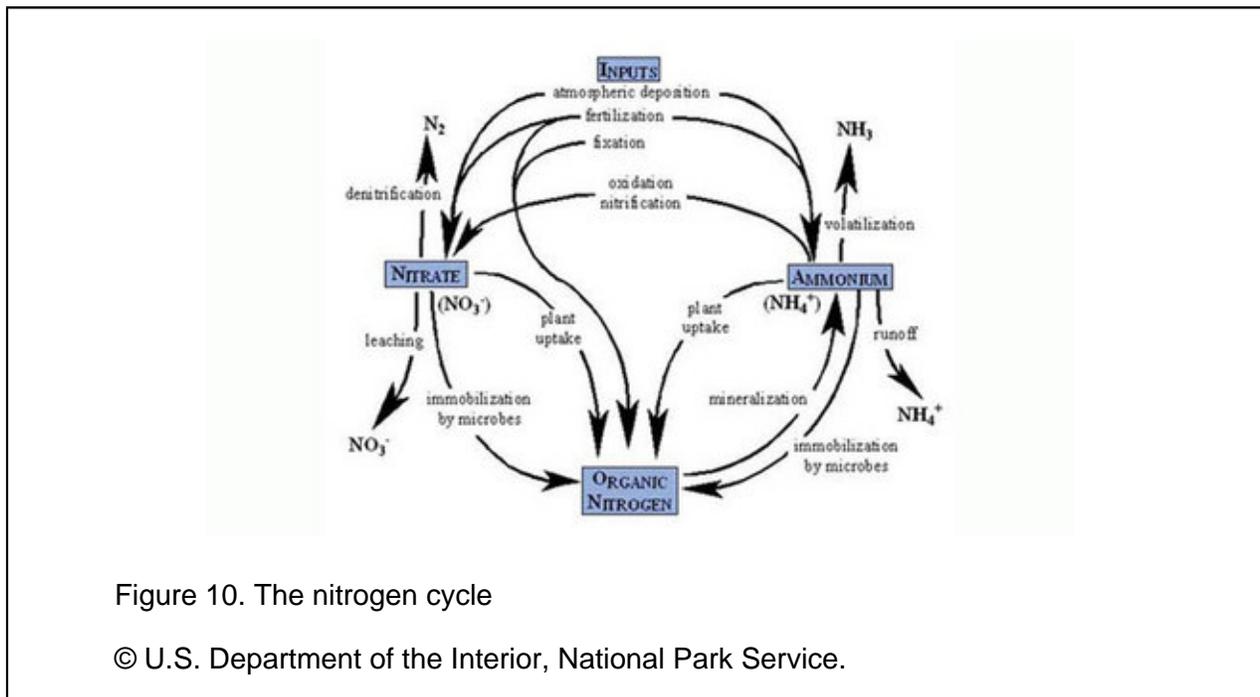
Currently, it is not clear what mechanisms are responsible for high absorption of carbon by land ecosystems. One hypothesis suggests that higher atmospheric CO<sub>2</sub> concentrations have increased the rates at which plants carry out photosynthesis (so-called [CO<sub>2</sub> fertilization](#)), but this idea is controversial. Controlled experiments have shown that elevated CO<sub>2</sub> levels are only likely to produce short-term increases in plant growth, because plants soon exhaust available supplies of important nutrients such as nitrogen and phosphorus that also are essential for growth.

Nitrogen and phosphorus are two of the most essential mineral nutrients for all types of ecosystems and often limit growth if they are not available in sufficient quantities. (This is why the basic ingredients in plant fertilizer are nitrogen, phosphorus, and potassium, commonly abbreviated as NPK.) A slightly expanded version of the basic equation for photosynthesis shows how plants use energy from the sun to turn nutrients and carbon into organic compounds:



Because atmospheric nitrogen (N<sub>2</sub>) is inert and cannot be used directly by most organisms, microorganisms that convert it into usable forms of nitrogen play central roles in the nitrogen cycle. So-called [nitrogen-fixing](#) bacteria and algae convert ammonia (NH<sub>4</sub>) in soils and surface waters into

nitrites ( $\text{NO}_2$ ) and nitrates ( $\text{NO}_3$ ), which in turn are taken up by plants. Some of these bacteria live in **mutualistic** relationships on the roots of plants, mainly legumes (peas and beans), and provide nitrate directly to the plants; farmers often plant these crops to restore nitrogen to depleted soils. At the back end of the cycle, decomposers break down dead organisms and wastes, converting organic materials to inorganic nutrients. Other bacteria carry out **denitrification**, breaking down nitrate to gain oxygen and returning gaseous nitrogen to the atmosphere (Fig. 10).



Human activities, including fossil fuel combustion, cultivation of nitrogen-fixing crops, and rising use of nitrogen fertilizer, are altering the natural nitrogen cycle. Together these activities add roughly as much nitrogen to terrestrial ecosystems each year as the amount fixed by natural processes; in other words, anthropogenic inputs are doubling annual nitrogen fixation in land ecosystems. The main effect of this extra nitrogen is over-fertilization of aquatic ecosystems. Excess nitrogen promotes algal blooms, which then deplete oxygen from the water when the algae die and decompose (for more details, see Unit 8, "Water Resources"). Additionally, airborne nitrogen emissions from fossil fuel combustion promote the formation of ground-level ozone, particulate emissions, and acid rain (for more details, see Unit 11, "Atmospheric Pollution").

Phosphorus, the other major plant nutrient, does not have a gaseous phase like carbon or nitrogen. As a result it cycles more slowly through the biosphere. Most phosphorus in soils occurs in forms that organisms cannot use directly, such as calcium and iron phosphate. Usable forms (mainly

orthophosphate, or  $\text{PO}_4$ ) are produced mainly by decomposition of organic material, with a small contribution from weathering of rocks (Fig. 11).

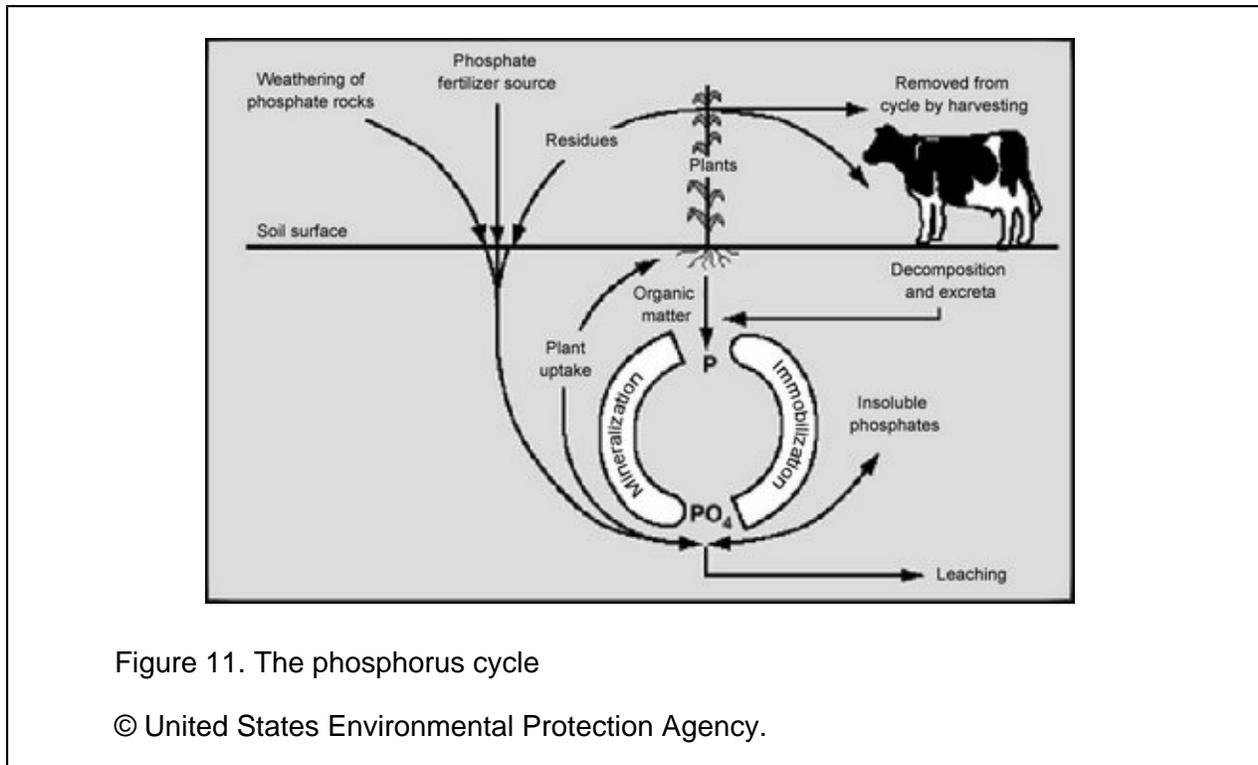


Figure 11. The phosphorus cycle

© United States Environmental Protection Agency.

The amount of phosphate available to plants depends on soil pH. At low pH, phosphorus binds tightly to clay particles and is transformed into relatively insoluble forms containing iron and aluminum. At high pH, it is lost to other inaccessible forms containing calcium. As a result, the highest concentrations of available phosphate occur at soil pH values between 6 and 7. Thus soil pH is an important factor affecting soil fertility.

Excessive phosphorus can also contribute to over-fertilization and eutrophication of rivers and lakes. Human activities that increase phosphorus concentrations in natural ecosystems include fertilizer use, discharges from wastewater treatment plants, and use of phosphate detergents (for details, see Unit 8, "Water Resources").

## 5. Population Dynamics

Every organism in an ecosystem divides its energy among three competing goals: growing, surviving, and reproducing. Ecologists refer to an organism's allocation of energy among these three ends throughout its lifetime as its **life history strategy**. There are tradeoffs between these functions: for

example, an organism that spends much of its energy on reproduction early in life will have lower growth and survival rates, and thus a lower reproductive level later in life. An optimal life history strategy maximizes the organism's contribution to population growth.

Understanding how the environment shapes organisms' life histories is a major question in ecology. Compare the conditions for survival in an unstable area, such as a flood plain near a river that frequently overflows its banks, to those in a stable environment, such as a remote old-growth forest. On the flood plain, there is a higher chance of being killed early in life, so the organisms that mature and reproduce earlier will be most likely to survive and add to population growth. Producing many offspring increases the chance that some will survive. Conversely, organisms in the forest will mature later and have lower early reproductive rates. This allows them to put more energy into growth and competition for resources.

Ecologists refer to organisms at the first of these two extremes (those adapted to unstable environments) as **r-selected**. These organisms live in settings where population levels are well below the maximum number that the environment can support—the **carrying capacity**—so their numbers are growing exponentially at the maximum rate at which that population can increase if resources are not limited (often abbreviated as *r*). The other extreme, organisms adapted to stable environments, are termed **K-selected** because they live in environments in which the number of individuals is at or near the environment's carrying capacity (often abbreviated as *K*).

Organisms that are r-selected tend to be small, short-lived, and opportunistic, and to grow through irregular boom-and-bust population cycles. They include many insects, annual plants, bacteria, and larger species such as frogs and rats. Species considered pests typically are r-selected organisms that are capable of rapid growth when environmental conditions are favorable. In contrast, K-selected species are typically larger, grow more slowly, have fewer offspring and spend more time parenting them. Examples include large mammals, birds, and long-lived plants such as redwood trees. K-selected species are more prone to extinction than r-selected species because they mature later in life and have fewer offspring with longer gestation times. Table 3 contrasts the reproductive characteristics of an r-selected mammal, the Norway rat, to those of a K-selected mammal, the African elephant.

**Table 3.** Reproduction in r-selected and K-selected species.

Feature	Norway rat (r-selected)	African elephant (K-selected)
Reaches sexual or reproductive maturity	3-4 months	10-12 years
Average gestation period	22-24 days	22 months
Time to weaning	3-4 weeks	48-108 months
Breeding interval (female)	Up to 7 times per year	Every 4 to 9 years

Feature	Norway rat (r-selected)	African elephant (K-selected)
Offspring per litter	2-14 (average 8)	1 average, 2 high

Many organisms fall between these two extremes and have some characteristics of both types. As we will see below, ecosystems tend to be dominated by r-selected species in their early stages with the balance gradually shifting toward K-selected species.

In a growing population, survival and reproduction rates will not stay constant over time. Eventually resource limitations will reduce one or both of these variables. Populations grow fastest when they are near zero and the species is uncrowded. A simple mathematical model of population growth implies that the maximum population growth rate occurs when the population size (N) is at one-half of the environment's carrying capacity, K (i.e., at  $N = K/2$ ).

In theory, if a population is harvested at exactly its natural rate of growth, the population will not change in size, and the harvest (yield) can be sustained at that level. In practice, however, it can be very hard to estimate population sizes and growth rates in the wild accurately enough to achieve this maximum sustainable yield. (For more on over-harvesting, see Unit 9, "Biodiversity Decline.")

## 6. Regulation of Ecosystem Functions

A key question for ecologists studying growth and productivity in ecosystems is which factors limit ecosystem activity. Availability of resources, such as light, water, and nutrients, is a key control on growth and reproduction. Some nutrients are used in specific ratios. For example, the ratio of nitrogen to phosphorus in the organic tissues of algae is about 16 to 1, so if the available nitrogen concentration is greater than 16 times the phosphorus concentration, then phosphorus will be the factor that limits growth; if it is less, then nitrogen will be limiting. To understand how a specific ecosystem functions, it thus is important to identify what factors limit ecosystem activity.

Resources influence ecosystem activity differently depending on whether they are essential, substitutable, or complementary. Essential resources limit growth independently of other levels: if the minimum quantity needed for growth is not available, then growth does not occur. In contrast, if two resources are substitutable, then population growth is limited by an appropriately weighted sum of the two resources in the environment. For example, glucose and fructose are substitutable food sources for many types of bacteria. Resources may also be complementary, which means that a small amount of one resource can substitute for a relatively large amount of another, or can be complementary over a specific range of conditions.

Resource availability serves as a so-called "bottom-up" control on an ecosystem: the supply of energy and nutrients influences ecosystem activities at higher trophic levels by affecting the amount of energy that moves up the food chain. In some cases, ecosystems may be more strongly influenced by so-called "top-down" controls—namely, the abundance of organisms at high trophic levels in the ecosystem (Fig. 12). Both types of effects can be at work in an ecosystem at the same time,

but how far bottom-up effects extend in the food web, and the extent to which the effects of trophic interactions at the top of the food web are felt through lower levels, vary over space and time and with the structure of the ecosystem.



Figure 12. Predators impose top-down control on ecosystems

Many ecological studies seek to measure whether bottom-up or top-down controls are more important in specific ecosystems because the answers can influence conservation and environmental protection strategies. For example, a study by Benjamin S. Halpern and others of food web controls in kelp forest ecosystems off the coast of Southern California found that variations in predator abundance explained a significant proportion of variations in the abundance of algae and the organisms at higher trophic levels that fed on algae and plankton. In contrast, they found no significant relationship between primary production by algae and species abundance at higher trophic levels. The most influential predators included spiny lobster, Kellet's whelk, rockfish, and sea perch. Based on these findings, the authors concluded that "[e]fforts to control activities that affect higher trophic levels (such as fishing) will have far larger impacts on community dynamics than efforts to control, for example, nutrient input, except when these inputs are so great as to create anoxic (dead) zones" (footnote 4).

Drastic changes at the top of the food web can trigger **trophic cascades**, or domino effects that are felt through many lower trophic levels. The likelihood of a trophic cascade depends on the number of trophic levels in the ecosystem and the extent to which predators reduce the abundance of a trophic level to below their resource-limited carrying capacity. Some species are so important to an entire ecosystem that they are referred to as **keystone species**, connoting that they occupy an ecological

niche that influences many other species. Removing or seriously impacting a keystone species produces major impacts throughout the ecosystem.

Many scientists believe that the reintroduction of wolves into Yellowstone National Park in 1995, after they had been eradicated from the park for decades through hunting, has caused a trophic cascade with results that are generally positive for the ecosystem. Wolves have sharply reduced the population of elk, allowing willows to grow back in many riparian areas where the elk had grazed the willows heavily. Healthier willows are attracting birds and small mammals in large numbers.

"Species, like riparian songbirds, insects, and in particular, rodents, have come back into these preferred habitat types, and other species are starting to respond," says biologist Robert Crabtree of the Yellowstone Ecological Research Center. "For example, fox and coyotes are moving into these areas because there's more prey for them. There's been an erupting trophic cascade in some of these lush riparian habitat sites."

## 7. Ecological Niches

Within ecosystems, different species interact in different ways. These interactions can have positive, negative, or neutral impacts on the species involved (Table 4).

**Table 4.** Relationships between individuals of different species.

Type of interaction	Effect of interaction	Examples
Competition	Both species are harmed (population growth rates are reduced).	Oak trees and maple trees competing for light in a forest, wading birds foraging for food in a marsh
Predation Parasitism	One species benefits, one is harmed.	Predation: wolf and rabbit Parasitism: flea and wolf
Mutualism	Both species benefit. Relationship may not be essential for either.	Humans and house pets, insect pollination of flowers
Commensalism	One species benefits, one is not affected.	Maggots decomposing a rotting carcass

Type of interaction	Effect of interaction	Examples
Amensalism	One species harms another (typically by releasing a toxic substance), but is not affected itself.	Allelopathy (plants that produce substances harmful to other plants): rye and wheat suppress weeds when used as cover crops, broccoli residue suppresses growth of other vegetables in the same plant family

Each species in an ecosystem occupies a niche, which comprises the sum total of its relationships with the biotic and abiotic elements of its environment—more simply, what it needs to survive. In a 1957 address, zoologist George Evelyn Hutchinson framed the view that most ecologists use today when he defined the niche as the intersection of all of the ranges of tolerance under which an organism can live (footnote 5). This approach makes ecological niches easier to quantify and analyze because they can be described as specific ranges of variables like temperature, latitude, and altitude. For example, the African Fish Eagle occupies a very similar ecological niche to the American Bald Eagle (Fig. 13). In practice it is hard to measure all of the variables that a species needs to survive, so descriptions of an organism's niche tend to focus on the most important limiting factors.



Figure 13. African fish eagle

Courtesy Wikimedia Commons. Public domain.

The full range of habitat types in which a species can exist and reproduce without any competition from other species is called its **fundamental niche**. The presence of other species means that few species live in such conditions. A species' **realized niche** can be thought of as its niche in practice—the range of habitat types from which it is not excluded by competing species. Realized niches are usually smaller than fundamental niches, since competitive interactions exclude species from at least some conditions under which they would otherwise grow. Species may occupy different realized niches in various locations if some constraint, such as a certain predator, is present in one area but not in another.

In a classic set of laboratory experiments, Russian biologist G.F. Gause showed the difference between fundamental and realized niches. Gause compared how two strains of **Paramecium** grew when they were cultured separately in the same type of medium to their growth rates when cultured together. When cultured separately both strains reproduced rapidly, which indicated that they were adapted to living and reproducing under the same conditions. But when they were cultured together, one strain out-competed and eventually eliminated the other. From this work Gause developed a fundamental concept in community ecology: the **competitive exclusion principle**, which states that if two competitors try to occupy the same realized niche, one species will eliminate the other (footnote 6).

Many key questions about how species function in ecosystems can be answered by looking at their niches. Species with narrow niches tend to be specialists, relying on comparatively few food sources. As a result, they are highly sensitive to changes in key environmental conditions, such as water temperature in aquatic ecosystems. For example, pandas, which only eat bamboo, have a highly specialized diet. Many endangered species are threatened because they live or forage in particular habitats that have been lost or converted to other uses. One well-known case, the northern spotted owl lives in cavities of trees in old-growth forests (forests with trees that are more than 200 years old and have not been cut, pruned, or managed), but these forests have been heavily logged, reducing the owl's habitat.

In contrast, species with broad niches are generalists that can adapt to wider ranges of environmental conditions within their own lifetimes (i.e., not through evolution over generations, but rather through changes in their behavior or physiologic functioning) and survive on diverse types of prey. Coyotes once were found only on the Great Plains and in the western United States, but have spread through the eastern states in part because of their flexible lifestyle. They can kill and eat large, medium, or small prey, from deer to house cats, as well as other foods such as invertebrates and fruit, and can live in a range of habitats, from forests to open landscapes, farmland, and suburban neighborhoods (footnote 7).

Overlap between the niches of two species (more precisely, overlap between their resource use curves) causes the species to compete if resources are limited. One might expect to see species constantly dying off as a result, but in many cases competing species can coexist without either being eliminated. This happens through **niche partitioning** (also referred to as resource partitioning), in which two species divide a limiting resource such as light, food supply, or habitat.

## 8. Evolution and Natural Selection in Ecosystems

As species interact, their relationships with competitors, predators, and prey contribute to natural selection and thus influence their evolution over many generations. To illustrate this concept, consider how evolution has influenced the factors that affect the foraging efficiency of predators. This includes the predator's search time (how long it takes to find prey), its handling time (how hard it has to work to catch and kill it), and its prey profitability (the ratio of energy gained to energy spent handling prey). Characteristics that help predators to find, catch, and kill prey will enhance their chances of surviving and reproducing. Similarly, prey will profit from attributes that help avoid detection and make organisms harder to handle or less biologically profitable to eat.

These common goals drive natural selection for a wide range of traits and behaviors, including:

- **Mimicry** by either predators or prey. A predator such as a praying mantis that blends in with surrounding plants is better able to surprise its target. However, many prey species also engage in mimicry, developing markings similar to those of unpalatable species so that predators avoid them. For example, harmless viceroy butterflies have similar coloration to monarch butterflies, which store toxins in their tissues, so predators avoid viceroy butterflies.
- Optimal foraging strategies enable predators to obtain a maximum amount of net energy per unit of time spent foraging. Predators are more likely to survive and reproduce if they restrict their diets to prey that provide the most energy per unit of handling time and focus on areas that are rich with prey or that are close together. The Ideal Free Distribution model suggests that organisms that are able to move will distribute themselves according to the amount of food available, with higher concentrations of organisms located near higher concentrations of food. Many exceptions have been documented, but this theory is a good general predictor of animal behavior.
- Avoidance/escape features help prey elude predators. These attributes may be behavioral patterns, such as animal herding or fish schooling to make individual organisms harder to pick out. Markings can confuse and disorient predators: for example, the automeris moth has false eye spots on its hind wings that misdirect predators (Fig. 14).

- Features that increase handling time help to discourage predators. Spines serve this function for many plants and animals, and shells make crustaceans and mollusks harder to eat. Behaviors can also make prey harder to handle: squid and octopus emit clouds of ink that distract and confuse attackers, while hedgehogs and porcupines increase the effectiveness of their protective spines by rolling up in a ball to conceal their vulnerable underbellies.
- Some plants and animals emit noxious chemical substances to make themselves less profitable as prey. These protective substances may be bad-tasting, antimicrobial, or toxic. Many species that use noxious substances as protection have evolved bright coloration that signals their identity to would-be predators—for example, the black and yellow coloration of bees, wasps, and yellowjackets. The substances may be generalist defenses that protect against a range of threats, or specialist compounds developed to ward off one major predator. Sometimes specialized predators are able overcome these noxious substances: for example, ragwort contains toxins that can poison horses and cattle grazing on it, but it is the exclusive food of cinnabar moth caterpillars. Ragwort toxin is stored in the caterpillars' bodies and eventually protects them as moths from being eaten by birds.



Figure 14. Automeris moth

© D.H. Jansen and Winnie Hallwachs, janzen.sas.upenn.edu.

Natural selection based on features that make predators and prey more likely to survive can generate predator-prey "arms races," with improvements in prey defenses triggering counter-improvements in predator attack tools and vice versa over many generations. Many cases of predator-prey arms races have been identified. One widely known case is bats' use of echolocation to find insects. Tiger moths respond by emitting high-frequency clicks to "jam" bats' signals, but some bat species have overcome these measures through new techniques such as flying erratically to confuse moths or sending echolocation chirps at frequencies that moths cannot detect. This type of pattern involving two species that interact in important ways and evolve in a series of reciprocal genetic steps is called **coevolution** and represents an important factor in adaptation and the evolution of new biological species.

Other types of relationship, such as competition, also affect evolution and the characteristics of individual species. For example, if a species has an opportunity to move into a vacant niche, the shift may facilitate evolutionary changes over succeeding generations because the species plays a different ecological role in the new niche. By the early 20th century, large predators such as wolves and puma had been largely eliminated from the eastern United States. This has allowed coyotes, who compete with wolves where they are found together, to spread throughout urban, suburban, and rural habitats in the eastern states, including surprising locations such as Cape Cod in Massachusetts

and Central Park in New York City. Research suggests that northeastern coyotes are slightly larger than their counterparts in western states, although it is not yet clear whether this is because the northeastern animals are hybridizing with wolves and domestic dogs or because they have adapted genetically to preying on larger species such as white-tailed deer (footnote 9).

## 9. Natural Ecosystem Change

Just as relationships between individual species are dynamic, so too is the overall makeup of ecosystems. The process by which one natural community changes into another over a time scale of years to centuries is called **succession**. Common succession patterns include plant colonization of sand dunes and the regrowth of forests on abandoned farmland (Fig. 15). While the general process is widely recognized, ecologists have offered differing views of what drives succession and how to define its end point. By analyzing the natural succession process, scientists seek to measure how stable ecosystems are at different stages in their trajectory of development, and how they respond to disturbances in their physical environment or changes in the frequency at which they are disturbed.

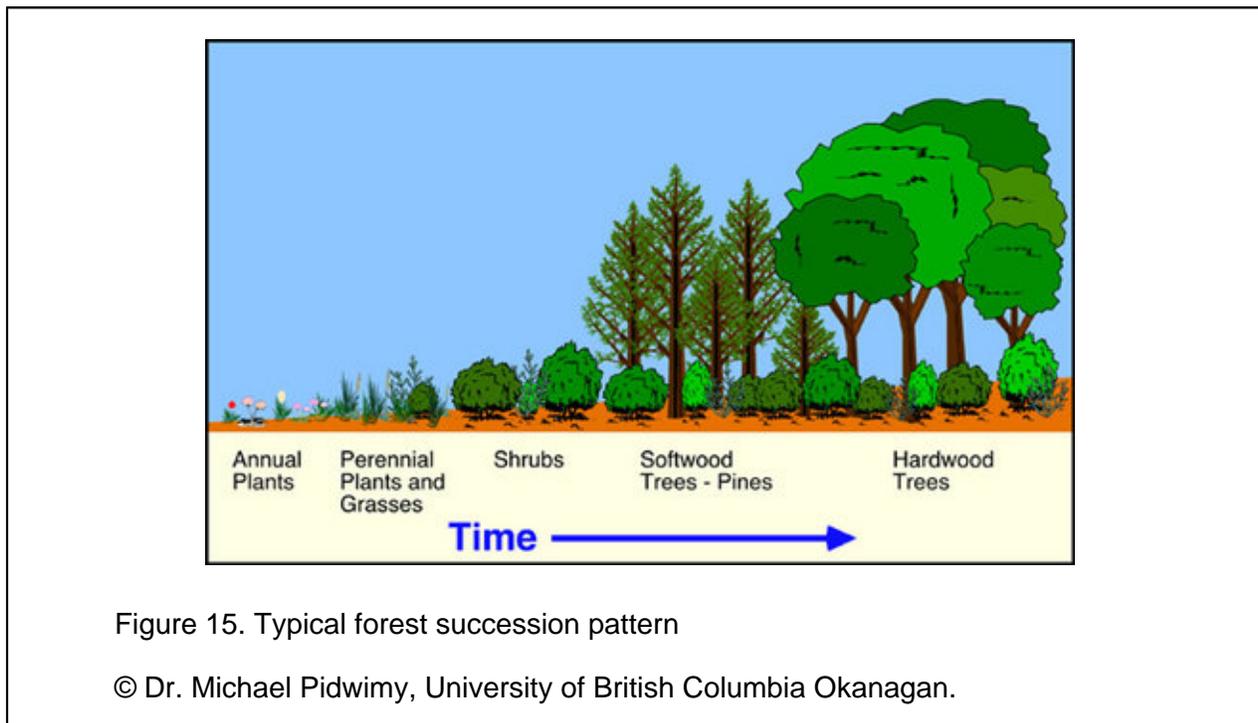


Figure 15. Typical forest succession pattern

© Dr. Michael Pidwimy, University of British Columbia Okanagan.

In the early 20th century, plant biologist Frederic Clements described two types of succession: primary (referring to colonization of a newly exposed landform, such as sand dunes or lava flows after a volcanic eruption) and secondary (describing the return of an area to its natural vegetation following a disturbance such as fire, treefall, or forest harvesting). British ecologist Arthur Tansley distinguished

between autogenic succession—change driven by the inhabitants of an ecosystem, such as forests regrowing on abandoned agricultural fields—and allogenic succession, or change driven by new external geophysical conditions such as rising average temperatures resulting from global climate change.

As discussed above, ecologists often group species depending on whether they are better adapted for survival at low or high population densities (r-selected versus K-selected). Succession represents a natural transition from r- to K-selected species. Ecosystems that have recently experienced traumatic extinction events such as floods or fires are favorable environments for r-selected species because these organisms, which are generalists and grow rapidly, can increase their populations in the absence of competition immediately after the event. Over time, however, they will be out-competed by K-selected species, which often derive a competitive advantage from the habitat modification that takes place during early stages of primary succession.

For example, when an abandoned agricultural field transitions back to forest, as seen in Figure 15, sun-tolerant weeds and herbs appear first, followed by dense shrubs like hawthorn and blackberry. After about a decade, birches and other small fast-growing trees move in, sprouting wherever the wind blows their lightweight seeds. In 30 to 40 years, slower-spreading trees like ash, red maple, and oak take root, followed by shade-tolerant trees such as beech and hemlock.

A common observation is that as ecosystems mature through successional stages, they tend to become more diverse and complex. The number of organisms and species increases and niches become narrower as competition for resources increases. Primary production rates and nutrient cycling may slow as energy moves through a longer sequence of trophic levels (Table 5).

**Table 5.** Characteristics of developing and mature ecosystems.

Ecosystem attributes	Developmental stages	Mature stages
	<b>Energetics:</b>	
Production/respiration	More or less than 1	Approaching 1
Production/biomass	High	Low
Food chains	Linear	Web-like
	<b>Community structure:</b>	
Niches	Broad	Narrow
Species diversity	Low	High
Nutrient conservation	Poor; detritus unimportant	Good; detritus important
Nutrient exchange rates	Rapid	Slow
Stability	Low	High

Many natural disturbances have interrupted the process of ecosystem succession throughout Earth's history, including natural climate fluctuations, the expansion and retreat of glaciers, and local factors such as fires and storms. An understanding of succession is central for conserving and restoring ecosystems because it identifies conditions that managers must create to bring an ecosystem back into its natural state. The Tallgrass Prairie National Preserve in Kansas, created in 1996 to protect 11,000 acres of prairie habitat, is an example of a conservation project that seeks to approximate natural ecosystem succession. A herd of grazing buffalo tramples on tree seedlings and digs up the ground, creating bare patches where new plants can grow, just as millions of buffalo maintained the grassland prairies that covered North America before European settlement (footnote 10).

## 10. Further Reading

Paul A. Colinvaux, **Why Big Fierce Animals Are Rare: An Ecologist's Perspective** (Princeton, NJ: Princeton University Press, 1979). A survey of major questions in ecology, including why every species has its own niche.

Chris Reiter and Gina C. Gould, "Thirteen Ways of Looking at a Hedgehog," **Natural History**, July/August 1998. Hedgehogs' spines are unique adaptations, but they have thrived in many regions for millions of years because they are generalists in terms of climate zones and diet.

ReefQuest Centre for Shark Research, "Catch As Catch Can," [http://www.elasmo-research.org/education/topics/b\\_catch.htm](http://www.elasmo-research.org/education/topics/b_catch.htm). Contrary to their popular image as mindless eating machines, great white sharks' foraging strategies are selective and efficient.

## Footnotes

1. One important exception is microbes, which are more diverse in temperate areas; see Unit 9, "Biodiversity Decline," for details.
2. Cornelia Dean, "In An Antarctic Desert, Signs of Life," **New York Times**, February 3, 1998, p. F1.
3. U.S. Geological Survey, "Mineral Substances in the Environment," <http://geology.er.usgs.gov/eastern/environment/environ.html>.
4. Benjamin S. Halpern, Karl Cottenie, and Bernardo R. Broitman, "Strong Top-Down Control in Southern California Kelp Forest Ecosystems," **Science**, May 26, 2006, pp. 1230–32.
5. G.E. Hutchinson, "Concluding Remarks," **Cold Spring Harbor Symposia on Quantitative Biology** 22 (1957), pp. 415–27.
6. G.F. Gause, **The Struggle for Existence** (Baltimore: Williams and Wilkins, 1934).
7. Matthew E. Gompper, **The Ecology of Northeast Coyotes**, Working Paper No. 17 (New York, NY: Wildlife Conservation Society, July 2002), [http://www.wcs.org/media/file/Ecology\\_of\\_NE\\_Coyotes.pdf](http://www.wcs.org/media/file/Ecology_of_NE_Coyotes.pdf).

8. S.D. Fretwell and H.J. Lucas, "Ideal Free Distribution," **Acta Biotheory** 19 (1970), pp. 16–21.
9. Matthew E. Gompper, **The Ecology of Northeast Coyotes**, Working Paper No. 17 (New York, NY: Wildlife Conservation Society, July 2002), [http://www.wcs.org/media/file/Ecology\\_of\\_NE\\_Coyotes.pdf](http://www.wcs.org/media/file/Ecology_of_NE_Coyotes.pdf), pp. 17–20.
10. For more information, see <http://www.nps.gov/tapr/index.htm>

## Glossary

**bioaccumulation** : The increase in concentration of a chemical in organisms that reside in environments contaminated with low concentrations of various organic compounds.

**biomes** : Broad regional areas characterized by a distinctive climate, soil type, and biological community.

**carbon dioxide fertilization** : Increased plant growth due to a higher carbon dioxide concentration.

**carrying capacity** : The number of individuals an environment can support without significant negative impacts to the given organism and its environment.

**coevolution** : Simultaneous evolution of two or more species of organisms that interact in significant ways.

**competitive exclusion principle** : The hypothesis stating that when organisms of different species compete for the same resources in the same habitat, one species will commonly be more successful in this competition and exclude the second from the habitat.

**denitrification** : Process of reducing nitrate and nitrite, highly oxidised forms of nitrogen available for consumption by many groups of organisms, into gaseous nitrogen, which is far less accessible to life forms but makes up the bulk of our atmosphere.

**fundamental niche** : The full range of environmental conditions (biological and physical) under which an organism can exist.

**gross primary productivity (GPP)** : The rate at which an ecosystem accumulates biomass, including the energy it uses for the process of respiration.

**K-selected** : Those species that invest more heavily in fewer offspring, each of which has a better chance of surviving to adulthood.

**keystone species** : A single kind of organism or a small collection of different kinds of organisms that occupy a vital ecological niche in a given location.

**latitudinal biodiversity gradient** : The increase in species richness or biodiversity that occurs from the poles to the tropics, often referred to as the latitudinal gradient in species diversity.

**life history strategy** : An organism's allocation of energy throughout its lifetime among three competing goals: growing, surviving, and reproducing.

**mimicry** : Evolving to appear similar to another successful species or to the environment in order to dupe predators into avoiding the mimic, or dupe prey into approaching the mimic.

**mutualistic** : Refers to an interaction between two or more distinct biological species in which members benefit from the association. Describes both symbiotic mutualism (a relationship requiring an intimate association of species in which none can carry out the same functions alone) and nonsymbiotic mutualism (a relationship between organisms that is of benefit but is not obligatory: that is, the organisms are capable of independent existence).

**net primary productivity (NPP)** : The rate at which new biomass accrues in an ecosystem.

**niche partitioning** : The process by which natural selection drives competing species into different patterns of resource use or different niches. Coexistence is obtained through the differentiation of their realized ecological niches.

**nitrogen fixing** : The conversion of nitrogen in the atmosphere ( $N_2$ ) to a reduced form (e.g., amino groups of amino acids) that can be used as a nitrogen source by organisms.

**primary producers** : Organisms that produce organic compounds from atmospheric or aquatic carbon dioxide, principally through the process of photosynthesis. Primary production is distinguished as either net or gross. All life on earth is directly or indirectly reliant on primary production.

**r-selected** : Species with a reproductive strategy to produce many offspring, each of whom is, comparatively, less likely to survive to adulthood.

**realized niche** : The ecological role that an organism plays when constrained by the presence of other competing species in its environment

**species richness** : A type of approach to assessing biodiversity that examines the distribution of all resident terrestrial vertebrates: amphibians, reptiles, birds, and mammals.

**succession** : A fundamental concept in ecology that refers to the more or less predictable and orderly changes in the composition or structure of an ecological community.

**trophic cascades** : Occur when predators in a food chain suppress the abundance of their prey, thereby releasing the next lower trophic level from predation (or herbivory if the intermediate trophic level is an herbivore). Trophic cascades may also be important for understanding the effects of removing top predators from food webs, as humans have done in many places through hunting and fishing activities.

**trophic level** : A feeding level within a food web.

## Unit 5 : Human Population Dynamics



Street Market in Mumbai, India. Courtesy of David E. Bloom

### Overview

What factors influence human population growth trends most strongly, and how does population growth or decline impact the environment? Does urbanization threaten our quality of life or offer a pathway to better living conditions? What are the social implications of an aging world population? Discover how demographers approach these questions through the study of human population dynamics.

### Sections:

1. Introduction
2. Mathematics of Population Growth
3. Determinants of Demographic Change
4. World Population Growth Through History
5. Population Growth and the Environment
6. Urbanization and Megacities
7. Other Consequences of Demographic Change
8. Demographic Convergence and Human Lifespan Trends
9. Further Reading



## 1. Introduction

Human population trends are centrally important to environmental science because they help to determine the environmental impact of human activities. Rising populations put increasing demands on natural resources such as land, water, and energy supplies. As human communities use more resources, they generate contaminants, such as air and water pollution and greenhouse gas emissions, along with increasing quantities of waste.

Population interacts with several other factors to determine a society's environmental impact. One widely-cited formula is the "I = PAT" equation, proposed by Paul R. Ehrlich and John P. Holdren in 1974 (footnote 1).

Environmental Impact = Population x Affluence (or consumption) x Technology

For generations people have tried to estimate Earth's **carrying capacity**, or the maximum population that it can support on a continuing basis. This is a slippery undertaking. Estimates of human carrying capacity over the past four centuries have varied from less than one billion people to more than one trillion, depending on how the authors defined carrying capacity. Some studies cast the issue solely in terms of food production, others as the availability of a broader set of resources.

In fact, the question depends on assumptions about human preferences. What standard of living is seen as acceptable, and what levels of risk and variability in living conditions will people tolerate? Many of these issues are not just matters of what humans want; rather, they intersect with physical limits, such as total arable land or the amount of energy available to do work. In such instances nature sets bounds on human choices (footnote 2).

Measuring Earth's carrying capacity at the global level obscures the fact that resources are not allocated equally around the world. In some areas such as the Sahel in West Africa (the transition zone between the Sahara desert and more humid woodlands to the south), population growth is putting heavy stresses on a fragile environment, so food needs are outstripping food production (Fig. 1). Other regions have better balances between populations and resources.



Figure 1. Gully erosion from over-cultivation, Sahel, West Africa

Courtesy United States Geological Survey, National Center for Earth Resources Observation Systems International Program.

Demography, the science of human population (or more specifically, the study of population structure and processes), draws together research from a number of disciplines, including economics, sociology, geography, public health, and genetics. In addition to the environmental impacts of population growth, population science also considers questions such as:

- How does population growth or decline influence economic and social well-being?
- Does population growth enhance or diminish economic growth?
- What impact does population growth have on poverty?
- Do specific aspects of population growth, such as age structure or sex imbalance, have bigger impacts on economic development and environmental quality than other aspects?
- What are the social and economic implications of population redistribution, through, for example, rural to urban or international migration?

This unit discusses basic population dynamics, including birth and death rates and factors that influence demographic change. It then summarizes the history of world population growth and projections through mid-century, with a focus on rising urbanization and the aging of the global population. Next we examine the environmental, economic, and institutional implications of population growth and some actions that governments can take to maximize benefits from population growth and limit harmful impacts. Finally, we consider whether nations' demographic patterns are becoming more similar, in spite of their different historic, cultural, and economic legacies, taking note of some regions that do not fit this general pattern.

## 2. Mathematics of Population Growth

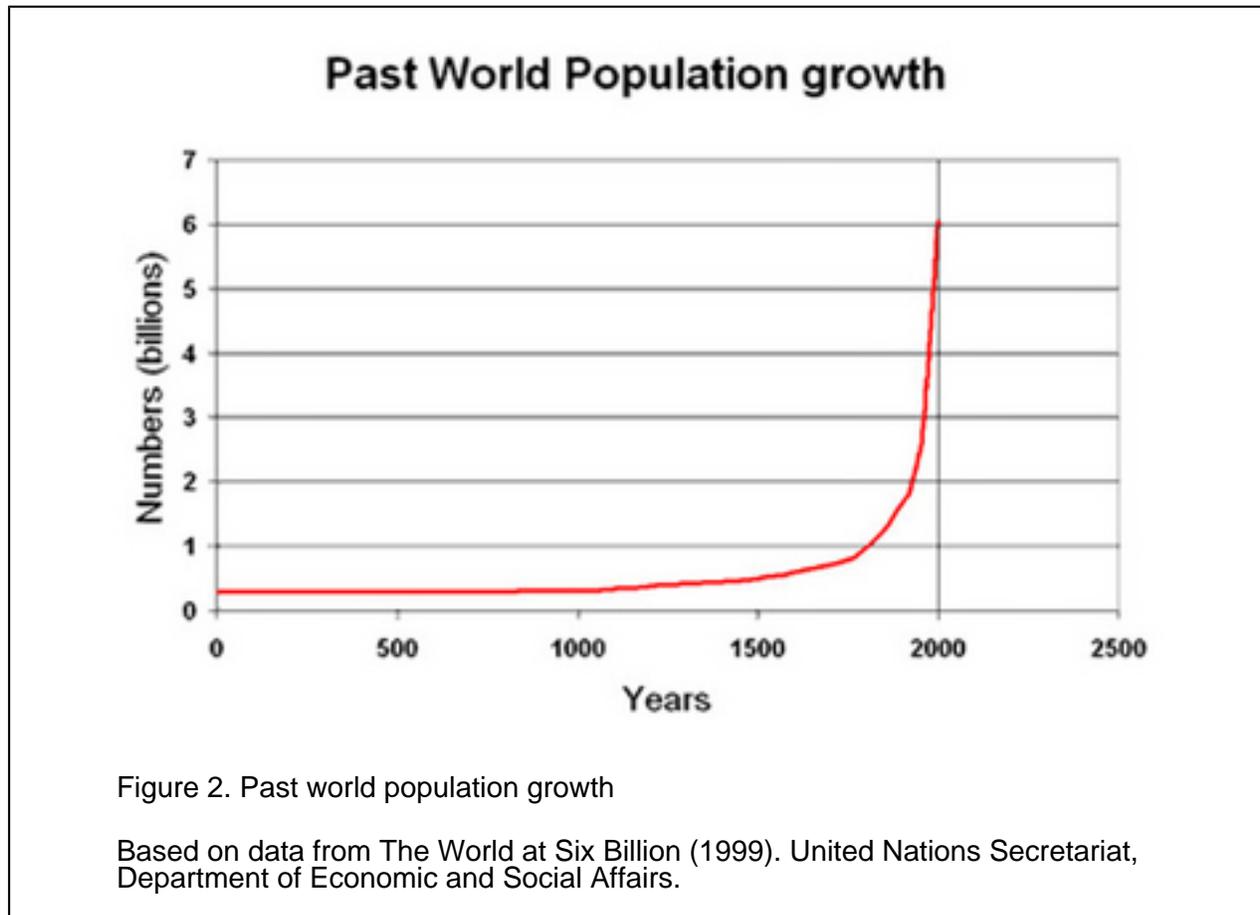
Population experts can make demographic predictions with more confidence than many other social scientists. Several basic truths apply to the demographics of all human societies:

- Everyone who is alive one year from now will be one year older at that time than s/he is today.
- Ages 15 to 49 are humans' prime childbearing years, biologically speaking (although resource constraints and social and political factors shape childbearing decisions differently from one country to another).
- Human mortality is relatively high among infants, children, and adults over age 60, and relatively low at other parts of the life cycle.

Putting these observations together, population analysts can develop a reasonably accurate map of how a society's population size, births, deaths, and age structure are likely to evolve in the next several decades.

Birth and death rates are the most important determinants of population growth; in some countries, net migration is also important in this regard. To calculate population growth rates, demographers take the difference between births and deaths in a given time period, add the net number of migrants (which for the world as a whole is 0), and divide that number by the total population. For example, there are now about 136 million births and 58 million deaths worldwide annually, adding a net of 78 million new inhabitants to a global population of 6.7 billion, a growth rate of nearly 1.2 percent.

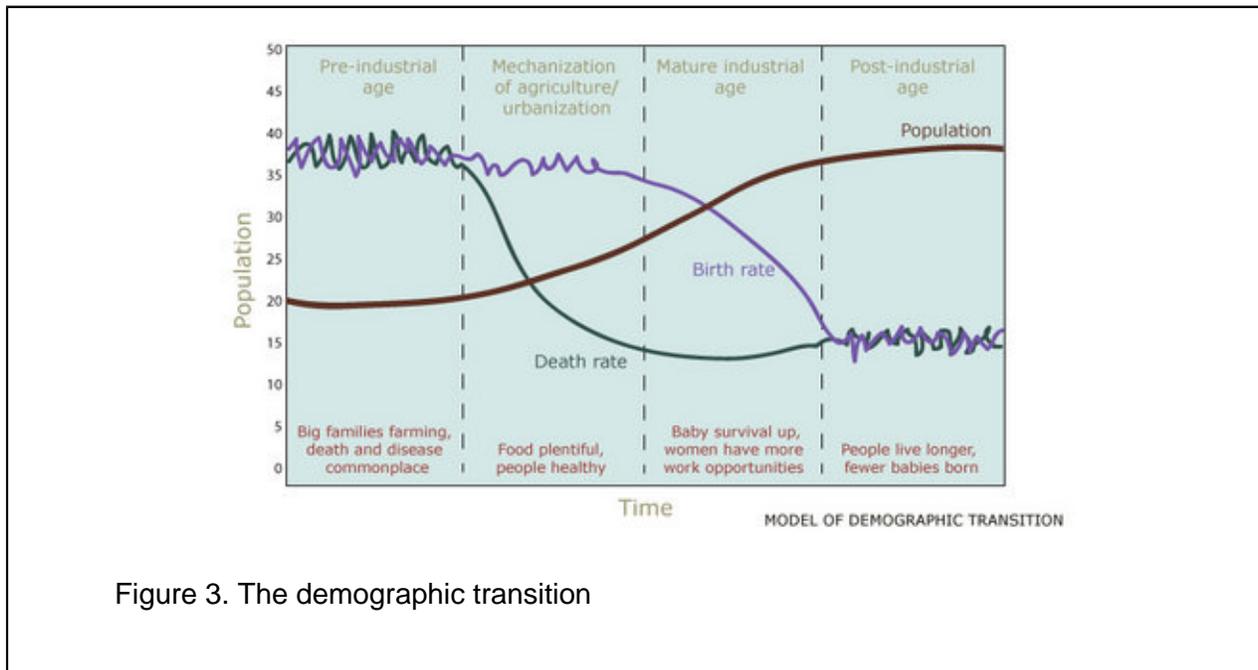
Until the mid-19th century birth rates were only slightly higher than death rates, so the human population grew very slowly. The industrial era changed many factors that affected birth and death rates, and in doing so, it triggered a dramatic expansion of the world's population (Fig. 2).



How did industrialization alter population growth rates so sharply? One central factor was the mechanization of agriculture, which enabled societies to produce more food from available inputs. (For more information about increasing agricultural productivity, see Unit 7, "Agriculture.") As food supplies expanded, average levels of nourishment rose, and vulnerability to chronic and contagious diseases declined over succeeding generations. Improvements in medical care and public health services—which took place more in urban than in rural areas—also helped people to live longer, so death rates fell. After several decades of lower mortality, people realized that they did not have to have so many children to achieve their desired family size, so birth rates began to fall as well.

In addition, **desired** family size tended to decrease. As women found many more opportunities to enter the labor force, they were less inclined to devote resources to childrearing rather than paid work, and the jobs they had were not conducive to having children beside them as they worked. The costs of raising children also increased, as slightly wealthier families living in urban areas faced higher expenses for a larger array of physical and social necessities.

This phased reduction in death and birth rates is a process called the **demographic transition**, which alters population growth rates in several stages (Fig. 3).



Because death rates fall before birth rates, population growth initially speeds up (a phase sometimes referred to as the mortality transition), adding a large cohort of young people to society. This group in turn will have children, although probably fewer per family than their parents did, and because this group of childbearing-age people is large, population will continue to grow in absolute numbers even though on a per-capita basis birth rates will decline—a phenomenon that demographers call the fertility transition. **Population momentum** (i.e., continued population growth after a fall in birth rates) accounts for a significant portion of world population growth today even though the global fertility rate has declined from about 5 children born per woman in 1950 to a little over 2.5 in 2006.

Developed nations have passed through the demographic transition, and most developing countries are at some point in the process today. As a result, a "bulge," or baby-boom, generation, distinctly larger than those preceding or following it, is moving through the age structure of the population in nearly all countries. These large cohorts create both opportunities and challenges for society. Expanded work forces can help nations increase their economic output, raising living standards for everyone. They also can strain available resources and services, which in turn may cause shortages and economic disruption. (For more details, see section 7, "Other Consequences of Demographic Change.")

The demographic transition is a well-recognized pattern, but it has shown many variations from country to country. We cannot predict when specific demographic changes will occur in particular countries, and it is hard to specify precisely which factors will shape a given society's path. Looking forward, a major question for the 21st century is what happens after the demographic transition, and whether some countries in areas such as western Europe, where birth rates are very low, will start striving to raise fertility (footnote 3). More important in terms of environment and health, however, is the question of how to help countries that are lagging on the transition path.

### 3. Determinants of Demographic Change

What factors drive population growth rates? One major indicator and determinant of demographic change is **fertility**, which demographers express as the total fertility rate, the number of births that can be expected to occur to a typical woman in a given society during her childbearing years. Fertility is a function of a woman's **fecundity** (her physiological ability to conceive and bear children (footnote 4)) and of social, cultural, economic, and health factors that influence reproductive choices in the country in question. The most important non-physical factors influencing a country's total fertility rate include relationship status (the fraction of women who are married or in a relationship that exposes them to the possibility of becoming pregnant); use of contraception; the fraction of women who are infecund—for example, because they are breastfeeding a child; and the prevalence of induced abortion (footnote 5).

Fertility levels are lower in developed countries than in developing nations because more women in developed countries work outside of the home and tend to marry later and to use contraception and abortion to delay or prevent childbearing. Nevertheless, fertility rates in nearly all countries have been falling since the 1950s (Fig. 4). Most of the exceptions are in Central and Western Africa.

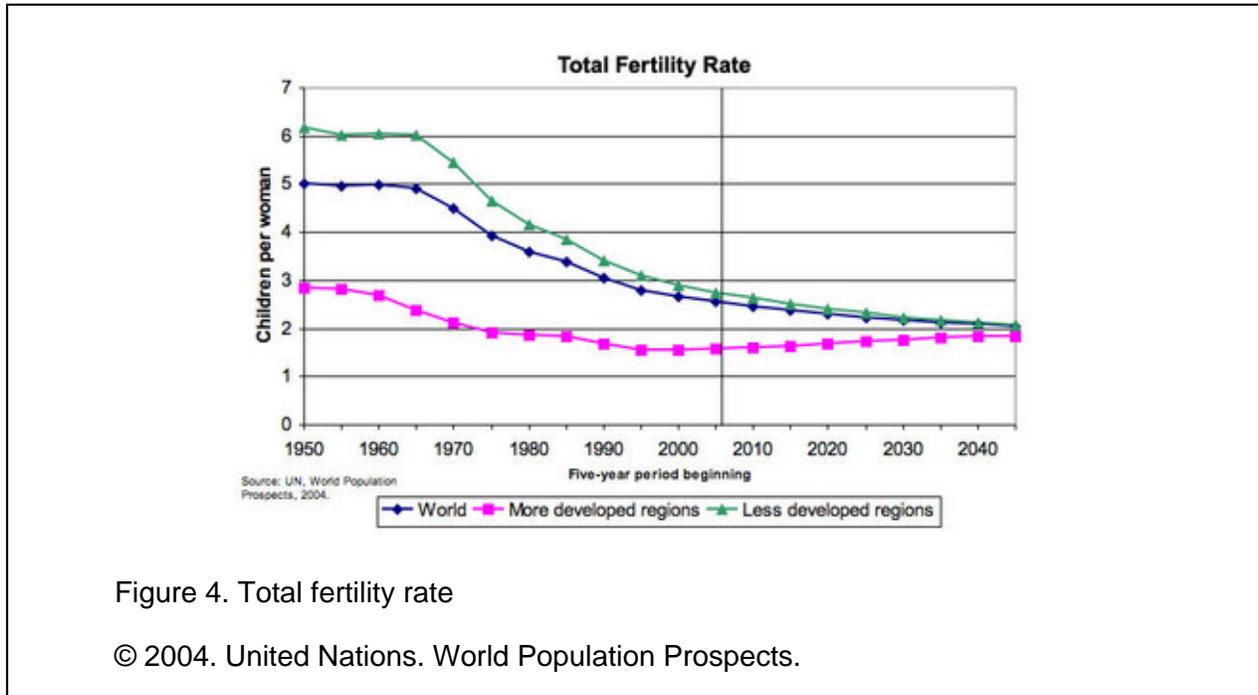


Figure 4. Total fertility rate

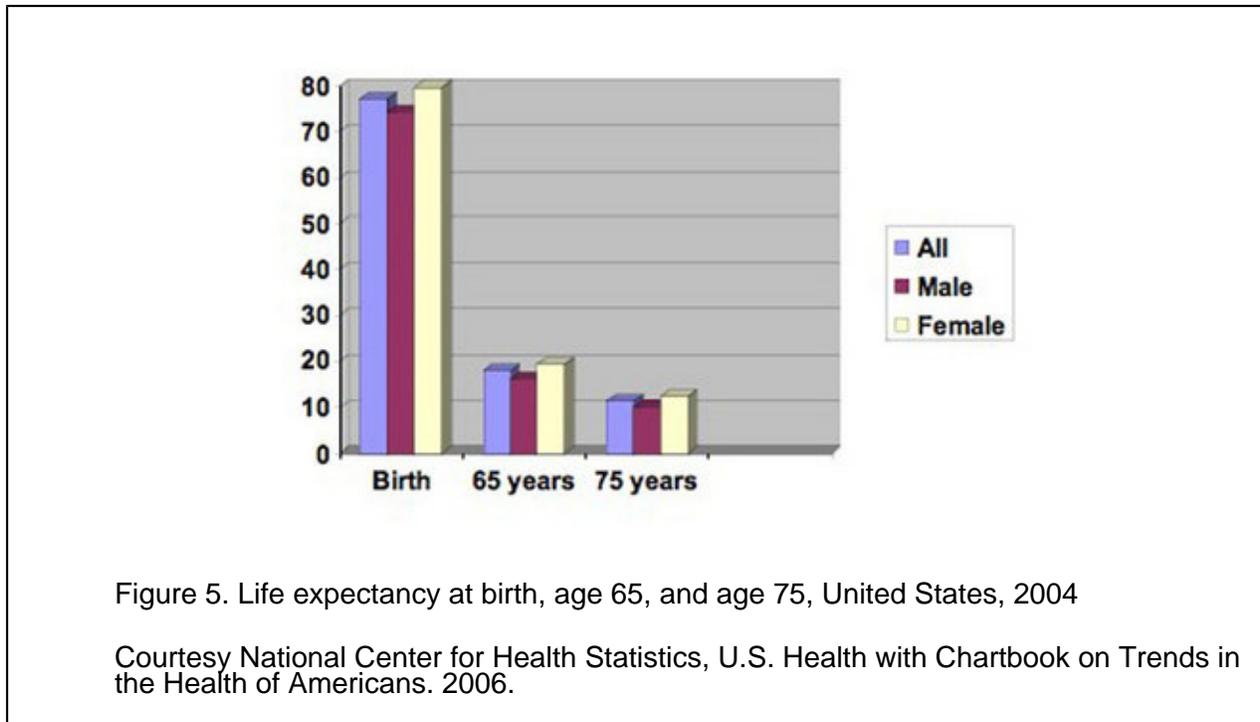
© 2004. United Nations. World Population Prospects.

Fertility patterns can vary widely within countries. Racial and ethnic minorities may have higher fertility rates than the majority, and families with low incomes or low levels of education typically have more children than those that are affluent or well-educated. Women who work outside the home generally have fewer children than those who stay home, and rural families have more children than city dwellers. In 2006, the number of births per 1,000 people worldwide averaged 21, with extremes ranging from a low of 8 or 9 (mainly in northern and western Europe and some former Soviet republics) to 50 or more in a few west African nations (footnote 6).

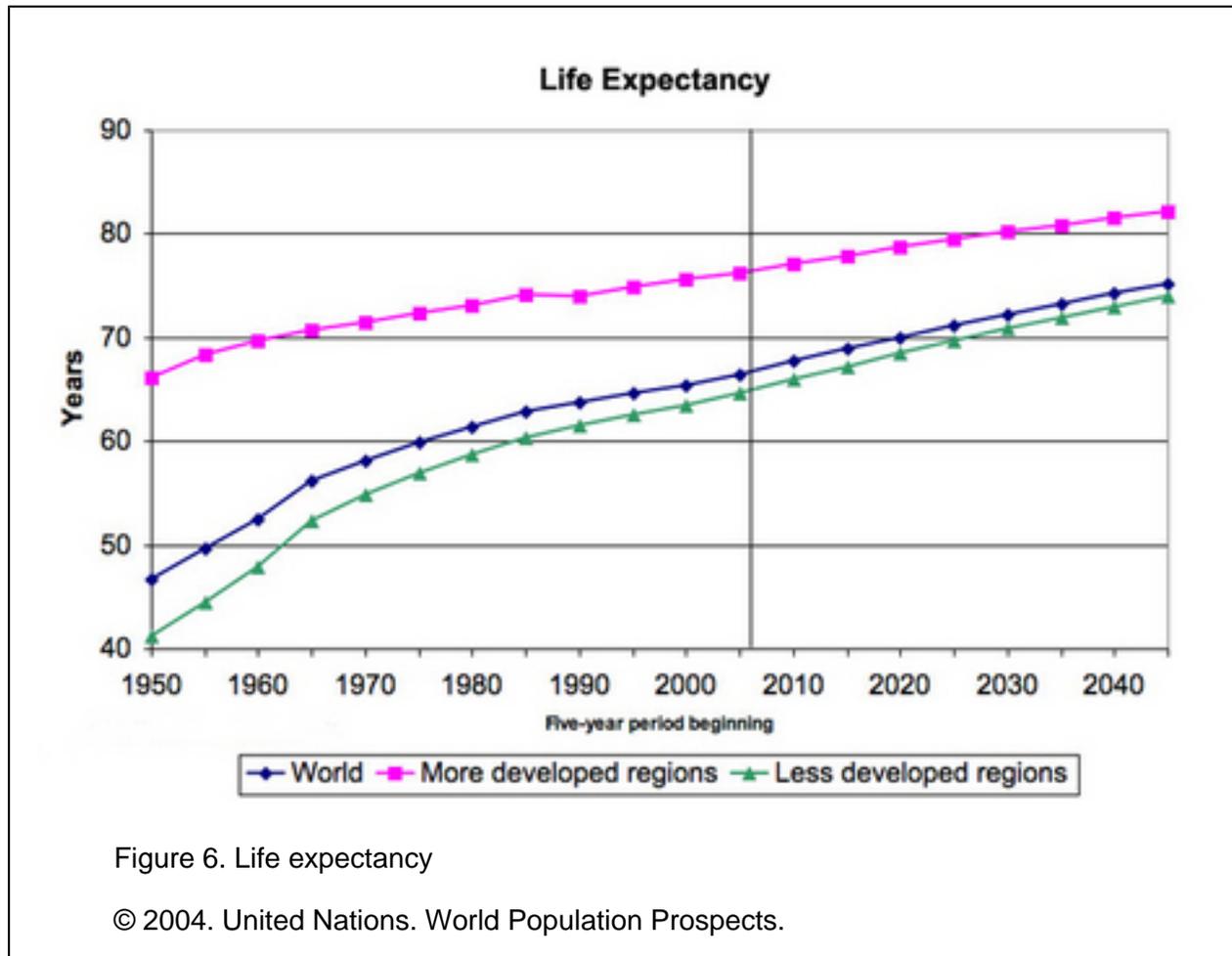
**Mortality** is the second major variable that shapes population trends. A population's age structure is an important factor influencing its death rate. Death rates are highest among infants, young children, and the elderly, so societies with many elderly people are likely to have more deaths per 1,000 people than those where most citizens are young adults. Developed countries with good medical services have more people in older age brackets than developing countries, so the developed societies can have higher death rates even though they are healthier places to live overall.

To assess longevity in a society, demographers calculate **life expectancy**—the age that a newborn would, on average, live to, assuming she were subject to a particular set of age-specific mortality rates—usually those prevailing in a particular year. The probability that a child will die at a given age drops through childhood and adolescence after she passes through the vulnerable early years, then starts to rise gradually in mid-life. Figure 5 shows remaining life expectancy at birth, 65 years of age, and 75 years of age in the year 2000 for people in the United States. Americans who were age 65 or

75 by 2000 had already survived many common causes of death, so they could expect at that point to live to an older age than would a baby born in that year (if life expectancy did not change during the baby's lifetime).



Life expectancy is trending upward around the world, but a substantial gap remains between developing and developed countries (Fig. 6). In 2006, life expectancies at birth ranged from the mid-30s in some African countries to the high 70s or low 80s in the United States, Australia, Japan, and some European countries (footnote 7).



What factors raise life expectancy? Because of the way in which it is calculated, life expectancy serves as a measure of the general health of the population, which depends on the satisfaction of many basic human needs such as adequate nutrition, clean water and sanitation, as well as access to medical services like vaccinations. Addressing these requirements reduces the incidence of many preventable illnesses. For example, nutritional deficiencies cause common illnesses like scurvy and pellagra, while dirty water and poor sanitation spread infectious agents such as cholera and typhoid. (For more details, see Unit 6, "Risk, Exposure, and Health," and Unit 8, "Water Resources.")

New threats to health are continually emerging, and often are spread across international borders through trade and human or animal migration. Recent examples that are severe enough to affect life expectancy in large areas include the HIV/AIDS pandemic and potentially avian flu and multi-drug-resistant malaria and tuberculosis. Researchers are also gaining new insight into existing threats, such as indoor air pollution from combustion of primitive biomass fuels like crop waste and

dung. Exposure to these pollutants is a major factor contributing to infant mortality and lower life expectancy in developing countries (Fig. 7). Environmental investments, such as providing cleaner energy sources and upgrading sewage treatment systems, can significantly improve public health.



Figure 7. Indian girls making tea, village of Than Gaon

Courtesy Wikimedia Commons. Creative Commons License.

Another step that increases life expectancy is creating a public health infrastructure that can identify and respond quickly to disease outbreaks, famines, and other threats. When severe acute respiratory syndrome (SARS) emerged as a disease that might cause an international epidemic, the U.S. Centers for Disease Control and Prevention (CDC) launched an emergency response program that required health departments to report suspect cases to CDC for evaluation, developed tests to identify the SARS virus, and kept health care providers and the public informed about the status of the outbreak. The United States and many other countries also reported their SARS cases to the World Health Organization. These types of close surveillance and preventive steps to control infections can help prevent diseases from spreading widely.

The third major factor that drives population trends is [migration](#), which includes geographic population shifts within nations and across borders. Migration is less predictable over long periods than fertility or mortality, since it can happen in sudden waves—for example, when refugees flee a war—or slowly over many years. Immigration often changes host nations' or regions' ethnic mixes and strains social services. On the positive side, it can provide needed labor (both skilled and unskilled). For source

countries, however, immigration may drain away valuable talent, especially since educated and motivated people are most likely to migrate in search of opportunities.

#### 4. World Population Growth Through History

Human population has grown very slowly for most of its existence on earth. Scientists currently estimate that modern human beings (**Homo sapiens**) evolved roughly 130,000 to 160,000 years ago. Many threats, from diseases to climate fluctuations, kept life expectancy short and death rates high in pre-industrial society, so it took until 1804 for the human population to reach one billion. From that point forward, however, population growth accelerated very quickly (Table 1).

**Table 1.** World population milestones.

World population reached:	Year	Time to add 1 billion
1 billion	1804	
2 billion	1927	123 years
3 billion	1960	33 years
4 billion	1974	14 years
5 billion	1987	13 years
6 billion	1999	12 years

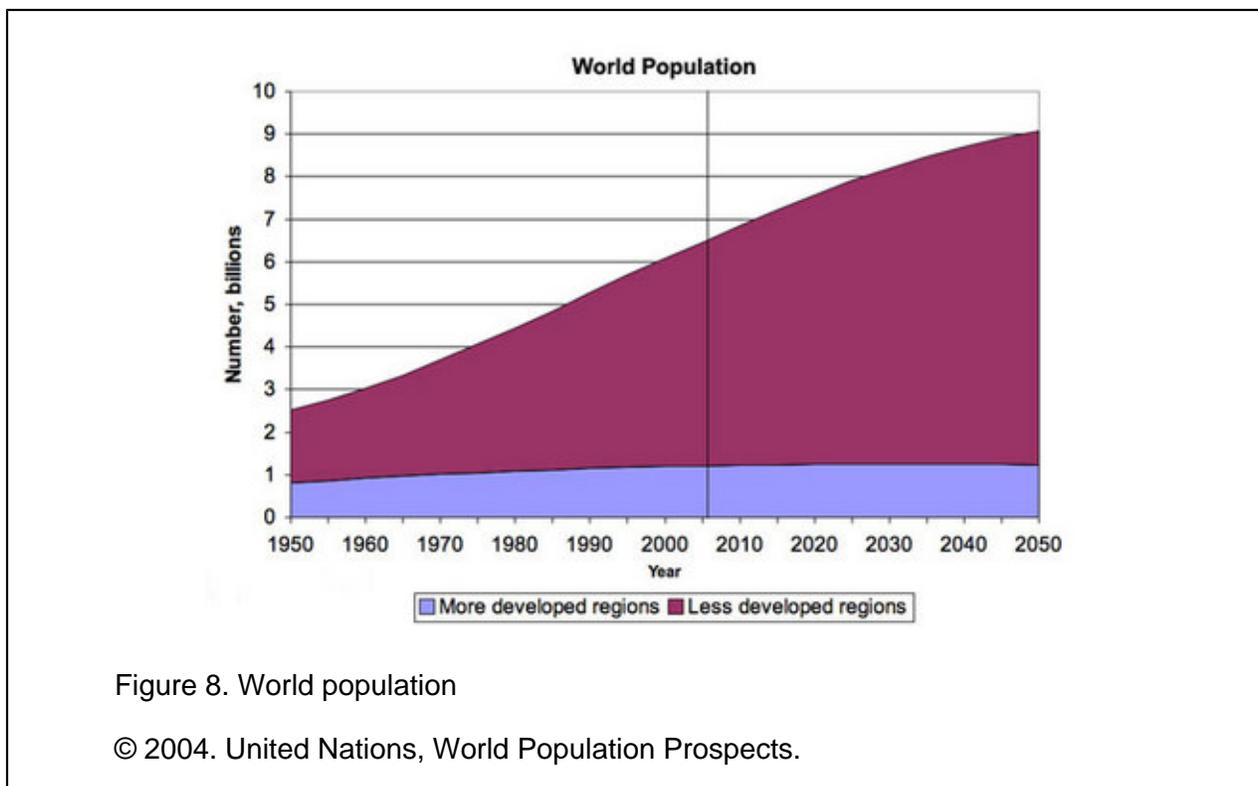
Through the early decades of the Industrial Revolution, life expectancies were low in western Europe and the United States. Thousands of people died from infectious diseases such as typhoid and cholera, which spread rapidly in the crowded, filthy conditions that were common in early factory towns and major cities, or were weakened by poor nutrition. But from about 1850 through 1950, a cascade of health and safety advances radically improved living conditions in industrialized nations. Major milestones included:

- improving urban sanitation and waste removal;
- improving the quality of the water supply and expanding access to it;
- forming public health boards to detect illnesses and quarantine the sick;
- researching causes and means of transmission of infectious diseases;
- developing vaccines and antibiotics;
- adopting workplace safety laws and limits on child labor; and

- promoting nutrition through steps such as fortifying milk, breads, and cereals with vitamins.

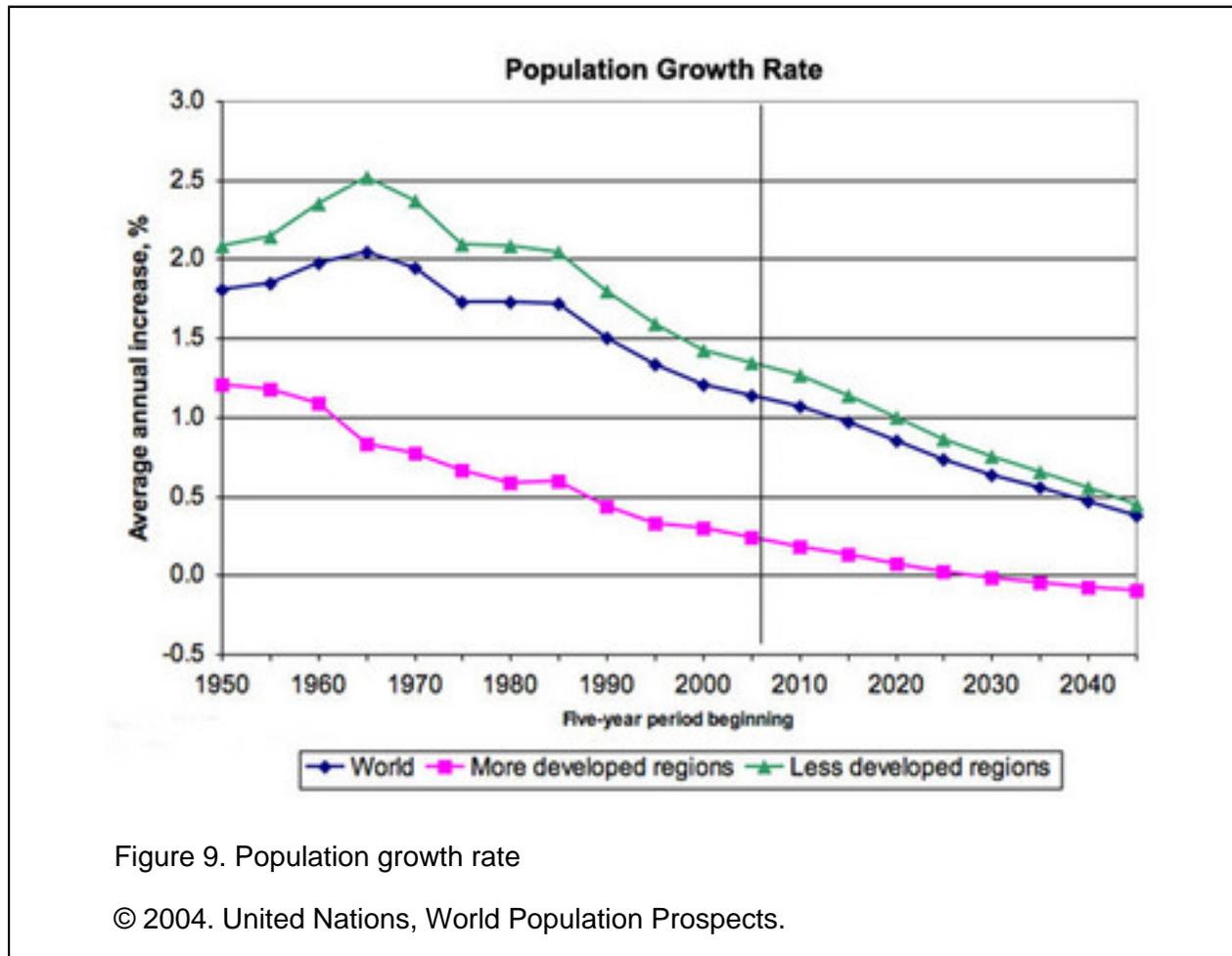
By the mid-20th century, most industrialized nations had passed through the demographic transition. As health technologies were transferred to developing nations, many of these countries entered the mortality transition and their population swelled. The world's population growth rate peaked in the late 1960s at just over 2 percent per year (2.5 percent in developing countries).

Demographers currently project that Earth's population will reach just over nine billion by 2050, with virtually all growth occurring in developing countries (Fig. 8). Future fertility trends will strongly affect the course of population growth. This estimate assumes that fertility will decline from 2.6 children per woman in 2005 to slightly over 2 children per woman in 2050. If the rate falls more sharply, to 1.5 children per woman, world population would be 7.7 billion in 2050, whereas a slower decline to 2.5 children per woman would increase world population to 10.6 billion by 2050.



Many people interpret forecasts like this to mean that population growth is out of control. In fact, as noted above, world population growth rates peaked in the late 1960s and have declined sharply in the past four decades (Fig. 9). The world's total population is still rising because of population momentum stemming from large increases that occurred in developing countries in the 1950s and early 1960s. But fertility rates are falling as many developing countries pass through the demographic transition,

thanks to factors that include lower infant mortality rates; expanding rights, education, and labor market opportunities for women; and increased access to family planning services.



World population growth in the 21st century will be different from previous decades in several important ways. First, humans are living longer and having fewer children, so there will be more older people (age 60 and above) than very young people (age zero to four). Second, nearly all population growth will take place in urban areas. Third, fertility rates will continue to decline (footnote 8).

All of these trends will affect nations' economic development. (On urbanization, see section 6, "Urbanization and Megacities.") Senior citizens can be active and productive members of society, but they have many unique needs in areas ranging from medical care to housing and transportation. Growing elderly populations will strain social services, especially in countries that do not have well-developed social safety nets to guarantee adequate incomes for older citizens. In countries that have "Pay As You Go" social security programs, increasing ratios of older to younger people may create

budget imbalances because fewer workers are paying funds into the system to support growing numbers of retirees.

As societies age, demand for younger workers will increase, drawing more people into the labor force and attracting immigrants in search of work. Declining fertility rates allow more women to work outside of the home, which increases the labor supply and may further accelerate the demographic transition (Fig. 10).



Figure 10. Woman supervising 25 employees at the Vegetable Dehydrates Factory, Parwan Province, Afghanistan

Courtesy Jeremy Foster, United States Agency for International Development.

As fertility rates fall, some countries have already dropped below **replacement level**—the number of children per woman that keeps population levels constant when births and deaths are considered together over time (assuming no net migration). Replacement-level fertility requires a total fertility rate of about 2.1 to offset the fact that some children will die before they reach adulthood and have their own families (in a society with higher mortality rates, replacement-level fertility would require more births) (footnote 9). Total fertility rates in most European and some Asian and Caribbean countries currently range from about 1.2 to 1.8, well below replacement level.

Some observers argue that declining fertility rates in both industrialized and developing countries will lead to a "birth dearth," with shrinking populations draining national savings and reducing tax revenues. However, societies can transition successfully from high mortality and fertility to low mortality and fertility with sound planning. Promoting good health standards (especially for children), expanding education, carefully opening up to international trade, and supporting older citizens



through retirement are all policies that can help to offset the negative impacts on society of an aging population (footnote 10).

## 5. Population Growth and the Environment

Many people (including national leaders) worry that population growth depletes resources and can trigger social or economic catastrophe if it is not contained. As discussed in the preceding section, most of the projected population growth during this century will take place in developing nations. These countries have faced many challenges in recent decades, including low levels of education, poor health standards, poverty, scarce housing, natural resource depletion, wars, and economic and political domination by other countries. In Sub-Saharan Africa industrial development has stalled and most workers still make a living from subsistence agriculture.

Countries in this situation generally have devoted less energy to addressing environmental issues than their wealthier neighbors, so these problems have intensified. Especially in the poorest countries, therefore, future population growth is likely to make environmental deterioration worse (although it does not automatically follow that countries with low population growth rates will have cleaner environments).

However, the relationship between population and the environment is complex. As noted in section 1, human societies' impacts on the environment are a function of three major, interconnected elements: population size, affluence or consumption, and technology. An expanded version of the IPAT equation separates technology into two factors: resource-intensity (how many resources are used to produce each unit of consumption) and waste-intensity (how much waste each unit of consumption generates), and also considers the sensitivity of the environment (footnote 11).

Societies' environmental impacts take two major forms. First, we consume resources such as land, food, water, soils, and services from healthy ecosystems, such as water filtration through wetlands. (For more on ecosystem services, see Unit 9, "Biodiversity Decline.") Over-consumption uses up or severely depletes supplies of non-renewable resources, such as fossil fuels, and depletes renewable resources such as fisheries and forests if we use them up faster than they can replenish themselves (Fig. 11).



Figure 11. Land conversion for grazing in the Amazon rainforest

Courtesy National Aeronautics and Space Administration, Goddard Space Flight Center.

Second, we emit wastes as a product of our consumption activities, including air and water pollutants, toxic materials, greenhouse gases, and excess nutrients. Some wastes, such as untreated sewage and many pollutants, threaten human health. Others disrupt natural ecosystem functions: for example, excess nitrogen in water supplies causes algal blooms that deplete oxygen and kill fish. (For more on these pollutants, see Unit 8, "Water Resources"; Unit 10, "Energy Challenges"; Unit 11, "Atmospheric Pollution"; and Unit 12, "Earth's Changing Climate.")

Rising population growth rates in the 1950s spurred worries that developing countries could deplete their food supplies. Starting with India in 1951, dozens of countries launched family planning programs with support from international organizations and western governments. As shown above in Figure 4, total fertility rates in developing countries declined from six children per woman to three between 1950 and 2000. National programs were particularly effective in Asia, which accounted for roughly 80 percent of global fertility decline from the 1950s through 2000 (footnote 12). It is important to note, however, that this conclusion is controversial. Some researchers have argued that **desired** fertility falls as incomes grow—and that family planning has essentially no independent influence (footnote 13).

These programs sought to speed the demographic transition by convincing citizens that having large numbers of children was bad for the nation and for individual families. Generally they focused on educating married couples about birth control and distributing contraceptives, but some programs

took more coercive approaches. China imposed a limit of one child per family in 1979, with two children allowed in special cases (Fig. 12).



Figure 12. Poster advertising China's one-child policy, 1980s

Artist Zhou Yuwei. Courtesy of the International Institute of Social History Stefan R. Landsberger Collection, <http://www.iisg.nl/~landsberger>.

In some parts of China the one-child policy reportedly has been enforced through methods including forced abortions and sterilizations. Forced sterilizations also occurred in India in the 1970s. These policies have spurred some Indian and Chinese families to practice selective abortion and infanticide of female babies, since boys are more valued culturally and as workers. Population sex ratios in both countries are skewed as a result. In 2005 there were 107.5 males per 100 females in India and 106.8 males per 100 females in China, compared to a worldwide average of 101.6 males per 100 females. Females slightly outnumber males on every continent other than Asia (footnote 14).

Large societies consume more resources than small ones, but consumption patterns and technology choices may account for more environmental harms than sheer numbers of people. The U.S. population is about one-fourth as large as that of China or India, but the United States currently uses far more energy because Americans are more affluent and use their wealth to buy energy-intensive goods like cars and electronics. But China and India are growing and becoming more affluent, so their environmental impacts will increase because of both population size and consumption levels in the next several decades. For example, in 2006 China surpassed the United States as the world's largest emitter of carbon dioxide ( $\text{CO}_2$ ), the main greenhouse gas produced as a result of human activities (footnote 15).

Economies tend to become more high-polluting during early stages of economic development because they first adapt inexpensive technologies that are relatively inefficient—for example, simple manufacturing systems and basic consumer goods such as cars. As income rises and technologies diffuse through society, consumers start to value environmental quality more highly and become more able to pay for it.

Some analysts have argued that developing countries can skip the early stage of industrialization through "leapfrogging"—deploying advanced, clean technologies as soon as they are fielded in developed nations, or even earlier. For example, some developing countries have skipped past installing telephone poles and wires and moved straight to cell phones as a primary communication system. If fast-growing nations like China and India can leapfrog to clean technologies, they can reduce the environmental impacts of their large and growing populations (Fig. 13). However, many new technologies will not flow easily across borders in the absence of special efforts. Developed countries and international financial institutions can promote technology transfer to reduce the environmental impacts of growth in developing countries.



Figure 13. Youths installing solar panels to power a rural computer center, São João, Brazil

Courtesy United States Agency for International Development.

## 6. Urbanization and Megacities

The year 2007 marked a new milestone for human population growth: for the first time, more people worldwide were living in cities than in rural areas (footnote 16). This trend is driven by several factors: people migrating from rural regions to cities; rural areas being reclassified as urban areas because of population growth; and urban populations growing and expanding their boundaries to incorporate land that was formerly classified as rural.

Urbanization is a predictable outcome of industrialization and the demographic transition. As nations shift to mechanized agriculture they can produce more food with less labor, so there are fewer work opportunities in rural areas. More investment flows into industry and service businesses, which concentrate in urban areas near customers and infrastructure such as highways and telecommunications services. As these sectors become the main arenas of economic activity, people who once might have spent their lives in rural areas move to cities in search of higher-paying work and better living standards.

Nearly all of the urban growth projected in this century will occur in developing countries. About three-quarters of the world's urban population lives in small or intermediate-sized cities, which have fewer than 1 million inhabitants or 1 to 5 million inhabitants, respectively (footnote 17). Most urban dwellers will continue to live in small and medium cities, but there also will be major growth in megacities (more than 10 million people) and metacities (more than 20 million people). In the 1950s only New York and Tokyo were large enough to qualify as megacities, but population experts project that six cities—São Paulo, Mexico City, Mumbai (formerly Bombay), Delhi, Dhaka, and Tokyo—will have topped the 20 million mark by 2015, with New York, Jakarta, and Lagos close behind (Fig. 14).



Figure 14. The world's largest cities in 2015

Urban growth can contribute to sustainable development if it is managed effectively. Because cities concentrate economic activities and large numbers of people close together, the unit cost of providing basic infrastructure and services like piped water, roads, and sewage treatment is lower than in rural areas. Governments can make cities more efficient and livable by investing in public transportation systems and clean energy sources, and by planning ahead for growth so that they are able to provide basic services when populations expand.

But city life can also be dirty, unhealthy, and dangerous. Many people moving to cities, especially in the developing world, end up living in slums, just like earlier migrants to places like Manchester, England, and New York City's notorious Five Points slum in the 19th century. In 2007 the number of slum dwellers worldwide exceeds one billion, about one-third of all city residents, with more than 90 percent of slum dwellers residing in developing countries (footnote 18). Slums are a large and entrenched sector of many cities in the developing world. Some, like Brazil's favelas and South Africa's townships, have become sightseeing attractions for adventurous tourists.

Urban poverty can be as severe as rural poverty for people in slum neighborhoods who do not have access to the benefits of city life. The United Nations defines a slum as "a contiguous settlement where the inhabitants are characterized as having inadequate housing and basic services" such as drinking water and sanitation (footnote 19, Fig. 15). Slum dwellers typically live in crowded conditions without durable shelter or reliable access to safe drinking water or proper toilets. Many are not protected by tenants' rights, so they can easily be evicted or forced out and become homeless.



Figure 15. Shanty town in Manila beside Manila City Jail

Courtesy Mike Gonzalez. Wikimedia Commons, GNU Free Documentation License.

People who live in slums have lower life expectancies than their neighbors in more affluent areas, and more slum residents are killed or sickened by environmental hazards like indoor air pollution and water-borne or water-related diseases. (For more about these threats, see Unit 6, "Risk, Exposure, and Health," and Unit 8, "Water Resources.") Ironically, many slum dwellers use less energy and resources and generate less waste than their upscale neighbors, but the poor live in dirtier areas and receive fewer resources and services, so they bear the burdens generated by higher-income consumers.

The scale of urban poverty, already a pressing issue in many developing countries, may become even worse, as most population growth in the 21st century will happen in cities. Ameliorating the conditions described above is essential for sustainable human development. Some countries, including Brazil, Cuba, Egypt, South Africa, Sri Lanka, Thailand, and Tunisia, have reduced or limited slum growth. These governments have made serious political commitments to upgrading slum neighborhoods, improving housing, giving more people access to clean water and sanitation, preventing more "informal" settlements (shanty neighborhoods), and investing in services like education and transportation that benefit poor communities. And after decades of focusing on rural communities, international aid organizations are paying increasing attention to urbanization.



*"As I walk through the slums of Africa, I find it hard to witness children suffering under what can only be described as an urban penalty. I am astonished at how women manage to raise their families under such appalling circumstances, without water or a decent toilet. The promise of independence has given way to the harsh realities of urban living mainly because too many of us were ill prepared for our urban future. Many cities are confronting not only the problems of urban poverty, but the very worst of environmental pollution. From Banda Aceh to New Orleans, whole communities are being wiped out through no fault of the innocent victims."*

Anna Tibaijuka, Executive Director, UN-HABITAT(Worldwatch Institute,  
**State of the World 2007: Our Urban Future)**

## 7. Other Consequences of Demographic Change

The environmental consequences of population growth are a subset of broader interactions between population and national economic development. Scholars have debated for many years whether population growth helps or hinders economic growth. And the issue of economic growth (which is affected in various ways by demographic change) is relevant to the environment because people in wealthier nations are generally healthier, more interested in environmental quality, and better able to afford cleaner technologies than those in poor countries (although, as discussed above, they also tend to choose high-consumption lifestyles and generate relatively large amounts of waste and air, land, and water pollution per capita).

It may seem intuitively obvious that large family size is associated with poverty. Indeed, family planning efforts in the 1960s were spurred by worries that rapidly growing populations would lead to slower economic growth in developing countries. However, some scholars have argued that population growth helps the economy by stimulating innovation and providing bigger markets. Recent studies confirm that high fertility tends to slow economic growth and keep poor families poor, while declines in fertility reduce poverty (footnote 20). They also cite another important factor: the age structure of the population.

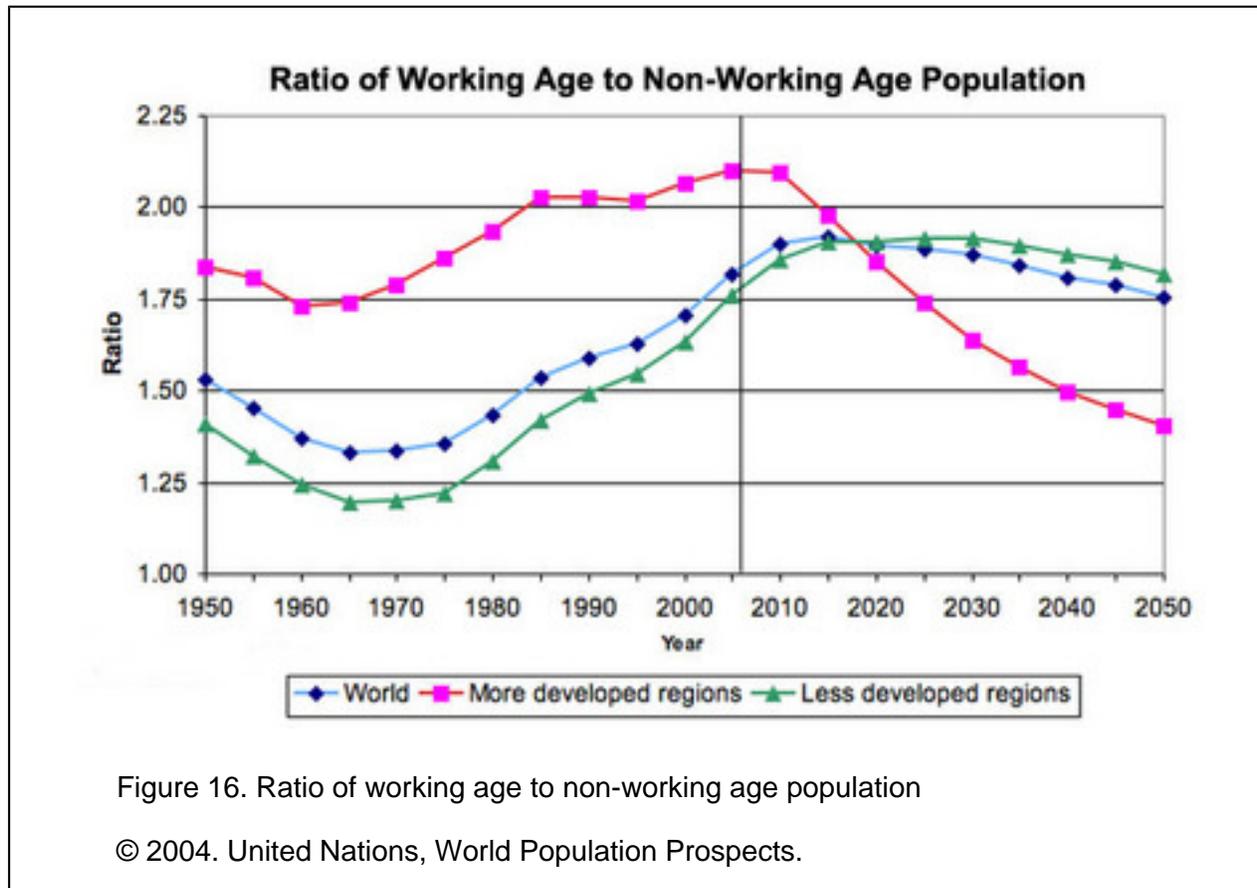
All societies can be divided into dependents (people who are too young or too old to work, so they depend on their families or on pensions for support) and workers who generate economic activity, generally defined as people ages 15 to 64. The ratio of dependents to the working-age population is called the **dependency ratio**. For example, as shown below in Table 2, there are 81 dependents (75 children and 6 elderly people) for every 100 working-age people in Africa. Developing countries tend to have higher dependency ratios than developed nations, and children account for a larger fraction of dependents in developing countries.

**Table 2.** Dependency ratios by region, 2005. (Because of rounding, Child + Old-age does not necessarily equal Total.)

Region	Total (Dependents per 100 working-age people)	Children per 100 working-age people	Old-age per 100 working-age people
<b>World</b>	55	44	11
<b>Africa</b>	81	75	6
<b>Latin America/ Caribbean</b>	57	47	10
<b>Oceania</b>	54	36	16
<b>Asia</b>	52	43	10
<b>North America</b>	49	31	18
<b>Europe</b>	47	23	23

Dependency ratios are key influences on economic growth. Nations with high dependency ratios spend large shares of their resources taking care of dependents, while those with lower ratios are able to devote more resources to investment in physical capital, technological progress, and education. When countries lower their fertility rates, they reduce the child component of the dependency ratio, which lightens the financial burden on wage earners and frees up more women to enter the work force.

Countries that reduce fertility rates have an important opportunity to reap a [demographic dividend](#). As discussed above in section 2, the demographic transition produces a "bulge" generation—a large cohort of people who are born after mortality rates fall but before fertility rates decline in response. Many developing countries are currently at this stage, with large numbers of people at or near working age and relatively few older dependents (Fig. 16).



Nations that have a particularly high ratio of working-age people to dependents can quickly build up capital and increase national per-capita income. Economists estimate that this demographic dividend accounted for roughly 20 to 40 percent of East Asia's economic boom between 1965 and 1990 (footnote 21). But the dividend does not pay out automatically. To earn it, nations must invest in education to train the large generation of young workers, and then manage their economies so that conditions are stable and workers can find rewarding jobs. Countries will not be able to reap the demographic dividend if they fail to create productive work opportunities.

The window of opportunity to earn a demographic dividend lasts for at least several decades, with the time depending largely on how quickly national fertility rates fall. As the boom generation matures and starts to retire from the work force, the dependency ratio goes up again since fewer workers follow in the wake of the boom cohort. In developed countries, the population of older citizens is growing more quickly than the population of workers. But nations that provide strong support for elderly citizens and that encourage workers to save for retirement may reap a second, longer-lasting demographic dividend spurred by such savings (footnote 22).

## 8. Demographic Convergence and Human Lifespan Trends

As the graphs in this unit illustrate, nations of the world appear to be converging demographically in several ways. The gaps between developed and developing countries for major indicators such as fertility rates and life expectancies have narrowed significantly in the past half-century and are projected to become smaller in coming decades. In 1950 life expectancy in developing countries as a whole lagged that of developed countries by 25 years. That gap has decreased to 12 years and is projected to be 8 years by 2050. Some developing countries have fertility and mortality rates that are lower than those in developed nations.

This trend indicates that the demographic transition is a widespread phenomenon and that life will continue to improve for people in poor countries as fertility rates decline further—especially if their governments adopt policies that recognize changing demographic realities, as discussed in section 7. But [demographic convergence](#) is not automatic. Some countries, including many in sub-Saharan Africa, are stuck in high fertility/low-growth traps, and others have hit speed bumps on the road toward longer, healthier lives. The HIV/AIDS pandemic has dramatically reduced life expectancy across Africa (Fig. 17). Life expectancy for men in Russia fell by more than seven years during the economic crisis and political instability that followed the breakup of the Soviet Union. Wars have left lasting marks on population distribution in some countries.

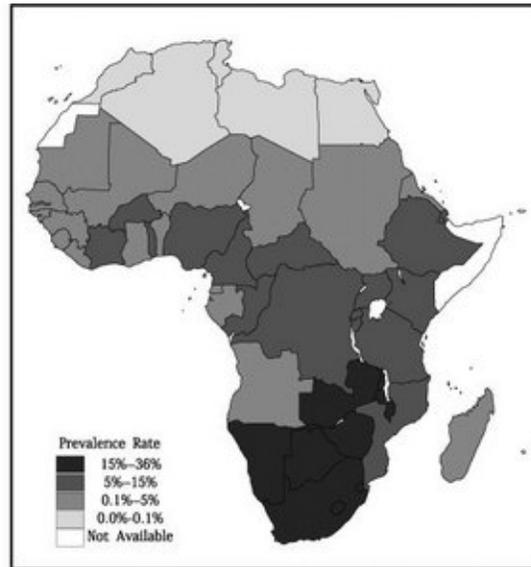


Figure 17. Adult HIV/AIDS rates in Africa, 2000

Courtesy United States Institute for Peace, <http://www.usip.org>.

Economist Jeffrey Sachs argues that extremely poor nations lag behind the rest of the world for several reasons. Impoverished areas such as sub-Saharan Africa, Central Asia, and the highlands of South America have climates and land resources that are poorly suited to large-scale agriculture. They also are economically and geographically isolated, and malaria is widespread in Africa.

Sachs and his colleagues estimate that if wealthy countries doubled their foreign aid spending from \$80 billion to \$160 billion per year, the world's poorest nations could cut poverty in half by 2015 and eliminate it by 2025. Key investments would address core environmental needs such as clean drinking water and sanitation, along with health, education, and food production (footnote 23). In sum, for very poor countries, population is just one of a set of issues that must be addressed to jump-start economic development.

Other experts contend that massive aid plans conceived by foreign experts and imposed from the top down by international agencies and wealthy donor nations have produced very poor returns and done little to reduce global poverty. Economist William Easterly writes of Sachs and other antipoverty advocates, "Poor people die not only because of the world's indifference to their poverty, but also because of ineffective efforts by those who do care." The right approach, in Easterly's view, is to

focus on smaller-scale tasks (such as delivering specific drugs to control specific diseases), relying on local channels and providers to the greatest extent possible (footnote 24).

In contrast to the situation in poor nations, most people in wealthy countries are living longer, healthier lives than at any time in history. This trend raises its own issues. For example, most of the acute illnesses that killed many people a century ago, such as tuberculosis, tetanus, and poliomyelitis, have been brought under control; one major killer, smallpox, has been eliminated. Most deaths in developed countries are now caused by chronic diseases such as cancer, heart disease, stroke, and chronic lower respiratory diseases such as emphysema. Tobacco use causes more deaths in the United States each year than HIV, alcohol use, illegal drug use, motor vehicle injuries, suicides, and murders combined (footnote 25).

Many chronic diseases develop slowly, and many are linked to personal choices such as diet. This means that effective public health programs must increasingly focus on long-term prevention and reducing risky behaviors such as smoking (Fig. 18).



Figure 18. Warning on a pack of British cigarettes

Courtesy Wikimedia Commons. Creative Commons Attribution ShareAlike 2.0 License.

If we bring chronic diseases under control, could humans live even longer than they do today? On average, 1 out of 10,000 people in developed countries lives beyond age 100; the longest documented human life was that of a French woman who died in 1997 at age 122. There is great scientific interest in exploring the limits of the human life span, although no agreement on a best means for extending life or how long humans could live under optimum conditions. Most scholars



believe that, absent major wars or unforeseen epidemics, life expectancy will increase during this century, to at least 85 in today's wealthy industrial countries, and perhaps to as high as 100.

## 9. Further Reading

Jeffrey Sachs, **The End of Poverty: Economic Possibilities For Our Time** (New York: Penguin, 2006). Economist Jeffrey Sachs offers a plan to eliminate extreme poverty around the world by 2025, focusing on actions to improve the lives of the world's one billion poorest citizens.

John Bongaarts, "How Long Will We Live?" **Population and Development Review**, vol. 32, no. 4 (December 2006), pp. 605–628. A look at the factors that have increased life expectancy in high-income countries since 1800 and at prospects for continued gains.

Joel E. Cohen, "Human Population Grows Up," **Scientific American**, September 2005, pp. 48–55. In the next 50 years, Earth's human population will be larger, slower-growing, more urban, and older than in the 20th century, with significant implications for sustainability.

"The Economics of Demographics," (whole issue) **Finance & Development**, vol. 43, no. 3, September 2006. A detailed look at policy adjustments that can help world leaders cope with demographic change.

Mike Davis, "Slum Ecology," **Orion**, March/April 2006. Living conditions in urban slums invert the principles of good urban planning: houses stand on unstable slopes, people live next to polluted and toxic sites, and open space is scarce or lacking.

Malcolm Gladwell, "The Risk Pool," **New Yorker**, August 28, 2006. Population age structures and dependency ratios explain Ireland's recent economic boom and the woes of many U.S. corporate pension plans.

Paul Harrison and Fred Pearce, **AAAS Atlas of Population and the Environment** (Berkeley: American Association for the Advancement of Science and University of California Press, 2000), <http://atlas.aaas.org/index.php?sub=intro>. An online source of information on the relationships between human population and the environment, with text, maps, and diagrams.

## Footnotes

1. J.P. Holdren and P.R. Ehrlich, "Human Population and the Global Environment," *American Scientist*, vol. 62 (1974), pp. 282–92.

2. Joel E. Cohen, **How Many People Can the Earth Support?** (New York: Norton, 1995), pp. 212–36, 261–62.

3. Dudley Kirk, "Demographic Transition Theory," **Population Studies**, Vol. 50, No. 3 (November 1996), pp. 381–87.

4. In popular usage, "fertility" means what demographers call "fecundity." This chapter uses "fertility" as demographers do.
5. Joseph A. McFalls, Jr., "Population: A Lively Introduction," **Population Bulletin**, December 2003, p. 5.
6. Population Reference Bureau, **2006 World Population Data Sheet**, <http://www.prb.org/pdf06/06WorldDataSheet.pdf>, pp. 5, 9.
7. *Ibid.*, pp. 5–10.
8. Joel E. Cohen, "Human Population Grows Up," **Scientific American**, September 2005, pp. 48–55.
9. A technical note: 2.1 is the long-run replacement level when the baby-boom generation has aged and the overall age structure has stabilized. Before then, a population can continue to grow with the total fertility rate at or below 2.1, depending on its age structure, a manifestation of the concept of population momentum described earlier.
10. David E. Bloom and David Canning, "Booms, Busts, and Echoes," **Finance & Development**, September 2006, p. 13.
11. Paul Harrison and Fred Pearce, **AAAS Atlas of Population and Environment** (Berkeley: American Association for the Advancement of Science and University of California Press, 2000), p. 7.
12. John C. Caldwell, James F. Phillips, and Barkat-e-Khuda, "The Future of Family Planning Programs," **Studies in Family Planning**, Vol. 33, No. 1, March 2002, p. 2.
13. Lant Pritchett, "Desired Fertility and the Impact of Population Policies," **Population and Development Review**, Vol. 1, No. 20, March 1994, pp. 1–55.
14. United Nations, Department of Economic and Social Affairs, **World Population Prospects: The 2006 Revision**, Population Database, <http://esa.un.org/unpp/index.asp?panel=2>.
15. Netherlands Environmental Assessment Agency, "China Now No. 1 in CO Emissions; USA In Second Position," press release, June 19, 2007.
16. United Nations Human Settlements Programme (UN-HABITAT), **State of the World's Cities 2006/7** (London: Earthscan, 2006), p. viii.
17. *Ibid.*, p. 5.
18. UN-HABITAT, **State of the World's Cities 2006/7**, p. 5.
19. United Nations Statistics Division, [http://unstats.un.org/unsd/cdb/cdb\\_dict\\_xrxx.asp?def\\_code=487](http://unstats.un.org/unsd/cdb/cdb_dict_xrxx.asp?def_code=487).
20. Nancy Birdsall and Steven W. Sinding, "How and Why Population Matters: New Findings, New Issues," in Nancy Birdsall, Allen C. Kelley, and Steven W. Sinding, eds., **Population Matters** (Oxford University Press, 2003), p. 14.



21. "Banking the 'Demographic Dividend,'" **Rand Policy Brief**, RB-5065-WFHF-DLPF-RF (2002); David E. Bloom and Jeffrey Williamson, "Demographic Transitions and Economic Miracles in Emerging Asia," **World Bank Economic Review**, Vol. 12, No. 3 (1998), pp. 419–55.
22. Ronald D. Lee and Andrew Mason, "What is the Demographic Dividend?" **Finance & Development**, September 2006, pp. 16–17.
23. Jeffrey D. Sachs, "Can Extreme Poverty Be Eliminated?" **Scientific American**, September 2005, pp. 56–65.
24. William Easterly, **The White Man's Burden: Why the West's Efforts To Aid the Rest Have Done So Much Ill and So Little Good** (New York: Penguin, 2006), p. 7.
25. U.S. Centers for Disease Control and Prevention, "Tobacco-Related Mortality," fact sheet, September 2006.

## Glossary

**carrying capacity** : The number of individuals an environment can support without significant negative impacts to the given organism and its environment.

**demographic convergence** : When the gaps narrow between developed and developing countries for major indicators such as fertility rates and life expectancies.

**demographic dividend** : A rise in the rate of economic growth due to a rising share of working age people in a population.

**demographic transition** : The pattern of population growth exhibited by the now-developed countries during the 19th and early 20th centuries.

**dependency ratio** : The ratio of non-workers (children and retirees) to workers in a human population: the higher the ratio, the greater the dependency load.

**fecundity** : A measure of the capacity of an organism to produce offspring.

**fertility** : A measure of reproduction: the number of children born per couple, person, or population.

**life expectancy** : Term usually used at birth, indicating the average age that a newborn can be expected to attain.

**migration** : When living organisms move from one biome to another. It can also describe geographic population shifts within nations and across borders.

**mortality** : The loss of members of a population through death.

**population momentum** : The impetus for continued expansion of the number of people in a country when the age structure is characterized by a large number of children. Even if birth control efforts are



effective in the adult community and the number of new births per person decreases, the number of people in the country expands as the large population of children reach reproductive age.

**replacement level** : The number of children per woman necessary to keep population levels constant when births and deaths are considered together over time; estimated to be an average of 2.1 children for every woman.

## Unit 6 : Risk, Exposure, and Health



Tar Creek Superfund Site.

### Overview

We are exposed to numerous chemicals every day from environmental sources such as air and water pollution, pesticides, cleaning products, and food additives. Some of these chemicals are threats to human health, but tracing exposures and determining what levels of risk they pose is a painstaking process. How do harmful substances enter the body, and how do they damage cells? Learn how dangers are assessed, what kind of regulations we use to reduce exposures, and how we manage associated human health risks.

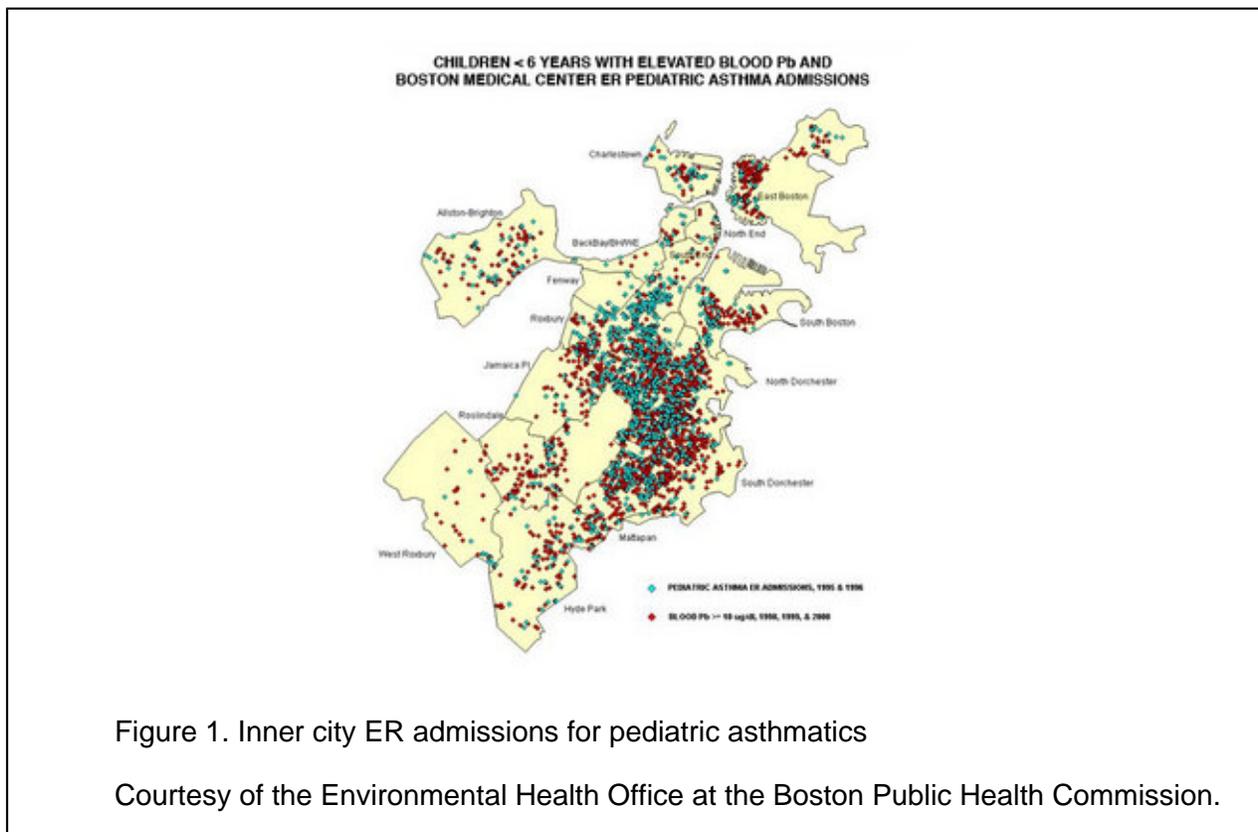
### Sections:

1. Introduction
2. Risk Assessment
3. Measuring Exposure to Environmental Hazards
4. Using Epidemiology in Risk Assessment
5. Cancer Risk
6. Other Risks
7. Benefit-Cost Analysis and Risk Tradeoffs
8. Risk Perception
9. The Precautionary Principle
10. Major Laws
11. Further Reading

## 1. Introduction

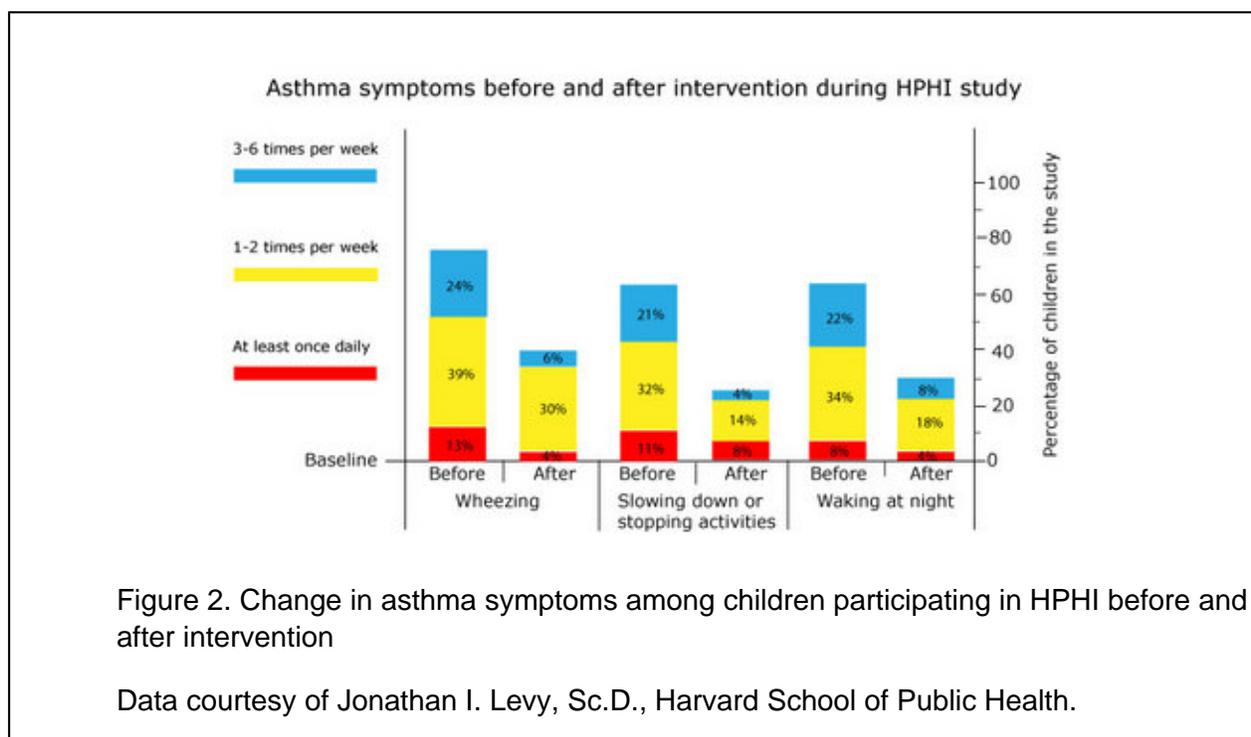
Kayla is a normal teenager except she has asthma, a chronic condition of the airways that makes it difficult for her to breathe at times. Allergens such as pollens, dust mites, cockroaches, and air pollution from cigarettes, gas stoves, and traffic make asthmatics' airways swell so that only limited amounts of air can pass through and respiration becomes a struggle akin to breathing through a tiny straw.

Growing up poor and black in Boston, Kayla is part of an epidemic that has seen the asthma prevalence rate for children rise from 3.6 percent in 1980 to 5.8 percent in 2005 (footnote 1). Asthma incidence has risen in many industrialized countries around the world (Fig 1), but it is much more common among children living in inner cities. Children like Kayla living in Roxbury and Dorchester, Massachusetts, are five times more likely to be hospitalized for asthma than children living in wealthier white sections of Boston.



Starting in 2001, the Healthy Public Housing Initiative (HPHI), a collaboration between Harvard, Tufts, and Boston University, worked with the Boston housing authority and tenant organizations to

conduct test interventions aimed at reducing the suffering of children with asthma. HPHI reduced allergen exposures by thoroughly cleaning apartments, educating mothers about pest controls, implementing integrated pest management (discussed in Unit 7, "Agriculture"), and providing dust-mite reducing mattresses. Symptoms decreased and quality of life measurements improved for Kayla and other asthmatic children living in three public housing developments during a year of follow-up assessments after the interventions (Fig. 2) (footnote 2).



We are exposed to environmental contaminants from conception to our last breath. Some of these materials are naturally-occurring substances such as dust, pollen, and mold, while others are manmade chemicals used for numerous industrial and commercial purposes. As of 2006, the U.S. Environmental Protection Agency (EPA) estimated that about there were 15,000 chemicals in commerce (footnote 3).

Some contaminants have been demonstrated to have harmful effects on various human organs, such as the reproductive or respiratory systems, or on functions such as fetal development. Based on evidence from toxicological, ecological, and [epidemiological](#) studies, health experts suspect many more contaminants of being possible risks to humans. The EPA screens chemicals that it believes are the greatest potential threats to human health and the environment, but most of the chemical compounds that are already in wide use today have been subject to little or no toxicological testing. Virtually none has been tested for potential as [endocrine disruptors](#).

In complex modern societies, the most critical environmental health challenge is defining a balance between the social and economic benefits that materials and technologies provide on one hand and risks to public health on the other hand. Numerous materials, from food additives to pesticides to manufacturing inputs, have valuable uses but may also threaten the health of the general public or smaller high-risk groups. In many cases such threats can be managed by setting usage guidelines or limiting exposure. In extreme cases they may require taking materials off of the market. Tetraethyl lead, asbestos, DDT, and PCBs are some examples of widely used substances that have been proven harmful (Fig. 3).

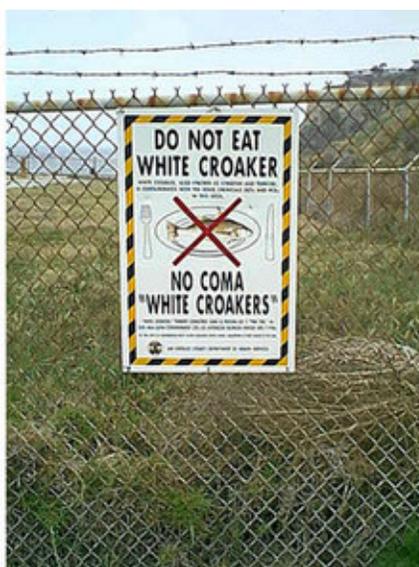


Figure 3. Warning sign, Palos Verdes Peninsula, California

Courtesy United States Environmental Protection Agency.

Health experts approach these tradeoffs by using **risk assessment** to systematically evaluate scientific, engineering, toxicological, and epidemiological information on specific environmental hazards. Next they use this factual analysis to develop strategies, such as standards, regulations, and restrictions, that reduce or eliminate harm to people and the environment, a process referred to as **risk management**. Risk management takes into consideration both the benefits and the costs of controlling or eliminating hazards. It weighs the strength of the scientific evidence along with the social and economic implications of controlling or not controlling environmental risks.

This process has limitations. Epidemiological studies cannot establish causal relationships between exposure and harm. Most toxicological studies carried out in laboratories use artificially high doses



to evoke responses within reasonable time periods, whereas real exposures to environmental contaminants often involve low-level exposures over very long time frames. And real exposures almost always involve mixtures of contaminants, such as heavy metals in mine drainage. The time course of exposures and doses is complex, both for individuals and for the population at large: levels, frequency, and intensity of exposure all can affect toxicity.

*"We have very good ideas of what individual toxicants can do to people. However, you cannot predict what the ultimate human health impacts might be from simply knowing what the individual toxicants can do. Mixtures can interact in ways that are unforeseen and give you toxic ramifications that are much greater than what can be predicted from the single exposures. On the other hand, in some mixtures toxicants can cancel each other out. So this has to be studied well and properly to understand what the real risks are."*

Howard Hu, University of Michigan/Harvard University

Genetic variability in the population adds to the uncertainty of risk assessment. Interactions between humans' genetic makeup and their environment take many forms, including characteristics that either protect individuals from specific risks or make them more susceptible. Both inherited genetic traits and environmental exposures can create genetic susceptibilities, which can then be transferred from one generation to another.

To be effective, risk management must take these uncertainties and sources of variability into account in developing strategies. Managing risks also involves political and philosophical issues. Governments have often acted regardless of the actual magnitude of a risk because of risk perceptions on the part of special interest groups or the general public.

This unit describes the risk assessment process and the central role of epidemiology—studying associations between exposure, risk factors, and outcomes. It then shows how public health experts use evidence to assess cancer and noncancer risks associated with environmental exposures. Next we look at the challenge of balancing risks and benefits and of assigning economic value to proposed environmental actions. The unit concludes with a discussion of the [Precautionary Principle](#), a sometimes-controversial approach to managing health and environmental risks with incomplete knowledge, and with brief summaries of relevant laws and regulations.

## 2. Risk Assessment

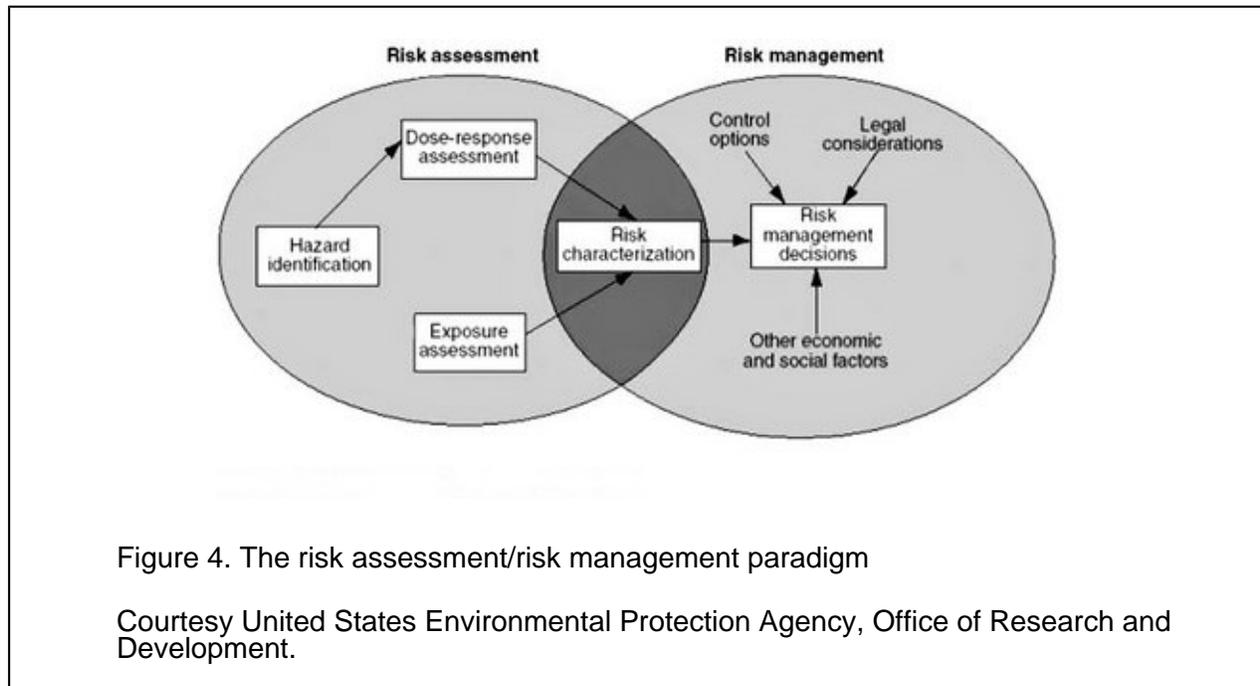
Risk assessment is the process of establishing risks to humans and the environment from chemicals, radiation, technologies, or other contaminants and agents that can affect health and well-being. It is part of a broader process called risk analysis that also includes developing policies to manage risks once they are identified and quantified.



As summarized by the Society for Risk Analysis, a professional association of experts, "Risk analysis uses observations about what we know to make predictions about what we don't know. Risk analysis is a fundamentally science-based process that strives to reflect the realities of Nature in order to provide useful information for decisions about managing risks . . . . [It] seeks to integrate knowledge about the fundamental physical, biological, social, cultural, and economic processes that determine human, environmental, and technological responses to a diverse set of circumstances" (footnote 4).

Health and environmental experts use risk analysis to assess many types of threats, from infectious agents to noise pollution. The process has several components (Fig. 4).

- Risk assessment: Scientists identify hazards, determine dose-response relationships, and estimate actual or projected exposures. These steps lead to an estimate of overall risk to the general population or target groups.
- Risk management: Experts develop options for limiting estimated risk. Unlike risk assessment, which is based on scientific findings, risk management takes political and economic factors into account along with technical considerations.
- Risk communication: Policy makers discuss the problem and options for addressing it with the public, then incorporate the feedback that they receive into their decisions. As discussed below in section 7, "Benefit-Cost Analysis and Risk Tradeoffs," effective risk communication helps to ensure that decisions will be broadly acceptable.

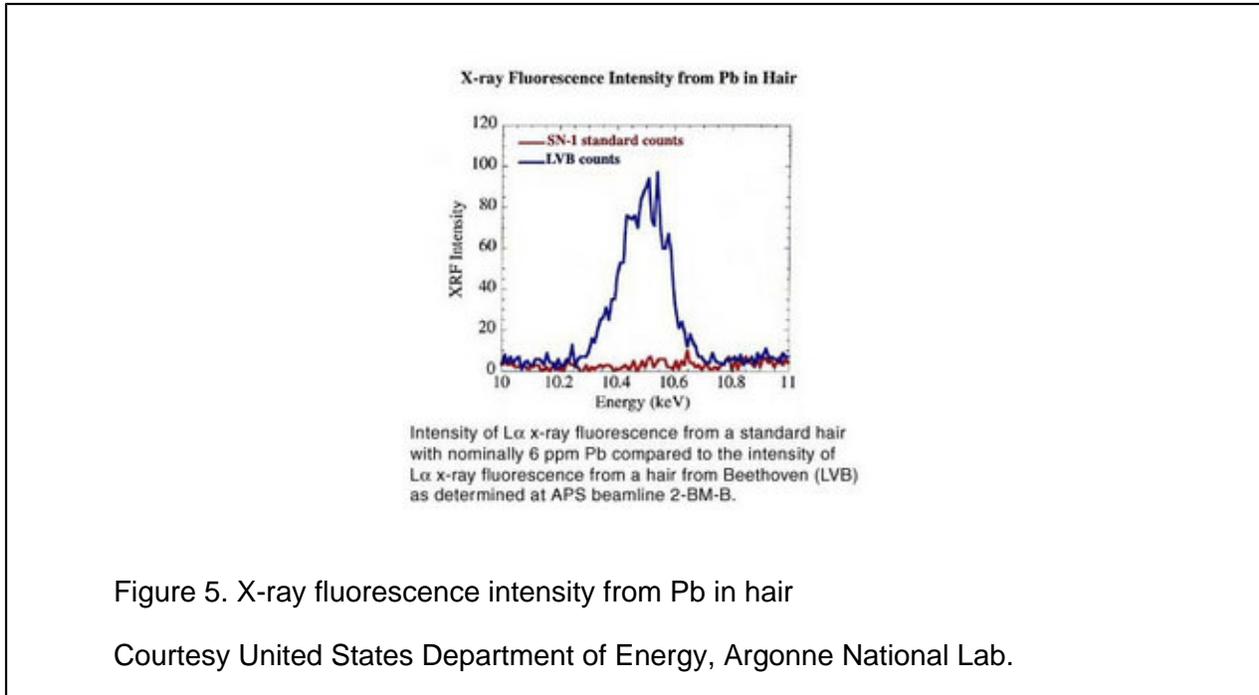


Risk assessment has been in use since the 1950s but has become more sophisticated and accurate over the past several decades, due in large part to increasing interest from government regulators. In the 1960s and 1970s, federal authority to regulate threats to health, safety, and the environment expanded dramatically with the creation of new oversight agencies such as the Environmental Protection Agency (EPA) and the Occupational Safety and Health Administration (OSHA), along with adoption of numerous laws regulating environmental hazards. At the same time, improved testing methods and better techniques for detecting contaminants made it easier to study relationships between exposure and health effects.

These developments made it easier in some ways to protect public health and the environment, since regulators at the new agencies had broad mandates for action and abundant data about potential threats. But regulators had to allocate their resources among many competing issues, so they needed tools to help them focus on the most dangerous risks. Former EPA administrator William K. Reilly recalls, "Within the space of a few years, we went to the possibility of detecting not just parts per million but parts per billion and even, in some areas, parts per quadrillion . . . . That forces you to acknowledge that what you need is some reasonable method for predicting levels of real impact on humans so that you can protect people to an adequate standard" (footnote 5).

As an illustration of the power of modern analytical methods, Figure 5 shows results from a X-ray analysis of a strand of composer Ludwig van Beethoven's hair performed in the year 2000 by the U.S. Department of Energy's Argonne National Laboratory. The experiment found lead levels of about 60

parts per million in Beethoven's hair, compared to less than six parts per million for an average U.S. human hair today, indicating that some of Beethoven's lifelong illnesses may have been due to lead poisoning.



Risk analysis gave scientists and regulators a way to sort through the vast amounts of health information provided by methods like that illustrated in Fig. 5, compare relative risks from various contaminants, and set priorities for action. By the mid-1970s a number of federal agencies were carrying out risk assessments, each using its own procedures and standards.

To address concerns about inconsistencies among agencies, Congress requested a study from the National Academy of Sciences, which in 1983 published a seminal report, **Risk Assessment in the Federal Government: Managing the Process** (often referred to as the "Red Book") (footnote 6). This study provided a general framework for cancer risk assessment and recommended developing uniform risk assessment guidelines for agencies. Although no government-wide guidelines have been produced, EPA has produced numerous assessments of human health risks from exposure to substances such as air pollutants and drinking water contaminants. The Office of Management and Budget, which oversees U.S. regulatory policies, requires EPA and other federal agencies to submit comprehensive risk assessments and benefit-cost analyses along with proposed rule makings and regulations.

Following a model outlined in the Red Book, environmental risk assessments typically include four steps.

- Hazard identification: Determining whether or not exposure to an agent causes health problems. Researchers often address this question by testing the agent to see whether it causes cancer or other harmful effects in laboratory animals.
- Dose-response assessment: Characterizing the relationship between receiving a dose of the agent and experiencing adverse effects. Analysts often have to extrapolate from high laboratory doses to low actual doses and from laboratory animals to humans.
- Exposure assessment: Measuring or estimating how often humans are exposed to the agent, for how long, and the intensity of exposure. This can involve methods such as asking subjects about their lifestyles and habits; taking environmental samples; and screening subjects' blood, urine, hair, or other physical samples to measure concentrations of the agents in their bodies (Fig. 6).
- Risk characterization: Combining exposure and dose-response assessments to estimate health impacts on subjects.

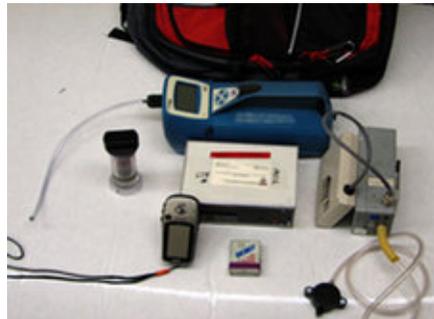


Figure 6. Backpack system for measuring exposure to fine particulate air pollution

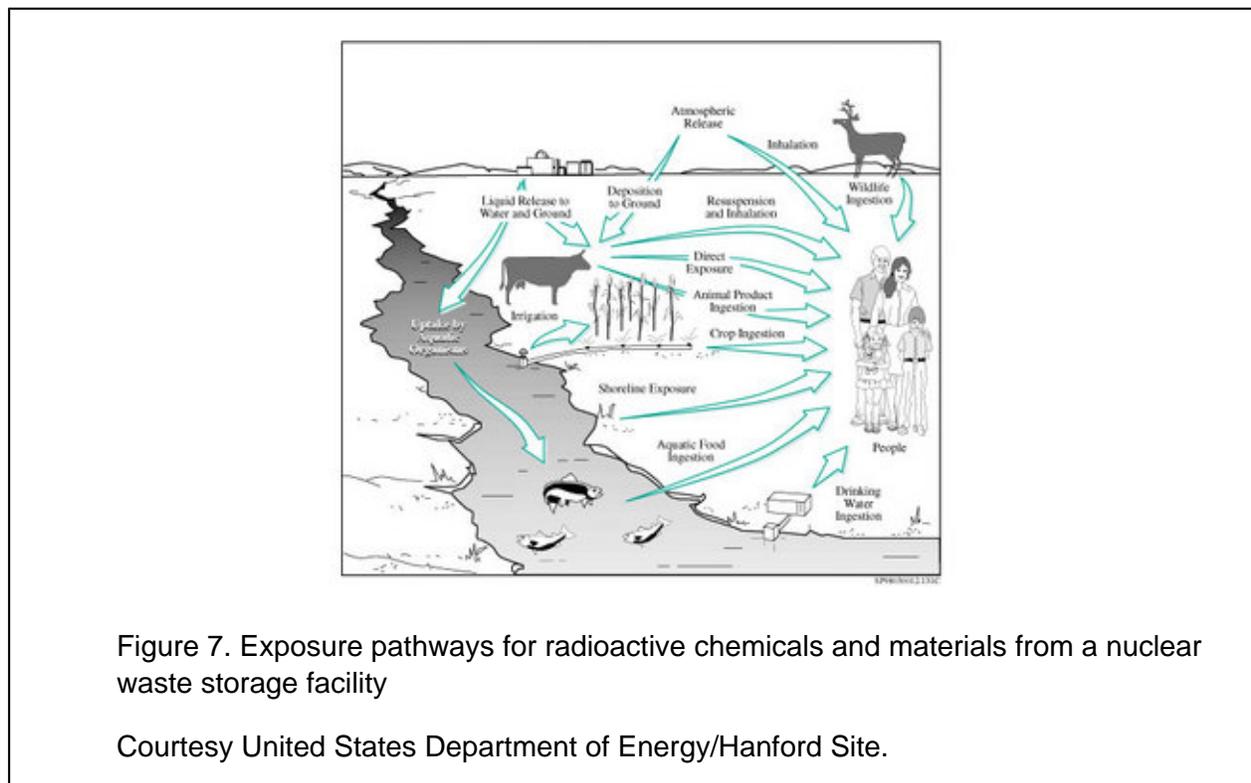
© John Spengler, Harvard School of Public Health.

### 3. Measuring Exposure to Environmental Hazards

Many hazardous materials are present in our environment, but some are more likely to cause actual harm than others. Humans come into contact with harmful agents in many ways. For example, we may inhale gases and particulates as we breathe, eat fruit that carries pesticide residues, drink polluted water, touch contaminated soils, or absorb radiation or chemical vapors through our skin. In each case, risk analysts want to measure several variables.

- **Exposure:** Contact between a contaminant and the exterior of an exposed person's body (skin and openings into the body such as mouth, nostrils, and cuts or breaks in the skin).
- **Intake or uptake:** The processes through which contaminants cross the boundary from outside to inside the body. Intake refers to processes like ingestion and inhalation that physically move the agent through an opening in the outer body, such as the mouth, nose, or a skin puncture. Uptake involves absorption of agents through the skin.
- **Dose:** The amount of contaminant that is inhaled or ingested into an exposed person's body or applied to the skin (potential dose), and the fraction of this dose that is absorbed and becomes available to impact biologically significant sites inside the body (internal dose) ().

Exposure assessments describe how frequently contact occurs, how long it lasts, its intensity (i.e., how concentrated the contaminant is), and the route by which contaminants enter the body (Fig. 7). They may also estimate dose, although if there is a known relationship between exposure to a specific hazard and how the body responds, a study may simply estimate the target group's exposure and use existing knowledge to calculate the average dose members have received.



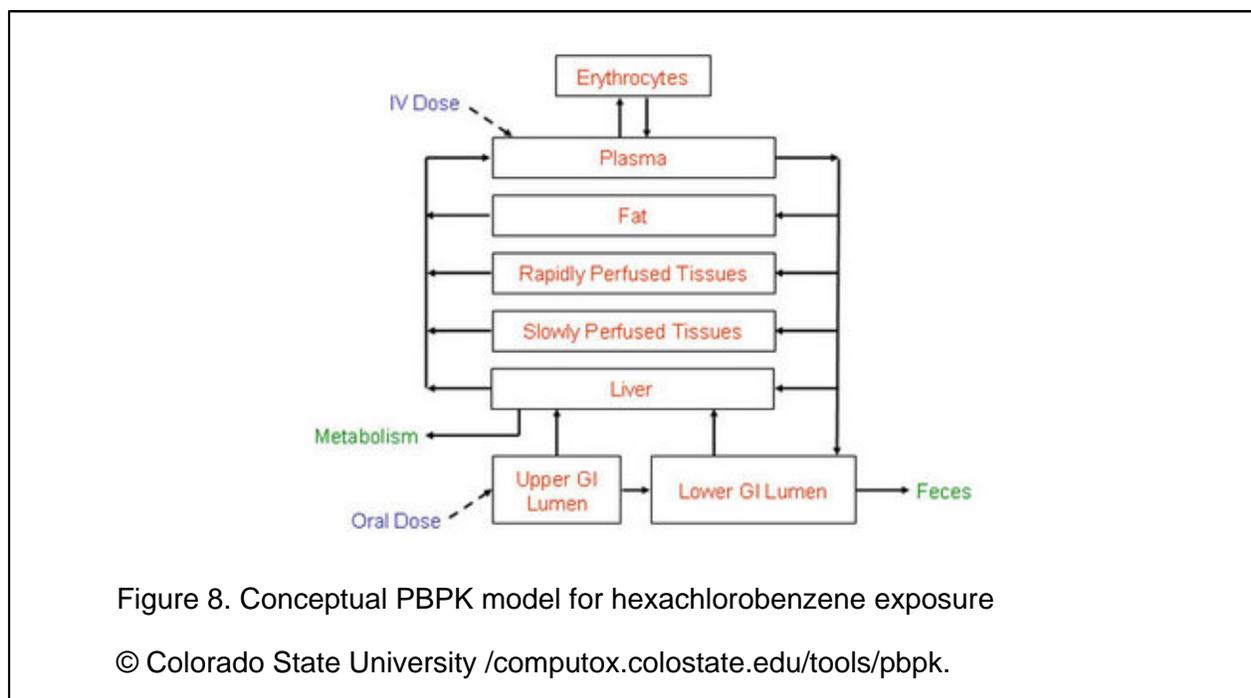
As this summary indicates, exposure assessment is a painstaking multi-step process that requires a lot of data. Researchers need to know the contaminant's physical and chemical properties, the form in which it occurs locally, the medium by which it comes into contact with humans, and how concentrated it is within that medium. They also need to know the demographics of the exposed population, major routes of exposure for that group, and relevant behavior and lifestyle issues, such as how many people smoke cigarettes or filter their tap water. And calculating the human impact of contact with hazardous agents requires detailed knowledge of physiology and toxicology.

Even when people ingest a contaminant or absorb it through their skin, much analysis is required to determine how they may be affected. Once an internal dose of a chemical is absorbed into the bloodstream, it becomes distributed among various tissues, fluids, and organs, a process called **partition**. Depending on the contaminant's physical and chemical properties, it can be stored, transported, metabolized, or excreted. Many contaminants that are highly soluble in water are excreted relatively quickly, but some, such as mercury, cadmium, and lead, bind tightly to specific organs. Agents that are not highly soluble in water, such as organochlorine insecticides, tend to move into fatty tissues and accumulate.

The portion of an internal dose that actually reaches a biologically sensitive site within the body is called the **delivered dose**. To calculate delivered doses, researchers start by mapping how

toxic substances move through the body and how they react with various types of tissues. For example, combustion of diesel fuel produces a carcinogenic compound called 1,3-butadiene. When humans inhale this colorless gas, it can pass through the alveolar walls in the lungs and enter the bloodstream, where it binds readily to lipids and is likely to move to other parts of the body. Experimental studies have shown that subjects who ate ice cream with a high fat content a few hours before inhaling low concentrations of 1,3-butadiene had reduced levels of the compound in their exhaled breath, demonstrating that more of the gas could partition to the lipid fraction of the body.

The delivered dose is the measurement most closely related to expected harms from exposure, so estimating delivered doses is central to exposure assessment. The most common methods are measuring blood concentrations or using PBPK (Physiologically-Based Pharmacokinetic) models. This approach simulates the time course of contaminant tissue concentrations in humans by dividing the body into a series of compartments based on how quickly they take up and release the substance. Using known values for physical functions like respiration, it estimates how quickly the agent will move through a human body and how much will be stored, metabolized, and excreted at various stages. Figure 8 shows a conceptual PBPK model (without calculated results) for intravenous or oral exposure to hexachlorobenzene, a synthetic pesticide.



Even when it relies on techniques like PBPK modeling, exposure assessment requires analysts to make assumptions, estimates, and judgments. Scientists often have to work with incomplete data. For example, in reconstructing exposures that have already taken place, they have to determine

how much of a contaminant may have been ingested or inhaled, which can be done by interviewing subjects, analyzing their environment, or physical testing if exposure is recent enough and the contaminant leaves residues that can be measured in blood, hair, or other biological materials. Some contaminants are easier to measure precisely in the environment than others, and relevant conditions such as weather and soil characteristics may vary over time or across the sample area.

To help users evaluate their results, exposure assessments include at least a qualitative description (plus quantitative estimates in some cases) of uncertainty factors that affect their findings. Addressing uncertainty ultimately makes the process of risk analysis stronger because it can point out areas where more research is needed and make a individual study's implications and limitations clear. As the EPA states in its current exposure assessment guidelines, "Essentially, the construction of scientifically sound exposure assessments and the analysis of uncertainty go hand in hand" (footnote 8).

#### 4. Using Epidemiology in Risk Assessment

When scientists perform risk analyses, the best source of information on specific contaminants' health effects is data from epidemiologic studies. Epidemiologists analyze how health-related events are distributed in specific human populations—who gets sick with what illnesses, when, and where. By comparing groups with different illness rates and looking at demographic, genetic, environmental, and other differences among these groups, epidemiologists seek to determine how and why certain groups get sick. These studies are designed to inform public health policies and help prevent further harm.

Epidemiologists may consider many possible determinants to explain patterns of illness, including physical, biological, social, cultural, and behavioral factors. In each case they seek to explain associations between certain exposures, risk factors or events, and illnesses or outcomes. Over the past half-century epidemiological studies have documented linkages between smoking and lung cancer, intravenous drug use and HIV/AIDS infection, and poor indoor air quality and health problems, to cite just a few examples.

To explore these associations, analysts have two basic study design options. **Cohort studies** follow a group of individuals who share some common characteristic such as age, place of residence, or exposure to a hazard, and study the frequency of illness in this group to see how strongly certain risk factors are associated with becoming sick. Researchers may also follow a control group that does not share the common factor with the cohort that is the study's subject. Whether they involve one group or two, cohort studies start with exposures and follow subject through time to find the outcomes.

For example, scientists have studied survivors of the Hiroshima and Nagasaki bombings to see how atomic bomb radiation exposure affects cancer rates in survivors and the incidence of genetic effects in survivors' children. Researchers in the Framingham Heart Study, launched in 1948, have assessed over 10,000 participants from Framingham, Massachusetts, spanning several generations to identify major risk factors for cardiovascular disease (Fig. 9). Many epidemiologic studies focus on

workplace exposures, which are generally higher and more frequent than other human exposures to environmental contaminants and therefore are more likely to show associations between exposure and illness.



Figure 9. Four generations from one family participating in the Framingham Heart Study and associate studies

© Tobey Sanford.

In contrast, **case-control studies** enroll a group of people who already have the disease of interest (the case group) and a group of people who do not have the disease but match the case group members as closely as possible in other ways (the control group). Researchers then work backwards to identify risk factors that may have caused the case group to get sick, and compare the groups to test how strongly these risk factors are associated with illness. Case-control studies start with the outcome and look backward to explain its causes.

In an early example of a case-control study, anesthesiologist John Snow investigated an 1854 cholera epidemic in London by mapping where victims lived, then marking the sites of public water pumps on the map (Fig. 10). Unlike area health authorities, Snow believed that contaminated water was a source of infection. Pump A, the Broad Street Pump, lay at the center of a cluster of cholera cases. Snow determined through interviews that other nearby pumps, which he labeled B and C, were used much less frequently than the Broad Street pump, and that all of the local cholera patients had consumed water from Pump A. Accordingly, Snow concluded that Pump A was the source of the infection. When he convinced local officials to remove the pump handle, cholera cases (which were already declining) stopped (footnote 9).



Figure 10. Snow's original map (shows cases of cholera around water pumps)

Courtesy Wikimedia Commons. Public Domain.

Each of these approaches has strengths and weaknesses. Cohort studies let researchers see how outcomes develop over long periods of time, but they require large groups to make the findings statistically significant and are expensive to administer. Case-control studies are a more effective way to study rare diseases, since researchers can select members of the exposed group instead of waiting to see which members of a cohort contract the disease, and are quicker and less expensive than cohort studies. However, since they usually look backward in time to reconstruct exposures, results may be skewed by incomplete data or participants' biased recollections.

Even if an exposure and a disease are associated, researchers cannot automatically assume that the exposure causes the disease. In 1965, pioneering British epidemiologist and statistician A.B. Hill proposed nine criteria for citing causal relationships between environmental threats and illness.

- Strength: Groups exposed to the threat have much higher rates of illness than unexposed groups.
- Consistency: The association is detectable consistently in different places, times, and circumstances by different observers.



- Specificity: The association is limited to well-defined groups, particular situations, and specific illnesses.
- Temporality: It is clear over time that the threat occurs first and leads to the outcome.
- Biological gradient: A consistent relationship exists between the size of dose and the scale of response.
- Plausibility: The proposed causal relationship makes biological sense.
- Coherence: The relationship does not conflict seriously with existing historical and scientific knowledge of the disease.
- Experiment: An experimental step (such as shutting down the Broad Street Pump) produces results that support the existence of a causal relationship.
- Analogy: The association is similar to documented causal relationships between threats and diseases .

What if the risk comes from a chemical that has not been studied yet, or has only been studied in a few small groups? In such cases analysts use information from animal toxicology studies, which can measure associations between contaminants and health effects in thousands of animal subjects quickly and inexpensively (relatively speaking—major animal studies can take several years and cost millions of dollars).

But animal data also has its drawbacks. Toxicology studies typically use large doses to produce a measurable response quickly, while environmental exposures usually occur at low levels over long periods of time, so analysts have to extrapolate from high study doses to low real-world doses. They also have to extrapolate from observed results in animals to expected results in humans, which assumes that a contaminant will affect humans in the same way. However, epidemiology and animal studies can inform each other. For example, if epidemiologic studies show that workers in a specific industry are developing cancer at higher than normal rates, researchers may carry out animal studies to see whether a specific material that those workers use causes illness.

## 5. Cancer Risk

Cancer is a major focus of environmental risk analysis for several reasons. First, it is a leading cause of death in developed countries that have passed through the demographic transition and brought other threats such as infectious disease and malnutrition under control (for more details, see Unit 5, "Human Population Dynamics"). Various types of cancer account for 25 percent or more of yearly

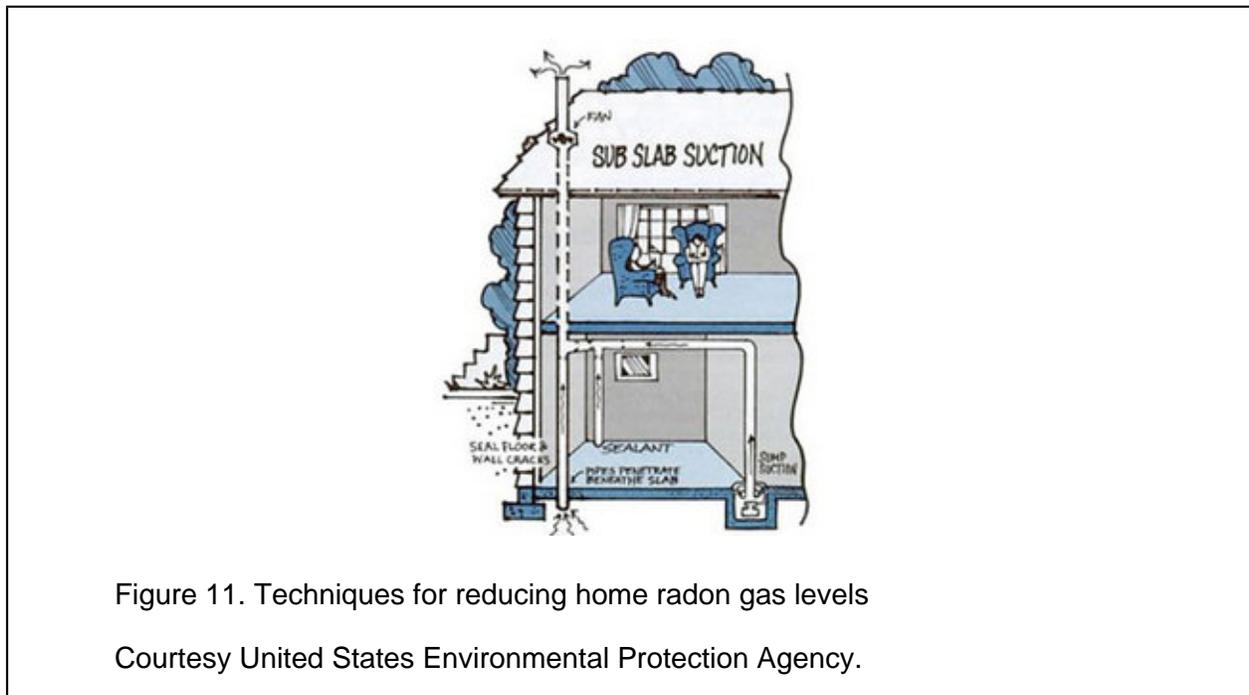


deaths in the United States and other industrialized nations. Cancer rates are also increasing in the developing world.

Second, environmental exposures broadly defined account for a substantial fraction of cancers—at least two-thirds of all cases in the United States, according to the National Institutes of Health (footnote 11). This estimate includes all influences outside the body, including many lifestyle choices such as smoking and eating a high-fat diet. Tobacco use alone causes about one-third of all annual U.S. cancer deaths, while inactivity and obesity together cause an estimated 25 to 30 percent of several major types of cancer (footnote 12).

In contrast, the narrower category of exposure to environmental pollutants causes about 5 percent of annual U.S. cancer deaths (footnote 13). However, these risks are not spread equally across the population. They have higher impacts on heavily-exposed groups—for example, workers in industries that use known or possibly carcinogenic substances or communities that draw their drinking water from a contaminated source. Environmental exposures also can cause gene alterations that may lead to cancer over time.

Risk analyses have led to bans or use restrictions on carcinogens such as benzene (a solvent), asbestos (an insulating fiber), and a number of pesticides, and have contributed to the development of guidelines and workplace standards that minimize exposure to other known or suspected carcinogens. Figure 11 shows one example, an illustration from an EPA brochure on reducing radon gas levels in houses. Exposure to radon, a natural byproduct of radioactive elements decaying in surrounding soil, causes an estimated 20,000 lung cancer deaths in the United States annually.



The Environmental Protection Agency and other regulators quantify cancer risks as probabilities—the number of excess individual lifetime cases of cancer (beyond those that could be expected to occur on average in the population) that will occur in response to a specific exposure. For example, in 1999 EPA estimated that the added cancer risk from polychlorinated biphenyl (PCB) pollution in the upper Hudson River was one additional case of cancer for every 1,000 people who ate one meal per week of fish caught in that section of the river (footnote 14). As this approach suggests, not everyone exposed to a hazard becomes ill, but exposure increases the likelihood of suffering harmful effects.

EPA's traditional classification system for carcinogens combines human data, animal data, and other supporting evidence to characterize the weight of evidence regarding whether a substance may cause cancer in humans (Table 2). However, these rankings are based on levels of certainty that agents may cause cancer, not on relative levels of risk from one substance versus another, so other materials not currently classified as carcinogens may be equally hazardous. Some materials are classified as possible or probable carcinogens because they have not been studied thoroughly enough yet to make a determination about whether they cause cancer in humans (footnote 15).

One of the most controversial issues in cancer risk assessment is whether the dose-response relationship for all carcinogens is linear. Most risk analyses assume that the answer is yes—in other words, that exposure to any amount of a carcinogen produces some risk of cancer, with risk increasing in proportion to the size of the dose. Under this approach, risk is estimated using the equation



Risk = LADD x CSF

where risk is the unitless probability of an individual developing cancer, LADD is the lifetime average daily dose per unit of body weight (milligrams per kilogram of body weight per day), and CSF is the cancer slope factor, or the risk associated with a unit dose of a carcinogen, also called the cancer potency factor (mg/kg-day)<sup>-1</sup>. The CSF usually represents an upper bound estimate of the likelihood of developing cancer, based on animal data (footnote 16).

Assuming a linear dose-response relationship has major implications for regulating carcinogens because it indicates that even very low exposure levels can be hazardous and thus may need to be controlled. However, cancer research findings over the past several decades indicate that some carcinogens may act in non-linear ways. For example, radon damages the DNA and RNA of lung cells, but the long-term risk associated with exposure to radon is much higher for smokers than for non-smokers, even if their exposures are the same. Another chemical, formaldehyde CSF, is under review by EPA because it has been shown that before animals exposed to high doses developed cancer, they developed ulcerations in their mucous membranes. This observation suggests that lower concentrations of formaldehyde CSF, a water soluble compound, had a different potency factor than higher concentrations.

Further complicating the issue, juvenile test animals are more susceptible to some cancer causing compounds than adult animals of the same species. EPA's cancer risk guidelines now reflect this difference. On the other hand, it is understood that the human body's ability to repair damaged DNA diminishes with age. Age-dependent cancer slope factors are not available for the hundreds of suspected cancer causing compounds, so the unit risk factors are assumed to apply uniformly over a lifetime, except where observations support a different risk for infants and children.

These questions can influence what type of model scientists use to calculate dose-response relationships for carcinogens, or even whether carcinogens are treated similarly to non-cancer endpoints with presumed population thresholds (as described below). A common model for dose-response for carcinogens is the so-called one-hit model, which corresponds to the simplest mechanistic explanation of cancer—that a single exposure to a dose as small as a molecule would have a non-zero probability of changing a normal cell into a cancer cell. Researchers typically use this model to analyze pollutants that are hypothesized to operate under this mode of action or as a default model in the absence of mechanistic evidence.

In contrast, multi-stage models (of which the one-hit model is a special case) assume that a cell passes through several distinct phases that occur in a certain order as it becomes cancerous. It is hard to determine empirically which model is more appropriate, so this choice relies on understanding the mode of action of the compound. Because CSF values are sensitive to these assumptions, EPA's newest carcinogen risk guidelines (issued in 2005) focus on finding a point in the range of observed data, called a point of departure, which is less sensitive to model choice. For compounds that are direct mutagens or with substantial background processes, linearity is assumed below the point of departure, while non-linear approaches are used if suggested by the mode of action.



## 6. Other Risks

Environmental contaminants cause many harmful effects in addition to cancer, such as toxicity, birth defects, reduced immune system function, and damage to other organs and physical systems. For noncarcinogens, researchers assume that a threshold exists below which no harmful effects are likely to occur in humans. To quantify these values, scientists first seek to identify the so-called **no observable adverse effects level (NOAEL)**, which is the highest exposure among all available studies at which no toxic effect was observed. Next they divide the NOAEL by one or more uncertainty factors, typically ranging from 10 to 1,000, based on the quality of the data that was used to measure the NOAEL and on how close the NOAEL is to estimated human exposures.

From these calculations, EPA sets **reference doses** for ingestion and reference concentrations for inhalation that represent levels at which humans can be exposed to chemicals for specific periods of time without suffering adverse health effects. These limits are fairly conservative because they incorporate uncertainty factors and assume that people may be exposed daily or constantly throughout their lives. Box 1 shows EPA's core health assessment figures for noncarcinogenic effects of paraquat, a widely-used and highly toxic herbicide.

Regulators also set limits for specific types of exposures. For example, the EPA establishes guidelines for pesticide residues in food, and the Agency for Toxic Substances and Disease Registry establishes minimal risk levels (MRLs) for acute, intermediate, and chronic exposure to contaminants at hazardous waste sites.

The EPA's peer-reviewed assessments of human health effects (both cancer and non-cancer) from exposure to chemicals are available through the agency's Integrated Risk Information System (IRIS) (footnote 17). These reports include descriptive and quantitative information on specific chemicals that cause cancer and other chronic health effects. Analysts can use this information along with exposure information to characterize public health risks from specific chemicals in specific situations and to design risk management programs.

The state of California has developed a similar list in compliance with Proposition 65, a 1986 ballot measure that required the state to publish a list of chemicals known to cause cancer, birth defects, or reproductive harm (footnote 18). Chemicals can be listed in three ways: if they are shown to cause cancer, birth defects, or reproductive harm by either of two state expert committees; if they are so identified by EPA, certain other U.S. regulatory agencies, or the International Agency for Research on Cancer; or if a state or federal agency requires them to be labeled as causing these effects (substances in this category are mainly prescription drugs).

Companies that do business in California must provide "clear and reasonable" warning before knowingly and deliberately exposing anyone to a listed chemical, unless exposure is low enough to pose no significant health risks. They also are barred from discharging listed chemicals into drinking water sources. The intent of Proposition 65 is to increase awareness about the effects of exposure to listed chemicals, enable Californians to reduce their exposure, and give manufacturers an incentive



to find substitutes for listed chemicals. The law has led to removal of many toxic substances from commerce, including faucets and tableware that contained lead.

## 7. Benefit-Cost Analysis and Risk Tradeoffs

Why are so many hazardous materials widely used in technology and commerce? Simply put, they also deliver benefits. For example, lead was used for decades as a gasoline additive in the United States (and is still used in developing countries) because it reduces "knocking," or pinging in the engine from premature fuel combustion. In many cases the full human health impacts of such materials were not known at the time when they entered use but only became clear years later, when they were common ingredients of commercial products.

When risk analysis shows that a material poses serious human health risks, policy makers often carry out formal economic analyses of risk reduction options. This involves setting an economic value on lives saved and injuries or illnesses avoided through policy actions, so that decision makers can compare these health benefits to the cost of proposed regulations. Most major environmental laws do not require use of cost-benefit analysis. For example, the Clean Air Act directs regulators to set national air quality standards that scientific evidence indicates will protect public health. One exception, the Safe Drinking Water Act, was amended in 1996 to require cost-benefit analysis of new standards.

Currently the federal Office of Management and Budget requires U.S. government agencies to do cost-benefit analyses of regulations that are expected to have economic impacts (positive or negative) of \$100 million or more—some 50 to 100 rules annually (footnote 19).

One widespread method for monetizing health benefits is called [hedonic valuation](#)—analyzing what people are willing to pay to live in an unpolluted area or willing to accept as a salary premium for working in a risky industry. Economists often calculate these values by looking at what workers earn in high-risk industries compared to less-dangerous fields (Fig. 12) or by comparing housing prices in polluted and clean areas. This method is also called the revealed-preference approach, on the assumption that society strikes balances between risks and benefits that are reflected in economic decisions.



Figure 12. Commercial king crab fisherman, Alaska

© Alaska Division of Community and Business Development.

In a survey of more than 30 risk premium studies conducted in U.S. workplaces between 1974 and 2000, W. Kip Viscusi and Joseph Aldy found that the average calculated value of a statistical life (VSL) was about \$7 million. One way to think about this figure is to imagine a population of 1 million people who are considering a regulation that would result on average in one fewer death from cancer each year. If each member of the group is willing to pay \$7 per year as a cost of imposing that regulation, the value of a statistical life in that society can be said to be \$7 million. This figure measures the collective value placed on reducing a generalized risk, not the value of any actual person's life. EPA guidelines recommend using a value of \$6.2 million for regulatory impact analyses, while some other agencies use lower values (footnote 20).

Analysts also monetize the benefits of regulations by measuring costs that those regulations can be expected to avoid, such as medical bills, lost wages due to illness and disability, and special aid programs for children born with birth defects due to exposure. Table 3 lists health effects considered by EPA in a 2006 regulatory impact analysis in support of national limits for fine particulate air pollution (some effects were not quantified because of limitations in data or methods).

**Table 3.** Human health effects of particulate air pollution

Quantified and Monetized Effects	Unquantified Effects
Premature mortality, based on cohort study estimates	Low birth weight
Bronchitis (chronic and acute)	Pulmonary function
Hospital admissions: respiratory and cardiovascular	Chronic respiratory diseases other than chronic bronchitis
Emergency room visits for asthma	Nonasthma respiratory emergency room visits
Nonfatal heart attacks	UVb exposure (may result in benefits or disbenefits)
Lower and upper respiratory illness	
Minor restricted-activity days	
Work loss days	
Asthma exacerbations (asthmatic population)	
Respiratory symptoms (asthmatic population)	
Infant mortality	

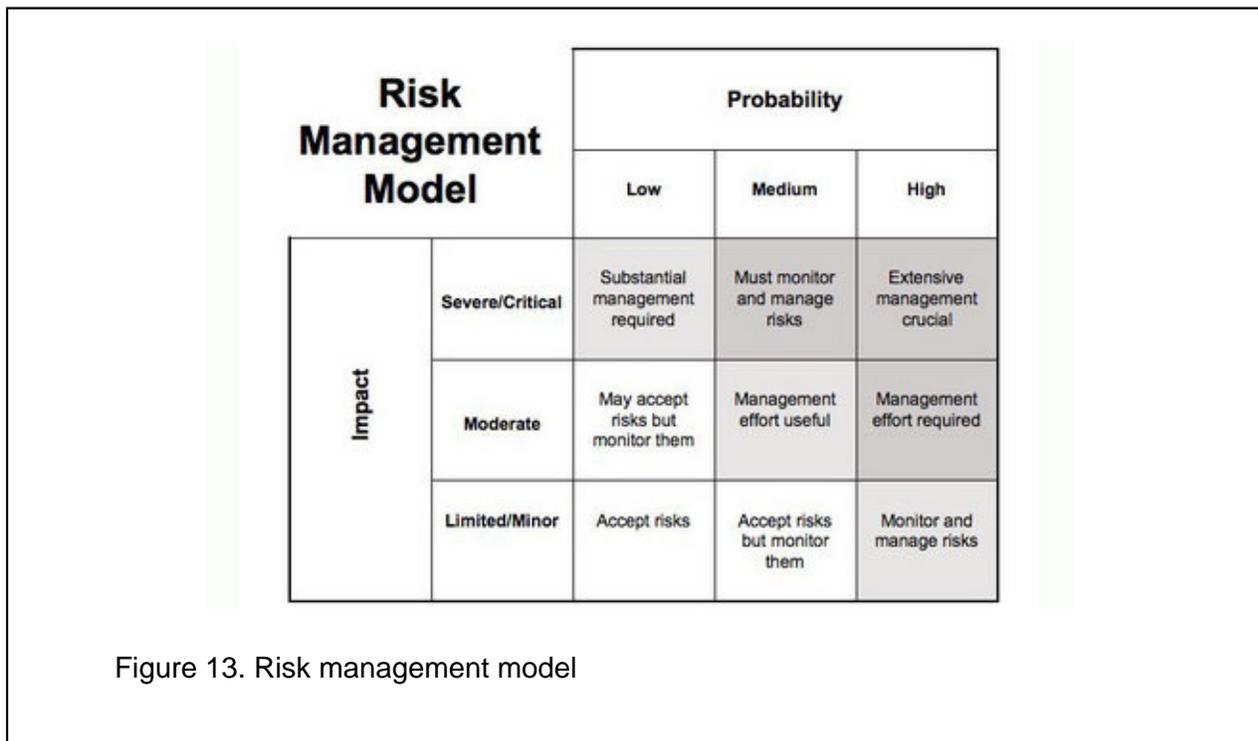
Cost-benefit analyses also set values on environmental impacts, such as improved visibility in scenic areas or protection of undeveloped land as wilderness. Sometimes monetizing these effects is straightforward because people pay for access to the resource and demand is likely to drop if the resource becomes less attractive. For example, researchers have assessed the economic impact of air pollution in national parks by measuring how sharply pollution events reduce visits to parks and calculating the resulting lost revenues, both at the park and in surrounding communities.

**Contingent valuation** is a less direct approach that involves asking people what they would theoretically be willing to pay for an environmental good. This method is often used to estimate demand for a resource for which a market does not currently exist. For example, if a power company proposes to dam a wild and scenic river to produce electricity, analysts might ask ratepayers whether they would be willing to pay higher rates for electricity from another, more expensive source to keep the river undeveloped. It can be hard to estimate accurate values with this method, which has generated a vast economic literature, but well-designed willingness-to-pay studies can provide reasonable indications of how highly the public values specific environmental benefits.

Many risk-management choices involve risk-risk tradeoffs—choosing between options that each may cause some harm. We make risk-risk tradeoffs every day. Some are personal choices, such as pursuing an intensive exercise program which has cardiovascular benefits but could lead to injuries. Others involve broad social regulations. For example, some environmental groups support an international ban on the insecticide DDT because of its toxic human and animal health effects, but

many public health agencies argue that this step would make it very difficult to control malaria in the developing world.

Regulators may consider many criteria when they confront risk-risk tradeoffs and have to decide which risks are and are not acceptable. Important factors include both the probability of a risk and whether its consequences would be negligible, moderate, or serious (Fig. 13).



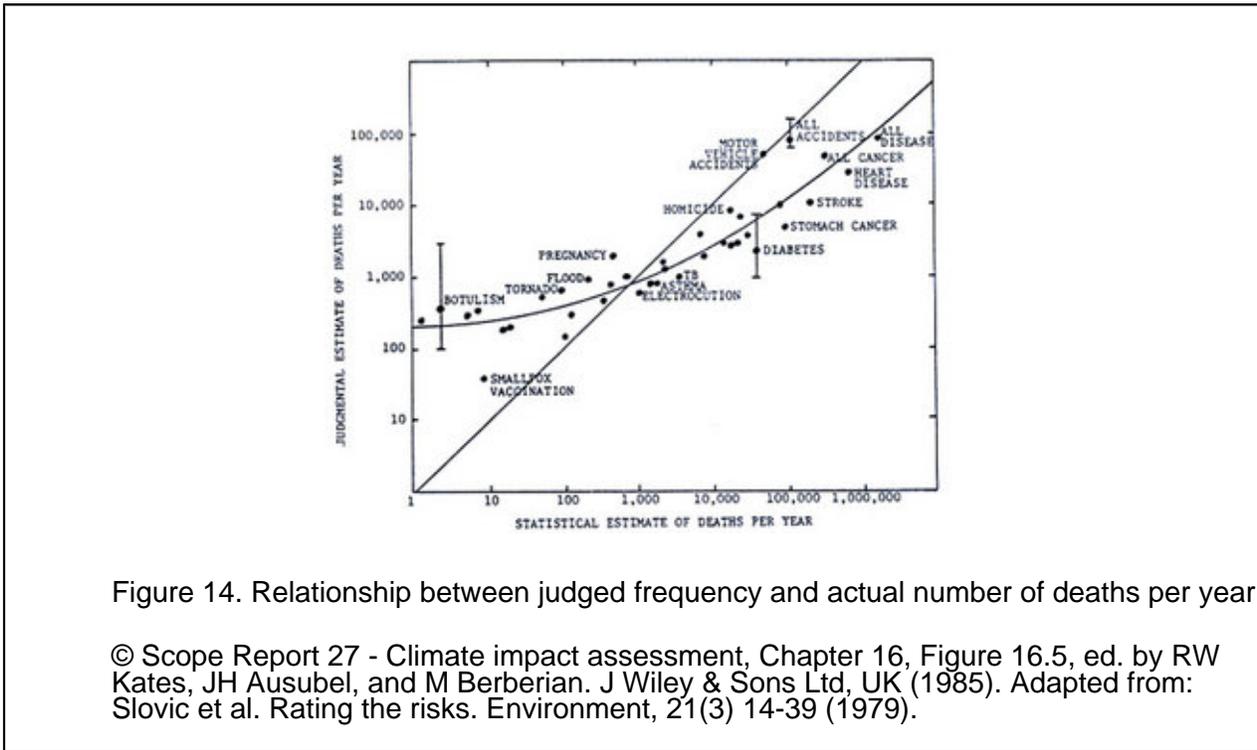
A high-consequence event, such as a plane crash or a radiation release at a nuclear power plant, can merit intensive regulation even if the probability of such accidents occurring is very low. Conversely, risks that have high probability but low consequences for the general public—for example, injuries from slipping on icy sidewalks—can be addressed through lower-level actions, such as passing local ordinances that require property owners to clear their sidewalks. Once officials decide what level of risk is involved, cost-benefit analysis may influence their choice of responses if it shows that one policy will produce much greater benefits relative to costs than another policy.

## 8. Risk Perception

Expert assessments and public perceptions of risk are not always the same. Decision makers need to understand factors that influence how people understand and interpret risk information for several reasons. First, public concerns may influence research and development priorities, such as which

chemicals to analyze in toxicity studies. Second, individual behavior choices are guided by risk avoidance, so if experts want people to avoid certain risks, they need to understand whether the public sees those actions as dangerous. If the public views a risky activity as benign, officials may have to develop public-education campaigns to change those perceptions. Current examples include labels warning about health risks on cigarette packages and alcoholic beverage containers.

Behavioral and social scientists have compared risk perceptions among many different groups, including scientists' views compared to those of laypersons, men compared to women, and differences among diverse ethnic and economic groups. One finding is that the general public overestimates the prevalence of some risks (such as those lying above the straight line in Fig. 14) and underestimates others (those lying below the line).



Laypeople judge risks differently from technical experts because they give greater weight to factors such as the potential for catastrophic damage, the likelihood of threats to future generations, and their own sense of whether they can control the risk. This can be seen in Table 4, which shows how technical experts and several sets of laypeople ranked the risk from a list of activities and technologies. Note, for example, that the expert group was much less worried about nuclear power but more worried about x-rays than laypeople. Both involve radiation exposure, but x-rays may have

seemed less risky to the non-specialists because the scale of an x-ray is much smaller than a nuclear reactor accident and because people usually have a choice about whether to undergo x-rays.

**Table 4.** Perceived risk for 30 activities and technologies

Activity or technology	League of Women Voters	College students	Active club members	Experts
<b>Nuclear power</b>	1	1	8	20
<b>Motor vehicles</b>	2	5	3	1
<b>Handguns</b>	3	2	1	4
<b>Smoking</b>	4	3	4	2
<b>Motorcycles</b>	5	6	2	6
<b>Alcoholic beverages</b>	6	7	5	3
<b>General (private) aviation</b>	7	15	11	12
<b>Police work</b>	8	8	7	17
<b>Pesticides</b>	9	4	15	8
<b>Surgery</b>	10	11	9	5
<b>Firefighting</b>	11	10	6	18
<b>Large construction</b>	12	14	13	13
<b>Hunting</b>	13	18	10	23
<b>Spray cans</b>	14	13	23	26
<b>Mountain climbing</b>	15	22	12	29
<b>Bicycles</b>	16	24	14	15
<b>Commercial aviation</b>	17	16	18	16
<b>Electric power (nonnuclear)</b>	18	19	19	9
<b>Swimming</b>	19	30	17	10
<b>Contraceptives</b>	20	9	22	11
<b>Skiing</b>	21	25	16	30
<b>X-rays</b>	22	17	24	7



Activity or technology	League of Women Voters	College students	Active club members	Experts
High school/college football	23	26	21	27
Railroads	24	23	29	19
Food preservatives	25	12	28	14
Food coloring	26	20	30	21
Power mowers	27	28	25	28
Prescription antibiotics	28	21	26	24
Home appliances	29	27	27	22
Vaccinations	30	29	29	25

Other factors can influence how both experts and laypeople perceive risks. Paul Slovic and other behavioral researchers have found that many Americans stigmatize certain industries, especially nuclear power and chemicals, which are widely viewed as repellent, disruptive, and dangerous. Conversely, scientists who work for industry tend to see chemicals as less threatening than do government and academic researchers (a phenomenon called affiliation bias). Ultimately, they argue, all groups bring their own assumptions to bear on discussions of risk.

Communicating risk information to the public is an important part of risk management. In the early decades of environmental regulation, public communication often took what critics called the "decide, announce, defend" approach: agencies developed policies and released their final results to the public and regulated industries. But since risk analysis involves many uncertainties, assumptions, and judgments, it requires policy makers to explain clearly how decisions are reached—especially if the issue involves risks that laypeople perceive differently from scientific experts.

Often effective risk communication means involving the public in the decision process, not just informing people at the end. Public involvement in risk decisions can take many forms. In early planning stages, it can help regulators identify the issues that citizens care most about, how much risk they will tolerate, and what they view as acceptable mitigation costs. Stakeholders may also take part in implementing decisions. For example, the Defense and Energy Departments have formed community advisory boards to help make decisions about cleaning up contaminated military bases and nuclear weapons production sites.

## 9. The Precautionary Principle

Under the basic risk analysis model, regulators quantify risks and compare the costs and benefits of various control options before they set limits on hazards. However, over the past several decades some governments have formally adopted a different approach called the Precautionary Principle as a guideline. This view holds that governments should not wait to limit contaminants in food, water, air, or commercial products until scientific studies have reduced uncertainties about exposure and effects.

Although the idea of "better safe than sorry" can be traced as far back in history as the Hippocratic Oath, the Precautionary Principle was first codified as an approach to environmental protection in West German national policies of the 1970s. References to a precautionary approach began to appear in international agreements in the 1980s and 1990s. The Wingspread Statement, a declaration drafted by government officials, attorneys, and labor and environmental advocates at an international conference in 1998, argued that existing environmental regulations (especially those based on risk assessment) did not protect human health and the environment well enough and that a new approach was required. "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. In this context the proponent of an activity, rather than the public, should bear the burden of proof," the statement asserted.

The Precautionary Principle has taken root most strongly in the European Union (EU). In 2000 the EU issued a communiqué stating that the principle applied "where scientific evidence is insufficient, inconclusive, or uncertain and preliminary scientific evidence indicates that there are reasonable grounds for concern that the potentially dangerous effects on the environment, human, animal, or plant health may be inconsistent with the high level of protection chosen by the EU." European regulators have invoked the Precautionary Principle to support steps such as banning imported beef treated with hormones and adopting the Restriction of Hazardous Substances Directive, which requires electronics manufacturers to remove lead, mercury, cadmium, and other hazardous substances from most of their products (Fig. 15).



Figure 15. Label indicating that a product complies with the EU's Restrictions of Hazardous Substances (RoHS) directive

© 2007. Image-Tek/www.image-tk.com.

The Precautionary Principle plays a much weaker role in U.S. environmental regulation, which generally assumes that some level of risk from exposure to contaminants is acceptable and sets controls intended to limit pollution to those levels. Unlike the EU, the United States does not require comprehensive product testing or labeling.

However, some U.S. laws take a precautionary approach in more limited areas. For example, new drugs must be tested before they can be sold, and the National Environmental Policy Act requires environmental impact assessments for any major projects that are federally funded, with an obligation to consider alternatives including no action. Some states and cities have adopted regulations that take a precautionary approach to policies such as using pesticides in schools or funding new technologies. For the most part, though, U.S. environmental laws require some scientific proof of harm as a basis for protective action.

## 10. Major Laws

The main U.S. law regulating exposure to hazardous materials is the Toxic Substances Control Act (TSCA), passed in 1976. The law authorizes the Environmental Protection Agency to regulate chemical hazards, from warning labels to outright bans. It also allows the EPA to require companies to perform extensive health and safety testing on chemicals before they can be marketed, to maintain



detailed records, and to report on how chemicals are used in commerce and industry. The EPA is required to take swift regulatory action if it finds that a chemical is likely to cause cancer, gene mutations, or birth defects.

There are important limitations to the EPA's ability to regulate the chemical industry under TSCA. First, the burden of proof falls more heavily on the EPA than on chemical manufacturers. The EPA has to have "substantial evidence" of "unreasonable risk" to require testing. Out of the tens of thousands of chemicals in commerce, the EPA has only banned a handful under TSCA. Second, the agency is required to analyze risks and benefits of all less burdensome regulatory alternatives before banning chemicals. The EPA also must evaluate the risk posed by substitute products.

Several other laws regulate specific classes of hazardous substances. The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) gives the EPA authority to control pesticides. All pesticides used in the United States must be registered with the EPA after they have been through health and safety testing, and users must take examinations to earn certification as applicators of pesticides. Under the Federal Food, Drug, and Cosmetic Act (FFDCA), the Food and Drug Administration regulates substances such as food additives and colorings, prescription drugs, and cosmetics.

In 1996 the U.S. Congress unanimously passed the Food Quality Protection Act (FQPA), which provides amendments to both FIFRA and FFDCA. Key provisions of FQPA under FFDCA include use of an additional 10-fold uncertainty factor to account for increased susceptibility of children and a requirement for regulators to consider aggregate exposures from multiple pathways (e.g., food, water, yards, pets, etc.) for pesticides with a common mechanism of toxicity (i.e., for organophosphates such as malathion and chlorpyrifos or for pyrethroids such as permethrin and resmethrin) in establishing allowable pesticide residue levels in food.

After three years of consideration, debate, and lobbying, the European Union's far-reaching regulation on chemicals, REACH, went into effect on June 1, 2007. REACH (Registration, Evaluation, Authorization, and Restriction of Chemicals) is an aggressive law that places priority on protecting health and the environment. The newly established European Chemicals Agency, located in Helsinki, will begin an 11-year process of registering some 30,000 chemical substances in use today. The agency will conduct evaluations, including risk management, to identify gaps in information about hazards, exposure pathways, and health and ecological impacts. REACH is designed to reduce harmful substances in products and the environment and to strongly encourage chemical producers and manufacturing companies to find alternative formulations, processes, and products.

The European market is important to the U.S. chemical industry, which exports some \$14 billion worth of products each year. U.S. manufacturers and the federal government opposed many aspects of REACH, but companies doing business with EU countries will have no choice but to comply. The U.S. chemical industry is already providing workshops and other assistance for producers to comply with REACH. Although this process is likely to be long and expensive, it will help to harmonize national regulations for the chemical industry—a positive development, since many hazardous chemicals are produced and distributed worldwide.



## 11. Further Reading

European Commission, Environment Directorate, "REACH," [http://ec.europa.eu/environment/chemicals/reach/reach\\_intro.htm](http://ec.europa.eu/environment/chemicals/reach/reach_intro.htm). An overview of the REACH regulation, including information on benefits and costs.

Dennis Paustenbach, ed., **Human and Ecological Risk Assessment: Theory and Practice** (New York: Wiley, 2002). A comprehensive textbook, including risks involving air, water, food, occupational exposures, and consumer products.

National Research Council, **Science and Judgment in Risk Assessment** (Washington, DC: National Academy Press, 1994). An exploration of how risk analysts make assumptions and deal with uncertainty, written to help the EPA make risk assessments more valid and credible by using scientific data more fully and making the limits of knowledge clear.

U.S. National Institutes of Health, National Library of Medicine, "Tox Town," <http://toxtown.nlm.nih.gov/>. An animated online guide to connections between chemicals, the environment, and public health, including common exposure locations, non-technical descriptions of chemicals, and links to scientific and health resources.

## Footnotes

1. Waltraud Eder, Markus J. Ege, and Erika von Mutius, "The Asthma Epidemic," **New England Journal of Medicine**, vol. 355 (2006), pp. 2226–2235.
2. Jonathan I. Levy et al., "A Community-Based Participatory Research Study of Multifaceted In-Home Environmental Interventions for Pediatric Asthmatics in Public Housing," **Social Science & Medicine**, vol. 63 (2006), pp. 2191–2203.
3. U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics, "New Chemicals and Existing Chemicals," <http://www.epa.gov/oppt/newchemicals/pubs/newvexist.htm>.
4. Society for Risk Analysis, "Principles for Risk Analysis," **RISK newsletter**, Third Quarter 2001.
5. U.S. Environmental Protection Agency, William K. Reilly oral history interview, <http://www.epa.gov/history/publications/reilly/20.htm>.
6. National Research Council, **Risk Assessment in the Federal Government: Managing the Process** (National Academy Press, 1983).
7. U.S. Environmental Protection Agency, Guidelines for Exposure Assessment, FRL-4129-5, 1992, pp. 16–17, <http://www.epa.gov/ncea/pdfs/guidline.pdf>.
8. EPA, Guidelines for Exposure Assessment, FRL-4129-5, 1992, p. 126, <http://www.epa.gov/ncea/pdfs/guidline.pdf>.



9. University of California, Los Angeles, Department of Epidemiology, "Broad Street Pump Outbreak," <http://www.ph.ucla.edu/epi/snow/broadstreetpump.html>.
10. Austin Bradford Hill, "The Environment and Disease: Association or Causation?" **Proceedings of the Royal Society of Medicine**, vol. 58 (1965), pp. 295–300, [www.edwardtufte.com/tufte/hill](http://www.edwardtufte.com/tufte/hill).
11. National Institutes of Health, **Cancer and the Environment**, NIH Publication No. 03-2039 (Washington, DC, August 2003), p. 1.
12. NIH, **Cancer and the Environment**, pp. 7–8.
13. Nancy Nelson, "The Majority of Cancers Are Linked to the Environment," **BenchMarks**, National Cancer Institute, June 17, 2004, <http://www.cancer.gov/newscenter/benchmarks-vol4-issue3>; Mayo Clinic, "Carcinogens In the Environment: A Major Cause of Cancer?" May 24, 2006.
14. U.S. Environmental Protection Agency, Region 2, "EPA Risk Assessments Confirm Exposure to PCBs in River May Increase Cancer Risk, Other Non-Cancer Health Hazards and Threaten Fish and Wildlife," press release, August 4, 1999.
15. U.S. Environmental Protection Agency, Technology Transfer Network, Air Toxics Website, "Risk Assessment for Carcinogens," <http://www.epa.gov/ttn/atw/toxsource/carcinogens.html>.
16. Pamela R.D. Williams and Dennis J. Paustenbach, "Risk Characterization," in Dennis J. Paustenbach, ed., **Human and Ecological Risk Assessment: Theory and Practice** (New York: Wiley, 2002), p. 325.
17. <http://www.epa.gov/iriswebp/iris/index.html>.
18. [http://www.oehha.ca.gov/prop65/prop65\\_list/Newlist.html](http://www.oehha.ca.gov/prop65/prop65_list/Newlist.html).
19. John D. Graham, Director, Office of Information and Regulatory Affairs, Office of Management and Budget, "Valuing Health: An OMB Perspective," remarks, February 13, 2003, [http://www.whitehouse.gov/omb/infoereg/rff\\_speech\\_feb13.pdf](http://www.whitehouse.gov/omb/infoereg/rff_speech_feb13.pdf).
20. W. Kip Viscusi and Joseph E. Aldy, "The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World," **Journal of Risk and Uncertainty**, Vol. 27, No. 1 (2003), pp. 5–76.

## Glossary

**case-control study** : One type of epidemiological study design used to identify factors that may contribute to a medical condition by comparing a group of patients who have that condition with a group of patients who do not.

**cohort study** : A study in which patients who presently have a certain condition and/or receive a particular treatment are followed over time and compared with another group who are not affected by the condition under investigation.



**contingent valuation** : A survey-based economic technique for the valuation of non-market resources, typically ecosystems and environmental areas and services. It involves directly asking people, in a survey, how much they would be willing to pay for specific environmental services. It is called “contingent” valuation, because people are asked to state their willingness to pay, contingent on a specific hypothetical scenario and description of the environmental service.

**delivered dose** : The portion of an internal dose that actually reaches a biologically sensitive site within the body.

**endocrine disruptors** : Chemical pollutants that have the potential to substitute for, or interfere with, natural hormones.

**epidemiology** : The science that deals with the incidence and distribution of human disease or disorders.

**hedonic valuation** : A method used to estimate economic values for ecosystem or environmental services that directly affect market prices.

**no observable adverse effects level (NOAEL)** : The level of exposure of an organism, found by experiment or observation, at which there is no biologically or statistically significant increase in the frequency or severity of any adverse effects in the exposed population when compared to its appropriate control.

**partition** : The division of a chemical into two or more compartments in an ecosystem or body parts of an organism.

**Precautionary Principle** : The belief that if a technology, chemical, physical agent, or human activity can be reasonably linked to adverse effects on human health or the environment, then controls should be implemented even if the problem or the cause-effect relationship is not fully understood; to wait for scientific certainty (or near certainty) is to court disaster.

**reference dose** : The U. S. Environmental Protection Agency's maximum acceptable oral dose (abbreviated RfD) of a toxic substance, most commonly determined for pesticides.

**risk analysis** : Identifying potential issues and risks ahead of time before these were to pose cost and/or schedule negative impacts.

**risk assessment** : An analytical study of the probabilities and magnitude of harm to human health or the environment associated with a physical or chemical agent, activity, or occurrence.

**risk management** : The human activity that integrates recognition of risk, risk assessment, development of strategies to manage it, and mitigation of risk using managerial resources.

## Unit 7 : Agriculture



Wheat Farm in Yaqui Valley, Mexico. Courtesy of Ivan Ortiz Monasterio

### Overview

Demographers project that Earth's population will peak during the 21st century at approximately ten billion people. But the amount of new cultivable land that can be brought under production is limited. In many nations, the need to feed a growing population is spurring an intensification of agriculture—finding ways to grow higher yields of food, fuel, and fiber from a given amount of land, water, and labor. This unit describes the physical and environmental factors that limit crop growth and discusses ways of minimizing agriculture's extensive environmental impacts.

### Sections:

1. Introduction
2. Earth's Land Resources
3. Key Inputs for Photosynthesis
4. Increasing Yields
5. Combating Pests and Disease
6. Livestock: Growing Animals
7. Genetic Improvement and Food Production
8. Agriculture and Energy
9. Sustainable Agriculture
10. Food for the Future
11. Further Reading

## 1. Introduction

Agriculture is the human enterprise by which natural ecosystems are transformed into ones devoted to the production of food, fiber, and, increasingly, fuel. Given the current size of the human population, agriculture is essential. Without the enhanced production of edible biomass that characterizes agricultural systems, there would simply not be enough to eat. The land, water, and energy resources required to support this level of food production, however, are vast. Thus agriculture represents a major way in which humans impact terrestrial ecosystems.

For centuries scholars have wrestled with the question of how many people Earth can feed. In 1798 English political economist Thomas Robert Malthus published what would become one of the most famous pamphlets in social science, **An Essay on the Principle of Population**. Malthus proposed that because population tended to increase at a geometric (exponential) rate, while food supplies could only grow at an arithmetic rate, all living creatures tended to increase beyond their available resources.

"Man is necessarily confined in room," Malthus argued. "When acre has been added to acre till all fertile land is occupied, the yearly increase of food must depend upon the melioration of the land already in possession. This is a fund; which, from the nature of all soils, instead of increasing must gradually be decreasing" (footnote 1). The resulting scarcity, he predicted, would limit human population growth through both "positive checks," such as poverty, diseases, wars, and famines, and self-imposed "negative checks," including late marriage and sexual abstinence.

In terms of global food production, however, Malthus has so far been proved wrong because his essay failed to take into account the ways in which agricultural productivity of cultivated lands, measured in terms of harvested (typically edible) biomass, could be enhanced. Agriculture involves the genetic modification of plant and animal species, as well as the manipulation of resource availability and species interactions. Scientific and technological advances have made agriculture increasingly productive by augmenting the resources needed to support photosynthesis and by developing plants and animals with enhanced capacity to convert such resources into a harvestable form. The outcome is that world food production has in fact kept up with rapid population growth. Gains have been especially dramatic in the past 50 years (Fig. 1).

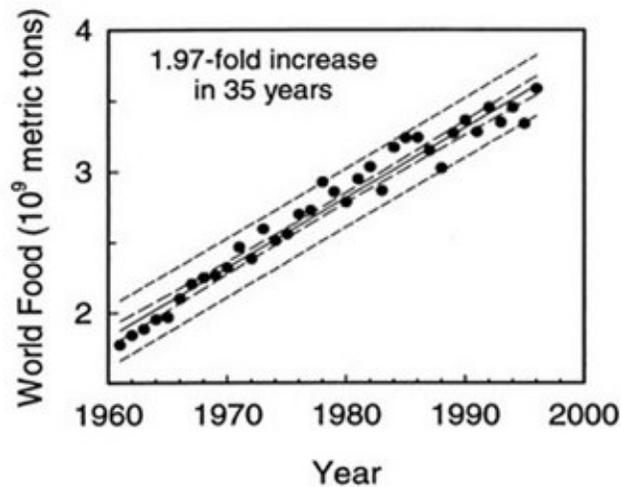


Figure 1. World food production, 1961–1996 (measured as the sum of cereals, coarse grains, and root crops)

© David Tilman.

But these gains carry with them serious environmental costs. Large-scale agriculture has reduced biodiversity, fragmented natural ecosystems, diverted or polluted fresh water resources, and altered the nutrient balance of adjacent and downstream ecosystems. Agriculture also consumes major amounts of energy and generates greenhouse gas emissions that contribute to global climate change. However, these negative impacts must be weighed against human demand for food, as well as the fact that agriculture is the primary livelihood for 40 percent of the human population. In some countries, more than 80 percent of the population makes a livelihood from farming, so increasing agricultural productivity not only makes more food available but also increases incomes and living standards.

The future impacts of agriculture will depend on many factors, including world demand for food, the availability and cost of resources needed to support high levels of productivity, and technological advances to make agriculture more efficient. Global climate change is expected to alter temperature, precipitation, and weather patterns worldwide, thus changing many fundamental conditions that guide current agricultural practice. (For more details, see Unit 12, "Earth's Changing Climate.")

## 2. Earth's Land Resources

How much of Earth's surface can be used for agriculture? The basic limits are temperature, topography, climate, soil quality, and available technologies, including scientific understanding of issues like plant and animal genetics. As technology improves over time, the zone where agriculture can be practiced successfully expands. For example, development of the horse collar in China allowed farmers to use livestock to pull ploughs and thus to farm in heavier soils than they could till by hand. Many social, political, and economic factors also shape agricultural land use, including land tenure patterns, population density, and environmental regulations.

As of the year 2000, about 37 percent of Earth's land area was agricultural land. About one-third of this area, or 11 percent of Earth's total land, is used for crops. The balance, roughly one-fourth of Earth's land area, is pastureland, which includes cultivated or wild forage crops for animals and open land used for grazing (footnote 2).

To help governments with land-use planning, the United Nations Food and Agricultural Organization (FAO) has developed a system called Agro-Ecological Zoning that characterizes land's suitability for agriculture based on physical parameters like climate, soil, and topography (footnote 3). Based on current soil, terrain, and climate data, FAO estimates that more than three-quarters of Earth's land surface is unsuitable for growing rain-fed crops (i.e., raising crops without irrigation). Most of the remainder is subject to some soil, terrain, and/or climate limitations. On average, only about 3.5 percent of Earth's surface is suitable for agriculture without any physical constraints (Fig. 2).

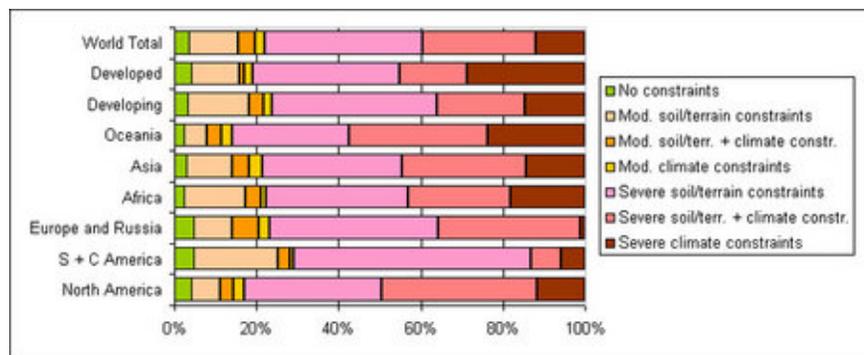


Figure 2. Distribution of climate and soil/terrain constraints by region

© 2000. International Institute for Applied Systems Analysis and Food and Agriculture Organization.

These physical constraints mean that not all farmland is equally productive, even with modern techniques and inputs. In areas where land is less productive, agriculture requires more techniques and inputs to address limitations such as poor soil quality. Less productive agricultural land generally has low market value, so in many countries farming must compete with other uses such as residential or commercial development or recreation. However, in areas that have received few modern inputs, such as many parts of Africa, fertilizer and other technologies can greatly increase productivity and raise the value of agricultural land.

In regions where productivity is rising faster than demand, such as the United States, the European Union, and Japan, land is being withdrawn from cultivation. These areas rely on agricultural intensification to keep output high as their farmed lands shrink. In contrast, land is being converted for agriculture in many parts of the developing world. Both trends are causes for concern. Agricultural intensification has serious environmental impacts, as we will see in the following sections, while land conversion is a major cause of deforestation. Clearing forests for agriculture alters ecosystems that provide important services such as sequestering carbon or absorbing floodwaters.

Another 30 percent of the world's land area is forested, with half of global forests managed at least partly for wood production (other forest functions may include land conservation or protecting indigenous plants and animals) (footnote 4). Forestry is generally a much less intense form of land use than agriculture because tree crops have longer rotation periods than agricultural commodities, so soils are less disturbed and fewer cultivation inputs like fertilizer are needed. However, some forestry practices—such as building roads through forest tracts and clear-cutting on hillsides where trees stabilize soil—can disrupt ecosystems on a scale comparable to farming.

### 3. Key Inputs for Photosynthesis

**Photosynthesis**, in which plants convert carbon dioxide and water into plant tissue, is the foundation for agricultural productivity. Before we consider how tools like plant breeding and fertilizer can increase farmers' output, we need to understand this basic process and the constraints on it, including the key roles played by water and nitrogen (N).

Drought is the biggest limit on agricultural productivity because plants need an enormous amount of water. When plants photosynthesize, they use energy from sunlight to convert carbon dioxide (CO<sub>2</sub>) and water into carbohydrates. As discussed in Unit 4, "Ecosystems," the basic equation for photosynthesis is:



From this equation it might appear that plants would need equal amounts of water and CO<sub>2</sub>, but the actual ratio is approximately 400 to 1. Where does all of this water go? More than 98 percent of a plant's water intake passes upward through the plant from roots to leaves and evaporates, exiting the leaf as water vapor through pores in the leaf surface called **stomata**. Movement of water from the soil to the atmosphere through the bodies of plants is called **transpiration**. This process serves

many important functions: it carries minerals from the soil to the leaves and prevents leaves from overheating. However, the principle reason that plants transpire is to allow uptake of  $\text{CO}_2$  from the atmosphere.

As water diffuses out of plants' leaves into the surrounding atmosphere,  $\text{CO}_2$  diffuses in (Fig. 3). The exchange ratio between  $\text{CO}_2$  intake and water loss is lopsided: diffusion of water molecules out of the leaf is much greater than diffusion of  $\text{CO}_2$  into the leaf. This happens because the outside atmosphere is much less moist than the interior of a plant: relative humidity is roughly 50 percent outside, compared to 100 percent at the center of leaves, so water diffuses easily out of plants. In contrast, the atmosphere is only about 0.037 percent  $\text{CO}_2$ , so there is a much smaller contrast between  $\text{CO}_2$  concentrations inside and outside the leaf.

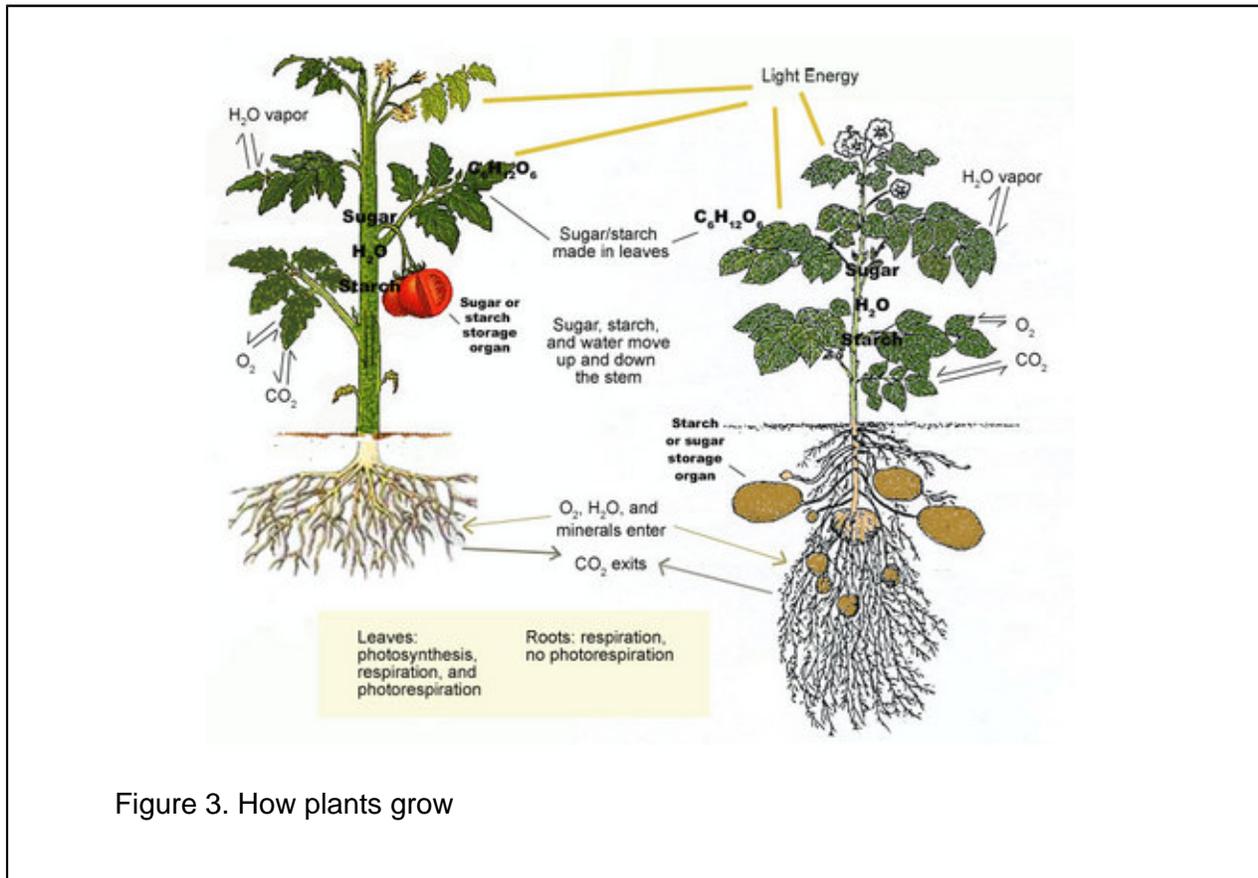


Figure 3. How plants grow

This simple relationship describing the diffusional exchange of water and  $\text{CO}_2$  explains why drought is the major factor limiting agricultural yields worldwide. Because the atmosphere is a very dilute  $\text{CO}_2$  source, plants need to maximize  $\text{CO}_2$  intake as long as it will not dry out their interiors. Stomata open

to promote gas exchange with the atmosphere when water is plentiful, and constrict or close when water is scarce. If stomata must close to conserve water, the plant will not have access to the  $\text{CO}_2$  it needs to photosynthesize. Therefore, to encourage growth it is essential to supply plants with enough water.

Many farming regions rely on irrigation to increase productivity and ensure consistent yields regardless of yearly fluctuations in rainfall. One-third of global food harvests come from irrigated areas, which account for about 16 percent of total world cropland. Every year, humans divert about 2,700 cubic kilometers of water (five times the annual flow of the Mississippi River) from the global water cycle for crops. Without irrigation, some countries such as Egypt would be able to support only very limited forms of agriculture, and grain production in northern China, northwest India, and the western Great Plains of the United States would fall sharply (Fig. 4).

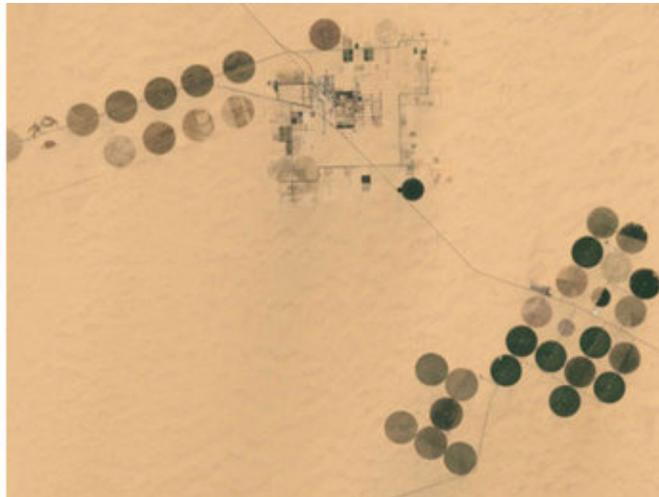


Figure 4. Irrigation in the heart of the Sahara

© National Aeronautics and Space Administration. Earth Observatory.

Nitrogen (N), which plants obtain from the soil, is another critical resource for photosynthesis. Natural levels of N availability frequently limit crop yields. Nitrogen is an essential component of proteins, including the enzyme ribulose-bisphosphate-carboxylase-oxygenase (abbreviated as RUBISCO), which catalyzes the incorporation of  $\text{CO}_2$  into an organic molecule. RUBISCO is thought to be the most abundant protein on Earth, with leaves typically being 2 percent (by dry weight) nitrogen. This is because RUBISCO, from a catalytic point of view, is one of the slowest enzymes known, reflecting an

inherent tradeoff between catalytic efficiency (speed) and selectivity (distinguishing between  $\text{CO}_2$  and  $\text{O}_2$ ).

Before dismissing RUBISCO as inefficient (and thus easily improved upon), it is important to realize the constraints under which it operates. From RUBISCO's point of view,  $\text{CO}_2$  and  $\text{O}_2$  are quite similar in many ways. RUBISCO is a very large molecule, along side of which  $\text{CO}_2$  and  $\text{O}_2$  appear quite similar in size. Furthermore, the two are both uncharged molecules that can react in a similar manner. However, the real challenge is that the ratio of  $\text{O}_2$  to  $\text{CO}_2$  in Earth's atmosphere is greater than 500:1. RUBISCO thus is forced to go slowly so that it can maintain a high selectivity for  $\text{CO}_2$ . Like the diffusional uptake of  $\text{CO}_2$ , which makes photosynthesis extremely water-intensive, this tradeoff for RUBISCO between speed and selectivity means that nitrogen plays an important role in natural and agricultural ecosystems.

Prior to World War I the main source of nitrogen fertilizer was organic manure from livestock animals. Explorers also sought out mineral deposits that could be exploited. Chile derived a major share of its gross domestic product from nitrate (saltpeter) mines from roughly 1880 through World War I. Prior to the exploitation of these naturally occurring mineral deposits, guano (seabird droppings) along the coasts of Chile and Peru and on Pacific islands, where seabirds feed on fish in nutrient-rich coastal waters, were a prized source (Fig. 5). In 1856 the U.S. Congress passed the Guano Islands Act, empowering U.S. citizens to take possession of unoccupied islands anywhere in the world that contained guano deposits if the islands were not under the jurisdiction of other governments (footnote 5).



Figure 5. Guano deposits on Gardner Pinnacles, Laysan Island, Hawaii, 1969

© Dr. James P. McVey, National Oceanic and Atmospheric Administration, Sea Grant.

In 1908 German chemist Fritz Haber developed the Haber-Bosch process for combining nitrogen and hydrogen gases at high temperatures to produce ammonia ( $\text{NH}_3$ ), which can be processed further into nitrate. The process was commercialized and developed on an industrial scale during World War I and World War II to make nitric acid for munitions. It also launched the fertilizer industry. Synthetic fertilizer entered widespread use after World War II, and the increased levels of nitrogen available to support plant growth boosted crop productivity in regions where farmers could afford synthetic fertilizers.

Producing nitrogen fertilizers requires substantial amounts of energy. Although Earth's atmosphere is about 80 percent nitrogen gas ( $\text{N}_2$ ), the triple bond of the dinitrogen molecule is so strong that only a small number of prokaryotic organisms can make use of it. Industrial production of N fertilizers takes place at high temperatures and pressures to crack this bond. In addition, the Haber-Bosch process involves oxidizing natural gas ( $\text{CH}_4$ ) over an inorganic catalyst to produce hydrogen gas.

Today nitrogen fertilizers are used on a vast scale. World nitrogen fertilizer consumption was approximately 80 million tons in 1999, with as much as 400 kilograms per hectare applied in areas of highly intensified agricultural production. To put this in perspective, the amount of atmospheric (gaseous) nitrogen incorporated in the production of synthetic fertilizers is of the same order as the amount that occurs globally through biological nitrogen fixation and lightning. In contrast, fossil fuel

combustion only releases about five percent of the carbon exchange that occurs naturally through photosynthesis and respiration.

Irrigation and fertilizer help farmers ensure that crops will have the basic inputs they need to grow, but these mainstays of modern agriculture can also cause serious environmental damages. In many regions, irrigation depletes normal river flows or contributes to salinization of agricultural lands (for more information, see Unit 8, "Water Resources"). Fertilizer that is not taken up by plant roots (especially nitrogen, which is extremely mobile in its most common form,  $\text{NO}_3^-$ ) can wash into nearby water bodies or into ground water, altering the species composition and nutrient balance of downstream ecosystems. This problem is most severe early in the growing season when plants are small and do not have enough root mass to keep water and nutrients from infiltrating into ground water. Mismanaged livestock manure (discussed in section 6) causes similar problems.

#### 4. Increasing Yields

Undisturbed ecosystems maintain themselves by cycling nutrients and other inputs, like water and energy, up through food webs. As discussed in Unit 4, "Ecosystems," these cycles are closed loops to a large extent. Substances change form as they move through ecosystems, but they are not destroyed or removed from the system.

Agriculture is fundamentally different from undisturbed ecosystems because harvesting crops removes material from the system. The product that can be harvested from an agricultural system, which is called its yield, represents a loss of materials such as water and nutrients from the system. Farmers can increase yields by adding energy and materials, by increasing the efficiency of energy conversion and allocation to the harvested product, or by reducing losses that occur during the growing process.

Agricultural yields have risen steadily throughout the history of human cultivation, with particularly steep increases throughout the 20th century. From 1961 through 1999, the FAO's aggregate crop production index increased at an average rate of 2.3 percent per year and world crop production per capita rose at an average annual rate of 0.6 percent (footnote 6). Global production of major cereal crops more than doubled during this period (Fig. 6).

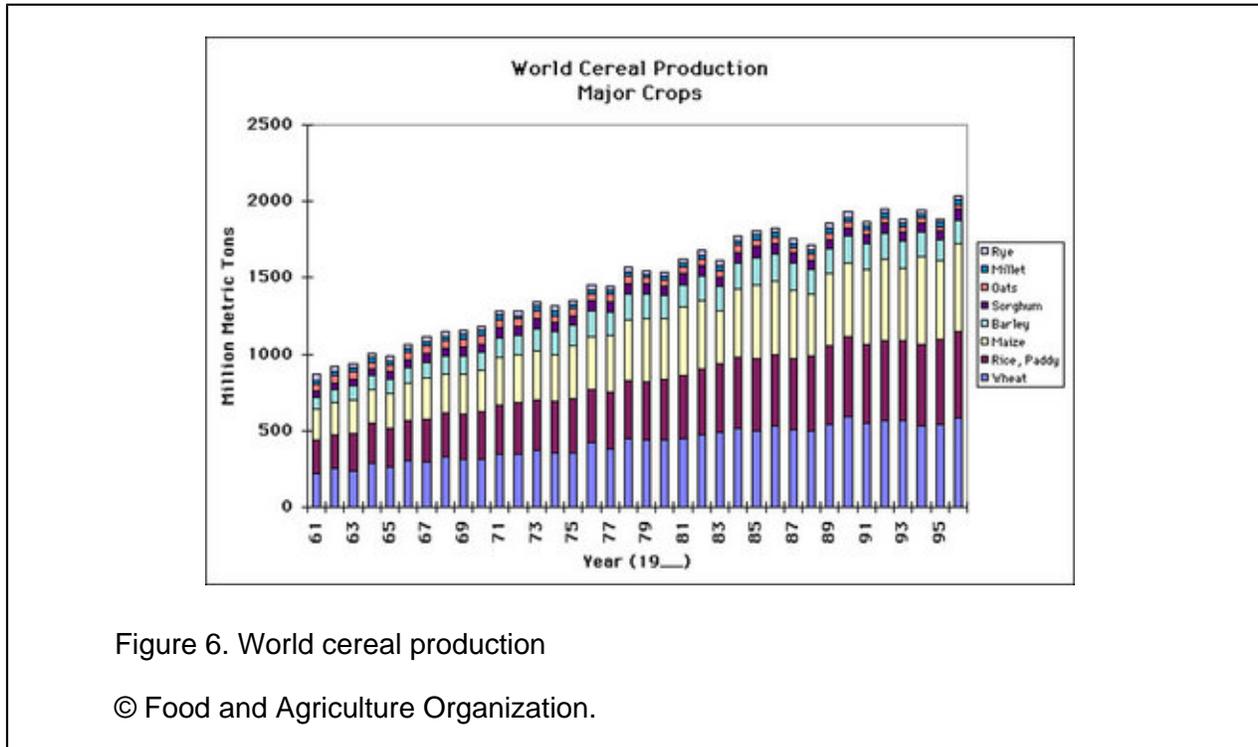


Figure 6. World cereal production

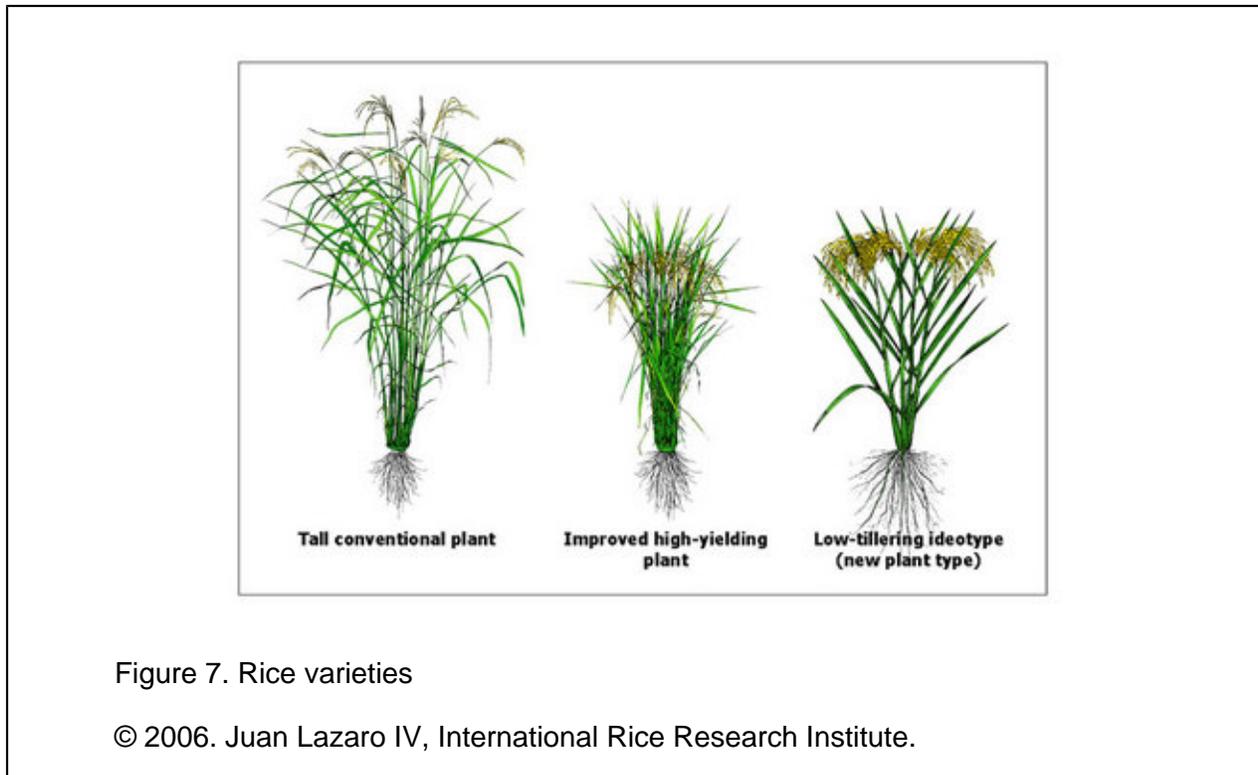
© Food and Agriculture Organization.

Productivity in agriculture is a measurement of farmers' total output per unit of land. If the gains shown in Figure 6 had come simply from bringing twice as much land under cultivation, they would not automatically signal that productivity was rising as long as farmers were using the same amount of inputs per acre. However, agriculture has become much more productive over time. In many parts of the world, modern farmers get far more product from each unit of land than their predecessors thanks to intensification—using more technological inputs per acre. In areas where such inputs are not available, such as Africa, output rates remain far below world averages.

Radical changes in agricultural inputs over the past century made this increase possible. Land and labor inputs have fallen drastically in industrialized countries, but technological advances such as large-scale irrigation, synthetic fertilizers, pesticides and herbicides, and capital investment (in the form of mechanization) have increased sharply. Scientific advances such as development of higher-yielding crop varieties have also contributed to increased productivity.

The most significant way in which scientists have produced bigger yields is by modifying plants so that they devote a larger proportion of their physical structures to producing biomass that is usable for food. This process is referred to as increasing the **harvest index** (the ratio of harvested biomass to total biomass). For example, growing deep root systems protects wild plants against drought, but this allocation strategy limits the amount of plant biomass available to make leaves, so they have fewer sugars from photosynthesis available to make seeds. Plant breeders selecting for higher-yielding

varieties might try to increase the harvest index by including plants that produce fewer roots and more seeds, or by developing dwarf or semi-dwarf strains. Figure 7 shows several modifications that have increased rice yields.



This approach was an important component of the "Green Revolution"—a 30-year transformation of agriculture in developing regions that started in the 1940s, when private foundations and national governments joined forces to distribute high-yielding crop varieties, synthetic fertilizer, irrigation, and pesticides to subsistence farmers in Asia and Latin America. By introducing semi-dwarf varieties of wheat and rice, researchers increased the crops' harvest indexes and reduced the problem of lodging (falling over before harvest due to excessive growth). This shift made it possible for farmers to apply higher levels of chemical fertilizers so that plants would photosynthesize at increased rates and produce more biomass. Scientists also developed these new strains to make them easier to harvest, more durable during transport, and longer-lasting in storage.

The Green Revolution helped world food production to increase at a rate faster than population growth from 1950 onward. However, these increases relied on synthetic fertilizer and irrigation because green revolution plant varieties were designed to produce high yields when supplied with high inputs of nitrogen and water. In other words, they were not inherently high-yielding plants (i.e., they were not able to use resources more efficiently than traditional varieties) and likely would have

done worse under "natural" conditions. Many varieties were highly susceptible to pests and diseases, so they also required heavy use of pesticides to thrive. Because the new plants were short, they were more susceptible to competition from weeds, so farmers also had to use herbicides to raise them.

As we will see in the next section, this strategy generated further complications for human health, non-target species, and the environment in the regions where it was applied. Conversely, because Green Revolution agriculture is capital-intensive and requires well-developed infrastructure systems for functions such as delivering irrigation water, it essentially bypassed sub-Saharan Africa.

## 5. Combating Pests and Disease

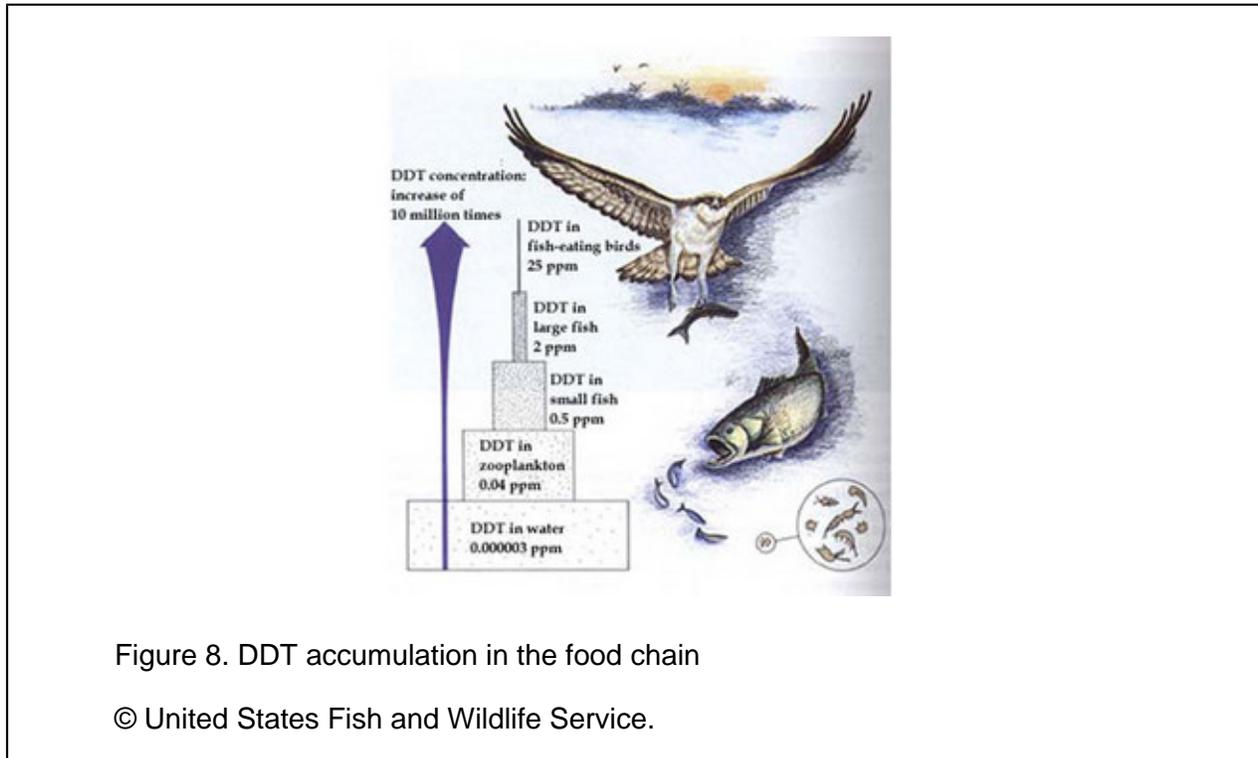
As agriculture became increasingly dependent on technological inputs throughout the 20th century, it also underwent a structural shift, particularly in developed countries. Instead of raising a diverse mix of crops, farmers increasingly planted large holdings of one or a few crop varieties that had been developed for high yields. **Monoculture** makes it easier to cultivate large acreages more efficiently, especially using mechanized equipment and chemical inputs. However, these artificial ecosystems are vulnerable to outbreaks of pests and pathogens because they do not have natural protection from genetic diversity and they are typically nutrient-rich, thanks to abundant fertilizer use. Moreover, many pest species have adapted to spread rapidly in ecosystems where recent disturbances, such as plowing, have eliminated natural predators (for background, see Unit 4, "Ecosystems").

Agricultural pests include insects, mammals such as mice and rats, unwanted plants (weeds), fungi, and microorganisms such as bacteria and viruses. Humans have controlled pests with naturally-occurring substances such as salt, sulfur, and arsenic for centuries, but synthetic pesticides, first developed during World War II, are generally more effective.

Many of the first pesticides that were widely used for agriculture were **organochlorines** such as DDT (dichloro diphenyl trichloroethane), aldrin, dieldrin, and heptachlor. These substances are effective against a range of insects and household pests, but in the 1950s and 1960s they were shown to cause human health effects including dizziness, seizures, respiratory illness, and immune system dysfunction. Most organochlorines have been banned in the United States and other developed countries but remain in use in developing countries.

In her 1962 book **Silent Spring**, biologist and author Rachel Carson drew wide-scale public attention to the environmental effects of pesticides. Carson described how actions such as spraying elm trees with broad-spectrum pesticides to prevent Dutch elm disease severely affected many other parts of local ecosystems (Box 1).

Bioaccumulation of DDT and other organochlorines drastically reduced populations of bald eagles and other large predatory birds that fed at the top of the food chain. The pesticides disrupted birds' reproductive systems and caused them to lay eggs with very thin shells that broke before young birds hatched (Fig. 8).



Organochlorines were replaced in the 1970s with other pesticides that were less toxic and more narrowly targeted to specific pests. However, many of these newer options still killed off pests' natural enemies, and when the insecticides were used repeatedly over time, pests became resistant to them through natural selection (many types of insects can develop through entire generations in days or weeks). Today hundreds of species of insects and weeds are resistant to major pesticides and herbicides.

In response some farmers have turned to methods such as releasing natural insect predators or breeding resistance into crops. For example, U.S. farmers can buy corn seeds that have been engineered to resist rootworms, corn borers, or both pests, depending on which are present locally, as well as corn that has been developed to tolerate herbicides. Others practice [integrated pest management \(IPM\)](#), an approach under which farmers consider each crop and pest problem as a whole and design a targeted program drawing on multiple control technologies, including pesticides, natural predators, and other methods.

In one notable case, Indonesia launched an IPM program in 1986 to control the brown planthopper, a notorious pest that lays its eggs inside rice plant stalks, out of range of pesticides. Outreach agents trained farmers to monitor their fields for planthoppers and their natural predators, and to treat outbreaks using minimal pesticide applications or alternative methods such as biological controls (Fig.

9). Over the following decade rice production increased by 15 percent while pesticide use fell by 60 percent. Yields on IPM lands rose from 6 to almost 7.5 tons of rice per hectare (footnote 7).



Figure 9. Gathering insects for identification during IPM training, Indonesia

© J.M. Micaud, Food and Agriculture Organization.

Plowing originally developed as a way to control pests (weeds), but created new issues in the process. Bare lands that have been plowed but have not yet developed crop cover are highly susceptible to erosion. The Dust Bowl that occurred in the United States in the 1930s was caused partly by poor agricultural practices. With support from the federal government, farmers plowed land that was too dry for farming across the Great Plains, destroying prairie grasses that held topsoil in place. When repeated droughts and windstorms struck the central and western states, hundreds of millions of tons of topsoil blew away. Today a similar process is taking place in northern China, where over-plowing and overgrazing are expanding the Gobi Desert and generating huge dust storms that scour Beijing and other large cities to the east.

Excessive plowing can also depress crop production by altering soil microbial communities and contributing to the breakdown of organic matter. To conserve soil carbon and reduce erosion, some farmers have turned to alternative practices such as no-till or direct-drill agriculture, in which crops are sown without cultivating the soil in advance. Direct drilling has been widely adopted in Australia,

and some 17.5 percent of U.S. croplands were planted using no-till techniques as of the year 2000 (footnote 8).

No-till agriculture enhances soil development and fertility. It is usually practiced in combination with methods that leave crop residues on the field, which helps to preserve moisture, prevent erosion, and increase soil carbon pools. However, no-till requires an alternative strategy for weed control and thus frequently involves substantial use of herbicides and chemical means to control other pests.

## 6. Livestock: Growing Animals

Many subsistence farmers in traditional societies raise livestock along with their crops, either for their families' use or for sale. But in industrialized nations, animal agriculture has been transformed in much the same way as crop production over the past century. Modern livestock farms are large and specialized and rely heavily on technology inputs. Like major plant crops, meat and dairy products are increasingly produced through a kind of monoculture in which farmers raise one or a few animal strains that have been bred to maximize output—hens that lay more eggs, dairy cows that produce more milk, or pigs that grow quickly and develop lean meat. Producers use technological inputs, such as antibiotics and hormone treatments, to make animals grow larger and more quickly.

To maximize efficiency, large-scale livestock farms confine animals indoors instead of letting them range outside (Fig. 10). Confining animals makes it easier to control the amount and type of feed they receive, administer medications and growth supplements, and artificially inseminate breeding females. But it also generates new management issues. Crowding stresses animals and promotes disease transmission, so many livestock farmers use antibiotics not only to treat sick animals but to prevent illnesses and promote growth. Many of these drugs are identical or similar to antibiotics used in human medicine, so their overuse threatens human health by promoting the development of drug-resistant bacterial strains that can infect humans through the food chain or via direct exposure to farm animals or wastes.

In addition, large farms accumulate massive quantities of animal waste. One cow can produce more than 40 pounds of manure per day. Manure liquefies when it is washed out of barns, so it is too heavy to transport economically over long distances. Many large farms store millions of gallons of manure onsite in tanks or lagoons (which may be lined or unlined, depending on local regulations), until it can be used on neighboring fields.



Figure 10. Confined hog production facility

© United States Geological Survey, Toxic Substances Hydrology Program.

When manure leaks or spills from storage, it sends large pulses of nutrients into local water bodies, causing algal blooms that deplete dissolved oxygen in the water and kill fish when they die and decompose. Nutrient pollution also occurs when manure is applied too heavily to farmland, so that plants cannot take up all of the available nitrogen and phosphate before the manure leaches into nearby rivers and streams. Excess nutrients, mainly from agricultural runoff, are a major cause of "dead zones" in large water bodies such as the Chesapeake Bay and the Gulf of Mexico (for details, see Unit 8, "Water Resources"). Manure also pollutes water with bacteria, hormones, and other chemical residues from animal feed.

Large livestock farms also generate air pollution from manure, dust, and greenhouse gases produced in the digestive systems of cattle and sheep. Many people who live near animal feeding operations complain about smells and suffer physical symptoms such as burning eyes, sore throats, and nausea. A 2003 National Research Council study found that livestock farms produce many air pollutants that are significant hazards at scales ranging from local to global (Table 1). However, the report concluded that more analysis was required to develop accurate measurements of these emissions as a basis for regulations and that the United States lacked standards for quantifying odor, which could be caused

by various combinations of hundreds of compounds (footnote 9). (For more details on emissions and health risks, see Unit 6, "Risk, Exposure, and Health"; Unit 11, "Atmospheric Pollution"; and Unit 12, "Earth's Changing Climate.")

**Table 1.** Potential importance of air emissions from animal feeding operations at different spatial scales.

Emission	Global, national, and regional importance	Local Importance (property line or nearest dwelling)	Primary effects of concern
<b>Ammonia (NH<sub>3</sub>)</b>	Major	Minor	Acid rain, haze
<b>Nitrous oxide (N<sub>2</sub>O)</b>	Significant	Insignificant	Global climate change
<b>Nitrogen oxides (NO<sub>x</sub>)</b>	Significant	Minor	Haze, acid rain, smog
<b>Methane (CH<sub>4</sub>)</b>	Significant	Insignificant	Global climate change
<b>Volatile organic compounds (VOCs)</b>	Insignificant	Minor	Quality of human life
<b>Hydrogen sulfide (H<sub>2</sub>S)</b>	Insignificant	Significant	Quality of human life
<b>Particulate matter (PM<sub>10</sub>)</b>	Insignificant	Significant	Haze
<b>Fine particulate matter (PM<sub>2.5</sub>)</b>	Insignificant	Significant	Health, haze
<b>Odor</b>	Insignificant	Major	Quality of human life

World demand for meat and dairy products is increasing, driven by population growth and rising incomes in developing countries. Because of this growth and the trend toward raising animals on large-scale farms, the FAO calls livestock farming "one of the top two or three most significant contributors to the most serious environmental problems, at every scale from local to global." According to FAO's estimates, livestock production generates 18 percent of world greenhouse gas emissions (more than the transport sector), accounts for 8 percent of world water use, and is probably the largest sectoral water pollution source (footnote 10).

With global meat and dairy production predicted to roughly double between 2000 and 2050, these environmental impacts will have to be drastically reduced just to keep agricultural pollution from worsening. And as we will see in section 8, "Agriculture and Energy," the fact that humans are eating at higher trophic levels by increasing their meat consumption makes agriculture more energy-intensive than it would be if people relied mainly on plant-based diets.

## 7. Genetic Improvement and Food Production

Farmers have manipulated the genetic makeup of plants and animals since the dawn of agriculture. Initially they used **selective breeding** to promote qualities that made breeds readily usable for agriculture, such as animals that domesticated well and plants that were easy to harvest. Next, breeders focused on varieties that could be grown outside of their native geographic range—for example, overcoming natural photoperiod requirements (the amount of daylight that plants need to flower). In the twentieth century, plant geneticists selected for traits that would allow plants to use high levels of water and nitrogen to increase yields. Similarly, animal breeders worked to increase the amount of meat or milk that various domestic animal lines produced.

Today classical agricultural breeding is a highly quantitative science that uses genetic markers (specific DNA sequences) to select for desired characteristics. This approach enables scientists to manipulate the genetic makeup of crops with substantial precision, as long as genetic variation exists for particular traits. Agricultural breeders also use biotechnology to move genes across taxonomic barriers, combining genetic material from species that would not cross-breed naturally. For example, Bt corn has been modified by inserting a gene from the bacterium **Bacillus thuringiensis** that kills harmful insects so that farmers do not need to use insecticide.

Since the mid-1990s, the U.S. Department of Agriculture has approved 63 genetically engineered (GE) crops for unrestricted sale, including strains of corn, soybeans, cotton, potatoes, wheat, canola, and papaya. Most of these crops have been developed to tolerate herbicides or resist insects or fungi, while others have been engineered for specific product qualities such as longer shelf life. Products under development include grains, field crops, fruits, vegetables, trees, and flowers designed to achieve desirable growing properties such as cold or drought resistance or efficient use of nitrogen (footnote 11). The extent to which such strategies will be able to enhance agricultural productivity, however, remains to be seen.

An alternative use of biotechnology that some supporters advocate is to develop crops with improved nutritional content to combat nutritional disorders. One widely-publicized product is golden rice, a rice variety into which several "trans" or foreign genes have been added so that the plant produces beta-carotene (vitamin A) in its grains (Fig. 11). Vitamin A deficiencies are widespread in societies that consume rice-based diets, causing thousands of cases of blindness and premature deaths among children in developing countries every year. Researchers are currently working to produce golden rice that contains the recommended daily allowance of vitamin A in a 100 to 200 gram serving, as well as to ensure the bioavailability of the beta-carotene contained within the modified rice grains. But not everyone is convinced by this approach: some experts argue that the same goals could be met more cheaply by promoting balanced, diverse diets in the target countries.



Figure 11. Conventional and golden rice

© 2007. Golden Rice Humanitarian Board.

In addition to questioning whether agricultural and nutritional goals might be more effectively met using more traditional approaches, critics have raised many concerns about GE foods, including potential harm to nearby ecosystems and the possibility that GE crops or animals will hybridize with and alter the genetic makeup of wild species. For example, over-planting Bt-resistant crops could promote increased Bt resistance among pests, while genes from GE crops could give wild plants qualities that make them more weedy and invasive. Although most of these effects will probably be benign, it is hard to predict when and where GE species could have harmful effects on surrounding ecosystems.

A 2002 National Research Council report concluded that genetically modified plants posed the same broad types of environmental risks as conventionally-produced hybrids, like the strains introduced during the Green Revolution. For example, both kinds of plants could spread into surrounding ecosystems and compete with local species. But the report noted that either type of plant could have specific traits that posed unique threats and accordingly called for case-by-case regulation of new GE strains. The committee also observed that future generations of GE plants are likely to have multiple introduced traits and forecast that these products will raise issues that cannot be predicted based on experience with early herbicide- and pest-resistant crops (footnote 12).

## 8. Agriculture and Energy

Agriculture consumes significant quantities of energy, especially in industrialized countries. Farmers use energy directly to heat and cool buildings, operate equipment, pump irrigation water, and transport products to market. Agriculture also consumes large quantities of fossil fuel indirectly as inputs for fertilizer (a prime ingredient of which is natural gas) and pesticides (made from petroleum and natural gas). Food processing and long-distance shipment consume additional energy.

U.S. energy use in agriculture has declined by more than 25 percent since the oil price shocks of the 1970, thanks partly to practices such as conservation tilling that require less working of soil. Changes in pesticide practices, including rapid adoption of GE crops that are resistant to insects, have lowered total use of pesticides. In addition, pesticides themselves have become much more sophisticated, so smaller quantities are required.

Nonetheless, some agricultural practices remain extremely energy-intensive—most notably, raising livestock on grain. In traditional farming systems, animals eat local forage and crop wastes that are not usable as food for humans. However, large-scale livestock farmers typically feed animals grains and other protein byproducts because the animals grow to market weight more quickly and it requires less land area than grazing, so meat production can take place closer to population centers (Fig. 12).

Raising animals on grain consumes fossil fuel as inputs for the pesticides and fertilizers needed to grow feed crops. Fattening one steer on corn to market weight can consume the equivalent of 35 gallons of oil (footnote 13). Some critics argue that this is an inefficient way to use food resources because animals convert only a fraction of the energy in their feed grain to growth (for background, see Unit 4, "Ecosystems"). According to one estimate it requires two kilograms of grain to produce one kilogram of poultry, four to make one kilogram of pork, and seven to produce a kilogram of beef (footnote 14).



Figure 12. Feedlot cattle

© United States Environmental Protection Agency.

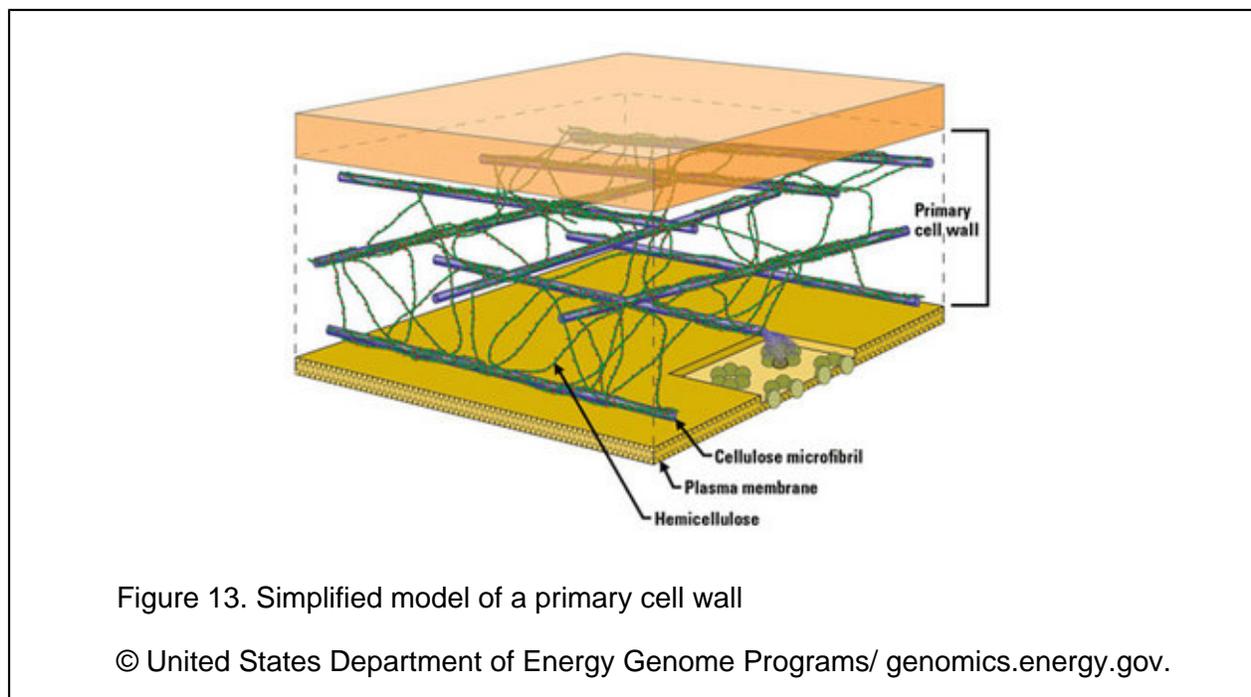
Other agricultural sectors could become energy resources in the coming decades. High oil and gas prices since the late 1990s have spurred worldwide interest in making liquid **biofuels** from plant sources such as forestry waste and fast-growing energy crops. Biofuels produce fewer atmospheric pollutants and greenhouse gases than fossil-based fuels when they are combusted, although the net effect in terms of CO<sub>2</sub> depends on energy use during production and subsequent processing of the crop.

Biofuels are valuable substitutes for imported oil in the transport sector because they can be used in most conventional engines with minor adjustments. They include **ethanol**, which is grain alcohol fermented from grain crops like corn (and soon from woody plants), and **biodiesel**, a natural version of diesel fuel made from oil crops such as soybean, sunflower, and rapeseed. (For more details, see Unit 10, "Energy Challenges.")

Several countries have made significant investments in biofuels. Most notably, all gasoline sold in Brazil is at least 25 percent ethanol made from local sugar cane. U.S. producers currently make about 4.5 billion gallons of ethanol per year from corn, equal to 3 percent of national gasoline consumption, with production scheduled to rise to 7.5 billion gallons per year by 2012. Most ethanol plants and fuel pumps are located in Midwestern corn-growing states.

Corn ethanol is the first type of ethanol to be commercialized in the United States because corn kernels and sugar cane juice are made up of simple carbohydrates that are easy to ferment, so the production process is relatively cheap. There is growing interest in making ethanol from the cell walls

of fast-growing plants such as switchgrass and willow and poplar trees, as well as corn stalks. These feedstocks are made up of complex polymers such as cellulose, hemicellulose, and lignin, which contain more energy (Fig. 13).



Corn ethanol has benefited U.S. farmers by increasing demand and driving up corn prices, but it only delivers modest environmental benefits. According to the U.S. Department of Energy, using corn ethanol only reduces greenhouse gas emissions by about 18 to 29 percent compared to gasoline because fertilizer and other inputs required to grow corn are made from fossil fuels. However, cellulosic ethanol could reduce greenhouse gas emissions by as much as 85 to 86 percent compared to gasoline (footnote 15).

Cellulosic plant materials are difficult to break down, and no method has been developed to date for fermenting lignin, so making cellulosic ethanol is more expensive and technically challenging than producing corn ethanol. Government agencies, universities, and private investors hope to commercialize cellulosic ethanol production in the United States as soon as 2012. If it develops into a large-scale industry, cellulosic ethanol could create new markets for farmers to grow energy crops that require fewer chemical inputs than corn and can be raised on land unsuited for food crops (Fig. 14). However, extending the footprint of agriculture in this way might also reduce biodiversity by converting more land into managed ecosystems.



Figure 14. Geographic distribution of potential biomass energy crops

© United States Department of Energy Genome Programs/ [genomics.energy.gov](http://genomics.energy.gov).

## 9. Sustainable Agriculture

Growing concern about agricultural intensification in developed countries and its negative environmental impacts spurred an alternative movement in the 1970s to promote what advocates called sustainable agriculture. This perspective drew inspiration from sources that included organic farming (raising crops and animals with minimal synthetic inputs), the international environmental movement, and development advocates who criticized the Green Revolution for relying too heavily on pesticides and fertilizer. Ecology is a central pillar of sustainable agriculture, which treats farmed areas first and foremost as ecosystems, albeit unique ecosystems that have been disturbed and simplified by harvesting.

Few people would argue against the concept of sustainable agriculture, but there is no universally-agreed definition of what it means. Agricultural economist Gordon Conway describes sustainability as "the ability of an agroecosystem [an agricultural ecosystem and its social and economic setting] to maintain productivity in the face of stress or shock." Farmers use countermeasures to respond to stresses and shocks. They may draw on resources that are internal to the system, such as plants' natural pest resistance, or on outside inputs like herbicides and fertilizers.

Internal inputs typically rely on natural resources. Figure 15 shows the re-emerging practice of green manuring—tilling fresh plant material into soil to improve its physical and biological qualities. Outside inputs may be equally useful, but they usually cost more and may alter farming systems in unexpected ways—for example, introducing new species that compete with established crops (footnote 16).



Figure 15. Chopping and disking mustard green manure, Washington state, 2003

© Washington State University Extension.

Other formulations of sustainable agriculture, including legislation passed by the U.S. Congress in 1990, present it as a compromise between several sets of social goals, including but not limited to environmental conservation. Producing enough food, fuel, and fiber to meet human needs is a major objective, along with improving environmental quality, using non-renewable resources efficiently, and ensuring that farmers can earn reasonable livings from their products (footnote 17). In terms of methods, sustainable agriculture typically stresses treating soil as an ecosystem and using methods to keep it healthy, such as retaining organic matter and preserving diverse communities of soil organisms.

Many people equate sustainable agriculture with [organic farming](#), which is practiced according to national legal standards in more than 60 countries, including the United States, the European Union, Britain, Canada, and Australia. Generally, organic standards bar the use of synthetic pesticides, herbicides, fertilizers, and genetically modified organisms for crop production and use of

antibiotics, hormones, and synthetic feeds for animals. Organic agriculture typically has less severe environmental impacts than intensive farming with synthetic inputs. On average, organic farming conserves biodiversity, improves the structure and organic content of soil, leaches less nitrate into water bodies, and produces much less pesticide pollution.

As of 2002–2003, about 4 percent of utilized agricultural land in the European Union and up to 4 percent of farmed land for certain crops in the United States was farmed organically (Fig. 16). Together, the United States and the E.U. account for 95 percent of global organic food sales.

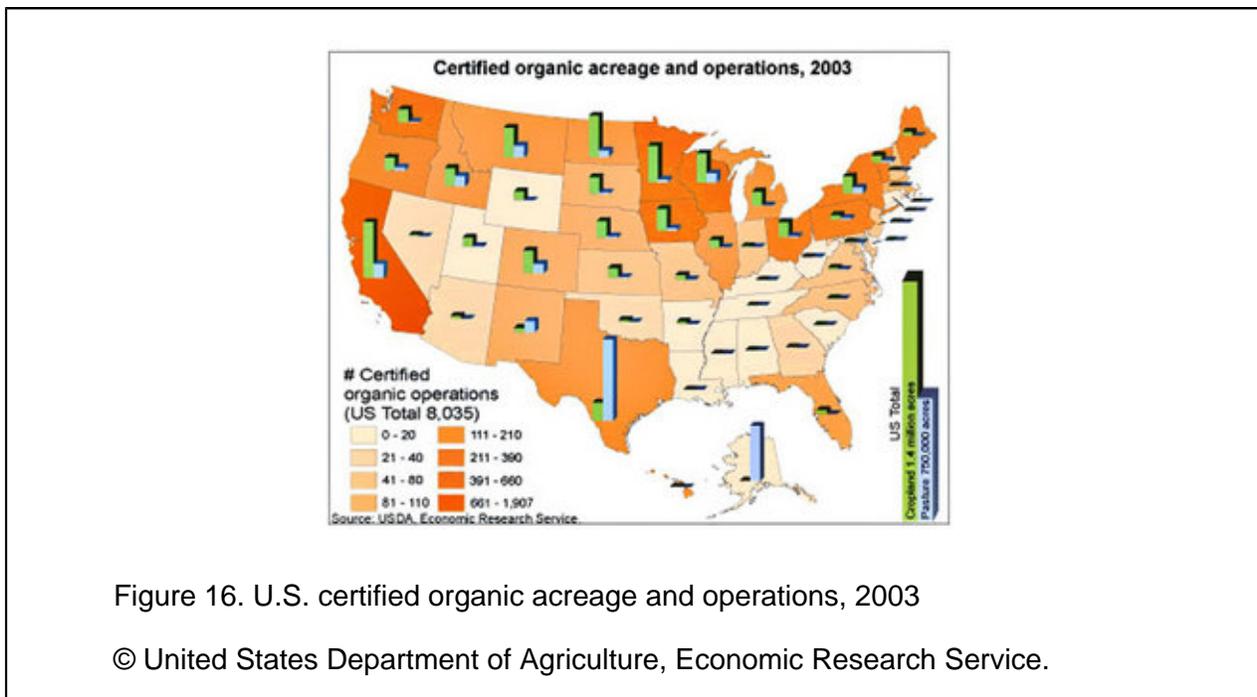


Figure 16. U.S. certified organic acreage and operations, 2003

© United States Department of Agriculture, Economic Research Service.

Organic farming is not without its drawbacks. Output from organic farms is typically lower than from conventional agriculture for at least several years after shifting to organic production, because it takes time to restore soil productivity naturally and establish beneficial insect populations. Organic agriculture is more labor-intensive than conventional farming, so production costs are higher and farmers must receive higher prices to make a profit. And transitioning to organic production takes several years, so it is too expensive and difficult for small-scale farmers without access to technical assistance and transition funding.

With world population projected to rise from 6.5 billion in 2006 to roughly 10 billion by 2050, and growing demand for meat in developing countries (which increases demand for grain as livestock feed), world grain production may have to double in coming decades. If nations take the intensive route to this goal, using even more fertilizer, pesticides, and irrigation, nutrient pollution and

freshwater depletion will increase well beyond current levels—the antithesis of sustainable agriculture (footnote 18).

One potential solution currently at the experimental stage is "precision agriculture"—using remote sensing to help farmers target fertilizer, herbicides, seeds, and water to exact locations on a field, so that resources are not over-applied or used where they are not needed. For example, satellite data could identify sectors within large cultivated fields that needed additional water or fertilizer and communicate the information to farmers driving machinery equipped with global positioning system receivers (reducing the need to apply inputs uniformly across entire fields) (footnote 19).

More broadly, agriculture will have to become more efficient in order to double world grain production without further degrading the environment. No single innovation will provide a complete solution. Rather, feeding the world sustainably is likely to require a combination of many technological inputs and sustainable techniques.

## 10. Food for the Future

Although the world produces enough total calories to feed all of its inhabitants today, more than 800 million people are undernourished, mainly in Africa and Asia (footnote 20). Food insecurity can have many causes, including poverty, wars, and diseases such as HIV/AIDS. However, in some areas soils and water supplies are so degraded that agricultural systems cannot produce enough food to meet human needs.

As the human population climbs towards its estimated peak of 10 billion, the environmental impacts of feeding so many people will increase. But this challenge also offers an opportunity: by making agriculture increasingly sustainable, we can meet the goal of feeding the world's population while reducing associated environmental problems such as water pollution. Reorienting agricultural systems is a complex task because the technical challenges are intertwined with social and economic issues such as land tenure and availability of foreign aid for developing countries. The magnitude of future agricultural effects on the environment will be influenced by many factors, including:

- **Actual demand for food.** Food demand will increase with population growth and rising income, which increases consumers' preference for animal protein.
- **Expansion of agricultural lands.** Agriculture will move into increasingly marginal areas because Earth's most fertile zones are already under cultivation, and will compete with other land uses such as urbanization.

- **Opportunities for increased yields.** Likely technological innovations include systems that increase availability of water and fertilizer; improved pesticides and biocontrols such as IPM; better soil conservation and management of microbial communities; and new crops that deliver increased yields under wider ranges of conditions and need fewer inputs than current strains.
- **Availability of water and chemical fertilizers.** The prices of these inputs are strongly affected by energy costs and competition for fresh water with other human activities.
- **Global climate change.** Variable weather is a major challenge for farmers because optimizing for high yields becomes more difficult as the range of potential weather conditions that might occur in any season increases. In the coming decades, global climate change is predicted to alter temperature and precipitation patterns in ways that could modify major elements of Earth's climate system (for details, see Unit 12, "Earth's Changing Climate").

## 11. Further Reading

Evans, L.T., **Feeding the Ten Billion: Plants and Population Growth** (Cambridge, UK: Cambridge University Press, 1998).

Halweil, Brian, "Can Organic Farming Feed Us All?" **Worldwatch**, May/June 2006.

King, John, **Reaching for the Sun: How Plants Work** (Cambridge, UK: Cambridge University Press, 1997).

Leigh, G.J., **The World's Greatest Fix: A History of Nitrogen and Agriculture** (Oxford, UK: Oxford University Press, 2004).

Pew Initiative on Food and Biotechnology, <http://pewagbiotech.org/>. Research and recommendations from a now-complete project on issues arising from advances in agricultural biotechnology.

Jennifer Thomson, **Seeds For the Future: The Impact of Genetically Modified Crops on the Environment** (Ithaca, NY: Cornell University Press, 2007).

## Footnotes

1. Thomas Robert Malthus, **An Essay on the Principle of Population: A View of its Past and Present Effects on Human Happiness; with an Inquiry into Our Prospects Respecting the Future Removal or Mitigation of the Evils which It Occasions**, Sixth Edition (London: John Murray, 1826), Book I, Chapter I, <http://www.econlib.org/library/Malthus/malPlong.html>.



2. United Nations Food and Agricultural Organization, Statistics Analysis Service, **Compendium of Agricultural-Environmental Indicators 1989–91 to 2000** (Rome, November 2003), p. 11.
3. For details, see <http://www.fao.org/AG/agl/agll/gaez/index.htm>.
4. United Nations Food and Agricultural Organization, **State of the World's Forests 2007** (Rome, 2007), pp. 64, 67.
5. 48 U.S.C., sections 1411–1419.
6. U.S. Department of Agriculture, Economics Research Service, **Agricultural Resources and Environmental Indicators, 2006 Edition/EIB-16**, p. 82.
7. Gordon Conway, **The Doubly Green Revolution: Food for All in the 21st Century** (Ithaca, NY: Cornell University Press, 1998), p. 215.
8. Carmen Sandretto, "Conservation Tillage Firmly Planted in U.S. Agriculture," **Agricultural Outlook** (USDA), March 2001, p. 5.
9. National Research Council, **Air Emissions from Animal Feeding Operations: Current Knowledge, Future Needs** (Washington, DC: National Academy Press, 2003), p. 6.
10. Henning Steinfeld et al., **Livestock's Long Shadow: Environmental Issues and Options** (Rome: United Nations Food and Agricultural Organization, 2006), pp. xx–xxii.
11. Jorge Fernandez-Cornejo et al., **The First Decade of Genetically Engineered Crops in the United States/EIB-11** (Washington, DC: U.S. Department of Agriculture, Economic Research Service, April 2006), pp. 6, 8.
12. National Research Council, **Environmental Effects of Transgenic Plants: The Scope and Adequacy of Regulation** (Washington, DC: National Academies Press, 2002), pp. 4–5, 14–15.
13. Michael Pollan, **The Omnivore's Dilemma: A Natural History of Four Meals** (New York: Penguin, 2006), pp. 83–84.
14. Polly Walker et al., "Public Health Implications of Meat Production and Consumption," **Public Health Nutrition**, vol. 8, no. 4 (2005), p. 351.
15. Michael Wang, **The Debate on Energy and Greenhouse Gas Emissions Impacts of Fuel Ethanol** (Argonne National Laboratory, August 3, 2005), <http://www.transportation.anl.gov/pdfs/TA/347.pdf>.
16. Conway, **The Doubly Green Revolution**, pp. 171–73.
17. Mary V. Gold, **Sustainable Agriculture: Definitions and Terms**, SRB 99-02 (U.S. Department of Agriculture, September 1999), [http://www.nal.usda.gov/afsic/AFSIC\\_pubs/srb9902.htm](http://www.nal.usda.gov/afsic/AFSIC_pubs/srb9902.htm).

18. David Tilman, "Global Environmental Impacts of Agricultural Expansion: The Need for Sustainable and Efficient Practices," **Proceedings of the National Academy of Sciences**, Vol. 96 (May 1999), p. 5995.

19. Doug Rickman et al., "Precision Agriculture: Changing the Face of Farming," **Geotimes**, November 2003.

20. United Nations Food and Agriculture Organization, **The State of Food Insecurity in the World 2006** (Rome, 2006), pp. 8–9.

## Glossary

**biodiesel** : A diesel-equivalent, processed fuel derived from biological sources (such as vegetable oils), that can be used in unmodified diesel-engine vehicles.

**biofuel** : Derived from biomass — recently living organisms or their metabolic byproducts, such as manure from cows.

**ethanol** : A flammable, colorless, slightly toxic chemical compound with a distinctive perfume-like odor. Also known as ethyl alcohol, drinking alcohol, or grain alcohol, in common usage it is often referred to simply as alcohol.

**harvest index** : The ratio of grain weight to total plant weight.

**integrated pest management (IPM)** : The use of a combination of the following to limit pest damage to agricultural crops: (1) agricultural practices (2) biological control agents (3) introduction of large numbers of sterile male insects (4) timed application of synthetic chemical pesticides and (5) application of pheromones and juvenile hormones.

**monoculture** : The growing of a single plant species over a large area.

**organic agriculture/farming** : A form of agriculture which avoids or largely excludes the use of synthetic fertilizers and pesticides, plant growth regulators, and livestock feed additives.

**organochlorines** : An organic compound containing at least one covalently bonded chlorine atom.

**photosynthesis** : A process in green plants and some bacteria during which light energy is absorbed by chlorophyll-containing molecules and converted to chemical energy (the light reaction). During the process, carbon dioxide is reduced and combined with other chemical elements to provide the organic intermediates that form plant biomass (the dark reaction). Green plants release molecular oxygen (O<sub>2</sub>), which they derive from water during the light reaction.

**selective breeding** : The process of developing a cultivated breed over time.

**stomata** : Tiny openings or pores, found mostly on the under-surface (epidermis) of a plant leaf, and used for gas exchange.



**transpiration** : The evaporation of water from aerial parts of plants, especially leaves but also stems, flowers, and fruits.

## Unit 8 : Water Resources



San Pedro River Valley, Arizona.

### Overview

Earth's water resources, including rivers, lakes, oceans, and underground aquifers, are under stress in many regions. Humans need water for drinking, sanitation, agriculture, and industry; and contaminated water can spread illnesses and disease vectors, so clean water is both an environmental and a public health issue. In this unit, learn how water is distributed around the globe; how it cycles among the oceans, atmosphere, and land; and how human activities are affecting our finite supply of usable water.

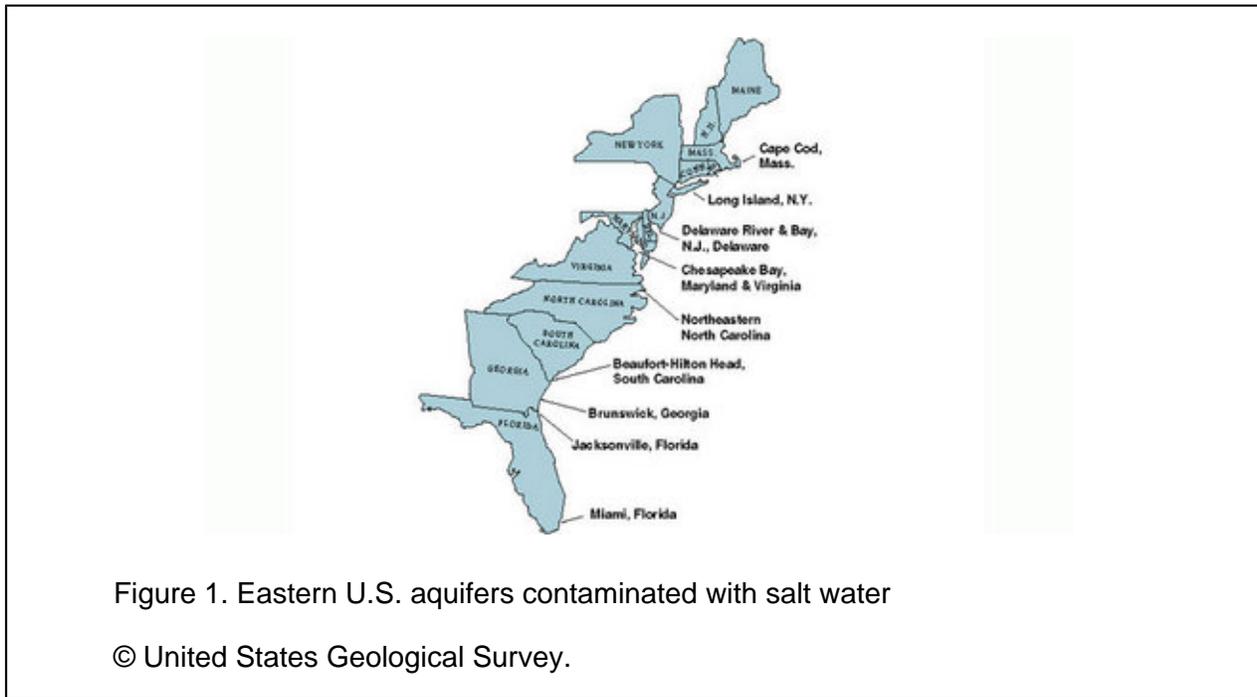
### Sections:

1. Introduction
2. The Global Water Cycle
3. Distribution of Freshwater Resources
4. Groundwater Hydrology: How Water Flows
5. World Demand for Water
6. Depletion of Freshwater Resources
7. Water Salinization
8. Water Pollution
9. Water-Related Diseases
10. Major Laws and Treaties
11. Further Reading

## 1. Introduction

Water resources are under major stress around the world. Rivers, lakes, and underground **aquifers** supply fresh water for irrigation, drinking, and sanitation, while the oceans provide habitat for a large share of the planet's food supply. Today, however, expansion of agriculture, damming, diversion, over-use, and pollution threaten these irreplaceable resources in many parts of the globe.

Providing safe drinking water for the more than 1 billion people who currently lack it is one of the greatest public health challenges facing national governments today. In many developing countries, safe water, free of pathogens and other contaminants, is unavailable to much of the population, and water contamination remains a concern even for developed countries with good water supplies and advanced treatment systems. And over-development, especially in coastal regions and areas with strained water supplies, is leading many regions to seek water from more and more distant sources (Fig. 1).



This unit describes how the world's water supply is allocated between major reserves such as oceans, ice caps, and groundwater. It then looks more closely at how groundwater behaves and how scientists analyze this critical resource. After noting which parts of the world are currently straining their available water supplies, or will do so in the next several decades, we examine the problems posed by salinization, pollution, and water-related diseases.

Scientists widely predict that global climate change will have profound impacts on the hydrologic cycle, and that in many cases these effects will make existing water challenges worse. As we will see in detail in Unit 12, "Earth's Changing Climate," rising global temperatures will alter rainfall patterns, making them stronger in some regions and weaker in others, and may make storms more frequent and severe in some areas of the world. Warming will also affect other aspects of the water cycle by reducing the size of glaciers, snowpacks, and polar ice caps and changing rates of evaporation and transpiration. In sum, climate change is likely to make many of the water-management challenges that are outlined in this unit even more complex than they are today.

At the same time, many current trends in water supply and water quality in Europe and North America are positive. Thirty years ago, many water bodies in developed countries were highly polluted. For example, on June 22, 1969, the Cuyahoga River in Cleveland, Ohio, caught fire when sparks ignited an oily slick of industrial chemicals on its surface. Today, the United States and western European countries have reduced pollution **discharges** into rivers and lakes, often producing quick improvements in water quality. These gains show that when societies make water quality a priority, many polluted sources can be made usable once again. Furthermore, in the United States water consumption rates have consistently declined over the last several decades.

## 2. The Global Water Cycle

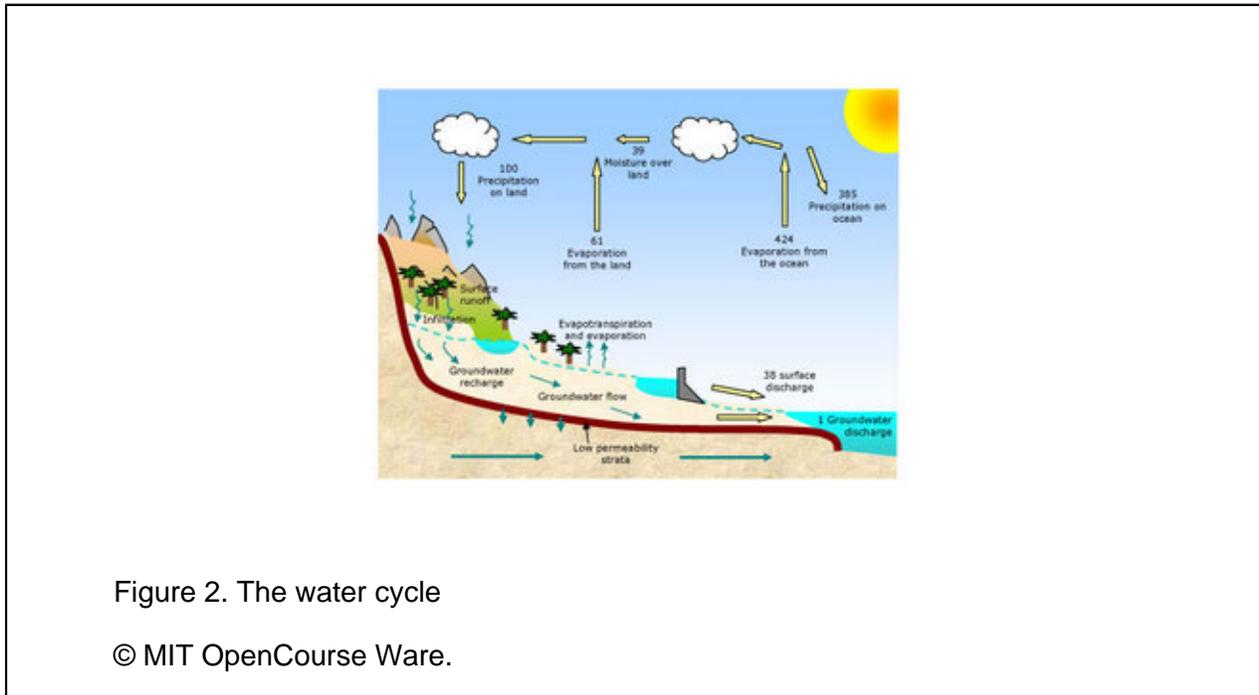
Water covers about three-quarters of Earth's surface and is a necessary element for life. During their constant cycling between land, the oceans, and the atmosphere, water molecules pass repeatedly through solid, liquid, and gaseous phases (ice, liquid water, and water vapor), but the total supply remains fairly constant. A water molecule can travel to many parts of the globe as it cycles.

As discussed in Unit 2, "Atmosphere," and Unit 3, "Oceans," water vapor redistributes energy from the sun around the globe through atmospheric circulation. This happens because water absorbs a lot of energy when it changes its state from liquid to gas. Even though the temperature of the water vapor may not increase when it evaporates from liquid water, this vapor now contains more energy, which is referred to as latent heat. Atmospheric circulation moves this latent heat around Earth, and when water vapor condenses and produces rain, the latent heat is released.

Very little water is consumed in the sense of actually taking it out of the water cycle permanently, and unlike energy resources such as oil, water is not lost as a consequence of being used. However, human intervention often increases the flux of water out of one store of water into another, so it can deplete the stores of water that are most usable. For example, pumping groundwater for irrigation depletes aquifers by transferring the water to evaporation or river flow. Our activities also pollute water so that it is no longer suitable for human use and is harmful to ecosystems.

There are three basic steps in the global water cycle: water precipitates from the atmosphere, travels on the surface and through groundwater to the oceans, and evaporates or transpires back to the

atmosphere from land or evaporates from the oceans. Figure 2 illustrates yearly flow volumes in thousands of cubic kilometers.



Supplies of **freshwater** (water without a significant salt content) exist because precipitation is greater than evaporation on land. Most of the precipitation that is not transpired by plants or evaporated, infiltrates through soils and becomes **groundwater**, which flows through rocks and sediments and discharges into rivers. Rivers are primarily supplied by groundwater, and in turn provide most of the freshwater discharge to the sea. Over the oceans evaporation is greater than precipitation, so the net effect is a transfer of water back to the atmosphere. In this way freshwater resources are continually renewed by counterbalancing differences between evaporation and precipitation on land and at sea, and the transport of water vapor in the atmosphere from the sea to the land.

Nearly 97 percent of the world's water supply by volume is held in the oceans. The other large reserves are groundwater (4 percent) and icecaps and glaciers (2 percent), with all other water bodies together accounting for a fraction of 1 percent. Residence times vary from several thousand years in the oceans to a few days in the atmosphere (Table 1).

**Table 1.** Estimate of the world water balance.

	Surface area (million km)	Volume (million km )	Volume (%)	Equivalent depth (m)	Residence time
<b>Oceans and seas</b>	361	1,370	94	2,500	~4,000 years
<b>Lakes and reservoirs</b>	1.55	0.13	<0.01	0.25	~10 years
<b>Swamps</b>	<0.1	<0.01	<0.01	0.007	1-10 years
<b>River channels</b>	<0.1	<0.01	<0.01	0.003	~2 weeks
<b>Soil moisture</b>	130	0.07	<0.01	0.13	2 weeks to 50 years
<b>Groundwater</b>	130	60	4	120	2 weeks to 100,000 years
<b>Icecaps and glaciers</b>	17.8	30	2	60	10 to 1,000 years
<b>Atmospheric water</b>	504	0.01	<0.01	0.025	~10 days
<b>Biospheric water</b>	<0.1	<0.01	<0.01	0.001	~1 week

Solar radiation drives evaporation by heating water so that it changes to water vapor at a faster rate. This process consumes an enormous amount of energy—nearly one-third of the incoming solar energy that reaches Earth's surface. On land, most evaporation occurs as transpiration through plants: water is taken up through roots and evaporates through stomata in the leaves as the plant takes in CO<sub>2</sub>. A single large oak tree can transpire up to 40,000 gallons per year (footnote 1). Much of the water moving through the hydrologic cycle thus is involved with plant growth.

Since evaporation is driven by heat, it rises and falls with seasonal temperatures. In temperate regions, water stores rise and fall with seasonal evaporation rates, so that net atmospheric input (precipitation minus evaporation) can vary from positive to negative. Temperatures are more constant in tropical regions where large seasonal differences in precipitation, such as monsoon cycles, are the main cause of variations in the availability of water. In an effort to reduce these seasonal swings, many countries have built reservoirs to capture water during periods of high flow or flooding and release water during periods of low flow or drought. These projects have increased agricultural production and mitigated floods and droughts in some regions, but as we will see, they have also had major unintended impacts on water supplies and water quality.



The hydrologic cycle is also coupled with material cycles because rainfall erodes and weathers rock. Weathering breaks down rocks into gravel, sand, and sediments, and is an important source of key nutrients such as calcium and sulfur. Estimates from river outflows indicate that some 17 billion tons of material are transported into the oceans each year, of which about 80 percent is particulate and 20 percent is dissolved. On average, Earth's surface weathers at a rate of about 0.5 millimeter per year. Actual rates may be much higher at specific locations and may have been accelerated by human activities, such as emissions from fossil fuel combustion that make rain and snowfall more acidic.

### 3. Distribution of Freshwater Resources

Freshwater accounts for only some 6 percent of the world's water supply, but is essential for human uses such as drinking, agriculture, manufacturing, and sanitation. As discussed above, two-thirds of global freshwater is found underground.

If you dig deeply enough anywhere on Earth, you will hit water. Some people picture groundwater as an underground river or lake, but in reality it is rarely a distinct water body (large caves in limestone aquifers are one exception). Rather, groundwater typically fills very small spaces (pores) within rocks and between sediment grains.

The water table is the top of the saturated zone (Fig. 3). It may lie hundreds of meters deep in deserts or near the surface in moist ecosystems. Water tables typically shift from season to season as precipitation and transpiration levels change, moving up during rainy periods or periods of little transpiration and sinking during dry phases when the rate of **recharge** (precipitation minus evaporation and transpiration that infiltrates from the surface) drops. In temperate regions the water table tends to follow surface topography, rising under hills where there is little discharge to streams and falling under valleys where the water table intersects the surface in the form of streams, lakes, and springs.

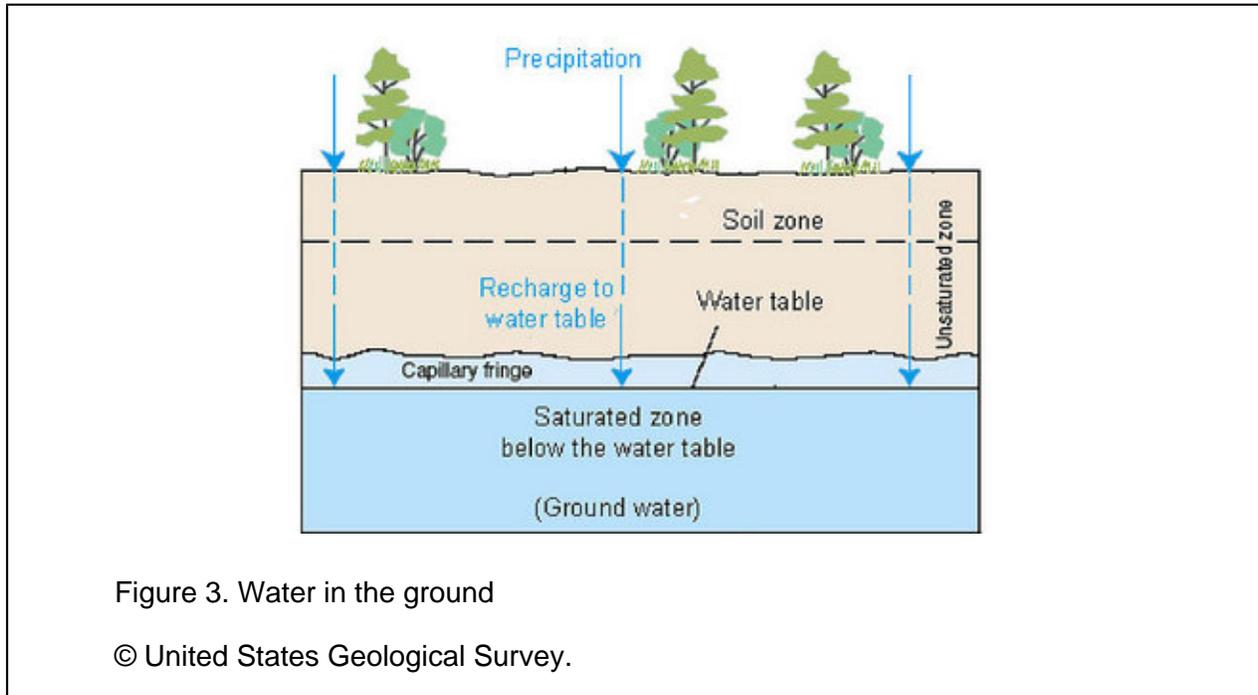


Figure 3. Water in the ground

© United States Geological Survey.

Above the water table lies the unsaturated zone, also referred to as the **vadose zone**, where the pores (spaces between grains) are not completely filled with water. Water in the vadose zone is referred to as soil moisture. Although air in the vadose zone is at atmospheric pressures, the soil moisture is under tension, with suctions of a magnitude much greater than atmospheric pressure.

This fluid tension is created by strong adhesive forces between the water and the solid grains, and by surface tension at the small interfaces between water and air. The same forces can be seen at work when you insert a thin straw (a capillary) into water: water rises up in the straw, forming a meniscus at the top. When the straw is thinner, water rises higher because the ratio of the surface area of the straw to the volume of the straw is greater, increasing the adhesive force lifting the water relative to the gravitational force pulling it down. This explains why fine-grained soils, such as clay, can hold water under very large suctions.

Water flows upward under suction through small pores from the water table toward plant roots when evapotranspiration is greater than precipitation. After a rainstorm, water may recharge the groundwater by saturating large pores and cracks in the soil and flowing very quickly downward to the water table.

Millions of people worldwide depend on groundwater stocks, which they draw from aquifers—permeable geologic formations through which water flows easily. Very transmissive geologic formations are desirable because water levels in wells decline little even when pumping rates are high, so the wells do not need to be drilled as deeply as in less transmissive formations and the

energy costs of lifting water to the surface are not excessive. Under natural conditions many aquifers are **artesian**: the water they hold is under pressure, so water will flow to the surface from a well without pumping.

Aquifers may be either capped by an impermeable layer (confined) or open to receive water from the surface (unconfined). Confined aquifers are often artesian because the confining layer prevents upward flow of groundwater, but unconfined aquifers are also artesian in the vicinity of discharge areas. This is why groundwater discharges into rivers and streams. Confined aquifers are less likely to be contaminated because the impermeable layers above them prevent surface contaminants from reaching their water, so they provide good-quality water supplies (Fig. 4).

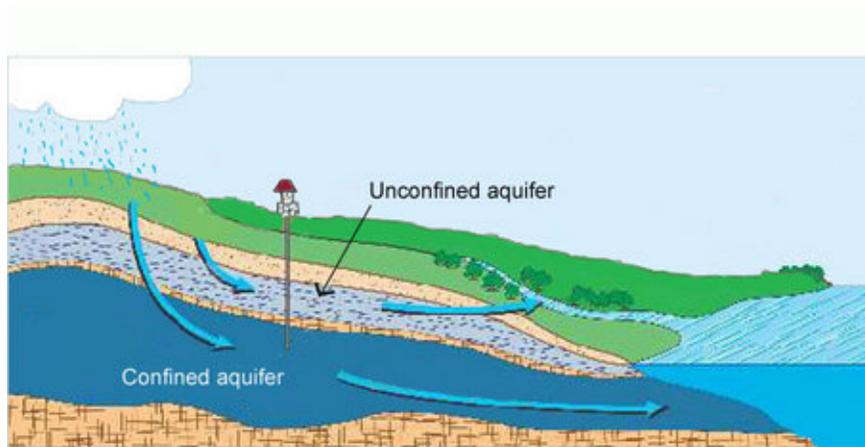


Figure 4. Confined and unconfined aquifer

Water has an average residence time of thousands to tens of thousands of years in many aquifers, but the actual age of a water sample collected from a particular well will vary tremendously within an aquifer. Shallow groundwater can discharge into streams and rivers in weeks or months, but some deep groundwater is millions of years old—as old as the rocks that hold the water in their pores. Because of this distribution of residence times in aquifers, contaminants that have been introduced at the surface over the last century are only now beginning to reach well depths and contaminate drinking water in many aquifers. Indeed, much of the solute load (salt and other contaminants) that has entered aquifers due to increased agriculture and other land use changes over the last several centuries has yet to reach discharge areas where it will contaminate streams and lakes (footnote 2).

Ice sheets and glaciers are not always thought of as freshwater sources, but they account for a significant fraction of world reserves. Nearly 90 percent of the water in icecaps and glaciers is in Antarctica, with another 10 percent in the Greenland ice sheet and the remainder in tropical and temperate glaciers. As discussed in Unit 1, "Many Planets, One Earth," and Unit 12, "Earth's Changing Climate," Earth's ice sheets constantly expand and contract as the planet's climate fluctuates. During warm periods ice sheets melt and sea levels rise, with the reverse occurring when temperatures fall. Water may remain locked in deep layers of polar ice sheets for hundreds of thousands of years.

Rivers contain a relatively small share of fresh water, but the flux of water down rivers is a large part of the global hydrologic cycle and they are centrally important in shaping landscapes. Their flow erodes solid sediment and carries it toward the sea, along with dissolved minerals. These processes shape land into valleys and ridges and deposit thick layers of sediment in flood plains. Over geologic time the erosion caused by rivers balances the uplift driven by plate tectonics. Much of Earth's freshwater flow passes through several of the planet's largest rivers: the Amazon carries 15 percent of total river flow on Earth, the Congo carries 3.5 percent, and rivers that flow into the Arctic Ocean carry 8 percent. The average residence time of water in rivers is less than a year.

#### 4. Groundwater Hydrology: How Water Flows

How does water move through the ground and interact with sediments and rock? Will an aquifer recharge slowly or quickly after water is withdrawn, and where will new groundwater come from? These questions are central for communities that need adequate drinking water, farmers tending crops and livestock, and engineers working to keep water supplies free of contaminants. For example, the 1986 trial recounted in the book and movie **A Civil Action** focused on town drinking wells in Woburn, Massachusetts, that were polluted with industrial chemicals suspected of causing cancer among residents. Plaintiffs asserted—and an investigation by the Environmental Protection Agency ultimately confirmed—that chemicals dumped by several local businesses had flowed through groundwater to the underlying aquifer and contaminated the wells (footnote 3).

The pore structure of soils, sediment, and rock is a central influence on groundwater movement. Hydrologists quantify this influence primarily in terms of:

- **porosity**: the proportion of total volume that is occupied by voids, like the spaces within a pile of marbles. Porosity is not a direct function of the size of soil grains—the porosity of a pile of basketballs is the same as a pile of marbles. Porosity tends to be larger in well sorted sediments where the grain sizes are uniform, and smaller in mixed soils where smaller grains fill the voids between larger grains. Soils are less porous at deeper levels because the weight of overlying soil packs grains closer together.

- **permeability**: how readily the medium transmits water, based on the size and shape of its pore spaces and how interconnected its pores are.

Materials with high porosity and high permeability, such as sand, gravel, sandstone, fractured rock, and basalt, produce good aquifers. Low-permeable rocks and sediments that impede groundwater flow include granite, shale, and clay.

Groundwater recharge enters aquifers in areas at higher elevations (typically hill slopes) than discharge areas (typically in the bottom of valleys), so the overall movement of groundwater is downhill. However, within an aquifer, water often flows upward toward a discharge area (Fig. 5). To understand and map the complex patterns of groundwater flow, hydrogeologists use a quantity called the **hydraulic head**. The hydraulic head at a particular location within an aquifer is the sum of the elevation of that point and the height of the column of water that would fill a well open only at that point. Thus, the hydraulic head at a point is simply the elevation of water that rises up in a well open to the aquifer at that point.

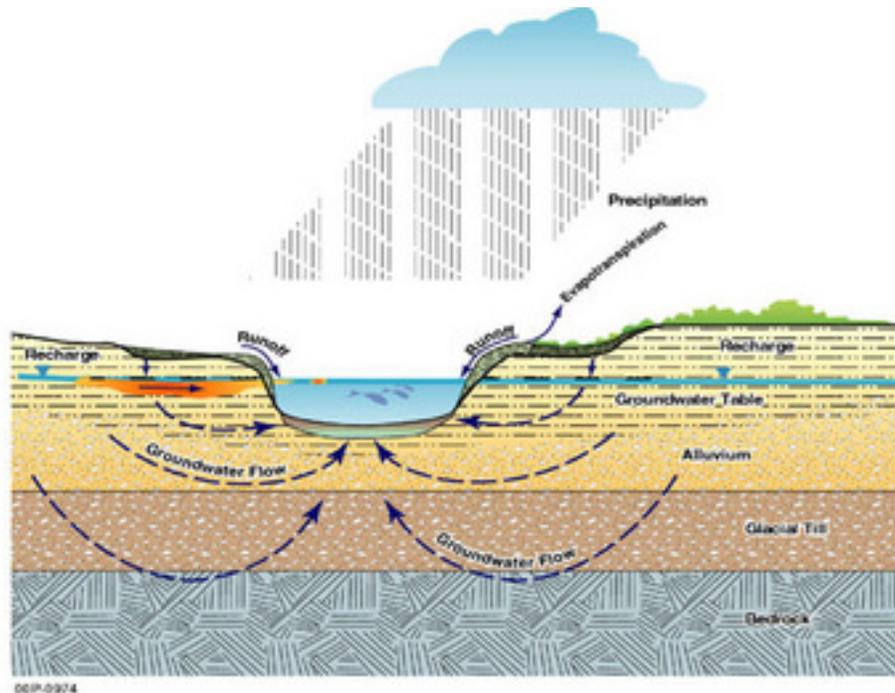


Figure 5. Groundwater flow under the Housatonic River, Pittsfield, Massachusetts

© United States Environmental Protection Agency.

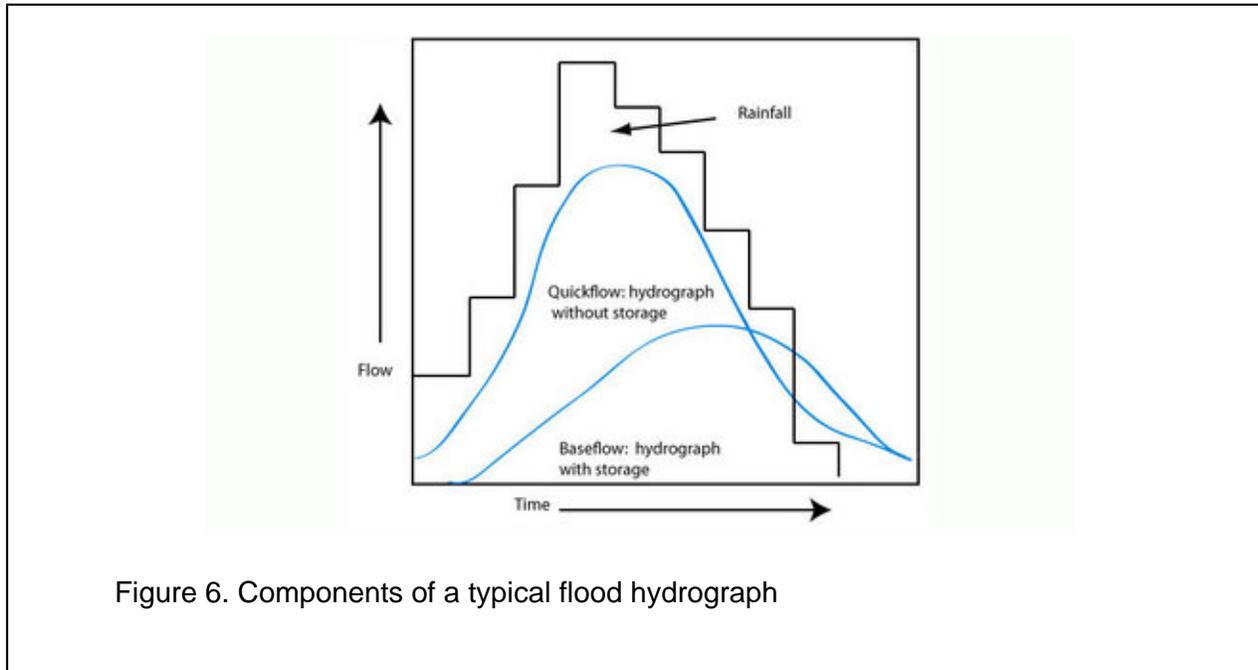


The height of water within the well is not the same as the distance to the water table. If the aquifer is under pressure, or artesian, this height may be much greater than the distance to the water table. Thus the hydraulic head is the combination of two potentials: mechanical potential due to elevation, like a ball at the top of a ramp, and pressure potential, like air compressed in a balloon. Because these are usually the only two significant potentials driving groundwater flow, groundwater will flow from high to low hydraulic head.

This theory works in the same way that electrical potential (voltage) drives electrical flow and thermal potential (temperature) drives heat conduction. Like these other fluxes, groundwater flux between two points is simply proportional to the difference in potential, hydraulic head, and also to the permeability of the medium through which flow is taking place. These proportionalities are expressed in the fundamental equation for flow through porous media, known as Darcy's Law.

The gradient in hydraulic potential may drive groundwater flow downward, upward, or horizontally. Hydrogeologists collect water levels measured in wells to map hydraulic potential in aquifers. These maps can then be combined with permeability maps to determine the pattern in which groundwater flows throughout the aquifer.

Depending on local rainfall, land use, and geology, streams may be fed by either groundwater discharge or surface runoff and direct rainfall, or by some combination of surface and groundwater. Perennial streams and rivers are primarily supplied by groundwater, referred to as baseflow. During dry periods they are completely supplied by groundwater; during storms there is direct runoff and groundwater discharge also increases. The hydrograph in Figure 6 shows flow patterns in a stream before, during, and after a storm with relative contributions from groundwater (baseflow) and surface water (quickflow, also referred to as storm flow).



## 5. World Demand for Water

How much water do humans use? The answer depends on where they live and on their socioeconomic status. Under primitive conditions a person will consume three to five gallons per day for drinking and subsistence farming. In a city where water is also used for cleaning, manufacturing, and sanitation, per capita use is around 150 gallons per day. In the United States, which has among the highest water consumption rates in the world, each person uses an average of 1,340 gallons of water per day. Table 2 shows how much water is required to produce common goods and services.

**Table 2.** Average water requirements.

Item	Gallons used
1 pound of cotton	2,000
1 pound of grain-fed beef	800
1 loaf of bread	150
1 car	100,000
1 kilowatt hour of electricity	25
1 pound of rubber	100

Item	Gallons used
1 pound of steel	25
1 gallon of gasoline	10
1 load of laundry	60
1 ten-minute shower	25-50

As discussed in Unit 2, "Atmosphere," and Unit 3, "Oceans," water resources are not distributed evenly in space or time around the world. Global circulation patterns create wet and dry climate zones, and in some regions seasonal or multi-annual climate cycles generate distinct wet and dry phases. As a result, some regions have larger freshwater endowments than others (Fig. 7).

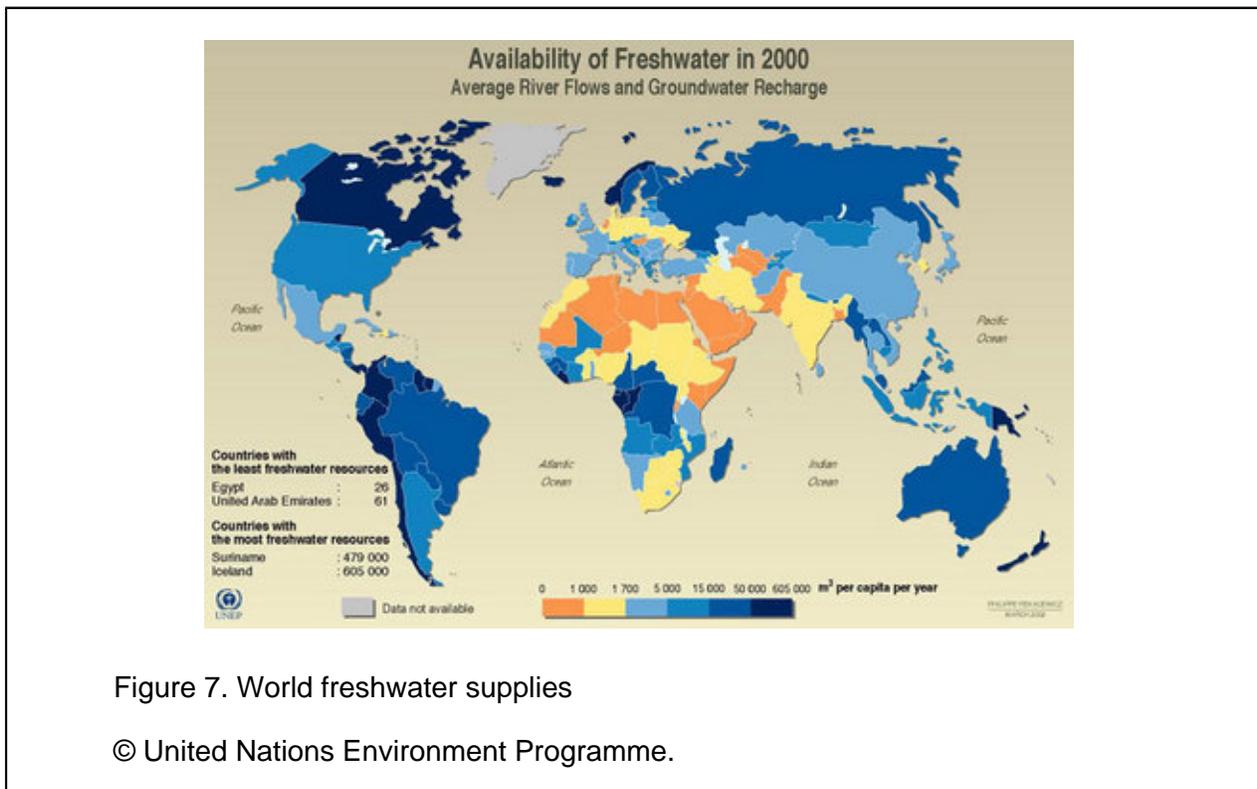
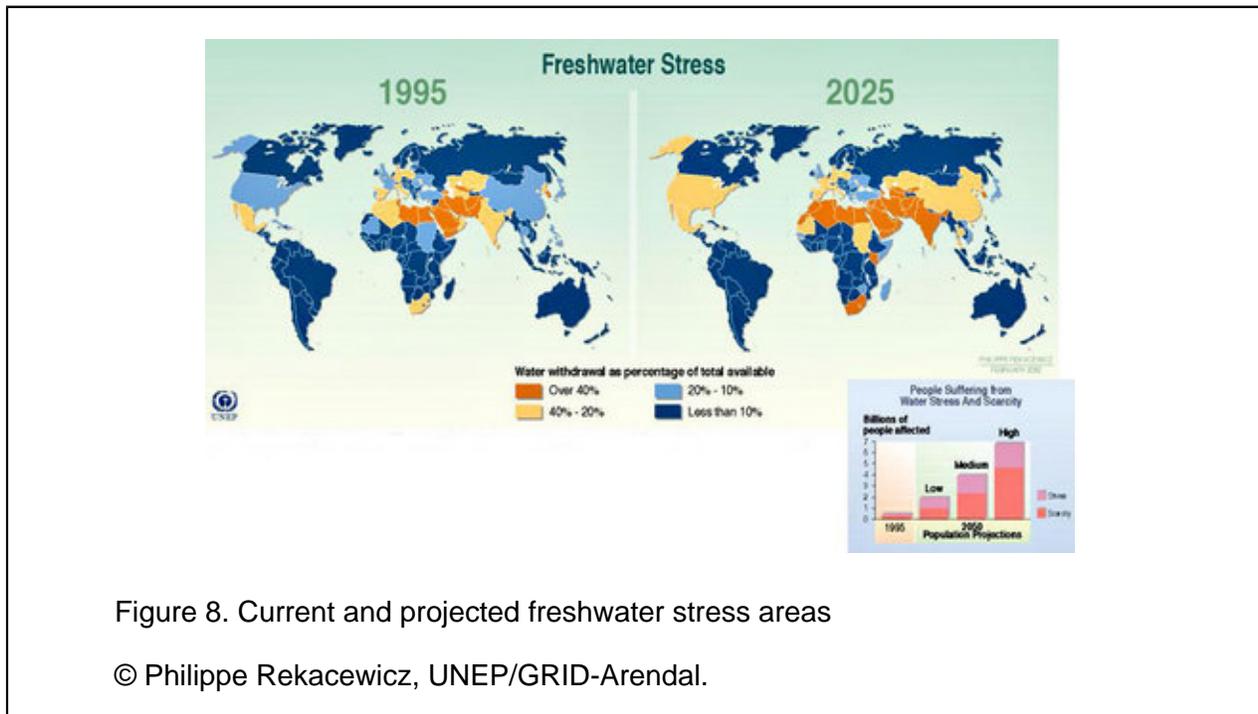


Figure 7. World freshwater supplies

© United Nations Environment Programme.

Although developed nations generally have more water available than many countries in Africa and the Middle East, some areas with good water endowments still are subject to "water stress" because they are withdrawing water from available supplies at extremely high rates (Fig. 8). High-intensity water uses in industrialized nations include agricultural production and electric power generation, which requires large quantities of water for cooling. In the United States electric power production accounts for 39 percent of all freshwater withdrawals (footnote 4), although almost all of this water is

immediately returned to the rivers from which it is withdrawn. Agriculture consumes much more water because irrigation increases transpiration to the atmosphere.

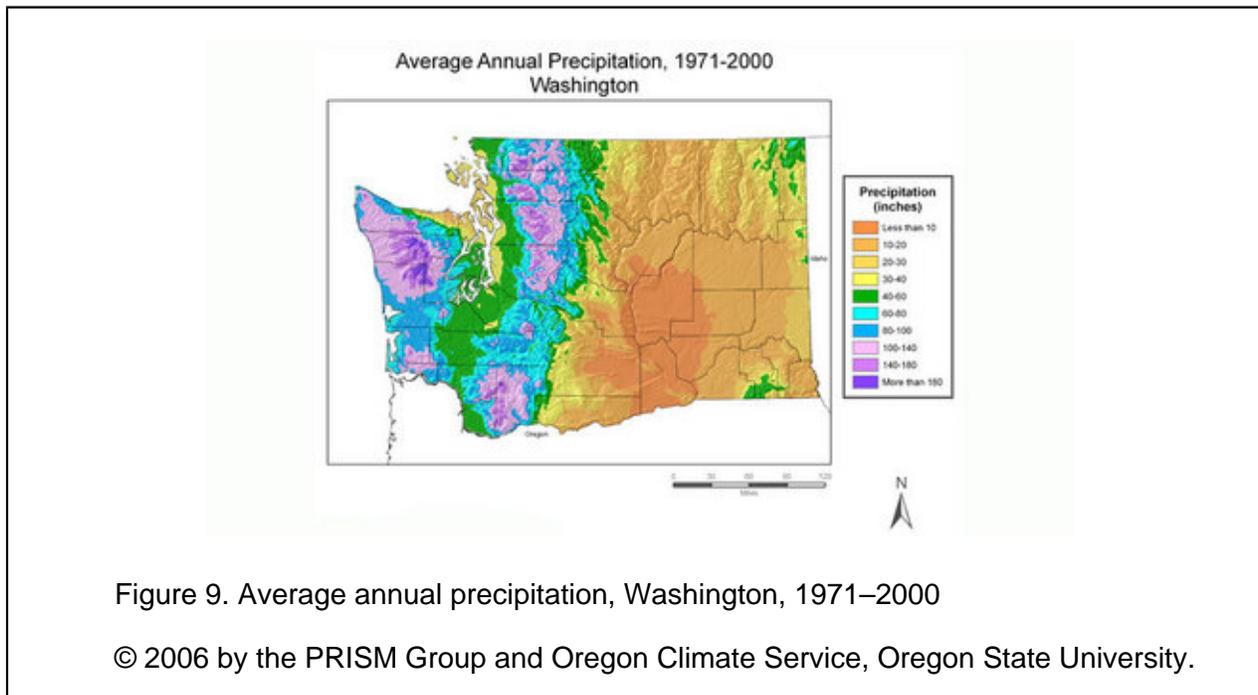


As of 2002, 1.1 billion people around the world (17 percent of global population) did not have access to safe drinking water and 2.6 billion people (42 percent of global population) lived without adequate sanitation. As a result, millions of people die each year of preventable water-related diseases. Most of the countries with inadequate supplies of safe drinking water are located in Africa, Asia, and the Pacific, but problems persist elsewhere as well. For example, many households lack adequate sewage treatment services in Eastern Europe. And inequity among water users is widespread: cities often receive better service than rural areas, and many poor communities in both rural and urban areas lack clean water and sanitation (footnote 5).

Although these challenges apply in many regions, it is hard to make broad generalizations about water resources at the global or national level; to paraphrase the famous saying about politics, all hydrology is local. The basic geologic unit that scientists focus on to characterize an area's water supply and water quality with precision is the [watershed](#) or [catchment area](#)—an area of land that drains all streams and rainfall to a common outlet such as a bay or river delta. Large watersheds, such as the Amazon, the Mississippi, and the Congo contain many smaller sub-basins (footnote 6).

To see why water issues are best studied at the watershed level, consider Washington State, which is divided centrally by the Cascade Mountains. West of the Cascades, Washington receives up to 160

inches of rainfall annually, and the mild, humid climate supports temperate rainforests near the Pacific coast. Across the Cascades, rainfall is as low as six inches per year in the state's semiarid interior where groundwater is pumped from deep within basalt formations to grow wheat (Fig. 9). Urban Seattle residents and ranchers in rural eastern Washington thus face very different water supply, runoff, and water quality issues.



Currently 10,000 to 12,000 cubic kilometers of freshwater are available for human consumption each year worldwide. In the year 2000 humans withdrew about 4,000 km<sup>3</sup> from this supply. About half of the water withdrawn was consumed, meaning that it was evaporated, transpired by plants, or contaminated beyond use, and so became temporarily unavailable for other users. The other 50 percent was returned to use: for example, some water used for irrigation drains back into rivers or recharges groundwater, and most urban wastewater is treated and returned to service.

Of the water withdrawn for human use, 65 percent went to agriculture, 10 percent to domestic use (households, municipal water systems, commercial use, and public services), 20 percent to industry (mostly electric power production), and 5 percent evaporated from reservoirs (footnote 7). About 70 percent of the water used for agriculture was consumed, compared to 14 percent of water used for domestic consumption and 11 percent of water used for industry.

Both population levels and economic development are important drivers of world water use. If current patterns continue, the World Water Council estimates that total yearly withdrawals will rise to more

than 5,000 km<sup>3</sup> by 2050 as world population rises from 6.1 billion to 9.2 billion. During the 20th century, world population tripled but water use rose by a factor of six (footnote 8). The United Nations and the international community have set goals of halving the number of people without adequate safe drinking water and sanitation by 2015. Meeting this target will require providing an additional 260,000 people per day with clean drinking water and an additional 370,000 people per day with improved sanitation through the year 2014, even as overall world demand for water is rising (footnote 9).

## 6. Depletion of Freshwater Resources

In many parts of the world people are extracting water from aquifers more quickly than the aquifers are replenished by recharge. In addition to draining aquifers, excessive groundwater pumping changes groundwater flow patterns around wells and can drain nearby rivers and streams. This happens because pumping changes the natural equilibrium that exists in an undeveloped aquifer with discharge balancing recharge.

When pumping starts, groundwater stores are depleted in the vicinity of the well, creating a cone of depression in the hydraulic head. If a new water source such as a river or stream is available close by, the well may capture (draw water from) that source and increase its recharge rate (Fig. 10) until this inflow matches the pumping rate. If no such source is available and pumping draws the water table down far enough, it will dry up the aquifer or deplete it so far that is it not physically possible or affordable to pump out the last stores of water.

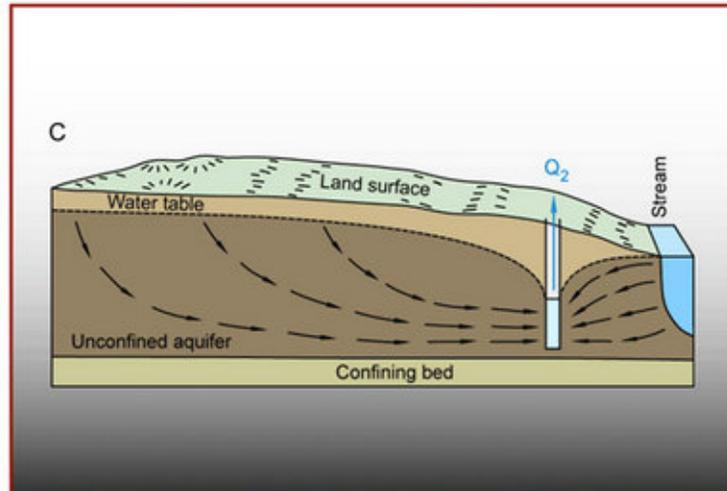


Figure 10. Effects of groundwater pumping

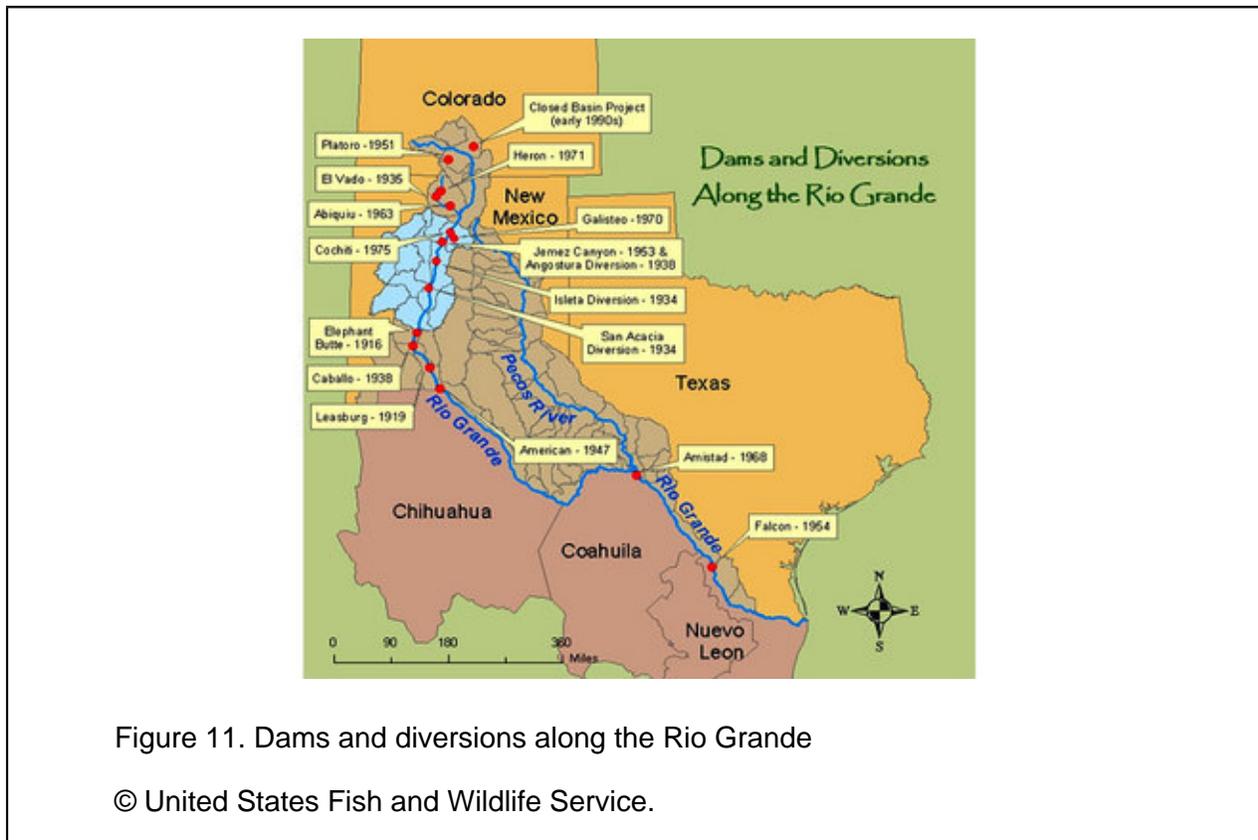
© United States Geological Survey.

Pumping quickly lowers the pressure within confined aquifers so that water no longer rises to the surface naturally. Fifty years ago artesian aquifers were common, but today they have become rare because of widespread groundwater withdrawals. In unconfined aquifers, air fills pores above the water table, so the water table falls much more slowly than in confined aquifers.

As aquifers are depleted, water has to be lifted from much greater depths. In some parts of the world, the energy costs of lifting groundwater from deep beneath the surface have become prohibitive. Overuse of groundwater can also reduce the quality of the remaining water if wells draw from contaminated surface sources or if water tables near the coast drop below sea level, causing salt water to flow into aquifers.

Serious groundwater depletion has occurred in major parts of North Africa, the Middle East, South and Central Asia, North China, North America, and Australia, along with other localized areas worldwide (footnote 10). In some cases, such as the Ogallala aquifer in the central United States, water tables are falling so low that wells can no longer produce water. In a draft plan issued in mid-2006, the Texas Water Development Board projected that the state's water supplies would fall by about 18 percent between 2010 and 2060, "primarily due to the accumulation of sediments in reservoirs and the depletion of aquifers," and that at the same time the state's population would more than double. If Texas did not implement the water management plan, the board estimated, water shortages could cost the state nearly \$100 billion by 2060 (footnote 11).

Many rivers around the globe have also been depleted by increasing water withdrawals. Some, such as the Colorado and Rio Grande, no longer reach the sea during much of the year because their flow levels have been reduced so drastically by dams and water diversion (Fig. 11). This overuse destroys estuaries at river mouths, which are important habitats and breeding grounds for fish and birds.



Under normal conditions, most rivers are gaining rivers: groundwater flows into the rivers because the local water table sits at a higher elevation than the river water. However, with excessive groundwater pumping, water tables slowly decline and natural discharge to the rivers is reduced, so river flow declines. Over the long term, groundwater extraction may greatly reduce river flows in many regions. This connection between water levels in aquifers and river flows complicates the process of estimating sustainable yield from aquifers. If users pump more water from an aquifer than the natural rate of recharge, the aquifer may draw water from adjoining rivers and increase its rate of recharge. However, by doing so it will reduce surface water flows.



*Almost every country in the world that uses groundwater as a resource is having troubles with it affecting surface water systems.*

Tom Maddock, University of Arizona

By regulating river flows to reduce floods and increase flows during dry periods, dams have major impacts on river ecosystems. Like forest fires, river floods play important ecological roles that we have only begun to appreciate and foster in recent decades. Among other services, floods scour out channels, deposit nutrient-rich sediments on flood plains, and help to replenish groundwater.

In regions where rivers have been channeled between levees to prevent flooding, they no longer deposit sediments and nutrients on surrounding lands. Scientists widely agree that damage from Hurricane Katrina in August 2005 was magnified because levees and canals around New Orleans had directed the Mississippi River's flow straight into the Gulf of Mexico for decades. Without fresh water and sediment from the Mississippi, southern Louisiana's wetlands degraded and subsided, reducing their ability to buffer the region against storms and flooding.

## 7. Water Salinization

When freshwater resources become saline, they can no longer be used for irrigation or drinking. Saline water is toxic to plants, and high sodium levels cause dry soils to become hard and compact and reduce their ability to absorb water. Irrigation water becomes toxic to most plants at concentrations above 1,300 milligrams/liter; for comparison, the salinity of seawater is about 35,000 mg/l (footnote 12). Salinity is not dangerous to humans, but water becomes nonpotable for human consumption at about 250 mg/l.

Groundwater extraction and irrigation can increase salt concentrations in water and soils in several ways. First, irrigation increases the salinity of soil water when evaporation removes water but leaves salt behind. This occurs when irrigation water contains some salt and irrigation rates are not high enough to flush the salt away. Saline water in the vadose zone can then contaminate surface water and soils. Irrigation has caused high salinity levels in areas including the cotton growing region near the Aral Sea in Central Asia, the lower reaches of the Colorado River, and California's Central Valley (Fig. 12).



Figure 12. Fields in central California suffering from severe salinization

© United States Department of Agriculture, Agricultural Research Service.

Irrigation can also cause salinization by raising the water table and lifting saline groundwater near the surface into the root zone. This occurs when irrigation efficiency is poor, so a large fraction of irrigation water infiltrates into the soil, and groundwater flow is slow. A similar problem occurs in some regions when trees are cut down, reducing transpiration and increasing the rate at which water flushes through the vadose zone. The increased infiltration flushes high concentrations of salt to the water table and lifts the water table toward the surface. This process has severely affected the Murray-Darling Basin in Australia.

A third type of salinization occurs in coastal areas, where excessive groundwater pumping draws seawater into aquifers and contaminates wells. In coastal aquifers freshwater floats on top of denser seawater. When this lens of freshwater is diminished by withdrawals, seawater rises up from below. Because world populations are increasing particularly rapidly in coastal regions, seawater intrusion is a threat in many coastal aquifers.

A recent analysis by scientists at the Institute of Ecosystem Studies found that salinity levels have also increased significantly in urban and suburban areas in the northeastern United States. The authors attributed this rise to two main factors: use of salts for de-icing roads in winter and increased levels of street paving. These trends deliver concentrated bursts of saline runoff to local water bodies after storms and floods. "As coverage by impervious surfaces increases, aquatic systems can receive increased and pulsed applications of salt, which can accumulate to unsafe levels in ground and surface waters over time," the authors observe (footnote 13).

## 8. Water Pollution

Many different types of contaminants can pollute water and render it unusable. Pollutants regulated in the United States under national primary drinking water standards (legally enforceable limits for public water systems to protect public health) include:

- **Microorganisms** such as cryptosporidium, giardia, and fecal coliform bacteria
- **Disinfectants and water disinfection byproducts** including chlorine, bromate, and chlorite
- **Inorganic chemicals** such as arsenic, cadmium, lead, and mercury
- **Organic chemicals** such as benzene, dioxin, and vinyl chloride
- **Radionuclides** including uranium and radium

These pollutants come from a wide range of sources. Microorganisms are typically found in human and animal waste. Some inorganic contaminants such as arsenic and radionuclides such as uranium occur naturally in geologic deposits, but many inorganic and most major organic pollutants are emitted from industrial facilities, mining, and agricultural activities such as fertilizer and pesticide application.

Sediments (soil particles) from erosion and activities such as excavation and construction also pollute rivers, lakes, and coastal waters. As discussed in Unit 3, "Oceans," availability of light is the primary constraint on photosynthesis in aquatic ecosystems, so adding sediments can severely affect productivity in these ecosystems by clouding the water. It also smothers fish and shellfish spawning grounds and degrades habitat by filling in rivers and streams (Fig. 13).

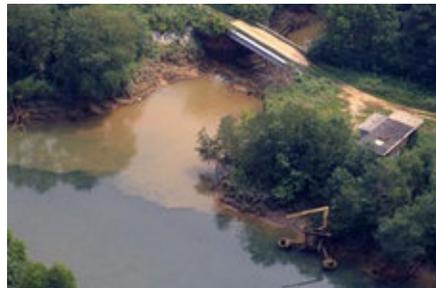
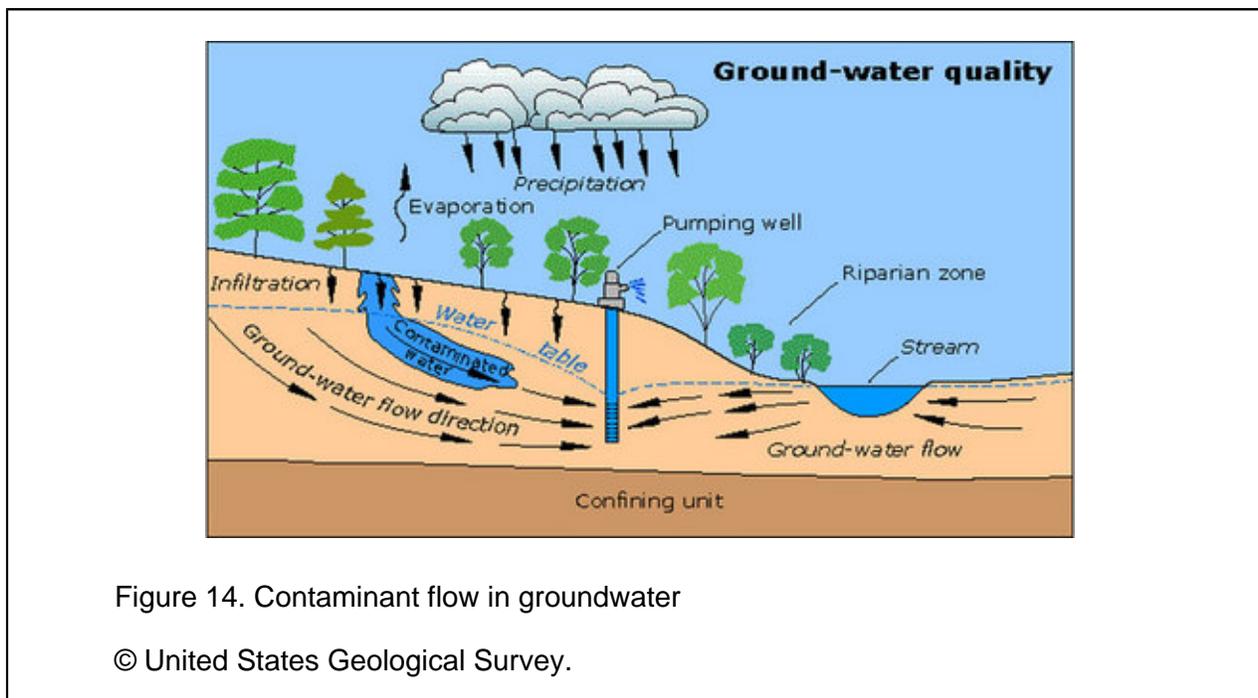


Figure 13. Sedimentation in Chattahoochee River, Atlanta, Georgia

© United States Geological Survey.

Water supplies often become polluted because contaminants are introduced into the vadose zone or are present there naturally and penetrate to the water table or to groundwater, where they move into wells, lakes, and streams. Many dissolved compounds can be toxic and carcinogenic, so keeping them out of water supplies is a central public-health goal. One critical question is how compounds of concern behave in water. **Non-aqueous phased liquids** (NAPLs) form a separate phase that does not mix with water and can reside as small blobs within the pore structure of aquifers and soils. Some, such as gasoline and diesel fuel, are lighter than water and will float on top. Others, including chlorinated hydrocarbons and carbon tetrachloride, are denser and will sink. Both types are difficult to remove and will slowly dissolve into groundwater, migrating downgradient as groundwater flows.

Other contaminants completely dissolve in water and, if they enter the aquifer at a single location (e.g., from a point source), are transported with flowing groundwater as plumes that gradually mix with native groundwater (Fig. 14). Over time, contaminated zones become larger but concentrations fall as the plume spreads. The paths that plumes follow can be extremely complex because of the complicated patterns of permeability within aquifers. Groundwater velocities are much higher through channels of high permeability, so these channels transport dissolved contaminants rapidly through the subsurface.



As a plume moves through groundwater, some contaminants in it may bind to soil particles, a process called **sorption**. High organic material and clay content in soils generally increases sorption because these particles are chemically reactive and have large surface areas. Sorption may prevent

contaminants from migrating: for example, in some spills containing uranium, the uranium has moved only a few meters over decades. However, contaminants like uranium can also adsorb to very small suspended particles called colloids that migrate easily through aquifers. Even if a contaminated plume is pumped out, sorbed contaminants may remain on the solid matrix to desorb later back into the groundwater, so sorption makes full cleanup of the contamination more expensive and time-consuming.

Water pollution is relatively easier to control when it comes from a **point source**—a distinct, limited discharge source such as a factory, which can be required to clean up or reduce its effluent. **Nonpoint source** pollution consists of diffuse, nonbounded discharges from many contributors, such as runoff from city streets or agricultural fields, so it is more challenging to control.

Approaches for controlling nonpoint source pollution include improving urban stormwater management systems; regulating land uses; limiting broad application of pesticides, herbicides, and fertilizer; and restoring wetlands to help absorb and filter runoff (Box 1). U.S. regulations are increasingly emphasizing limits on total discharges to water bodies from all sources (for details, see the discussion of Total Maximum Daily Loads below in Section 10, "Major Laws and Treaties").

Along with freshwater bodies, many coastal areas and **estuaries** (areas where rivers meet the sea, mixing salt and fresh water) are severely impacted by water pollution and sedimentation. Ocean pollution kills fish, seabirds, and marine mammals; damages aquatic ecosystems; causes outbreaks of human illness; and causes economic damage through impacts on activities such as tourism and fishing.

A 2000 National Research Council report cited nutrient pollution (excess inputs of nitrogen and phosphorus) as one of the most important ocean pollution problems in the United States (footnote 14). As discussed in Unit 3, "Oceans," and Unit 4, "Ecosystems," nutrient-rich runoff into ocean waters stimulates plankton to increase photosynthesis and causes "blooms," or population explosions. When excess plankton die and sink, their decomposition consumes oxygen in the water.

Since the beginning of the industrial age, human activities, especially fertilizer use and fossil fuel combustion, have roughly doubled the amount of nitrogen circulating globally, increasing the frequency and size of plankton blooms. This process can create **hypoxic** areas ("dead zones"), where dissolved oxygen levels are too low to support marine life—typically less than two to three milligrams per liter. Seasonal dead zones regularly appear in many parts of the world. One of the largest, in the Gulf of Mexico, covers up to 18,000 square kilometers each summer, roughly the size of New Jersey (Fig. 15), where river and groundwater flow deliver excess nutrients from upstream agricultural sources to the coast.

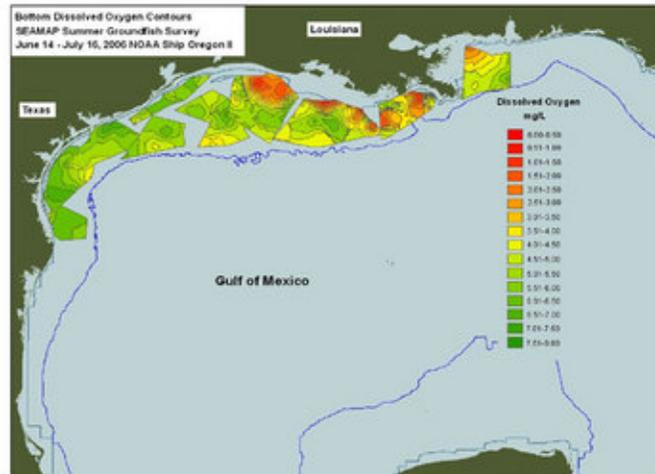


Figure 15. Gulf of Mexico Dead Zone, July 2006

© NOAA Satellite and Information Service, National Environment Satellite, Data, and Information Services.

## 9. Water-Related Diseases

More than 2 million people die each year from diseases such as cholera, typhoid, and dysentery that are spread by contaminated water or by a lack of water for hygiene. These illnesses have largely been eradicated in developed nations, although outbreaks can still occur. In 1993 an infestation of cryptosporidium, a protozoan that causes gastrointestinal illness, killed 110 people and sickened an estimated 400,000 in Milwaukee, Wisconsin. The city's water treatment system was in compliance with federal and state regulations at the time, but after the outbreak federal regulators increased testing requirements for turbidity (cloudiness) in drinking water, an indicator of possible contamination.

Water-related illnesses fall into four major categories:

- **Waterborne diseases**, including cholera, typhoid, and dysentery, are caused by drinking water containing infectious viruses or bacteria, which often come from human or animal waste.

- **Water-washed diseases**, such as skin and eye infections, are caused by lack of clean water for washing.
- **Water-based diseases**, such as schistosomiasis, are spread by organisms that develop in water and then become human parasites. They are spread by contaminated water and by eating insufficiently cooked fish.
- **Water-related insect vectors**, such as mosquitoes, breed in or near water and spread diseases, including dengue and malaria. This category is not directly related to water supply or quality.

As noted above, more than 1 billion people worldwide lack safe drinking water, mainly in developing countries. Conventional large-scale engineering projects that pipe water from central distribution systems can provide safe water at a cost of approximately \$500 per person. Small-scale approaches, such as drilling wells and chlorination, can reduce this cost to less than \$50 (Fig. 16).



Figure 16. Sodium hypochlorite solution for disinfecting water

© U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Infectious Diseases.

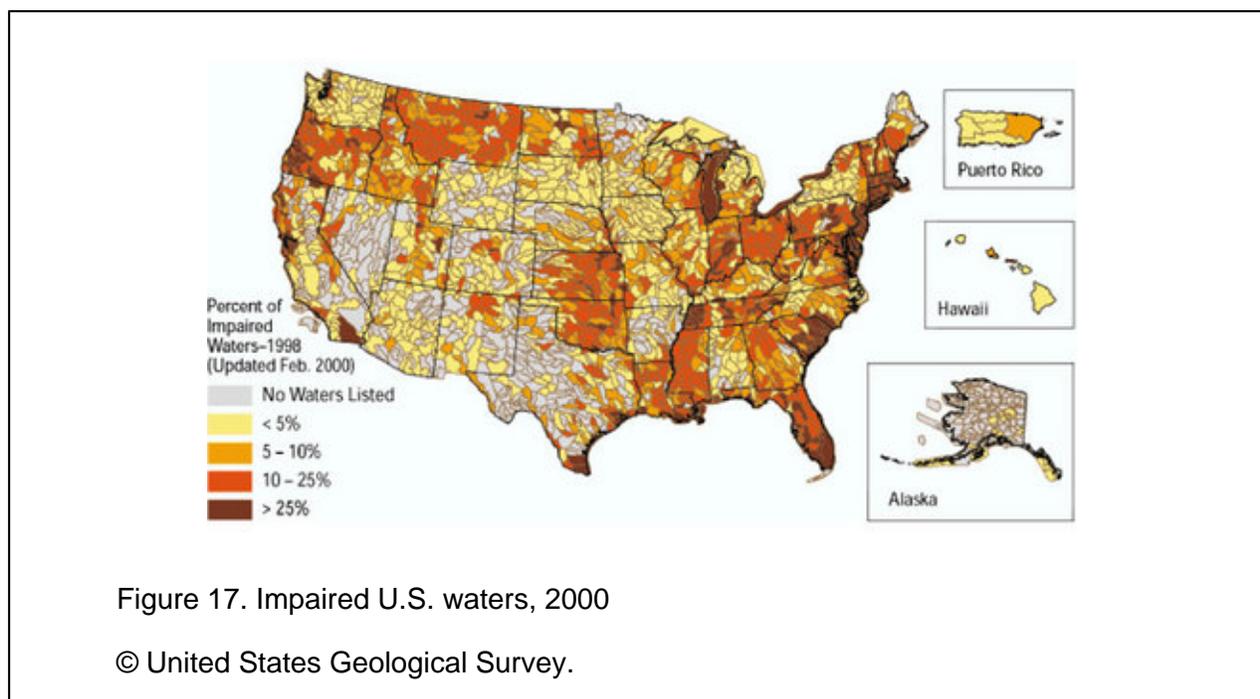
Scientists are still learning how many water-related diseases spread and how infectious agents behave. For example, until the 1970s cryptosporidium was not believed to infect humans, although it was recognized as a threat to animals. A 2003 World Health Organization report on water-related infectious diseases warned that "the spectrum of disease is altering and the incidence of many water-related microbial diseases is increasing." Processes such as urbanization and dam construction can

spread water-related diseases by creating new environments for infectious agents, and global climate change is expanding the range of mosquitoes and other disease vectors. However, advances in microbiology are enabling researchers to detect pathogens in water more quickly and to identify and characterize new infectious agents (footnote 15).

## 10. Major Laws and Treaties

The central U.S. law regulating water quality is the Clean Water Act (CWA), adopted in 1972. The Act initially focused on point sources, which it regulates through a national program that requires sources to obtain permits for any discharges of controlled pollutants into the nation's "navigable waters." It also increased federal aid to states and localities for sewage treatment facilities. One contentious issue in recent years has been CWA protection for wetlands. Developers have filed multiple lawsuits over issues such as whether isolated and seasonal wetlands fall under the definition of "navigable waters" and whether various types of dredging and filling constitute discharges into wetlands.

The CWA permitting system has substantially reduced water pollution from point sources in the United States, but nonpoint source pollution remains a serious problem. Since the mid-1990s the Environmental Protection Agency has increasingly focused on requirements in the CWA for states to identify "impaired" water bodies (those that remain polluted even after point sources install technical controls) and to develop **Total Maximum Daily Load (TMDL)** requirements for these systems. Some 20,000 water bodies across the nation fall under this heading (Fig. 17).





TMDLs represent the maximum levels of specific pollutants that can be discharged into impaired water bodies from all point and nonpoint sources, including a safety margin. Once states calculate TMDLs they must assign discharge limits to all sources and develop pollution reduction strategies (footnote 16). TMDLs are difficult and expensive to calculate because state regulators need extensive data on all polluters that are discharging into impaired water bodies and must quantify relative contributions from all sources to total pollution.

The Safe Drinking Water Act (SDWA), enacted in 1974, regulates contaminants in public water supplies, which serve about 90 percent of the U.S. population. The law sets mandatory limits on some 90 contaminants to protect public health and recommends voluntary standards for other substances that can affect water characteristics such as odor, taste, and color (footnote 17). The SDWA has significantly improved the quality of drinking-water supplies, but new issues are still emerging. For example, methyl tertiary butyl ether (MTBE), an additive widely used to improve combustion in gasoline, has contaminated public water supplies in many regions where gasoline has leaked from underground storage tanks. The EPA has issued a drinking water advisory for MTBE because small amounts can cause discoloration and odor that make water unpotable, but has not yet set a drinking water standard for MTBE even though the agency's Office of Research and Development calls MTBE "a possible human carcinogen." As of 2004, 19 states had acted independently to ban or limit use of MTBE (footnote 18).

At the international level, the United Nations Convention on the Law of the Sea (LOS Convention), finalized in 1982, creates a comprehensive framework for nations' use of the oceans. The convention outlines each country's rights and responsibilities within its territorial boundaries and in international waters for issues including pollution control, scientific research, resource management, and seabed mining. Coastal states have jurisdiction to protect the marine environment in their Exclusive Economic Zones (areas typically extending 200 miles outward from shore) from activities including coastal development, offshore drilling, and pollution from ships.

The United States is not among the 149 nations that have ratified the convention, which President Reagan refused to sign in 1982, citing restrictions on deep seabed mining that were later renegotiated to address U.S. concerns. Two expert commissions and many stakeholders have called for the United States to ratify the pact (footnote 19). The United States is a party to a number of other international treaties and agreements that regulate ocean activities, including agreements on dumping pollutants at sea, protecting the Arctic and Antarctic environments, regulating whaling, and protecting endangered species.

## 11. Further Reading

Peter H. Gleick, **The World's Water 2006–2007: The Biennial Report on Freshwater Resources** (Washington, DC: Island Press, 2006). Current information on water needs, trends, and policies worldwide.



John McPhee, "Atchafalaya," in **The Control of Nature** (New York: Farrar Strauss Giroux, 1989). A renowned journalist describes the technical challenges and environmental impacts of human efforts to manage the flow of the Mississippi River.

Sandra Postel, **Liquid Assets: The Critical Need to Safeguard Freshwater Ecosystems**, Worldwatch Paper 170 (Washington, DC: Worldwatch Institute, July 2005). An overview of the valuable functions performed by freshwater ecosystems and policy options for protecting them.

## Footnotes

1. U.S. Geological Survey, "The Water Cycle: Evapotranspiration," <http://ga.water.usgs.gov/edu/watercycleevapotranspiration.html>.
2. Bridget R. Scanlon et al., "Global Impacts of Conversions from Natural to Agricultural Ecosystems on Water Resources: Quantity versus Quality," **Water Resources Research**, vol. 43, W43407 (2007).
3. Jonathan Harr, **A Civil Action** (New York: Random House, 1995). For a retrospective on the case (including the hydrologic evidence) by a local journalist who covered it, see Dan Kennedy, "Take Two," *The Boston Phoenix*, January 1–8, 1998, [http://www.bostonphoenix.com/archive/features/98/01/01/DON\\_T\\_QUOTE\\_ME.html](http://www.bostonphoenix.com/archive/features/98/01/01/DON_T_QUOTE_ME.html).
4. Thomas J. Feeley and Massoud Ramezan, "Electric Utilities and Water: Emerging Issues and R&D Needs," [http://www.netl.doe.gov/technologies/coalpower/ewr/pubs/WEF%20Paper%20Final%20header\\_1.pdf](http://www.netl.doe.gov/technologies/coalpower/ewr/pubs/WEF%20Paper%20Final%20header_1.pdf), p. 2.
5. United Nations Environment Programme, *GEO: Global Environment Outlook 3* (London: Earthscan, 2002), pp. 150–177, <http://www.unep.org/geo/geo3/english/pdf.htm>.
6. For a map of the world's 114 largest watersheds, see World Conservation Union, "Watersheds of the World," <http://www.iucn.org/themes/wani/eatlas/html/gm1.html>.
7. World Water Council, "Water at a Glance," <http://www.worldwatercouncil.org/index.php?id=5>.
8. World Water Council, "Evolution of water withdrawals and consumption since 1900," <http://www.worldwatercouncil.org/index.php?id=5>.
9. World Health Organization, "Water, Sanitation, and Hygiene Links to Health: Facts and Figures," [http://www.who.int/water\\_sanitation\\_health/factsfigures2005.pdf](http://www.who.int/water_sanitation_health/factsfigures2005.pdf).
10. Leonard F. Konikov and Eloise Kendy, "Groundwater Depletion: A Global Problem," **Hydrogeology Journal**, vol. 13 (2005), p. 317.
11. Texas Water Development Board, 2007 Draft State Water Plan, [http://www.twdb.state.tx.us/publications/reports/State\\_Water\\_Plan/2007/Draft\\_2007SWP.htm](http://www.twdb.state.tx.us/publications/reports/State_Water_Plan/2007/Draft_2007SWP.htm), p. 3.



12. University of California Cooperative Extension, "Irrigation Management Water Quality FAQs," [http://ceimperial.ucdavis.edu/Custom\\_Program275/Water\\_Quality\\_FAQs.htm](http://ceimperial.ucdavis.edu/Custom_Program275/Water_Quality_FAQs.htm).
13. Sujay S. Kaushal et al., "Increased Salinization of Fresh Water in the Northeastern United States," **Proceedings of the National Academy of Sciences**, vol. 102, September 20, 2005, p. 13517.
14. National Research Council, **Clean Coastal Waters: Understanding and Reducing the Effects of Nutrient Pollution** (Washington, DC: National Academy Press, 2000).
15. World Health Organization, *Emerging Issues In Water and Infectious Disease* (Geneva, 2003), [http://www.who.int/water\\_sanitation\\_health/emerging/emerging.pdf](http://www.who.int/water_sanitation_health/emerging/emerging.pdf) (quote on page 7).
16. For an example, see New Jersey Department of Environmental Protection, "Seven Total Maximum Daily Loads for Total Coliform To Address Shellfish-Impaired Waters in Watershed Management Area 17, Lower Delaware Water Region," February 21, 2006, [http://www.state.nj.us/dep/watershedmgt/DOCS/TMDL/Coastal\\_Pathogen\\_TMDLs\\_WMA17.pdf](http://www.state.nj.us/dep/watershedmgt/DOCS/TMDL/Coastal_Pathogen_TMDLs_WMA17.pdf).
17. U.S. Environmental Protection Agency, "List of Drinking Water Contaminants & MCLs," <http://www.epa.gov/safewater/mcl.html>.
18. U.S. Environmental Protection Agency, "State Actions Banning MTBE (Statewide)," <http://www.epa.gov/mtbe/420b04009.pdf>.
19. U.S. Commission on Ocean Policy, **An Ocean Blueprint for the 21st Century** (2004), <http://www.oceancommission.gov/welcome.html>; Pew Commission on Ocean Policy, **America's Living Oceans: Charting a Course for Sea Change** (2003), [http://www.pewtrusts.org/pdf/env\\_pew\\_oceans\\_final\\_report.pdf](http://www.pewtrusts.org/pdf/env_pew_oceans_final_report.pdf).

## Glossary

- aquifers** : Underground formations, usually composed of sand, gravel, or permeable rock, capable of storing and yielding significant quantities of water.
- artesian** : Describes a confined aquifer containing groundwater that will flow upwards out of a well without the need for pumping.
- catchment area** : The area that draws surface runoff from precipitation into a stream or urban storm drain system.
- discharges** : Defined by the Clean Water Act as the addition of pollutants (including animal manure or contaminated waters) to navigable waters.
- estuaries** : Coastal waters where seawater is measurably diluted with freshwater; a marine ecosystem where freshwater enters the ocean.



**freshwater** : Water without significant amounts of dissolved sodium chloride (salt). Characteristic of rain, rivers, ponds, and most lakes.

**groundwater** : Water contained in porous strata below the surface of the Earth.

**hydraulic head** : The force per unit area exerted by a column of liquid at a height above a depth (and pressure) of interest. Fluids flow down a hydraulic gradient, from points of higher to lower hydraulic head.

**hypoxic** : Referring to a condition in which natural waters have a low concentration of dissolved oxygen (about 2 milligrams per liter, compared with a normal level of 5 to 10 milligrams per liter). Most game and commercial species of fish avoid waters that are hypoxic.

**non-aqueous phased liquids (NAPL)** : Organic liquids that are relatively insoluble in water and less dense than water. When mixed with water or when an aquifer is contaminated with this class of pollutant (frequently hydrocarbon in nature), these substances tend to float on the surface of the water.

**nonpoint source** : A diffuse, unconfined discharge of water from the land to a receiving body of water. When this water contains materials that can potentially damage the receiving stream, the runoff is considered to be a source of pollutants.

**permeability** : The ease with which water and other fluids migrate through geological strata or landfill liners.

**point source** : An identifiable and confined discharge point for one or more water pollutants, such as a pipe, channel, vessel, or ditch.

**porosity** : The total volume of soil, rock, or other material that is occupied by pore spaces. A high porosity does not equate to a high permeability because the pore spaces may be poorly interconnected.

**recharge** : A hydrologic process where water moves downward from surface water to groundwater. This process usually occurs in the vadose zone below plant roots, and is often expressed as a flux to the water table surface.

**sorption** : The physical or chemical linkage of substances, either by absorption or by adsorption.

**total maximum daily load** : The maximal quantity of a particular water pollutant that can be discharged into a water body without violating a water quality standard.

**vadose zone** : The area of the ground below the surface and above the region occupied by groundwater.

**watershed** : The area of land that drains into a lake or stream.

## Unit 9 : Biodiversity Decline



Buttress roots of a canopy tree in the rainforest of northern Queensland, Australia. Courtesy William Laurance

### Overview

Living species on Earth may number anywhere from 5 million to 50 million or more. Although we have yet to identify and describe most of these life forms, we know that many are endangered today by development, pollution, over-harvesting, and other threats. Earth has experienced mass extinctions in the past due to natural causes, but the factors reducing biodiversity today increasingly stem from human activities. In this unit we see how scientists measure biodiversity, how it benefits our species, and what trends might cause Earth's next mass extinction.

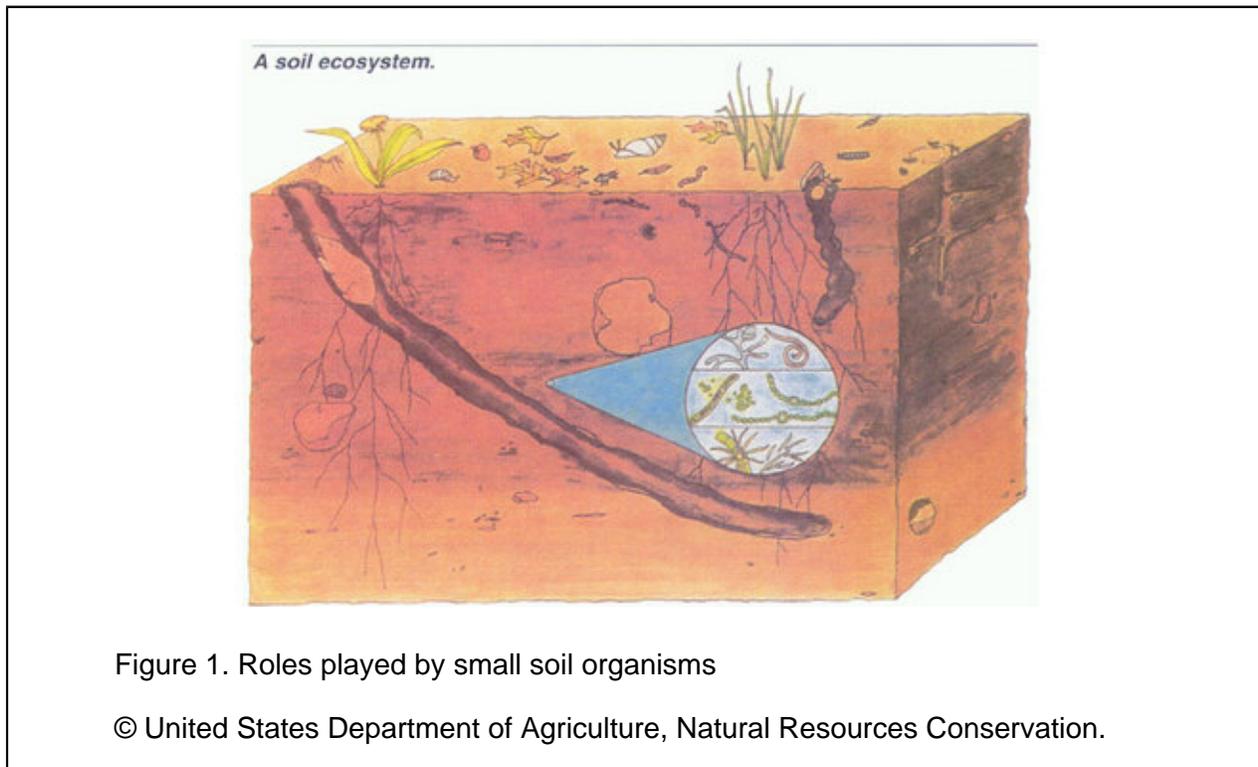
### Sections:

1. Introduction
2. Defining Biodiversity
3. Counting Species
4. Biodiversity Hotspots
5. Categories of Concern: Critically Endangered, Endangered, Vulnerable
6. A Sixth Mass Extinction?
7. Habitat Loss: Causes and Consequences
8. Invasion by Exotic Species
9. Other Drivers of Biodiversity Loss
10. Why Biodiversity Matters
11. Biodiversity in Your Back Yard
12. Major Laws and Treaties
13. Further Reading

## 1. Introduction

The term "biodiversity" was introduced in 1988 by evolutionary biologist E.O. Wilson, one of the leading experts in this field (footnote 1). Discussion of biodiversity has become commonplace in the past several decades. Although scientists are still trying to answer basic questions, including how many species there are on Earth, a broad trend is clear: extinctions are occurring today at an exceptionally high rate, and human activities are a major cause (footnote 2).

Is this a serious problem if millions of species remain? At times the question is posed this way—for example, when the fate of one seemingly obscure organism is at stake—but in fact there are important connections between biodiversity and the properties of ecosystems. For example, a tract of forest land can sustain more plants if it contains significant numbers of organisms that enhance soil quality, such as earthworms and microbes (Fig. 1). As we will see, a change in the status of one species can affect many others in ways that are not always predictable. Healthy ecosystems provide many important services to humans, although these functions are not always recognized or awarded economic value. If biodiversity erodes, we may lose some of these services permanently.



There also is an aesthetic case for maintaining biodiversity. We take it for granted that nature is attractive, but much of the world's appeal is rooted in the contrast between many types of species, whether the setting is a coral reef filled with tropical fish or a forest filled with autumn colors.

This unit explores how scientists define and measure biodiversity, and how biodiversity is distributed around the globe and divided among the various types of organisms. It then discusses factors that are impacting biodiversity, including habitat loss, invasion by alien species, and over-harvesting. Emerging threats to biodiversity from global climate change are addressed in Unit 12, "Earth's Changing Climate," and Unit 13, "Looking Forward: Our Global Experiment."

## 2. Defining Biodiversity

The basic currency of biodiversity is **species richness**—the number of species in a given habitat, or worldwide. By analyzing fossilized life forms, which date back as far as 3.5 billion years, scientists can estimate how many species were present during past eras and compare those numbers to the present range of life. As we will see in Section 3, "Counting Species," species biodiversity is at a peak today compared to past levels, but at the same time many scientists believe that the current rate of extinction is also alarmingly high.

Another important biodiversity indicator is the level of genetic diversity within a species, which may influence the species' future trajectory (Fig. 2).



Figure 2. Color variation in the Oldfield mouse (*Peromyscus polionotus*)

© Hopi Hoekstra, Harvard University.

Species with low genetic diversity may be less likely to survive environmental stresses because they have fewer genetic options when problems arise. Conversely, populations with high levels of genetic diversity may be more likely to survive environmental and other stresses.

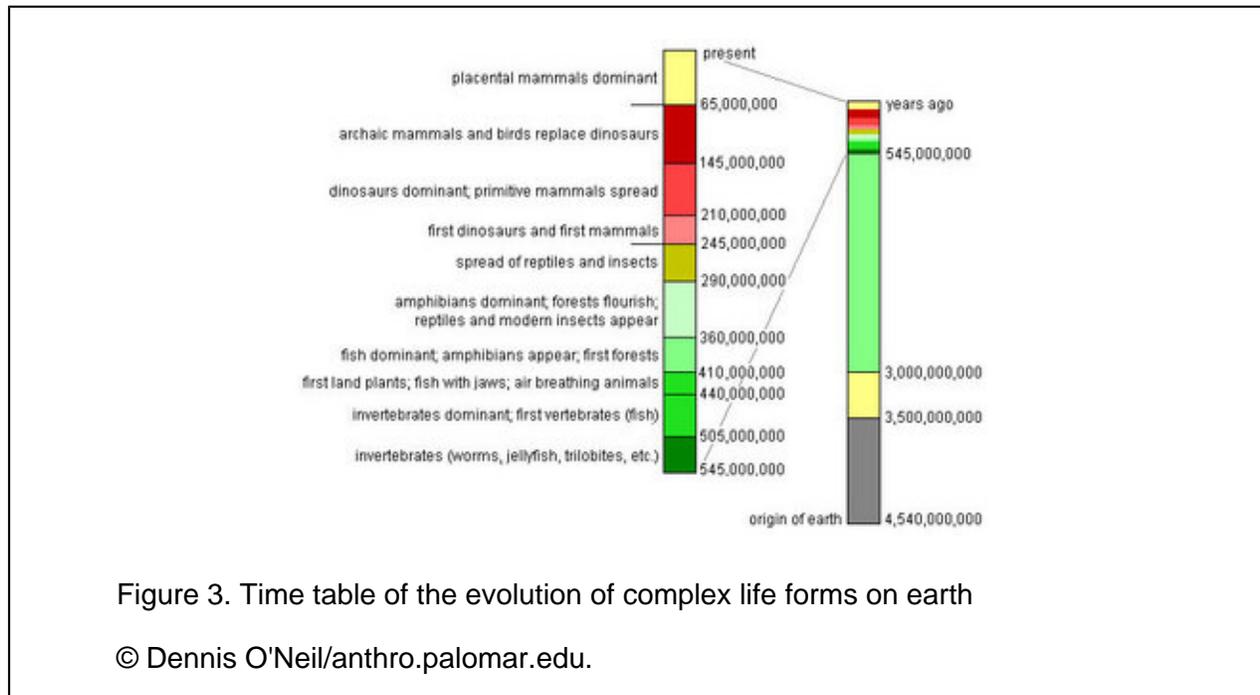
A historic example comes from Harvard Forest in Petersham, Massachusetts, where researchers have documented that an outbreak of an insect called the eastern hemlock looper caused an abrupt decline in the population of hemlock trees across the northeastern United States about 4,800 years ago. Hemlock numbers remained low for nearly 1,000 years after the blight struck but then recovered, possibly because the remaining individuals developed a resistance to the looper (footnote 3). If this theory is correct, it indicates that some hemlocks were genetically less susceptible to looper infestations than others and that natural selection slowly favored those trees in the years following the blight.

**Ecosystem** diversity is a third type of biodiversity. Life on Earth is distributed among many types of habitats, each of which provides a suitable living environment for specific kinds of organisms. These ecosystems range from tropical rainforests to hydrothermal vents on the ocean floor, where super-heated water bursts through cracks in the planet's crust (for more details, see Unit 4, "Ecosystems"). Many ecosystems are made up of species that have adapted to life under unusual conditions, such as Arctic sea ice communities (Box 1). The loss of these unique ecosystems can wipe out the many species that are highly specialized and unable to shift to other areas.

### 3. Counting Species

As discussed in Unit 1, "Many Planets, One Earth," life first appeared on Earth as early as 3.8 billion years ago. The earliest life forms were single-celled bacteria and archaea that harvested energy through chemical reactions before free oxygen began to accumulate in Earth's atmosphere. Early photosynthetic bacteria appeared about 3.5 billion years ago, but several billion years passed before multicellular organisms developed. This step took place around 600 million years ago, when Earth's atmosphere and oceans were accumulating increasing amounts of oxygen.

Early life forms were limited to the oceans until plants and animals evolved to live on land about 400 million years ago. Life on Earth became increasingly diverse as organisms adapted to environments on land, despite several waves of mass extinctions (Fig. 3).



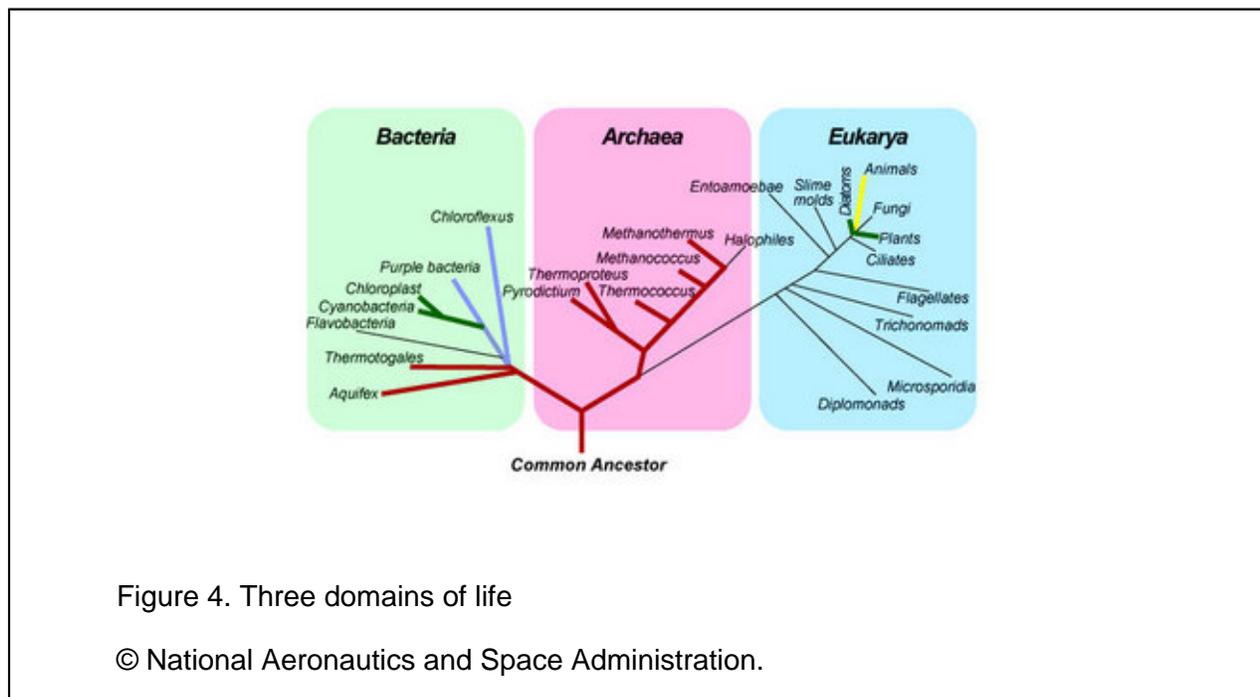
How do scientists estimate past and current numbers of species? Fossil records are key sources. By looking carefully at fossil records, scientists can use plant and animal fossils to trace species' evolution over time, estimate rates of **speciation** (the formation of new biological species), and assess how various organisms responded to known environmental changes in the past (footnote 4). It is important to note that fossil records are imperfect: not all organisms leave recognizable, well-preserved skeletons, so some species are easier to count than others. As a result, it is very difficult to make precise estimates of the number of species on earth at specific points in time, but the records do indicate trends in biodiversity levels.

Based on analyses of fossils, scientists estimate that marine biodiversity today is about twice the average level that existed over the past 600 million years, and that biodiversity among terrestrial organisms is about twice the average since life adapted to land about 440 million years ago. Fossil records also indicate that on average species exist for about 5 to 10 million years, which corresponds to an extinction rate of 0.1 to 1 species per million **species-years** (footnote 5).

Molecular phylogenetics is a newer tool for studying biodiversity. By measuring the degree of similarity between DNA, RNA, and proteins in the cells of closely related organisms, scientists can reconstruct these organisms' evolutionary histories and see how species are formed. For example, although it was long believed that fungi were closely related to plants, genetic analyses led by Mitchell Sogin of the Marine Biological Laboratory in Woods Hole, Massachusetts, have shown that fungi are more directly related to animals (footnote 6).

# The Habitable Planet

Our current understanding of biodiversity is uneven: for example, we know more about animals than we do about protists. As discussed in Unit 1, "Many Planets, One Earth," biologists classify life on Earth into three broad domains: Eukarya, Bacteria, and Archaea. The first group includes all animals and plants, as well as fungi and protists, while the latter two groups comprise different types of microbes (Fig. 4). None of these groups are descended from the others, and all have unique features.



The biodiversity of Bacteria and Archaea is poorly understood for several reasons. Scientists do not agree on how to define a species of bacterium and do not have accurate estimates of the total number of bacterial individuals. From a practical standpoint, counting species of bacteria and other microbes is harder than counting bird species because birds have more defining features, including color, songs, and shapes. Visual and behavioral differences are clearer to the human observer. And many microbes are found in extremely challenging habitats, such as vents in the ocean floor.

Finally, humans tend to see animal and plant species as more interesting and important than other, smaller life forms. A discovery of a new monkey species is apt to become major science news, but new species of fungi are routinely reported in specialized journals without attracting serious popular interest. And the scientific literature focuses heavily on mammals and birds at the expense of other groups such as invertebrates and insects (footnote 7).

Based on fossil records and the opinions of experts who study various groups of living organisms, biodiversity on Earth appears to be at a historical peak today. Some 1.5 million species have been identified and described, but estimates suggest that at least another 5 to 15 million species remain



to be catalogued (footnote 8). Insects and microorganisms are thought to account for large shares of these uncounted species. In July 2006, scientists involved in the Census of Marine Life (a ten-year project to measure ocean biodiversity) stated that there could be as many as five to ten million different types of bacteria in the oceans, some 10 to 100 times more than previously estimated. Many of these species may exist in relatively low numbers, but could have important ecological functions (footnote 9).

*"Just as scientists have discovered through ever more powerful telescopes that stars number in the billions, we are learning through DNA technologies that the number of marine organisms invisible to the eye exceeds all expectations and their diversity is much greater than we could have imagined."*

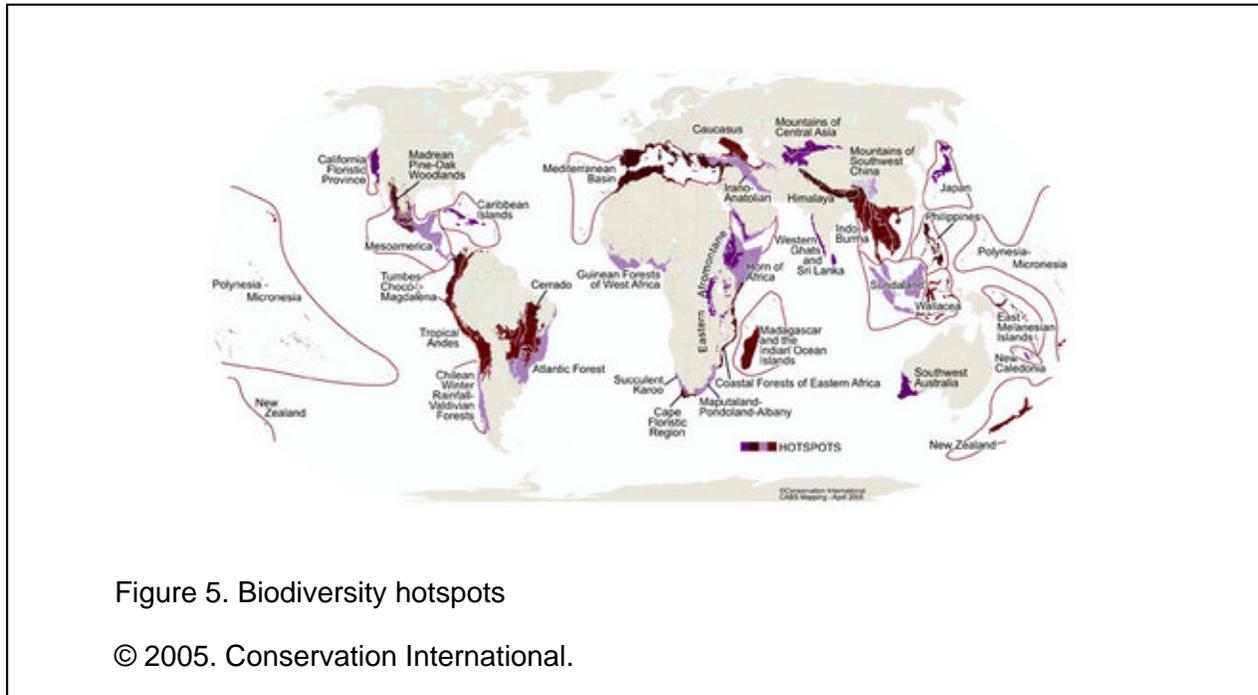
Mitchell L. Sogin, Marine Biological Laboratory (Woods Hole, MA)

#### 4. Biodiversity Hotspots

Some areas of the globe are richer in species than others. As discussed in Unit 4, "Ecosystems," a latitudinal biodiversity gradient exists for animals and plants, with more species found in tropical than in temperate or polar regions. Recent work suggests that microbial communities are more diverse in temperate zones (footnote 10).

Conservationists refer to areas especially rich in biodiversity as **hotspots**. Based on work by British ecologist Norman Myers, who first proposed the concept, the nonprofit group Conservation International defines hotspots as regions that have at least 1,500 species of vascular plants that are **endemic** (found only in that area) and that have lost at least 70 percent of their original habitat. Drawing the exact borders of hotspots can be difficult, but Myers and others define a hotspot as "a separate biota or community of species that fits together as a biogeographic unit"—in other words, a community of organisms that live in a geographically unified zone and interact with each other (footnote 11).

Conservation International identifies 34 such hotspots in tropical and temperate regions around the globe (Fig. 5). There are three hotspots on U.S. territory: the California Floristic Province, a Mediterranean climate zone that covers the state's west coast and much of its central region; the Caribbean islands, including the U.S. Virgin Islands and Puerto Rico; and small patches of the Madrean-Pine Oak Woodlands, mountainous forests extending from Mexico into southern New Mexico and Arizona (footnote 12).



Hotspots are a tool for setting conservation priorities. Because endemic species are found only in one place, protecting them requires preserving the areas in which they live. The 34 hotspots identified by Conservation International represent 2.3 percent of Earth's land surface but are home to at least 150,000 endemic plant species (50 percent of the world's total number of plant species) and nearly 12,000 terrestrial vertebrates (42 percent of the world's total number of terrestrial vertebrates).

What makes hotspots such rich biodiversity nodes? Many are located in moist tropical forests, the most diverse of Earth's major **biomes**. A number are islands or are physically bounded by deserts or mountain ranges, which has facilitated the evolution of endemic species by keeping populations in relative isolation and minimizing hybridization with other species. Most hotspots have widely varied topography, from lowlands to mountains, which produces a broad range of climatic conditions.

Advocates contend that hotspots should be protected because their loss could greatly accelerate what many scientists believe is an ongoing mass extinction (see Section 6, "A Sixth Mass Extinction?," below). Conversely, protecting them could save many of Earth's most threatened species. By definition, world hotspots have already lost at least 70 percent of their land area; many scientists warn that fragmenting them further will accelerate species loss because populations of endemic species will become smaller and more extinction-prone. This view is based on the theory of island biogeography, which is discussed further below in Section 7 ("Habitat Loss: Causes and Consequences").

About 10 percent of the original area of world biodiversity hotspots is currently protected as parks or reserves. Conservation International calls many of these areas "paper parks," where protection requirements are not enforced or where the covered zones have low biodiversity value (footnote 13). A 2004 study of how well the global network of protected areas preserved biodiversity found that many species were not covered by any habitat protections. "Global conservation strategies based on the recommendation that 10% (or other similar targets) of each country or biome be protected will not be effective because they are blind to the fact that biodiversity is not evenly distributed across the planet; by the same token, neither should protected areas be," wrote the authors (footnote 14).

## 5. Categories of Concern: Critically Endangered, Endangered, Vulnerable

What are the criteria for determining whether a species is in danger of extinction? To advise national governments, the multinational World Conservation Union (abbreviated IUCN for its formal title, the International Union for the Conservation of Nature) maintains a database of threatened species and subgroups and publishes the Red List, which catalogues species most at risk. For species in the highest risk groups—Critically Endangered, Endangered, and Vulnerable—IUCN weighs criteria including population size, geographic range, and how individuals are distributed, especially if the population is very small.

IUCN is working to improve its data on species, which currently is biased toward forests and other land ecosystems, emphasizes animals more strongly than plants, and does not cover microbes. Priority areas for the Union include better data on marine species and on arid and semi-arid ecosystems, which are expanding as a result of global climate change. Table 1 summarizes some estimates of threatened species (Critically Endangered, Endangered, and Vulnerable) from the 2006 Red List.

**Table 1.** Species classified as threatened by IUCN, 2006.

	Number of described species in IUCN database	Number of threatened species in 2006	Number threatened as % of described species
<b>Vertebrates</b>			
Mammals	5,416	1,093	20%
Birds	9,934	1,206	12%
Reptiles	8,240	341	4%
Amphibians	5,918	1,811	31%
Fishes	29,300	1,173	4%
<b>Subtotal</b>	<b>58,808</b>	<b>5,624</b>	<b>10%</b>

	Number of described species in IUCN database	Number of threatened species in 2006	Number threatened as % of described species
<b>Invertebrates</b>			
Insects	950,000	623	0.07%
Mollusks	70,000	975	1.39%
Crustaceans	40,000	459	1.15%
Others	130,200	44	0.03%
<b>Subtotal</b>	<b>1,190,200</b>	<b>2,101</b>	<b>0.18%</b>
<b>Plants</b>			
Mosses	15,000	80	0.53%
Ferns and allies	13,025	139	1.0%
Gymnosperms	980	306	31%
Dicotyledons	199,350	7,086	4%
Monocotyledons	59,300	779	1%
<b>Subtotal</b>	<b>287,655</b>	<b>8,390</b>	<b>3%</b>
<b>Others</b>			
Lichens	10,000	2	0.02%
Mushrooms	16,000	1	0.01%
<b>Subtotal</b>	<b>26,000</b>	<b>3</b>	<b>0.01%</b>

Other biological inventories cover different sets of organisms and offer different perspectives on which species are most highly threatened. For example, NatureServe ([www.natureserve.org](http://www.natureserve.org)) pools data from a network of natural heritage programs and estimates the number of threatened species in the United States to be far greater than the IUCN's estimates. "There is no single authoritative list of the world's endangered species, because we have yet to count and describe many living species," says Harvard University biologist Anne Pringle.

Scientific evidence is central to identifying endangered species. To determine whether a species is endangered or might become so, scientists collect data to answer questions including:

- Is the population growing, shrinking, or at a steady state, and why? How completely does it occupy its habitat? How is it being affected by competitors, parasites, harvesting, and hybridization with other species?

- Is the species' geographic range expanding or contracting? Is it fragmented into small areas? How many mature (breeding) individuals exist, and where are they located? How are external impacts on its habitat, such as pollution and development, expected to affect the species' range? How much habitat is needed to support a target population level?
- If the population is very small, is it expected to grow or contract? Will the number of mature individuals remain steady or fluctuate? Can they reach each other to breed?
- How biologically distinct is the species from other closely related organisms? Does the target group consist of one single species, or should it be reclassified as several distinct species?
- If a species is recovering from endangered status, what population size and distribution indicate that it no longer needs special protection?

These assessments draw on scientific fields including conservation biology, population ecology, biogeography, and genetics. Captive breeding programs have helped to preserve and reintroduce some species that were extinct in the wild, such as California condors. Recently scientists have successfully cloned several endangered varieties of cows and sheep, and some biologists advocate creating DNA libraries of genetic material from other endangered species. Others counter that cloning fails to address the root causes of the problem, including habitat loss and over-harvesting.

## 6. A Sixth Mass Extinction?

Species have appeared and disappeared continually throughout Earth's history, with extinctions occurring on average at a rate of 0.1 to 1 species per million species-years. At several points, however, this rate has risen sharply, producing five "mass extinction" events (Fig. 6). The most famous of these events, the Cretaceous-Tertiary (K-T) extinction 65 million years ago, is thought to have been caused at least in part by a giant asteroid that struck Earth off the coast of Mexico, causing tidal waves and climate-altering dust clouds. Dinosaurs, along with two-thirds of the other species on Earth, were killed off by the K-T extinction.

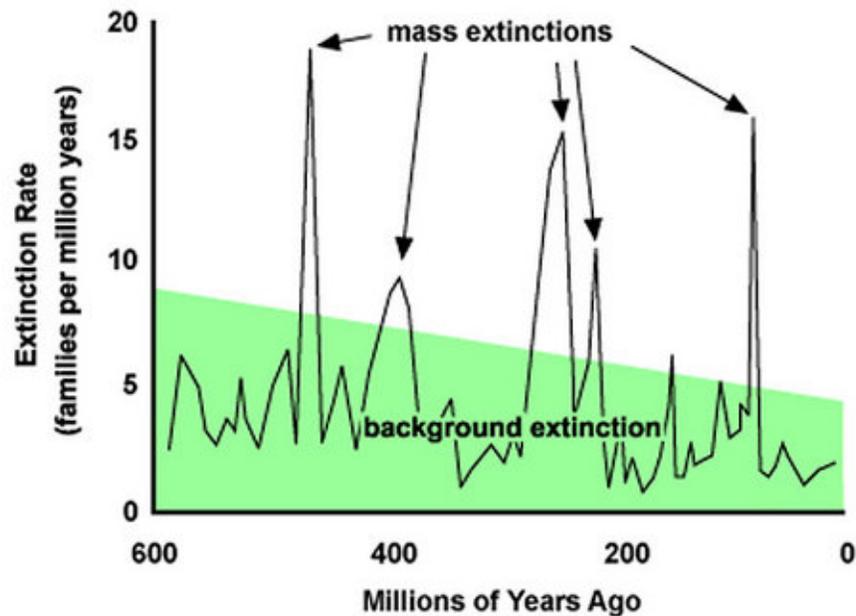


Figure 6. Earth's five mass extinctions

© University of California Museum of Paleontology's Understanding Evolution/  
[evolution.berkeley.edu](http://evolution.berkeley.edu).

Several lines of evidence suggest that Earth is experiencing a sixth mass extinction today. Estimates of extinction rates are imprecise for many reasons. In particular, they are extrapolated from a few well-known groups with relatively few species, such as birds and large mammals, to other groups for which there is little data—for example, fungi. However, there is wide agreement that current rates of extinction are at least several hundred times greater than historical background levels. Most scientific studies estimate that in the near term, extinction rates could rise by three to four orders of magnitude above past averages (footnote 15).

In addition to tracking extinction rates, scientists can look at population decline and habitat loss trends to estimate how quickly Earth's biodiversity levels are changing. To date, about 50 percent of the planet's natural habitats have been cleared for human use, and another 0.5 to 1.5 percent of nature is lost each year. Ongoing and current mass extinctions have been documented for many groups of organisms, including marine and freshwater fish, amphibians, and European farmland birds and macrofungi.



The current mass extinction is different from past events in several ways. First, it is happening much more quickly: each of the "Big Five" played out over thousands of years, but the current mass extinction is likely to be concentrated within 200 years. By the end of the 21st century, we may have lost two-thirds of the species on Earth (footnote 16). Second, past mass extinctions are thought to have been caused by natural phenomena such as the shifting of continents, comet or meteoroid impacts, or climate change independent of human influence, or some combination of these factors. In contrast, as we will see below, humans are causing the current mass extinction.

Despite concerns about a sixth mass extinction, new species are identified each year. For example, on a joint expedition to China and Nepal in early 2006, scientists from Conservation International and Disney's Animal Kingdom found new species that included a wingless grasshopper, a subspecies of vole, up to three new species of frogs, eight new species of insects, and ten new species of ants. These discoveries are evidence that we still know very little about Earth's biodiversity. In addition, new techniques for analyzing organisms' molecular structures have led scientists to reclassify some groups once viewed as single species into multiple species.

Sometimes it can be hard to determine the exact status of a rare species. In 2004, scientists from Cornell University and other institutions reported that they had seen and videotaped an ivory-billed woodpecker in Arkansas (Fig. 7). Ivory-bills had been presumed extinct since the 1930s, so this sighting caused great excitement but also spurred debate over whether the bird that was caught for a few seconds on film was in fact a more common type of woodpecker. Other researchers subsequently reported more than a dozen sightings and sound recordings of ivory-billed woodpeckers in Florida, but debate about whether ivory-bills still exist was ongoing as of late 2006. As the ivory-bill controversy shows, there is no definitive standard of proof for existence of a rare species, save perhaps a conclusive DNA sample—which may be impossible to get.



Figure 7. Watercolor painting of Ivory-billed Woodpeckers by John James Audubon  
Courtesy National Audubon Society, Inc., 700 Broadway, New York, NY 10003, USA.

## 7. Habitat Loss: Causes and Consequences

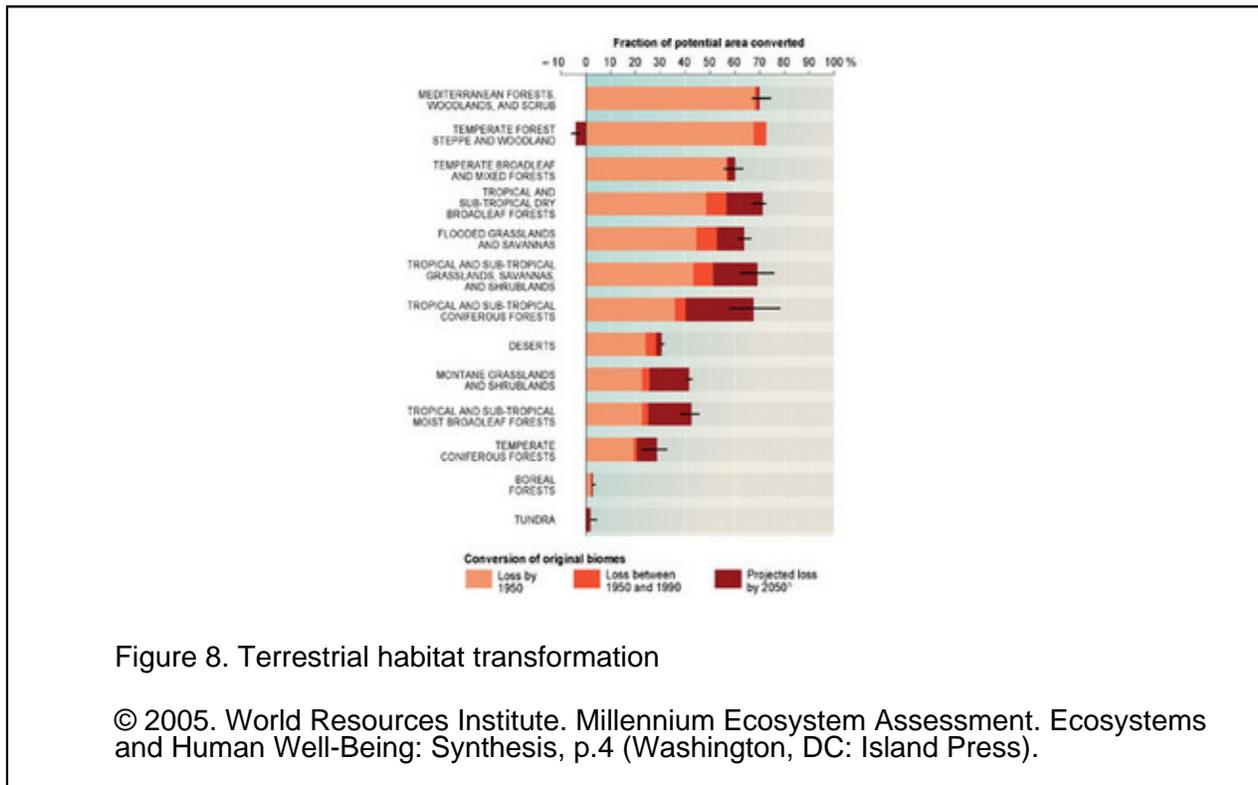
Most endangered species are threatened by multiple factors, but habitat loss is generally viewed as the largest single cause of biodiversity loss worldwide. When humans convert wild areas for agriculture, forestry, urban development, or water projects (including dams, hydropower, and irrigation), they reduce or eliminate its usefulness as a habitat for the other species that live there.

As discussed in Unit 4, "Ecosystems," a species' ecological niche is the sum of all of the ranges of tolerance under which it can survive in a specific ecosystem, including temperatures, climate, type of shelter, food sources it needs, and many other factors. Generalist species can adapt to many different types of living conditions, but more specialized organisms may not be able to adjust when their habitat is changed or disappears as a result of land development.

All forms of development alter natural ecosystems. Commercial forestry involves road-cutting through forests and the harvesting of trees that are important as shelter or food for some species. Dams change river flow patterns, dissolved oxygen levels, and water temperatures and may prevent fish from swimming upstream to spawn. Farmers clear land, withdraw large quantities of water from local sources, and introduce pesticides and chemical fertilizers to the environment. Ranching impacts land physically through grazing and generates air and water emissions from animal wastes. Urban

development clears land and paves it, which changes local water cycles by increasing surface runoff and reducing groundwater supplies. It also generates air and water pollution from industrial activities and transportation. (For more on these impacts, see Unit 7, "Agriculture," Unit 8, "Water Resources," and Unit 11, "Atmospheric Pollution.")

According to the Millennium Ecosystem Assessment (MA), a four-year, multinational analysis of the health of global ecosystems, cultivated land (including land used for livestock production and aquaculture) now covers one-quarter of Earth's terrestrial area. Mediterranean and temperate forests have been most heavily impacted by land conversion, but substantial conversion of tropical forests is also projected to occur by 2050 (Fig. 8). In contrast, boreal forests and tundra have experienced almost no conversion, although they are threatened by other forces, such as global climate change.



Land can become less suitable as habitat even if it is not directly converted to other uses. When actions such as suburban development and road-building carve large sectors of land into fragments, the undeveloped parcels may be too small or isolated to support viable populations of species that thrived in the larger ecosystems. This process, which is called [habitat fragmentation](#), reduces biodiversity by:

- Splitting populations into smaller groups, which may be less viable because it is harder for the isolated individuals within the groups to defend themselves or find mates
- Increasing crowding and competition within the fragments
- Reducing species' foraging ranges and access to prey and water sources
- Increasing friction between animals and humans as animals range into developed areas

Fragmentation of natural ecosystems intensifies **edge effects**, impacts that stem from the juxtaposition of two different ecosystems—for example, a meadow and a paved street. The edges of natural ecosystems are more susceptible to light, wind, and weather than interior areas, so they are less suitable habitat for species that live in sheltered areas. Edges also are vulnerable to invasive species.

"We're seeing mortality rates [of old-growth rainforest tree species] go through the ceiling as a consequence of edge effects. It's obvious that the ecology of the rain forest is being altered in a profound way by fragmentation," says Bill Laurence of the Smithsonian Tropical Research Institute, who studies edge effects in Panamanian rain forests.

The theory of island biogeography, developed by ecologists Robert MacArthur and E.O. Wilson to explain the uneven distribution of species among various islands, offers some insights into how habitat fragmentation affects local species. According to the theory, the number of species on an island is a balance between the rate of colonization by new species and the rate of extinction of existing species. An island's population will approach an equilibrium level where the two trends are balanced and the number of species remains stable. Large islands typically have more resources, so they can be expected to support larger equilibrium numbers of species. Accordingly, extinction rates should increase with habitat fragmentation because smaller habitat fragments support fewer species.

Island biogeography also suggests that habitat fragmentation will reduce the rate at which species colonize new areas because they have trouble crossing gaps in between the smaller sections of remaining habitat. For example, as illustrated in Figure 9, many animals are killed crossing highways that divide their ranges (and are at risk of losing genetic diversity if they cannot reach other local populations to breed).



Figure 9. Habitat fragmentation and species mobility

© United States Department of Transportation, Federal Highway Administration.

One solution that is attracting increasing interest is to create corridors of land linking separate habitat zones, making it easier for wildlife to move from one sector to another without injury or interference. Research has shown that these corridors increase wildlife movement between habitat zones, although it can be difficult to maintain access for animals when the corridors cross human structure such as highways and railroad tracks (footnote 17).

## 8. Invasion by Exotic Species

Throughout the history of commerce, humans have transported animals, plants, and other live organisms around the globe. In many cases these species were deliberately brought, either as food supplies for the journey or to create new crops and animal colonies on arrival. Some live cargo has come along uninvited—for example, termites embedded in the wood slats of shipping crates, or fingernail-sized zebra mussels that ships suck up in ballast water. Thousands of exotic species have been introduced to new habitats worldwide as a result of human trade and travel, but they all have something in common: the potential to thrive and multiply beyond all expectations in their new environments.

When species are moved to new locations that offers conditions for life similar to their native habitats, they may exploit vacant ecological niches and grow quickly, especially if they have no natural predators in their new settings. Figure 10 shows a shopping cart that was pulled from Great Lakes waters infested with exotic zebra mussels, which have covered nearly every inch of the cart's surface.



Figure 10. Colonization by zebra mussels, Great Lakes

© United States Environmental Protection Agency.

These **invasive species** are major threats to biodiversity because local species are not adapted to compete with them. One extreme case, the brown tree snake, was introduced to Guam after World War II (probably as a stowaway on military cargo planes) from its native range in the eastern Pacific. The snake has killed off nine of Guam's twelve forest bird species, half of its lizards, and possibly some of its bat species, and caused major damage to the island's poultry industry. In 2004, the U.S. Congress authorized up to \$75 million over five years to prevent brown tree snakes, which have traveled as far as Texas in cargo shipments, from becoming established in Hawaii and the U.S. mainland and to control their presence in Guam.

Plants can also become invasive. Spotted knapweed, a perennial that probably came to the United States from Eastern Europe or Asia a century ago in imported hay or alfalfa seed, has become established across Montana. The plants, each of which can produce up to 18,000 seeds annually, compete for water and nutrients with native bunch grasses and produce a toxin that damages other plants. This technique, in which plants compete by poisoning other species, is called allelopathy. For example, garlic mustard, a weed found across 30 states and Canada, suppresses the growth of native trees by killing the fungi that help the trees take up nutrients from soil.

Invasive species flourish because they have left their normal predators behind, so one way to control them is to import those enemies. For example, biologists have introduced eight of spotted knapweed's natural insect predators to Montana with mixed success. This strategy assumes that the imported predator can survive in the new environment, and that it will not become invasive itself. Cane toads were imported to Australia in the 1930s to control beetles that fed on sugar cane, but although they had been used successfully for this purpose in Hawaii and the Caribbean, in Australia the toads did not breed at the right time of year to eat cane beetle larvae. However, they did spread across most of Australia, and have outcompeted many other local frog species. Because they produce toxins in their bodies, the toads are also poisonous to predators such as fish and snakes, although some Australian birds and rodents are learning to eat only the non-toxic parts of the toads (footnote 18).

Invasion by exotic species threatens nearly 50 percent of the endangered species in the United States (footnote 19). Scientists are using remote sensing and geographic information systems to detect and map land cover changes and the spread of exotic plants, and also use high-speed computation and modeling to project how populations will grow. The yellow areas in Figure 11 show invasive salt cedar along the Rio Grande River.

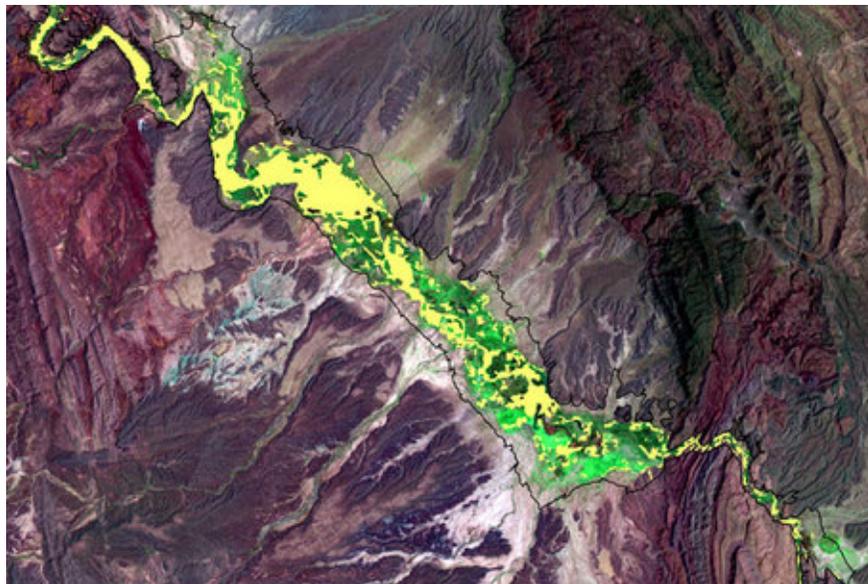


Figure 11. Landsat image of invasive salt cedar, 2002

© Center for Space Research, University of Texas.

It is important to note that many introduced species do not become invasive or have harmful impacts in their new settings. For example, major food crops and domestic animals have been traded worldwide and are rarely invasive. What makes a species likely to become invasive? The examples cited here point to several characteristics. Exotic species that reproduce quickly, can poison predators or competitors, and do not face natural predators in their new locales are well-positioned to spread and outcompete local species. As discussed in Unit 4, "Ecosystems," many r-adapted species have the capacity to become invasive pests because they flourish and reproduce quickly in unstable environments.

## 9. Other Drivers of Biodiversity Loss

Humans are harvesting many plants and animals at faster rates than the populations can maintain themselves (harvesting includes the use of animals and plants for food as well as other applications like medicines, trophies, and clothing). The declining yield from world fisheries shows how over-harvesting threatens biodiversity on a large scale.

According to the Food and Agriculture Organization (FAO), about half of ocean fish stocks are being exploited close to their maximum sustainable limits (i.e., close to levels of harvesting that will not reduce future yields). About one-fourth of global stocks are either overexploited, depleted, or recovering from depletion, and the remaining fourth are underexploited or moderately exploited. In several Atlantic and Pacific fisheries, traditional stocks have been depleted and fishermen are targeting other, less-valuable species (footnote 20).

In addition to depleting human food supplies, overfishing disrupts marine ecosystems and affects other species in the food chain. "Overfishing removes the most important and abundant consumers in a natural ecosystem," says Jeremy Jackson of the Smithsonian Tropical Research Institute. "Fish eat fish, but fish also eat seaweed. If they're not there, the seaweed grows 10 times or 100 times faster than corals, it grows over corals, smothers them, and kills them."

Hunting may also threaten animal and bird species. When North America was settled by European explorers, millions of passenger pigeons wintered in southern forests and migrated north to summer nesting grounds around the Great Lakes, darkening the skies as they flew. But large-scale hunting and forest clearing in the 19th century decimated pigeon numbers. The last known individual died in the Cincinnati Zoo in 1914 (Fig. 12).



Figure 12. Passenger pigeons, from John James Audubon's Birds of America

Courtesy National Audubon Society, Inc., 700 Broadway, New York, NY 10003, USA.

Today, major hunting and fishing organizations such as Ducks Unlimited and the Izaak Walton League take a more balanced approach and devote significant resources to protecting and conserving wildlife habitat. However, in developing countries where hunting is less strictly regulated, poaching and illegal trafficking threaten many wild species. International trade in wildlife generates billions of dollars annually and has depleted many species that are prized for uses including trophies, jewelry, food, exotic clothing, ingredients for medicine, and other uses. Many of these applications have encouraged wasteful and inhumane harvesting practices, such as cutting fins off of live sharks for use in shark-fin soup and throwing the maimed sharks back into the ocean to drown.

Pollution reduces biodiversity by either changing organisms' biological functions or altering the environmental conditions that they need to survive. For example, pesticides can affect the health and reproductive patterns of many species. Bald eagles, which declined to near-extinction in the early 1960s, are a well-known example. The chlorinated hydrocarbon pesticide DDT, which was used heavily in the 1940s and 1950s, bioaccumulated in eagles' fatty tissue and caused them to lay eggs with thin shells that broke before hatching. In 1963 there were 417 breeding pairs of bald eagles in the Lower 48 states; a federal ban on DDT and other protective measures helped increase this to more than 6,400 pairs by 2000. However, other chemicals currently in use may also act as endocrine disruptors in species including turtles, amphibians, and some fish.

Pollution is an important threat to aquatic ecosystems, where it can affect many environmental parameters. Agricultural runoff carries excess nutrients that cause algal blooms and deplete dissolved

oxygen levels, while siltation from logging and construction reduces available light. Mining generates toxic chemical wastes that can poison local water supplies.

Global climate change threatens biodiversity worldwide because it is modifying average temperatures and rainfall patterns, and thereby shifting climate zones. Ecologists have documented changes in the geographic ranges and breeding cycles of many species. However, some organisms that are highly adapted to specific conditions—such as the Arctic sea ice communities described above in Box 1—may not be mobile enough to find new habitats as their local climate conditions change. As a result, many scientists believe that climate change could increase current extinction rates. (For more details, see Unit 12, "Earth's Changing Climate.")

## 10. Why Biodiversity Matters

There are both practical and moral arguments for preserving biodiversity. From a practical point of view, humans derive many kinds of food, medicine, and fuel from plant and animal sources. All food crops were originally domesticated from wild species. More recently, researchers have derived anticancer drugs from the bark of the Pacific yew tree and from the Madagascar periwinkle flower. Scientists are currently working to isolate enzymes from a diversity of microbes that can digest plants for human use as transportation fuels. Many existing species that we know little about—especially insects, fungi, and microbes—may yield similarly important products.

Natural resources also provide jobs and income to many communities. Roughly 40,000 jobs were lost in Newfoundland in the 1990s when North Atlantic cod stocks collapsed due to over-fishing. Protecting biodiversity helps to ensure that crops, fisheries, and animal stocks will have enough genetic variety to survive natural disasters and diseases. When new diseases strike common crops, genetic stocks derived from wild relatives can be used to identify traits for resistance to the new infestations. This is the rationale for protecting areas that harbor wild relatives of domesticated plants and for creating stock centers that maintain seeds from wild relatives. For example, the C.M. Rick Tomato Genetics Resource Center at the University of California at Davis is a repository for wild and mutant strains of tomato and provides seed samples for research worldwide (footnote 21).

Healthy ecosystems also provide many other important services to human communities. These functions are so basic that they can go unnoticed, until they disappear. The American Dust Bowl occurred in the 1930s because repeated plowing and cultivation across the Great Plains broke up the cover of prairie grasses (viewed as weeds by farmers) that held the soil in place. Several years of extreme drought left the plains vulnerable to wind erosion, which blew topsoil off millions of acres and displaced thousands of farming families (Fig. 13).



Figure 13. Buried machinery in barn lot, Dallas, South Dakota, 1936

© United States Department of Agriculture.

In the past several decades, societies have begun to recognize the economic value of ecosystem services. For example, New York City signed an agreement in 1997 with state and federal agencies and 80 upstate communities to buy and protect lands in the Catskill and Delaware watersheds, which supply about 90 percent of the city's drinking water. By spending \$1.4 billion on land acquisition and related measures to reduce pollution in the target areas, New York avoided building a \$6–8 billion filtration plant to purify water from these sources (footnote 22).

Concerns about global climate change have increased awareness of the role that ecosystems play in sequestering carbon from the atmosphere and are spurring investment in programs to preserve this service. The World Bank's Prototype Carbon Fund (PCF), which invests in projects that reduce greenhouse gas emissions and promote sustainable development, is supporting ecosystem protection initiatives including native forest restoration in Brazil, soil conservation in Moldova, and

afforestation (planting new forests) on degraded agricultural land in Romania. Under procedures outlined in international climate change agreements, each of these projects will generate economic credits for reducing carbon dioxide emissions—a mechanism that effectively monetizes the benefit that the ecosystems provide by taking up atmospheric carbon dioxide in plants. By purchasing these credits, PCF will give local agencies a financial incentive to carry out the projects. The bank hopes to spur similar commitments from private investors that will help to create a market for carbon credits.

If ecosystems provide such valuable services, why do many communities exploit and damage them? First, valuing ecosystem services is a relatively new concept, and there are many different approaches to estimating those values. Second, local communities often have more to gain from quick exploitation than from conservation unless they receive special incentives, such as premium prices for sustainably-produced products. Third, existing incentives may reward communities that develop ecosystems instead of conserving them. For example, many U.S. communities encourage commercial development because it generates property taxes, even though this development reduces open space and increases traffic and pollution (footnote 23). The development pictured in Fig. 14 threatens the habitat of the Douglas County pocket gopher, which is endemic to the area.



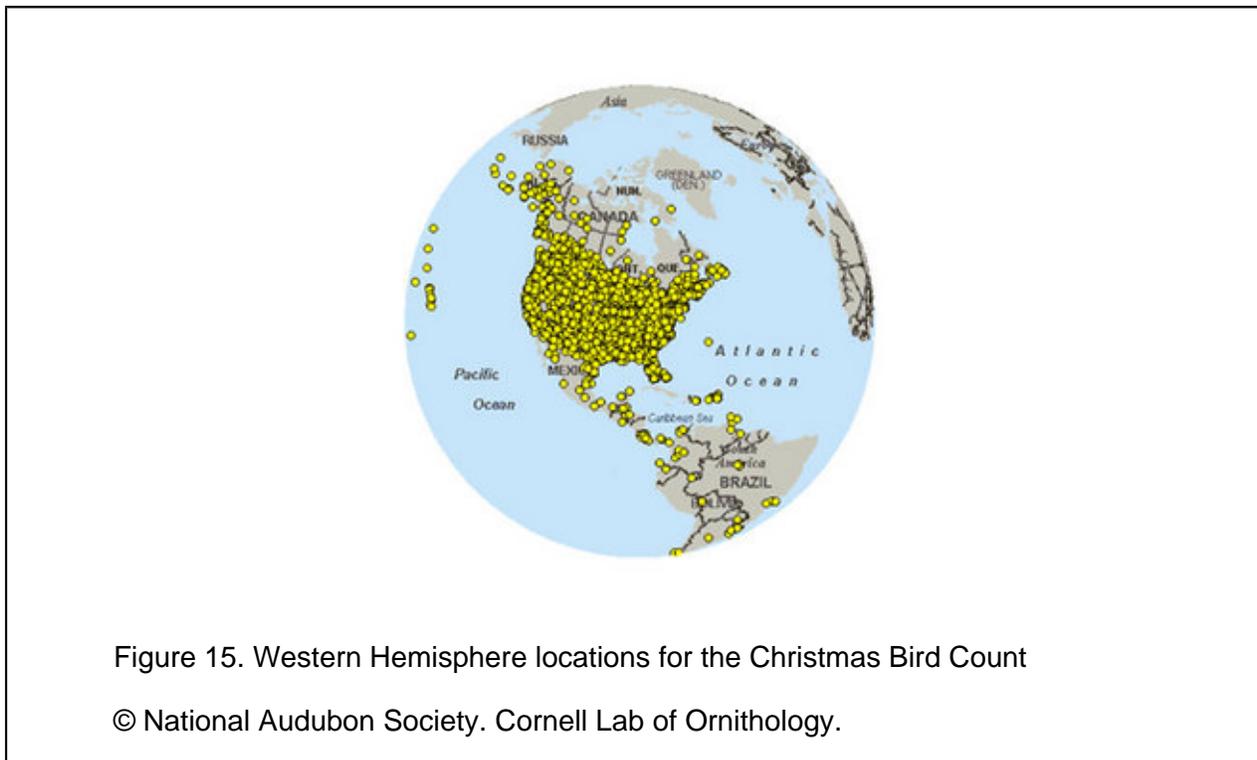
Figure 14. Suburban development in Douglas County, Colorado

© Center for Native Ecosystems.

Many advocates also make aesthetic and moral arguments for conserving biodiversity. Species richness adds to our enjoyment of nature, even at a simple level: most hikers would probably agree that a wild meadow, with its variety of plants, animals, and birds, is more interesting to visit than a cultivated field. Morally, the fact that speciation rates for many types of organisms are less than one per million years means that extinction is permanent, at least on human time scales: once a species is extinct, it will not be replaced for thousands or millions of years. In the words of biologists Rodolfo Dirzo and Peter Raven, "The loss of biodiversity is the only truly irreversible global environmental change the Earth faces today" (footnote 24).

## 11. Biodiversity in Your Back Yard

Biodiversity is not just an issue for scientists. Anyone who interacts with nature has a role to play in understanding and protecting biodiversity in her or his everyday environment. Observers contribute to many species monitoring projects—for example, the National Audubon Society's Christmas Bird Count (Fig. 15) and the annual Great Backyard Bird Count (GBBC), sponsored by Audubon and Cornell University on President's Day Weekend; Frogwatch USA, a frog and toad monitoring program sponsored by the National Wildlife Federation and the U.S. Geological Survey; and the Reef Environmental Education Foundation's fish and invertebrate monitoring programs (footnote 25).





Monitoring programs can help to protect biodiversity by increasing the amount of information that is available to scientists and policy makers. There are not enough trained scientists to monitor all species that are endangered or otherwise of interest, or the spread of invasive species, or the impacts of trends such as habitat fragmentation. In the United States, many agencies and organizations collect data on ecosystems, but often their data is not coordinated or integrated. In a 2006 report, the H. John Heinz Center identified ten key data gaps that impede effective reporting on the state of the nation's ecosystems. These gaps include:

- Reporting on species and communities at risk of extinction or loss
- Measuring the extent and impacts of non-native species
- Assessing the condition of plant and animal communities
- Assessing the condition of riparian areas and stream habitat

Monitoring programs that involve the public are not always subject to the same design criteria and quality controls as scientific field studies, but they can generate large data sets over broad geographic areas at a low cost. During the 2006 Great Backyard Bird Count, volunteers tallied 623 species and more than 7.5 million individual birds, both records for the decade-old event. And new species may be found anywhere, especially microbial species. For example, in 2003 researchers from the American Museum of Natural History found a new species of centipede in leaf litter in New York's Central Park, and in 2006 a graduate student discovered a new bacterium in a salt pond on Cape Cod (footnote 27).

## 12. Major Laws and Treaties

The main international agreement to protect biodiversity is the Convention on Biological Diversity, which was opened for signature at the United Nations Conference on Environment and Development in 1992 and currently has 188 members (the United States signed the convention in 1992 but has not ratified it). The Convention sets out three goals: conserving biodiversity, using its components sustainably, and sharing the benefits of genetic resources fairly. Because national governments set the rules that govern most uses of biodiversity, such as farming, forestry, and economic development, the Convention requires members to develop national strategies for measuring and conserving their biodiversity resources. Policies for protecting threatened species and restoring damaged ecosystems are also left to individual governments.

Threatened species are also protected by the Convention on International Trade in Endangered Species (CITES), which has 169 adherents. Under CITES, nations agree to set up licensing systems for all trade in species protected under the treaty and to apply specific restrictions outlined in the

convention. CITES protects some 5,000 species of animals (mammals, birds, reptiles, amphibians, fish, and invertebrates) and 28,000 species of plants (Fig. 16).



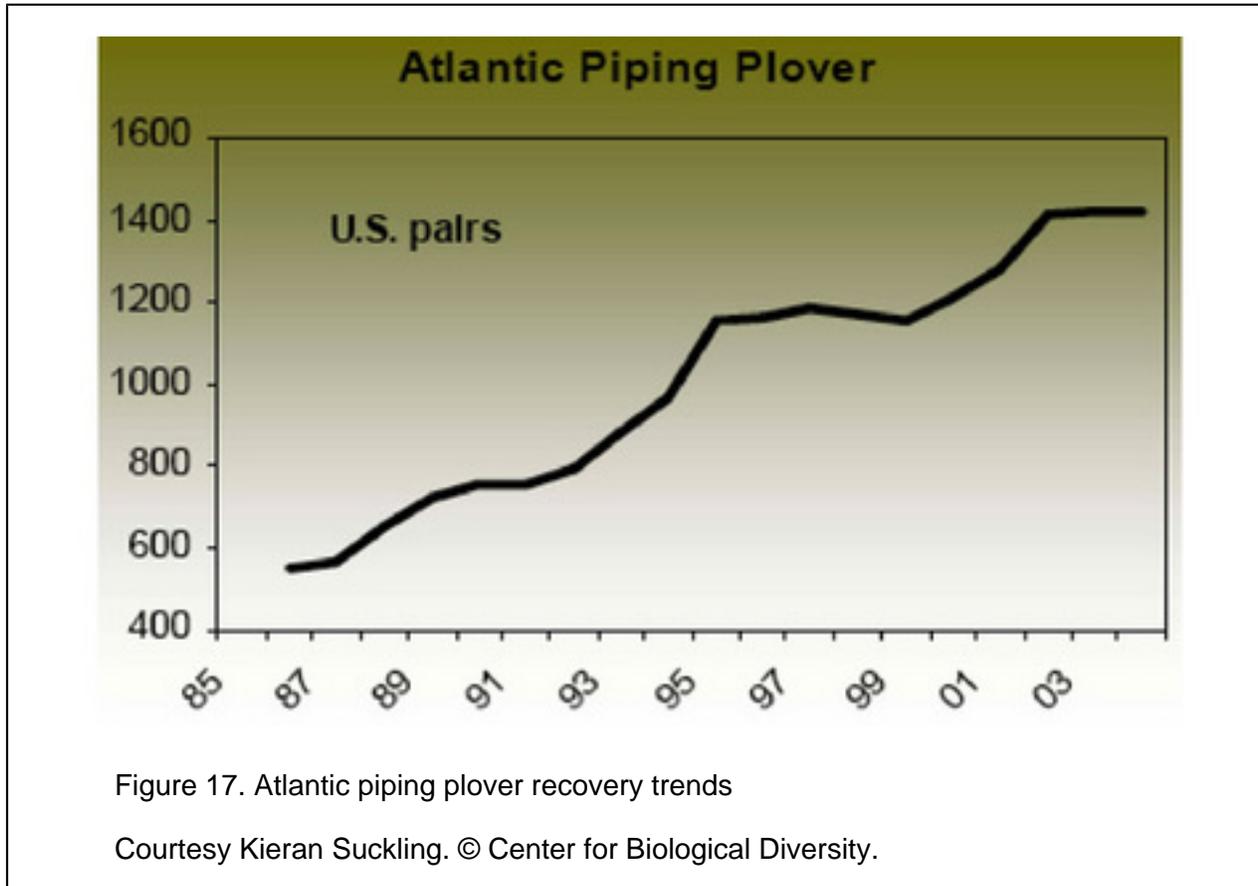
Figure 16. Tomato frog (*Dyscophus antongilii*). Listed on CITES Appendix I (threatened with extinction)

© Franco Andreone.

Many nations have passed domestic laws that protect endangered species, using frameworks similar to the IUCN Red List and focusing on the most threatened species. However, the Red List is generally accepted as the most complete global data source, even taking into account its gaps.

The U.S. Endangered Species Act (ESA), passed in 1973, seeks to protect species that are endangered (threatened with extinction throughout all or a significant portion of their ranges) or threatened (likely to become endangered throughout all or a significant portion of their range) within the foreseeable future. It does so by barring the "take" of listed species (including actions such as killing, harvesting, harassing, pursuit, and habitat alterations that kill or hurt wildlife, such as destroying nesting grounds) and any trade in those species without a federal permit. Federal agencies are required to designate "critical habitat" for listed species when it is judged to be "prudent and feasible," and actions such as development that would adversely impact critical habitat areas are prohibited.

Protection under the ESA has helped dozens of endangered species to recover and establish self-sustaining populations, including the bald eagle and green sea turtle. Figure 17 shows population trends for the Atlantic piping plover, which was listed as endangered in 1985.



Decisions taken under the ESA about whether to list or delist a species and how to define critical habitat often become highly controversial, with debate centering on two issues: the quality of the scientific analysis that provides a foundation for these actions and the economic tradeoffs involved in restricting development to protect species that may be present in very low numbers. These controversies underline the fact that the greatest current threat to biodiversity is human development.

In July 2006, 19 leading biodiversity experts from 13 nations issued a statement warning that the world is "on the verge of a major biodiversity crisis" and that governments and private actors needed to take the issue more seriously. Part of the problem, they stated, is that biodiversity is even more scientifically complex than issues such as stratospheric ozone depletion or climate change: "By definition, biodiversity is diverse: it spans several levels of biological organization (genes, species, ecosystems); it cannot be measured by simple universal indicators such as temperature and atmospheric CO<sub>2</sub> concentration; and its distribution and management are more local in nature" (footnote 28).



Accordingly, they argued, a need exists for an expert body to provide organized, coordinated, and internationally validated scientific advice on biodiversity issues, as the Intergovernmental Panel on Climate Change does for global climate change (for more details, see Unit 12, "Earth's Changing Climate"). France is sponsoring a consultation process aimed at designing such a panel.

### 13. Further Reading

Annenberg Media, *Rediscovering Biology: Molecular to Global Perspectives*, Units 4 ("Microbial Diversity") and 12 ("Biodiversity"), <http://www.learner.org/channel/courses/biology/index.html>. Two units from a professional development course for high school biology teachers, including video, online text, and supporting materials.

Center for Biological Diversity, "The Road to Recovery: 100 Success Stories for Endangered Species Day 2006," <http://www.esasuccess.org/reports/>. Profiles of species that have increased their populations under the U.S. Endangered Species Act.

Cornell Laboratory of Ornithology, **The Search for the Ivory-Billed Woodpecker**, <http://www.birds.cornell.edu/ivory>. News, evidence, and updates from the researchers who claimed to have found an ivory-bill in 2004.

Rolf Gradinger, "Arctic Sea Ice: Channels of Life," NOAA Ocean Explorer, [http://oceanexplorer.noaa.gov/explorations/02arctic/background/sea\\_ice/sea\\_ice.html](http://oceanexplorer.noaa.gov/explorations/02arctic/background/sea_ice/sea_ice.html). Explore the unique environment of Arctic sea ice communities.

University of California Museum of Paleontology, "The Phylogeny of Life," <http://www.ucmp.berkeley.edu/allife/threedomains.html>. Detailed information on the structure, ecology, and fossil records of major subgroups in the three domains of life.

<http://www.discoverlife.org>. An online guide to the world's flora and fauna, with information on each group's taxonomy, natural history, abundance, distribution, and ecology.

### Footnotes

1. E.O. Wilson, ed., **Biodiversity** (Washington, DC: National Academy of Sciences, 1988).
2. Rodolfo Dirzo and Peter H. Raven, "Global State of Biodiversity and Loss," **Annual Review of Environment and Resources**, vol. 28 (2003), pp. 154–160.
3. David R. Foster and John D. Aber, eds., **Forests in Time: The Environmental Consequences of 1,000 Years of Change in New England** (New Haven: Yale University Press, 2004), pp. 59–61.
4. For examples, see Field Museum, "Meet the Scientist," [http://www.fieldmuseum.org/biodiversity/scientist\\_department5.html](http://www.fieldmuseum.org/biodiversity/scientist_department5.html).
5. Dirzo and Raven, pp. 140–41.



6. Natalie Angier, "Animals and Fungi: Evolutionary Tie?" **New York Times**, April 16, 1993, p. A18.
7. J. Alan Clark and Robert A. May, "Taxonomic Bias in Conservation Research," **Science**, July 12, 2002, pp. 191–192.
8. Dirzo and Rave, "Global State of Biodiversity and Loss," pp. 141–42.
9. Census of Marine Life, "Ocean Microbe Census Discovers Diverse World of Rare Bacteria," July 31, 2006, [http://www.coml.org/medres/microbe2006/CoML\\_ICOMM%20Public\\_Release\\_07-31-06.pdf](http://www.coml.org/medres/microbe2006/CoML_ICOMM%20Public_Release_07-31-06.pdf).
10. Noah Fierer and Robert B. Jackson, "The Diversity and Biogeography of Soil Bacterial Communities," **Proceedings of the National Academy of Sciences**, vol. 103 (2006), pp. 626–31.
11. Norman Myers et al., "Biodiversity Hotspots for Conservation Priorities," **Nature**, vol. 403, February 24, 2000, p. 853.
12. For an interactive map with detailed descriptions of all 34 global hotspots, see Conservation International, [http://www.biodiversityhotspots.org/xp/Hotspots/hotspots\\_by\\_region/](http://www.biodiversityhotspots.org/xp/Hotspots/hotspots_by_region/).
13. Conservation International, "Protected Area Coverage in the Hotspots," [http://www.biodiversityhotspots.org/xp/Hotspots/hotspotsScience/conservation\\_responses/protected\\_area\\_coverage.xml](http://www.biodiversityhotspots.org/xp/Hotspots/hotspotsScience/conservation_responses/protected_area_coverage.xml).
14. Ana S. L. Rodrigues et al., "Effectiveness of the Global Protected Area Network in Representing Species Diversity," **Nature**, vol. 428, April 8, 2004, p. 642.
15. Andrew Balmford, Rhys E. Green, and Martin Jenkins, "Measuring the Changing State of Nature," **TRENDS in Ecology and Evolution**, vol. 18, no. 7, July 2003, p. 327.
16. Dirzo and Raven, p. 164.
17. Douglas J. Levey et al., "Effects of Landscape Corridors on Seed Dispersal by Birds," **Science**, vol. 309, July 1, 2005, pp. 146–48; Cornelia Dean, "Home on the Range: A Corridor for Wildlife," **New York Times**, May 23, 2006, p. F1.
18. "The Unwanted Amphibian," Frog Decline Reversal Project, Inc., <http://www.fdrproject.org/pages/toads.htm>.
19. David S. Wilcove et al., "Quantifying Threats to Imperiled Species in the United States," **Bioscience**, vol. 48, no. 8, August 1, 1998.
20. Food and Agriculture Organization, **The State of World Fisheries and Aquaculture 2004**, <http://www.fao.org/DOCREP/007/y5600e/y5600e00.htm>, p. 32.
21. C.M. Rick Tomato Genetics Research Center, <http://tgrc.ucdavis.edu/>.
22. U.S. Environmental Protection Agency, "New York City Watershed Partnership," June 2006, <http://www.epa.gov/innovation/collaboration/nyc.pdf>.



23. Andrew Balmford et al., "Economic Reasons for Conserving Wild Nature," **Science**, August 9, 2002.
24. Dirzo and Raven, pp. 137–67.
25. For details, see "The Great Backward Bird Count," <http://www.birdsource.org/gbbc/>; National Wildlife Federation, "Frogwatch USA," <http://www.nwf.org/frogwatchUSA/index.cfm>; and Reef Environmental Education Foundation, <http://www.reef.org/index.shtml>.
26. **Filling the Gaps: Priority Data Needs and Key Management Challenges for National Reporting on Ecosystem Condition** (Washington, DC: H. John Heinz Center, May 2006), p. 3.
27. "Central Park Survey Finds New Centipede," American Museum of Natural History, January 29, 2003; "Graduate Student Discovers an Unusual New Species," **Oceanus**, February 10, 2006.
28. Michel Loreau et al., "Diversity Without Representation," **Nature**, July 20, 2006, pp. 245–46.

## Glossary

**biomes** : Broad regional areas characterized by a distinctive climate, soil type, and biological community.

**ecosystem** : A level of organization within the living world that includes both the total array of biological organisms present in a defined area and the chemical-physical factors that influence the plants and animals in it.

**edge effect** : The observed increase in the number of different species along the margins of two contrasting environments in an ecosystem. This term is commonly used in conjunction with the boundary between natural habitats, especially forests, and disturbed or developed land.

**endemic** : Describing a disease or characteristic commonly found in a particular region or group of people; a disease constantly present at low levels in an area.

**habitat fragmentation** : A process of environmental change important in evolution and conservation biology. It can be caused by geological processes that slowly alter the layout of the physical environment or by human activity, such as land conversion, that can alter the environment on a much faster time scale.

**hotspots** : An informal expression designating specific areas as being contaminated with radioactive substances, having a relatively high concentration of air pollutant(s), or experiencing an abnormal disease or death rate.

**invasive species** : Refers to a subset of introduced species or non-indigenous species that are rapidly expanding outside of their native range.



**speciation** : The formation of two or more genetically distinct groups of organisms after a division within a single group or species. A group of organisms capable of interbreeding is segregated into two or more populations, which gradually develop barriers to reproduction.

**species richness** : A type of approach to assessing biodiversity that examines the distribution of all resident terrestrial vertebrates: amphibians, reptiles, birds, and mammals.

**species-years** : A way to measure extinction rate. There is approximately one extinction estimated per million species-years. This means that if there are a million species on Earth, one would go extinct every year, while if there was only one species, it would go extinct in one million years.

## Unit 10 : Energy Challenges



Drill rig for test site for sequestering carbon underground. Courtesy of Larry Hamill

### Overview

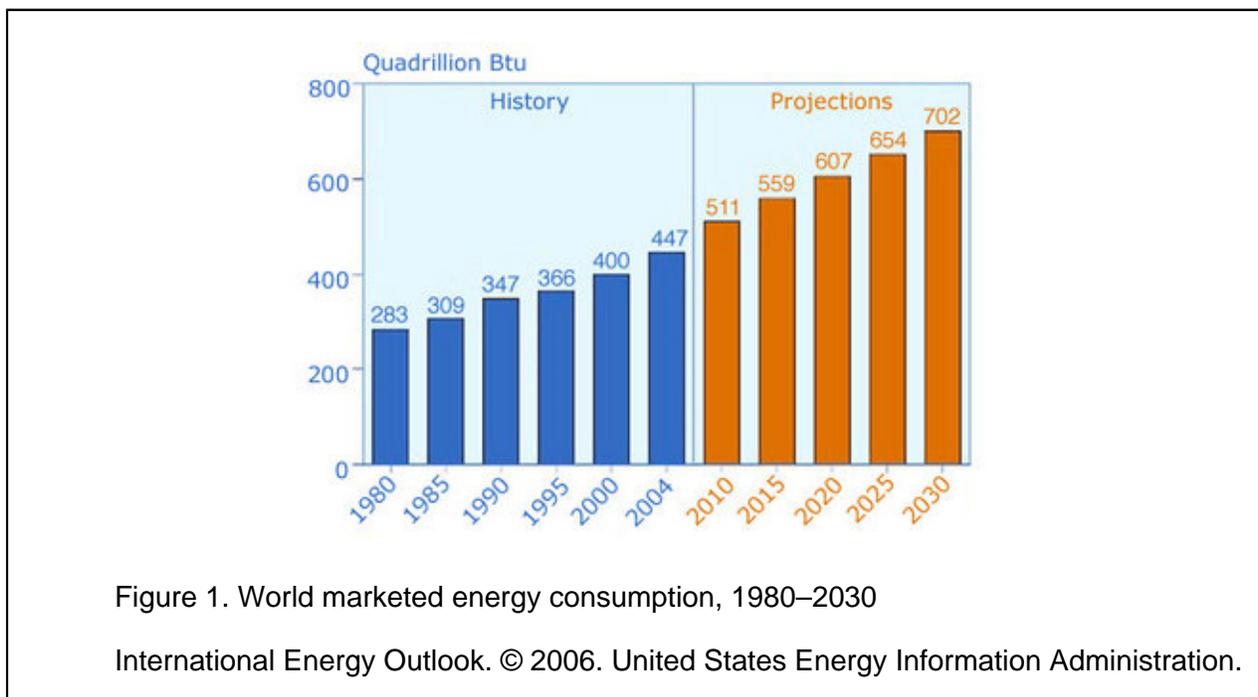
Industrialized nations rely on vast quantities of readily available energy to power their economies and produce goods and services. As populations increase in developing countries and citizens demand better standards of living, global energy consumption will continue to rise, along with demands for non-fuel mineral resources such as iron and steel. Learn about new technologies that can produce ample supplies of energy without some of the environmental costs linked to current energy resources.

### Sections:

1. Introduction
2. Thinking About Supply
3. Fossil Fuels: Coal
4. Fossil Fuels: Oil and Gas
5. Unconventional Fossil Fuels and Technologies
6. Nuclear Power
7. Biomass Energy and Feedstocks
8. Hydropower and Ocean Energy
9. Geothermal Energy
10. Wind Power
11. Direct Solar Energy
12. Hydrogen Power
13. Material Resources: Metals
14. Other Material Resources
15. Increasing End-Use Efficiency of Energy and Materials
16. Further Reading

## 1. Introduction

Industrialized nations rely on vast quantities of readily available energy to power their economies and produce goods and services. As populations increase in developing countries and their citizens demand better standards of living, global energy use will continue to rise, with developing nations accounting for a growing share of total world demand (Fig. 1).



Today most of the world's energy is derived from fossil fuels, which are **non-renewable resources** available only in limited supply. In contrast, many alternative sources of energy, such as wind, solar, and hydropower, are **renewable resources** because their supplies are refreshed faster than humans consume them. Human society has profited from exploiting energy sources, particularly since energy use became much more efficient during the Industrial Revolution. We are now deeply dependent on reliable, cheap sources of energy. However, it is important to note that energy consumption does not directly improve the human condition. Rather, what matters are the services that we generate using energy.



*"Customers don't want lumps of coal, raw kilowatt-hours, or barrels of sticky black goo. Rather, they want the services that energy provides: hot showers and cold beer, mobility and comfort, spinning shafts and energized microchips, baked bread and smelted aluminum. And they want these 'end uses' provided in ways that are secure, reliable, safe, healthful, fair, affordable, durable, flexible, and innovation friendly."*

Amory B. Lovins and L. Hunter Lovins, "Mobilizing Energy Solutions," **The American Prospect**, January 28, 2002

Modern societies also consume vast amounts of material resources, including metals, minerals, stone, chemicals, and fibers. In most cases, these materials are abundant enough that they can be considered either renewable or available in such quantities that we will not soon deplete them. The main concerns associated with material resources, therefore, are generally the costs and environmental impacts of extracting, transporting, and refining them.

Scientists who study energy and material resources seek to understand what types of resources are available and where they can be found, and to develop new technologies for locating, extracting, and exploiting them. Discovering new supplies and using more energy and materials is one way to derive more benefits. But we also can use these resources more efficiently, so that we obtain a rising amount of service from a constant level of inputs. Over the longer term, scientific and technological advances may enable societies to substitute new energy sources and material stocks for old ones. This typically happens when new resources perform as well as or better than current options and produce fewer negative impacts, such as pollution or health and safety risks.

But changing from one resource type to another involves more than simply discovering a new mineral deposit or developing a new technology. It also means altering the systems that produce, process, and distribute these resources. For example, major commercial energy fuels like coal, natural gas, and uranium are mined, cleaned, processed, refined, and delivered through complex, multi-stage systems that represent billions of dollars in infrastructure investments and complicated logistical interconnections (Fig. 2). Energy facilities typically operate for 30 to 50 years, so they cannot change to different resource or technology mixes overnight. Retiring them prematurely to replace them with something "better" is very costly even if the new plants are not more expensive than the old ones.



Figure 2. Offshore oil drilling platform, Gulf of Mexico

© United States Government Printing Office. Materials Management Service.

This unit describes the main energy sources available or under study today to meet world demand in the current century. It begins with fuels that have been commercialized and are in use on a large scale, including conventional fossil fuels (coal, oil, and natural gas) and nuclear power. We then consider alternatives such as non-conventional fossil fuels, various renewable energy sources, and hydrogen energy. As we will see, the viability of conventional and alternative energy resources depends largely on developing new technologies that will harness them more efficiently while mitigating their harmful environmental consequences—especially their contributions to air pollution and global climate change.

This unit also surveys major uses of non-fuel mineral (material) resources and their environmental impacts. It concludes with a discussion about using resources efficiently as a way to save money, extend limited supplies, and reduce environmental damages.

## 2. Thinking About Supply

Because energy and other mineral resources are so central to our lifestyle in developed countries, they often make news when prices rise and supplies tighten. At these times, national leaders and consumers alike want to know how much of the scarce resource remains, when it will run out, and what alternatives exist. These debates are sometimes cast in stark terms—asking, for example, when the world will run out of oil.

In reality, societies never use up nonrenewable resources completely or exploit the entire flow of renewable resources. Typically the best deposits and sites are found and exploited first, followed by other lower-quality sources as demand rises. As demand grows and a resource becomes scarce, its price rises. This reduces demand and gives explorers incentive to develop sources that are lower-quality and/or more expensive to exploit, and to improve technologies for locating, extracting, and processing the resource.

Rising prices also spur the development of substitutes that were uneconomic when the original resource was cheap. For example, as discussed later in this unit, high oil prices are driving significant investments today into fuel production from plant sources (Fig. 3). In the words of Sheikh Zaki Yamani, a former oil minister of Saudi Arabia, "The Stone Age did not end for lack of stone, and the Oil Age will end long before the world runs out of oil" (footnote 1). The race is between finding new supplies and exploiting them more efficiently on one hand and declining resource abundance and/or quality on the other.



Figure 3. Pump offering bio-based fuels, Santa Fe, New Mexico

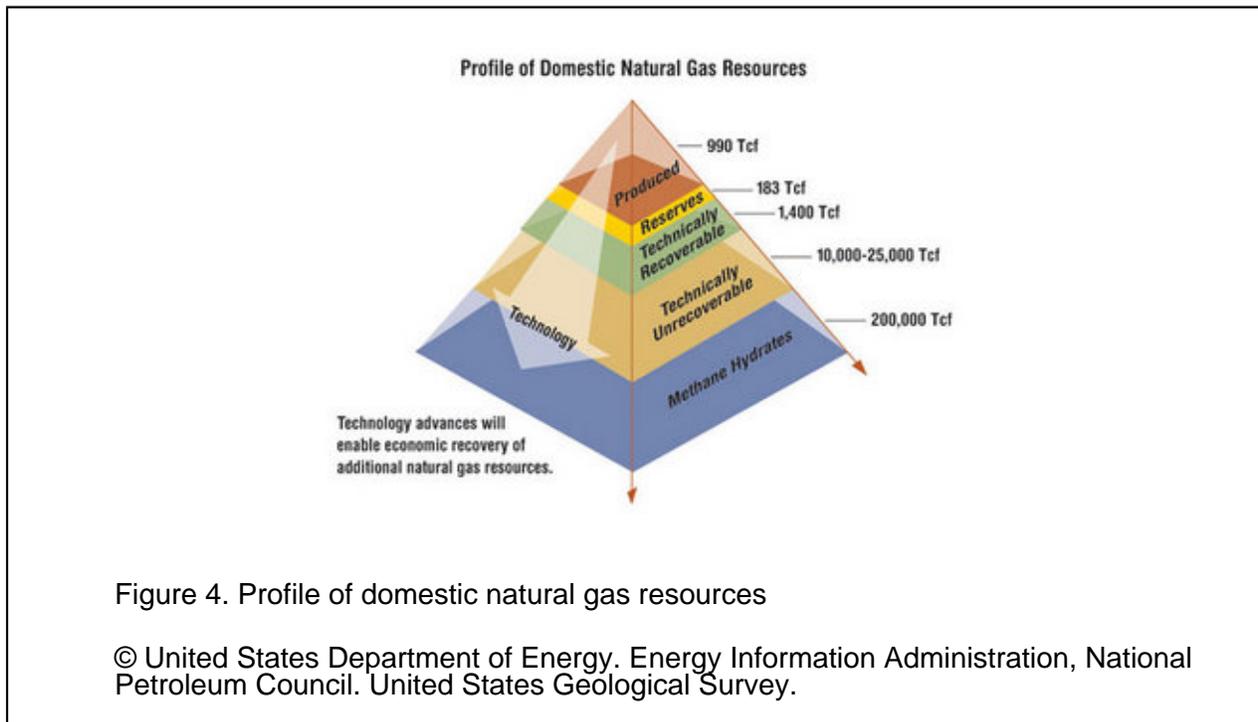
© Bensinger, Charles and Renewable Energy Partners of New Mexico.

The concepts of **stocks** and **flows** are important in thinking about resource supplies. A stock is the amount of material in a certain deposit or reservoir—for example, the total quantity of oil in a field that can be recovered with today's technology. Flow refers to the rate at which new material is added

to the stock (inflow) or removed from the stock (outflow). The net flow rate (inflow minus outflow) determines whether the stock grows, shrinks, or remains constant.

Non-renewable resources are limited by the size of their stock, but energy developers consider stocks on several levels. For example, total U.S. copper resources include all known copper deposits and those that are estimated or believed to exist, even if they cannot be economically found or extracted with today's technology. Reserves are the subset of this supply whose location is known or very likely based on geological evidence and that can be extracted profitably with current technology at current prices. A larger fraction, often referred to as unrecoverable or ultimately recoverable reserves, will require technical advances to locate and develop economically.

These categories are imprecise and shift as exploration and technology breakthroughs enable us to recover supplies that once were out of reach. Figure 4 shows current estimates of how many trillion cubic feet of natural gas the United States has in each of these categories—including sources such as [methane hydrates](#) (discussed further below) that cannot be exploited today but could become an important source in coming decades.

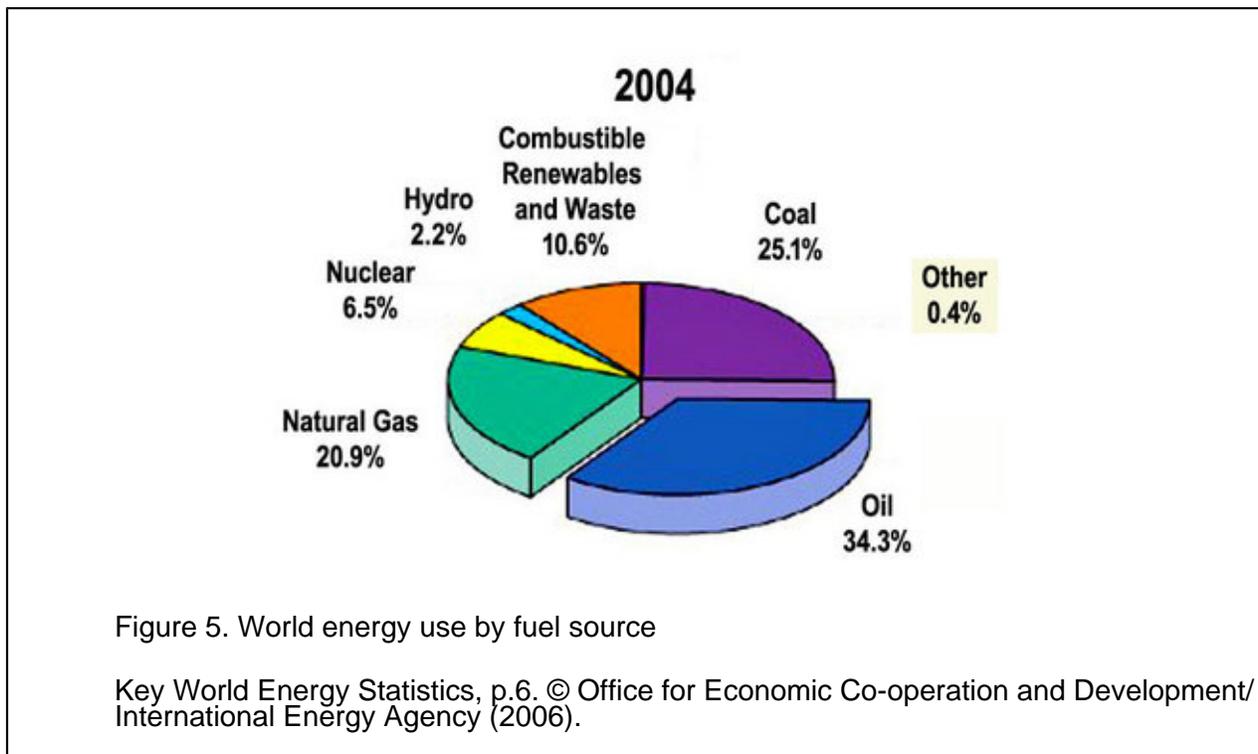


In contrast, use of renewable resources is limited by their flow rate, which can be divided into total flow and exploitable flow—the portion that can be practically exploited with current technology. The fraction of the total flow that is exploitable depends on the abundance of sites where the resource is sufficiently concentrated and close enough to the point of end-use to be harnessed economically—

a question that naturally depends in part on the state of the technology available for doing so. For example, the United States has good wind resources in the Great Plains states, but many of the windiest regions are far from major electricity demand centers, so the cost of building long-distance transmission lines affects decisions about where wind farms are built.

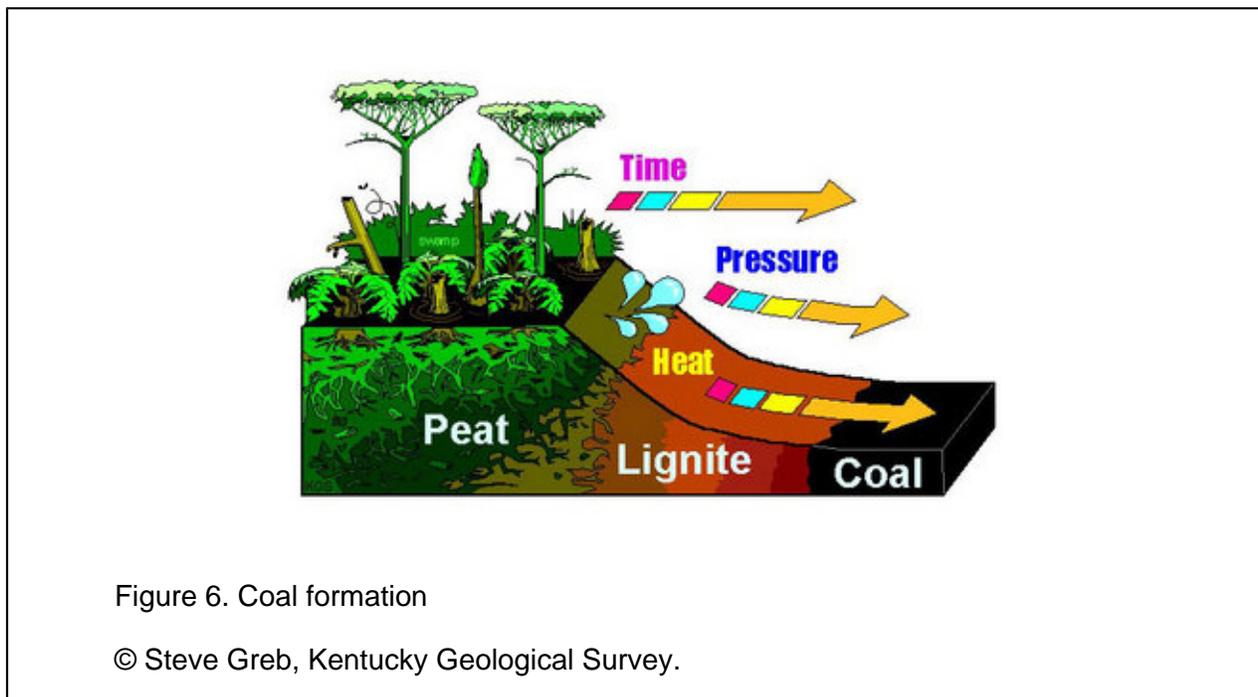
### 3. Fossil Fuels: Coal

Oil, natural gas, and coal, the traditional fossil fuels that have powered modern societies since the Industrial Revolution, remain the dominant world energy sources today. These fuels account for roughly 80 percent of world energy use and 86 percent of that of the United States, where the remainder comes from nuclear power and renewable fuels (Fig. 5).



Fossil fuels hold energy stored in plant tissues by photosynthesis millions of years ago. When these ancient plants and the animals that fed on them died, they were buried in sediments, where Earth's heat and compression from the weight of overlying rock eventually turned the deposits into coal, oil, and natural gas. Exploring for and extracting these fossil fuels requires an intimate knowledge of the Earth's structure and history, and employs many of today's geoscientists.

Coal, the first fossil fuel exploited by humans for energy on a large scale, is a carbonaceous rock formed from buried plants in ancient forests or swamps. These plant materials are initially converted to peat—a loose, brown, organically rich soil that itself is an important energy resource in some areas. As more rock layers press down on the buried deposits, geothermal energy heats the peat and reduces its oxygen and hydrogen content, converting it to coal (Fig. 6). As materials go through this process, known as thermal maturation, their energy content by weight increases.



Coal comes in several grades that reflect its thermal maturity and energy content:

- Brown coal (lignite), the first type of coal to form when plant matter is compacted, has an energy value of 9 to 17 million **British thermal units (Btu)** per ton. Because it has a low energy content, larger volumes are needed relative to higher-grade coals in order to generate the same amount of power.
- Sub-bituminous coal (16 to 24 million Btu/ton) and bituminous coal (19 to 30 million Btu/ton) are characteristically dark black and represent the most important coal grade for energy production (both direct heating and electricity generation) throughout the world.

- Anthracite coals are metallic gray and have a very high energy content, typically 22 to 28 million Btu per ton. Most readily accessible anthracite reserves in the eastern United States have been exhausted, and the remaining deposits are generally reserved for use in processing metals because of anthracite's high energy output and low volatile content.

Coal is extracted in both subsurface and surface strip mining operations. These processes have significant but different environmental impacts. Underground mining has relatively low immediate impact at the surface, but can cause ground subsidence when mineshafts collapse. Coal dust and methane gas (which is commonly found along with coal) raise significant risks of explosions. Worldwide, several thousand miners on average die each year in coal mining-related accidents.

In contrast, the impacts of strip mining—removing soils and **overburden** to extract shallow coal deposits—are highly visible at the surface. Strip mining operations generally leave permanent scars on the landscape. In its most extreme form, mountaintop removal, land is clear-cut and leveled with explosives to expose coal seams, with most of the removed overburden dumped into neighboring valleys (Fig. 7).



Figure 7. Mountaintop removal site, Kayford Mountain, West Virginia (2005)

© Vivian Stockman/www.ohvec.org. Flyover courtesy SouthWings.

Coal often contains a significant amount of sulfur, in either organic or metallic compounds, such as the mineral pyrite. When rain or groundwater comes in contact with coal, it produces sulfuric acid. Acid drainage from coal mines can pollute surrounding areas long after the mines are shut down. Many underground mines are dug to levels below the water table, so they flood easily after they

are abandoned. When this happens, contaminated water flows out of mines, lowering the pH of lakes, rivers, and streams and leaching toxic heavy metals from the ground. Runoff from abandoned mines is a major source of water pollution in states with large coal industries like West Virginia and Pennsylvania.

Beyond the mine, coal produces significant amounts of atmospheric pollution and greenhouse gas emissions when it is burned. Coal combustion generates sulfate and nitrogen emissions that contribute to acid deposition, regional haze, and smog. It also produces mercury, which accumulates in the fatty tissues of animals and fish and can harm humans who consume certain species. (For more on these issues, see Unit 11, "Atmospheric Pollution.") And coal is the most carbon-intensive of all fossil fuels, so it produces a disproportionate share of total greenhouse gas emissions from energy use. On average, coal contains roughly 30 percent more carbon per unit of energy than crude oil and 75 percent more than natural gas (footnote 2). (For more discussion of GHG emissions from energy consumption, see Unit 12, "Earth's Changing Climate.")

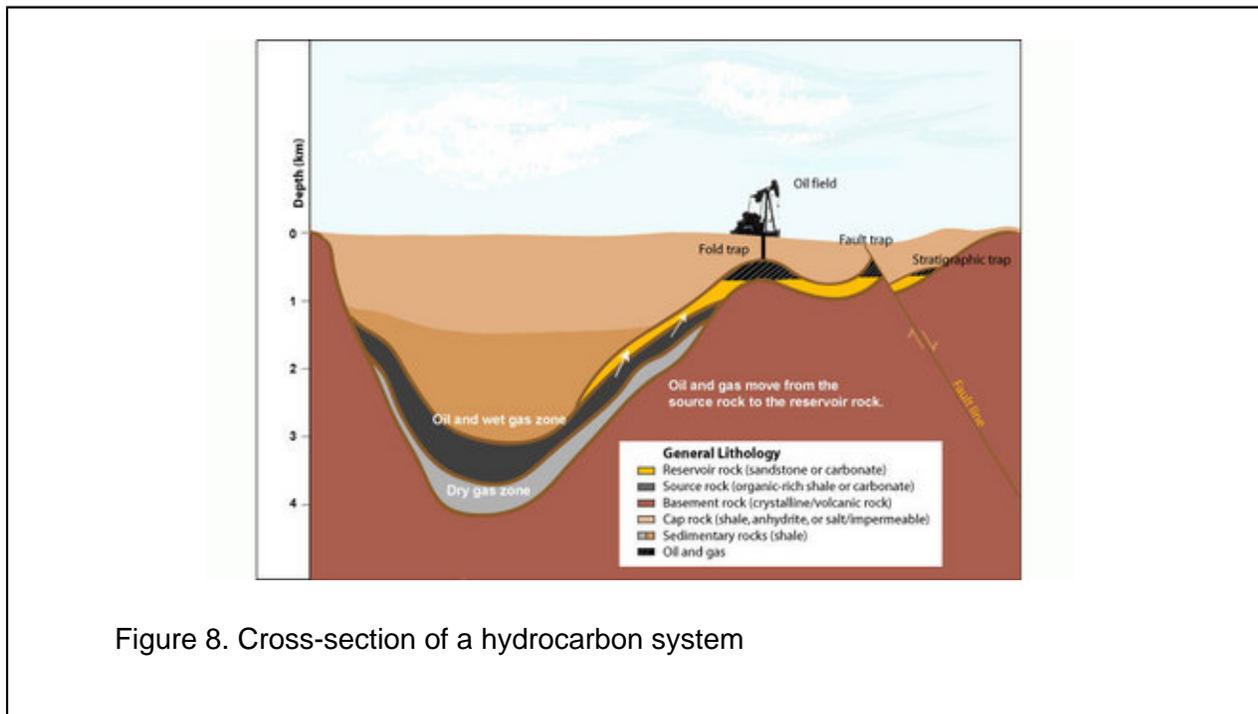
A variety of technologies exist to make coal cleaner to burn, including filtration systems that reduce particulate emissions, scrubbing systems to reduce hazardous sulfur and nitrogen emissions, and methods for removing mercury. Moreover, coal can be turned into a form of syngas (synthetic natural gas), which can be burned with smaller environmental impacts. Many of these technologies are proven and some are in use at modern power generation facilities. It is generally very costly to retrofit older power plants with these capabilities, however. Technologies that could capture a large part of the carbon dioxide from coal-burning power plants, for subsequent storage away from the atmosphere, are under intensive development but are certain to be expensive.

#### 4. Fossil Fuels: Oil and Gas

Oil and natural gas are formed primarily when marine organisms die and settle to the seafloor in anaerobic environments where they cannot rapidly decompose, then are buried by sediments and heated by Earth's [geothermal gradient](#) (the rate at which temperatures increase moving inward from the surface toward the planet's core). This process breaks down complex organic molecules into a viscous gel called kerogen, which evolves with further heating into [hydrocarbons](#)—organic compounds consisting of carbon and hydrogen atoms. Petroleum and natural gas, and the products that we derive from them, such as gasoline and diesel fuel, are composed of hydrocarbon molecules of various sizes and shapes.

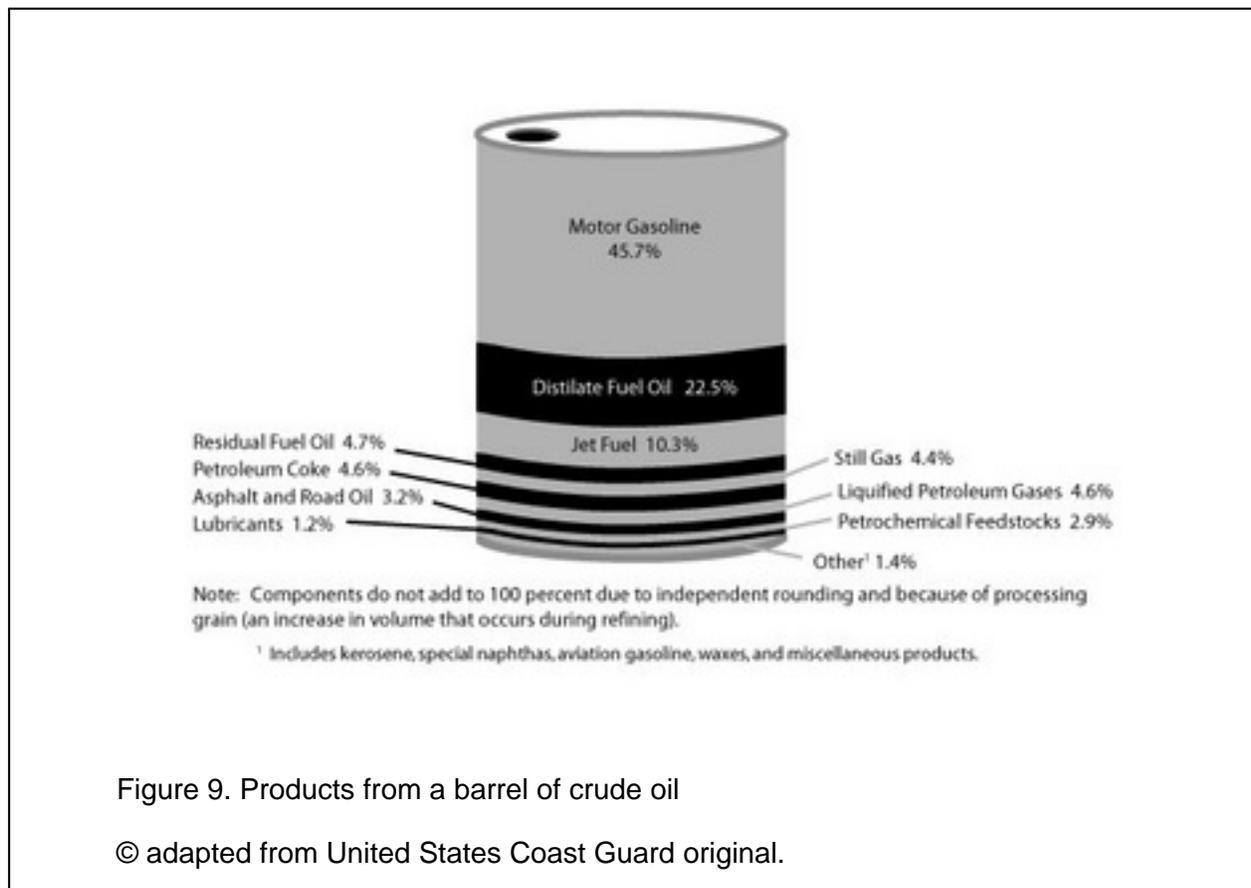
Most oil and gas forms in [source rocks](#) that represent ancient sea beds buried and heated to a certain temperature range, which corresponds to a depth range below the surface often called the oil or gas window. The oil window is generally between 1 and 6 kilometers deep, depending on the type of kerogen present and the geothermal gradient, so it is rare to produce oil from deeper levels. Gas can be produced and retained at much greater depths, such that some important gas deposits reside at depths greater than 10 km.

Oil and gas migrate out of source rocks into porous and permeable rocks called reservoirs and collect in traps that are often formed by faults or folded rocks. Reservoirs must be overlain by impermeable rocks called cap rocks or seals. The combination of a source rock, reservoir, trap, and cap rock is called a **hydrocarbon system**—the essential geologic elements that must be in place to yield a large oil or gas field (Fig. 8).



Developers tap these deposits by drilling wells into oil and gas reservoirs. In many cases, natural pressures drive the hydrocarbons to the surface. For certain heavy oils, or in fields where pressure has been depleted by production, oil must be pumped to the surface or driven from below by injecting water, natural gas, CO<sub>2</sub>, or steam into the reservoir. In many parts of the world, oil and gas exploration is pushing the frontiers of technology, with developers drilling wells more than seven miles below the surface, in deep water, or horizontally through reservoir rocks.

Refineries distill crude oil to produce a wide range of fuels, lubricants, and industrial chemicals. On average, about half of a standard barrel of oil (42 gallons) is converted to gasoline. Refined petroleum also yields kerosene, jet fuel, diesel fuel, home heating oil, and lubricants in varying proportions, depending on the original type of crude oil and the refining process (Fig. 9). Natural gas may also require processing to remove undesirable gases such as hydrogen sulfide and other impurities. In some cases this process can yield useful byproducts, such as sulfur, which is sold and used to generate fertilizer and for a wide range of other industrial purposes.



Oil and gas drilling can have adverse environmental impacts, from surface disturbance for construction of drilling pads and access roads to contamination of aquifers with drilling muds and fluids. Offshore drilling can cause spills and leaks that pollute ocean waters, either as a result of industrial accidents or through storm damage to drilling rigs. Transporting oil and gas from wells to processors to users also requires large infrastructures and creates environmental risks. Oil is shipped worldwide by pipelines and tankers, both of which are subject to spills. Most natural gas is currently transported via pipeline, but tanker shipment of liquefied natural gas (LNG) that has been chilled to -260°F represents a growing segment of the world market. LNG is re-gasified at receiving terminals and delivered by pipelines to end users.

Oil produces somewhat lower levels of CO<sub>2</sub>, sulfur dioxide, nitrogen oxide, and mercury emissions than coal when it is burned, but still contributes significantly to acid rain, photochemical smog, and global climate change. Natural gas combustion emits lower amounts of nitrogen oxide and CO<sub>2</sub> and virtually no sulfur dioxide or mercury. (For more details, see Unit 11, "Atmospheric Pollution," and Unit 12, "Earth's Changing Climate.")



Modern practices of drilling for and producing oil and gas attempt to minimize adverse environmental impacts. For example, co-produced waters are now generally re-injected or cleaned before disposal, and enhanced safety systems and procedures have made drilling and production accidents rare. Oil spills from tankers still pose a serious environmental hazard, but national governments have agreed on steps such as eliminating old tankers in favor of double-hulled designs by 2015 in an effort to further reduce these risks.

Nevertheless, because the United States has exploited many of its prime oil and gas reserves, exploration on land is now moving into environmentally sensitive regions, such as public lands that hold fossil fuel deposits but also are home to rare and endangered species. As a result, the environmental impacts of oil and gas exploration have become highly controversial in many parts of the western United States.

## 5. Unconventional Fossil Fuels and Technologies

Vast quantities of oil and gas are held in deposits where they cannot be produced by conventional drilling. When the world price of oil is high, it becomes more profitable to find ways of exploiting these unconventional fossil fuel resources. For example:

- **Tar sands** are deposits of heavy oil that are too thick, or viscous, to flow. These deposits form in the same manner as conventional deposits of oil, but microbes at shallow depths break down the light fraction of the oil. Energy developers extract oil from tar sands by injecting hot steam, which heats the sands and makes the tar less viscous so that it can be pumped out. Because generating the steam consumes large amounts of energy, tar sands are economically viable only when oil prices are high. The largest tar sand deposits in the world lie in the Athabasca region of Alberta, Canada, where production has increased rapidly in recent years (Fig. 10).



Figure 10. Athabasca tar sands, Alberta, Canada

© Suncor Energy Inc.

- **Oil shales** are tight source rocks that are not permeable enough to pump the oils out directly. Potential technologies for extracting shale oil include fracturing and igniting the shales, causing the kerogen to mature and migrating the light oil fraction to pumping stations. To date, however, only a very limited amount of oil has been recovered from shales in pilot studies. The United States has huge reserves of shale oil, which could extend our national oil supply by decades to a century if technologies are developed to harvest them economically.

- Natural gas can be extracted from coal, typically by pumping the gas directly from subsurface coal deposits (coal-bed methane). Alternatively, coal can be processed to produce syngas (coal **gasification**), which can also be used as a source of energy. Coal gasification offers tremendous promise, not only because it offers an opportunity to tap our domestic coal reserves but because pollutants and carbon dioxide can be removed from the gas before it is burned. Gasification technology has already been commercialized for industrial purposes but has not been pursued for energy to date because it is more expensive than using coal directly as fuel. However, spurred by rising oil prices and concerns about emissions from coal combustion, several U.S. companies have announced plans to build coal gasification power plants. If these plants include the capacity to capture and store carbon dioxide emissions (which would increase their capital costs still further), they could help to make coal a more environmentally acceptable energy source.
- Huge reserves of gas also occur trapped in ice within shallow sediments, both in permafrost and deep undersea environments. No affordable technology has been developed yet to harvest these broadly distributed methane hydrates or **clathrates**, but they may prove to be an important potential energy source because of their abundance and because natural gas burns more cleanly than other fossil fuels. These deposits are also under study because methane is a powerful greenhouse gas, so if they were to be vented to the atmosphere (for example, if frozen tundra thaws as Earth's surface temperature rises), they could substantially increase the rate of global climate change. Finding ways to develop methane hydrates and prevent uncontrolled releases thus would have both energy and climate change benefits.

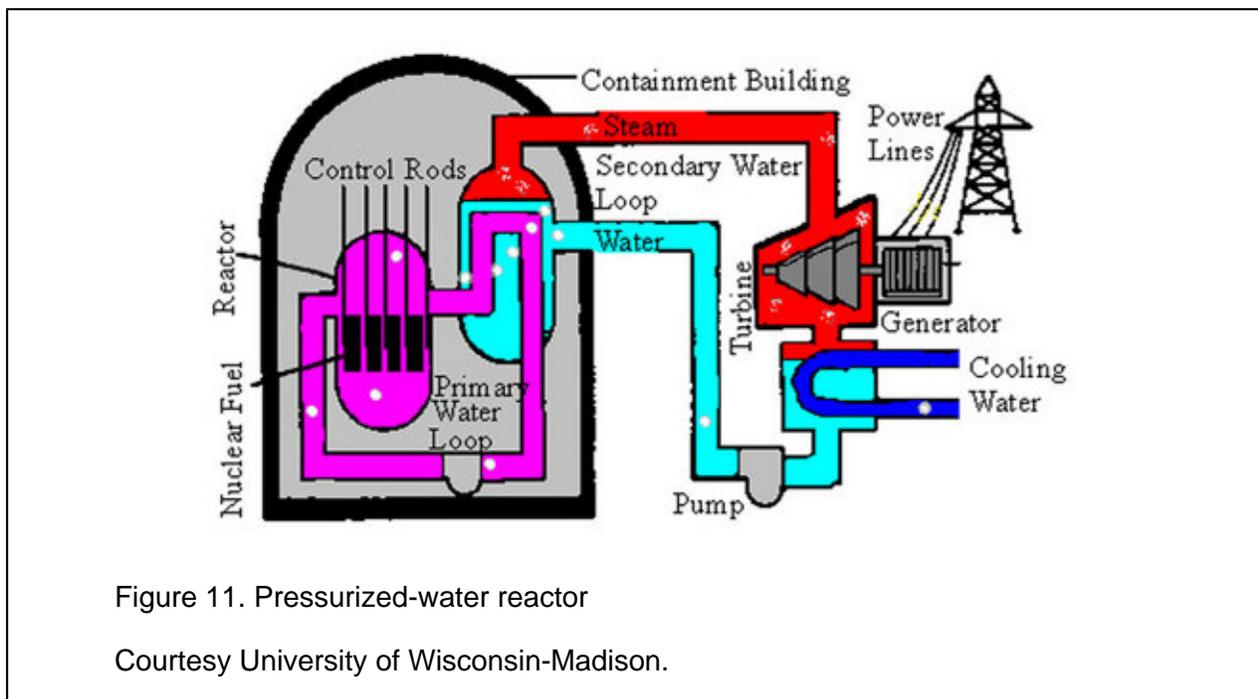
## 6. Nuclear Power

Nuclear energy, which generates about 17 percent of world electricity supplies (roughly 6 percent of total energy consumption), is produced by enhancing the radioactive decay of naturally **fissile** materials—elements whose atoms can be split by thermal (slow) neutrons, releasing energy. About 0.7 percent of natural uranium consists of the isotope uranium-235, which is fissile and is the most widely-used fuel in standard nuclear reactors. The remainder is the more stable uranium-238.

To exploit this energy source, companies mine uranium ore and, by a process called **uranium enrichment**, increase the concentration of U-235 to about 4 percent. Enriched uranium is formed into fuel rods or pellets, which are placed inside a nuclear reactor and bombarded by neutrons. This process causes U-235 atoms to split into two or more smaller atoms, called daughter products, and

releases large amounts of energy. This process also releases excess neutrons, which split other U-235 atoms, causing a nuclear fission chain reaction.

Operators control the rate of fission using control rods and moderators that absorb excess neutrons and by adjusting the reactor temperature, which affects the reaction rate. Energy generated in the reactor heats water, steam, or some other fluid, which is pumped from the reactor and used to produce steam that drives electric turbines (Fig. 11).



Nuclear power is a well-established method of electric power generation. Uranium is abundant, and a number of countries are making substantial investments in new nuclear power reactors. The United States has more than 100 licensed commercial nuclear power reactors, but no new reactor has been ordered since 1978, although interest has revived in recent years.

Major obstacles to the expansion of nuclear power worldwide include concerns about safety and high capital costs compared to other energy sources. Nuclear accidents at the Three Mile Island plant in Harrisburg, Pennsylvania, in 1979 and the Chernobyl reactor in Ukraine in 1986 convinced many people that nuclear power was unsafe. Chernobyl caused more than 30 deaths in the days immediately following the accident (from acute radiation exposure), and widespread exposure to radioactivity from the accident over a large part of the Northern hemisphere may ultimately lead to tens of thousands of deaths from cancer over a period of decades. Although both accidents were

largely results of human errors, and modern facilities have much more substantial safety procedures, these events demonstrated that nuclear accidents were possible.

Spent nuclear fuels remain highly radioactive for thousands of years, and finding appropriate sites to store radioactive waste is a highly contentious issue in virtually every nuclear nation. The United States is struggling to build and license a national repository at Yucca Mountain, Nevada, after decades of study (Fig. 12), but concerns persist about whether the site's complex geology can isolate nuclear waste from the environment until its radioactivity decays to background levels. This failure has forced many nuclear power stations to store their spent fuel onsite for years longer than owners planned and has undercut public support for new nuclear reactors in the United States.

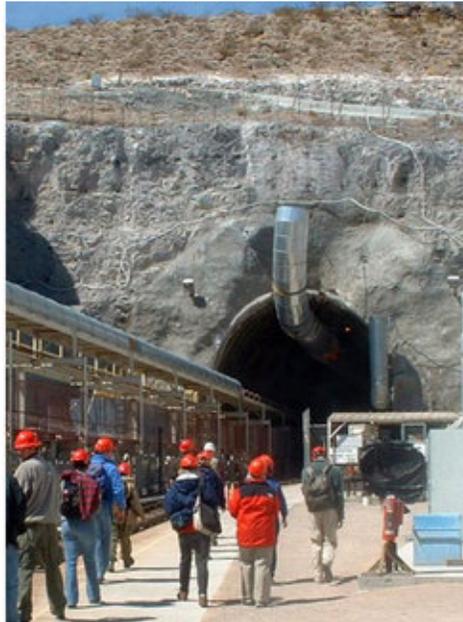


Figure 12. Main tunnel shaft, Yucca Mountain repository site

Courtesy Daniel Mayer, 2002. Wikimedia Commons, GNU General Public License.

In addition to these environmental impacts, nuclear power also raises security concerns because it produces two types of fissile material that can be used in nuclear weapons. First, as noted above, uranium fuel for commercial power reactors is enriched to a concentration of about 4 percent U-235. Although this low-enriched uranium is not usable for weapons, the same facilities can often enrich uranium to 90 percent U-235 or higher, and this highly enriched uranium is the easiest material from which to make a nuclear weapon.

Second, when nuclear fuel is irradiated in a reactor, a portion is converted to plutonium, which is also fissile and, given somewhat greater skill on the part of weapon-makers than is needed for a uranium bomb, can also serve in the weapon role. The process of plutonium production is enhanced in certain modern reactor designs called **fast breeder** systems, which use plutonium as an additional source of nuclear fuel and are now in use in several nuclear nations. Plutonium fuel cycles pose increased proliferation risks because plutonium can be stolen or diverted while it is being handled in bulk quantities during fuel processing.

## 7. Biomass Energy and Feedstocks

Biomass energy sources are all around us. They include many types of plants and plant-derived material, such as agricultural crops and wastes; trees, forestry wastes, and mill residues; animal wastes; and municipal solid waste. About 11 percent of world primary energy today is produced from biomass, but this comes mostly from low-technology applications in developing countries, such as burning crop waste, wood, or dung for home heating and cooking.

While these fuels might seem like low-impact energy sources, biomass burning in primitive stoves is very inefficient and generates large quantities of pollutants such as fine particulates, sulfur and nitrogen oxides, and carbon monoxide, which are especially harmful when released in poorly ventilated houses. Indoor air pollution is a serious health risk and causes many respiratory illnesses in developing countries (for more details, see Unit 6, "Risk, Exposure, and Health"). The inefficient combustion of biomass fuels in primitive stoves also means that much of their energy content is wasted. Programs to help users switch fuels or install more efficient stoves are among the cost-effective options to reduce health impacts from using traditional biofuels (Fig. 13).



Figure 13. A ceramic cook stove saves fuel in Myanmar

© G. Bizzarri, Food and Agriculture Organization of the United Nations.

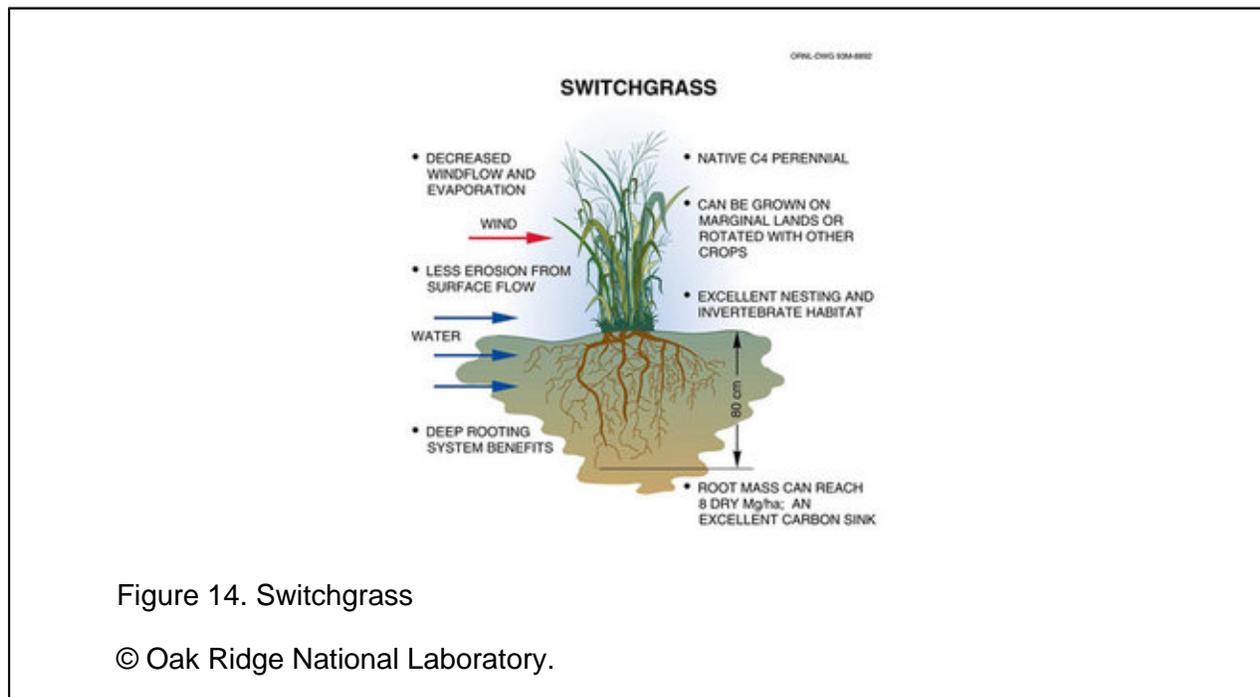
Advanced biomass technologies that use organic material cleanly and efficiently offer much greater opportunities. One of the fastest-growing biomass applications today is production of transportation fuels from plant sources. Notably:

- **Ethanol**, also known as ethyl alcohol or grain alcohol, can be fermented from sugars found in corn and other crops and added to conventional gasoline. As an additive, ethanol lowers reliance on conventional oil and increases the combustion efficiency of gasoline, reducing pollutant emissions. In Brazil, which has a sizeable ethanol industry based on sugar cane, all gasoline sold contains 25 percent alcohol, and over 70 percent of the cars sold each year can run on either ethanol or gasoline.
- **Biodiesel**, which is essentially vegetable oil, can also be derived from a wide range of plant sources, including rapeseed, sunflowers, and soybeans, and can be used in most conventional diesel engines. Because it burns more cleanly than its petroleum-based counterpart, biodiesel can reduce pollution from heavy-duty vehicles such as trucks and buses.

Both ethanol and biodiesel are viable sources of renewable energy that can reduce our dependence on conventional fossil fuels and reduce harmful emissions. However, growing biofuel crops—

especially corn for ethanol—requires major quantities of fossil fuel to manufacture fertilizer, run farm machines, and ship the fuel to market, so these biofuels do not always offer significant net energy savings over gasoline and diesel fuel. Growing corn is also water-intensive and removes significant levels of nutrients from soil (hence its high need for fertilizer). In addition, relying too heavily on these sources could mean diverting crops from the food supply at some point to produce energy.

Analysts generally agree that, at best, corn ethanol offers modest energy savings over gasoline, and that the real promise lies in making ethanol from cellulosic (woody) plants such as switchgrass, willows, and poplars (footnote 3). Some of these plants, notably switchgrass, also sequester large amounts of carbon, restore nutrients to soil, and can be used to stabilize land threatened by erosion (Fig. 14). Cellulosic plant tissue is tough and must be broken down before it can be fermented, but contains substantially more energy per unit than carbohydrates such as corn. Current research focuses on developing quick and efficient methods of breaking down cellulosic plant tissue for fermentation into fuel (footnote 4). Many experts believe that cellulosic ethanol will become a significant energy source in the United States in the next one to two decades.



Biomass can also be used to generate electricity. A number of U.S. power plants either run completely on biomass fuels or co-fire them with coal to reduce emissions. In most cases, biomass fuel is burned directly to boil water and turn steam turbines, but some advanced plants convert biomass fuels to gas by heating them in a low- or zero-oxygen environment. The resulting gas burns more efficiently than solid wood waste or plant material, thus extracting more energy from the fuel

with fewer pollutants (footnote 5). Many industrial facilities, especially in the pulp and paper industry, produce significant quantities of electric power using residual biomass fuels such as wood pulp that are generated from their own production facilities.

Human and animal wastes can also produce electricity. On farms, devices called anaerobic digesters use microbes to break down manure into organic solids and biogas, which typically contains about 60 percent methane and 40 percent CO<sub>2</sub>. The methane can be burned to generate electricity and reduce greenhouse gas emissions. Similarly, many large landfills collect the biogas that is generated by decomposition of buried organic waste and burn it to generate electricity.

One of the most important environmental benefits of using biofuels is that biomass energy is carbon-neutral—that is, using it does not increase long-term greenhouse gas levels in the atmosphere. Biomass fuels such as timber release carbon when they are burned, but this carbon was sequestered from the atmosphere when the original trees grew and would have been released when they died and decayed, so using biofuels simply completes the natural carbon cycle. In contrast, burning coal or oil releases carbon that was previously sequestered underground for millions of years and would have stayed there if it were not mined for energy, so it represents a net transfer of carbon from terrestrial sinks to the atmosphere. (For more on the carbon cycle, see Unit 2, "Atmosphere.") Biomass energy thus does not contribute to global climate change unless it is harvested more quickly than it regenerates—for example, when large forest areas are clear-cut.

## 8. Hydropower and Ocean Energy

Hydropower and ocean energy systems indirectly tap the Earth's solar energy flux, which drives the cycling of water between Earth's surface and the atmosphere and heats the upper layer of the oceans (for more details, see Unit 8, "Water Resources"). By damming rivers for hydropower or placing turbines in zones of the ocean areas with significant tides and currents, we can use the power of flowing water to generate electricity. Today hydropower generates about 17 percent of world electricity supplies (the same as nuclear energy, although hydropower is credited with only a third as much primary energy as nuclear because it does not generate two units of heat for every unit of electricity). Ocean power systems are still at an experimental stage.

Like biomass, hydropower is an established renewable energy technology that is widely used in many parts the world. The basic technology is simple: falling water flows through pipes, called penstocks, then turns turbine blades to spin a generator and produce electricity (Fig. 15). When excess power is available from the grid, some hydropower stations pump water up to storage reservoirs and hold it in reserve, then release it to generate power when demand rises.

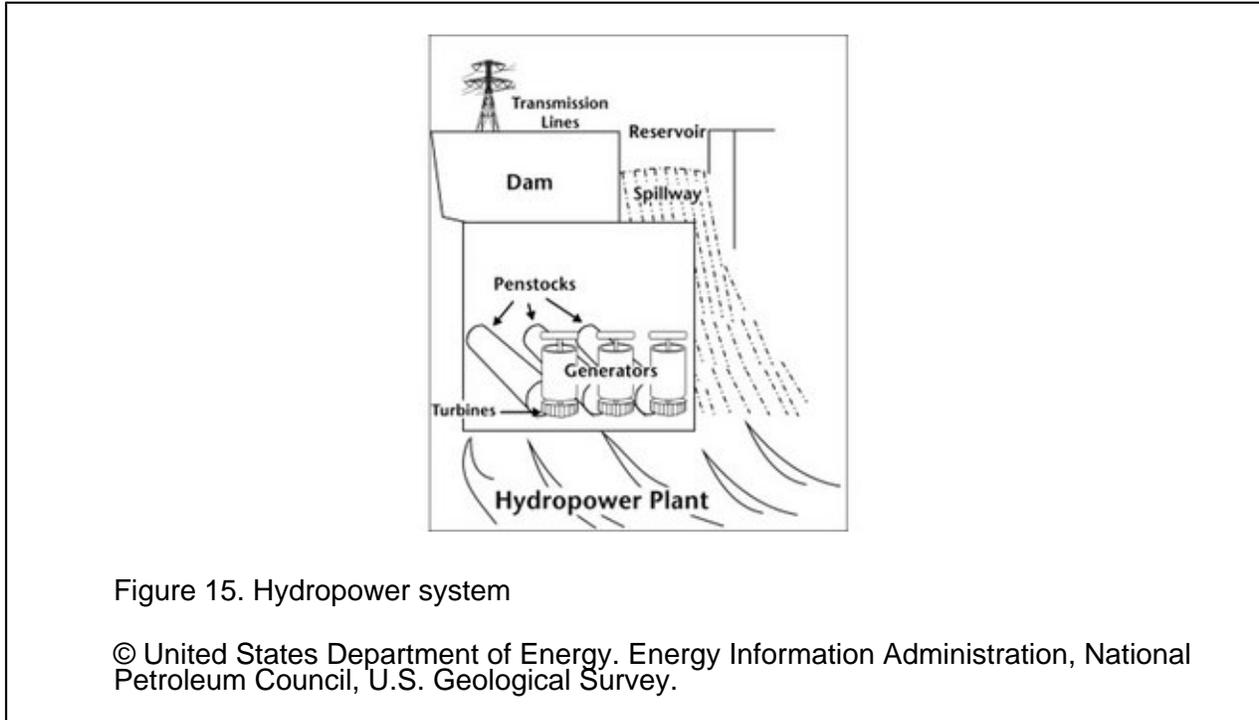


Figure 15. Hydropower system

© United States Department of Energy. Energy Information Administration, National Petroleum Council, U.S. Geological Survey.

Hydropower generates electricity without producing significant air pollution, except for emissions from building and maintaining dams. Large hydropower dams can also serve other purposes: for example, the reservoirs that develop where rivers are dammed can provide drinking water supplies, and many are used for fishing and boating. Some hydropower reservoirs, especially in dry regions like Africa, have become important habitats for birds.

In recent years, however, critics have drawn attention to hydropower's negative environmental impacts. A report issued in 2000 by an independent international commission catalogued ways in which large dams can harm ecosystems, such as:

- Killing plants and displacing animals, including endangered species, when reservoirs are flooded;
- Altering river flow rates, the quantity and character of sediments moving through the channel, and materials that make up stream beds and river banks;
- Modifying water parameters such as temperature and levels of nutrients and dissolved oxygen;

- Degrading downstream channels, floodplains, and deltas by reducing the transport of nutrients and sediments below dams; and
- Blocking migration of fish and other aquatic species up and downstream.

The report also noted that while hydropower does not generate greenhouse gas emissions as water spins electric turbines, reservoirs emit CO<sub>2</sub> and/or methane from rotting submerged vegetation and carbon inflows from the catchment area. Calculating how much a specific dam contributes to climate change depends on many factors, including whether the flooded land was previously a carbon source or sink and what land use changes result from building the dam and displacing people from the flooded area. On balance, however, it appears that warm, shallow tropical dams emit more GHGs than deep, cold dams at higher latitudes (footnote 6).

In spite of these negative aspects, hydropower is an attractive alternative to fossil fuels for many countries with good resources. In addition to their low pollutant emissions, hydropower plants provide dispatchable power: their output can be raised or lowered quickly to meet fluctuating levels of demand. Other renewable sources, such as wind and solar energy, produce energy intermittently when the wind blows or the sun shines, so they are not as responsive to daily market conditions.

At best, however, world hydropower capacity can be expanded only by a factor of 2 or 3 because a limited number of good sites remain available for development, mainly in Africa, Asia, and Latin America. In the United States, it is estimated that more than half of the hydropower generating capacity is already tapped, and most of the remaining potential dam sites would adversely affect sensitive environments. China's Three Gorges Dam, the largest hydroelectric project in the world, is scheduled to enter operation in 2009 (Fig. 16). The project will provide 18,000 megawatts of electricity-generating capacity—an important asset for China, which relies heavily on coal to meet its fast-growing energy needs. However, it has been criticized for flooding many rural valleys and displacing more than 1.5 million people.



Figure 16. Aerial view of the Three Gorges Dam, China

© National Aeronautics and Space Administration. Earth Observatory.

Ocean energy occurs in the form of tides, waves, currents, and heat. Tidal energy resources are modest on a global basis, and tapping them involves building major dams on inlets and estuaries that are prized for other purposes, so few tidal energy facilities have been developed. Harnessing waves and currents on a significant scale will involve designing turbine structures that are large, inexpensive, and can operate for long periods under the physical stresses and corrosive forces of ocean environments. For the most part, such systems are at the research stage today.

The largest but most experimental form of ocean energy is ocean thermal energy conversion, which taps heat stored in the ocean to generate electricity. This process runs warm surface seawater through several different types of systems that use the water's stored heat to turn a turbine, then cools the resulting steam or vapor with cold deep-seawater (footnote 7). Making this conversion work affordably on a large scale is technologically very difficult because it requires large structures and physical challenges associated with working in the ocean environment. It works most effectively in regions where there are large temperature differences between surface and deeper waters, mainly in the tropics. If ocean thermal energy conversion can be commercialized at some point, however, it could become an enormous new energy source.

## 9. Geothermal Energy

Geothermal power systems tap our planet's natural radioactive energy and the fact that temperature and pressure inside Earth increase with depth. Earth's geothermal gradient is steeper in some regions than others, generally because of volcanic activity or large natural deposits of naturally radioactive material in granitic rocks. Energy companies can drill a mile or more to tap underground reserves of steam and hot water, much in the same way as they drill for oil and natural gas.

Early geothermal plants used steam pumped directly from underground. Today, however, most geothermal power plants pump water down into wells, use subsurface heat to warm it, and return it to the surface to form steam, which drives electric turbines to generate electricity. Geothermal power has been an established technology since the early 20th century and is economically viable in geologically suitable sites, such as in the Geysers field in northern California or in Iceland, which produces most of its energy in this way.

Geothermal energy is considered a renewable resource because it draws from the essentially unlimited heat in the Earth's interior. Where resources are good, it produces reliable power with virtually no atmospheric pollutants or greenhouse gas emissions. In the United States, most of the best geothermal resources are located west of the Mississippi River (Fig. 17). It has proven difficult to extend this technology to areas with great demand for electricity, such as the eastern United States and much of Europe, because the local geology does not provide sufficiently high subsurface temperatures. Thus, geothermal power is a minor component of energy supply in most parts of the world today. In the future, it may prove viable in areas of tremendous geothermal potential to use excess geothermal energy to produce more transportable forms of energy—for example, by extracting hydrogen from sea water.

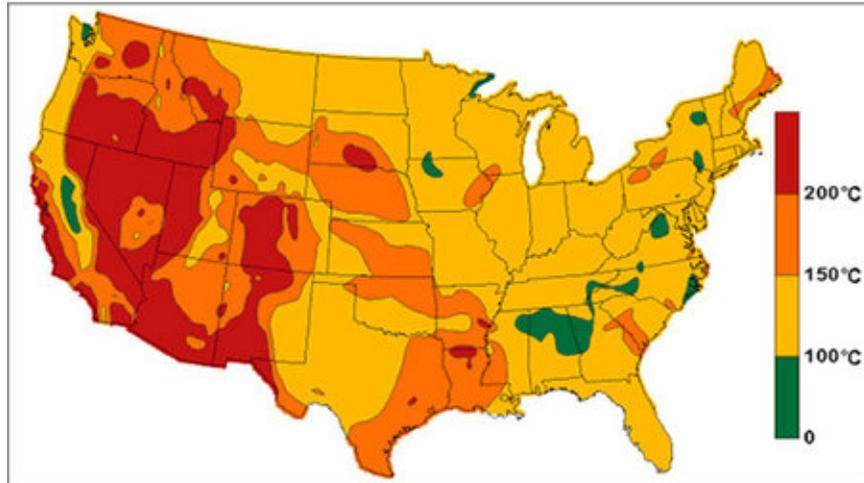


Figure 17. U.S. geothermal resources (estimated temperature at 6 kilometers depth)

© 2006. United States Department of Energy. Energy Efficiency and Renewable Energy.

## 10. Wind Power

Wind power also indirectly taps Earth's solar energy flux, which causes differences in atmospheric temperatures and pressures that drive winds (for more details, see Unit 2, "Atmosphere"). Wind turbines directly harness wind power and convert it to electricity. Today wind power generates less than 0.5 percent of the world's electricity supply, but it is growing rapidly and has several times the growth potential of hydropower. Wind power scales with the area swept by the blades (proportional to the square of the blade diameter) and the cube of the wind velocity. The best locations for wind farms thus are areas with consistently high winds where large turbines can be situated. Turbines can be located both on land and in shallow waters offshore.

Advances in wind turbine technology, such as larger and more durable towers, are making wind power more efficient and economically competitive relative to fossil-fuel-fired and nuclear-generated electricity. Several European countries have made major investments in wind energy, both for domestic power production and as a high-technology export. Most notably, Denmark generates 20 percent of its electricity from wind and is a major supplier of wind turbines around the world (Fig. 18).



Figure 18. Offshore wind turbines near the southwest coast of Denmark

© Sandia National Laboratory.

Wind turbines generate electricity without producing air pollutants or greenhouse gases. Concerns about the environmental impacts of wind energy center on finding appropriate sites for wind farms. Some critics argue that wind towers mar natural settings, such as ridge lines and coastal areas, while others worry that turbine blades will kill large numbers of birds and bats. Some early wind power installations, such as the turbines in California's Altamont Pass, had significant impacts on birds, but the industry has learned from these cases. Today, wildlife issues can usually be managed with careful siting processes and thorough environmental reviews. Replacing any significant fraction of fossil fuel consumption with wind power will require widespread siting of turbines, so resolving these concerns is a key step for expansion of wind energy.

## 11. Direct Solar Energy

Direct solar power systems typically generate electricity in one of two ways. **Photovoltaic** (PV) cells, the type commonly seen on homes and commercial buildings, use semiconducting materials such as silicon to produce electricity from sunlight: when light hits the cells, the material produces free electrons that migrate across the cell, creating an electric current. Alternatively, large-scale solar concentrating systems focus sunlight with mirrors to heat a liquid that boils water, creating steam to turn a turbine. Solar energy can also be used for residential water heating (replacing fossil fuels

or electricity) by circulating household water or a heat-carrying fluid through roof-mounted solar collection systems.

Concentrating solar power is best suited for power plants in areas with strong sunlight and clear skies, like the southwestern United States, while PV and solar hot water systems can be used in a wide range of climates and latitudes. PV cells are used widely as power supplies in electronic consumer goods such as hand-held calculators. However, because PV technologies for consumer applications have a maximum efficiency of about 15 percent, large expanses of PV cells are required to generate significant amounts of electricity. It would take more than 25 square kilometers of standard photovoltaic cells to generate the same amount of electric power as a large coal-fired power plant.

Residential PV systems are now available at major home-supply stores in several states that offer financial incentives to promote home solar power, including California and New Jersey. Some developers are integrating them into energy-saving home designs (Fig. 19). Making solar energy cost-competitive on a large scale will require further gains in efficiency and reductions in the cost of manufacturing PV cells.



Figure 19. Solar-powered housing complex, Watsonville, California

© Dan Coyro.

## 12. Hydrogen Power

As economies develop and mature, they tend to follow an energy path that moves from high carbon/low hydrogen fuels at early stages to fuels with higher hydrogen and lower carbon contents. Typically, nations use wood as their main primary fuel at a pre-industrial stage, shift to coal during industrialization, and then transition to oil and natural gas as their economies mature. This progression takes place because each new fuel is cleaner-burning and easier to distribute and store (once an infrastructure has been built to handle it) than its predecessor. The United States and western Europe are beginning to plan for perhaps the next stage on the decarbonization path—hydrogen—but this transition will require several decades to design and deploy systems for producing, transporting, and using hydrogen fuel.

Although it is sometimes called a "fuel of the future," hydrogen is more accurately described an energy carrier. Like electricity, pure hydrogen does not occur naturally in quantities worth harnessing to meet human energy needs: the main naturally occurring stocks of hydrogen are tied up in chemical compounds, most importantly water molecules ( $H_2O$ ) and hydrocarbons such as coal (approximately  $CH$ ), oil (approximately  $CH_2$ ), and methane ( $CH_4$ ). Stripping hydrogen from hydrocarbon fuels or obtaining it by splitting water using electricity or heat is not technically difficult, but in any of these approaches, more electricity or primary-fuel energy is used than the resulting hydrogen contains.

The benefit of paying this energy price to get hydrogen comes in the form of hydrogen's portability, storability, amenability to high-efficiency application not only in combustion engines but in fuel cells (discussed below), and low emissions. One application currently under research is the use of hydrogen fuel cells to power cars (Fig. 20).

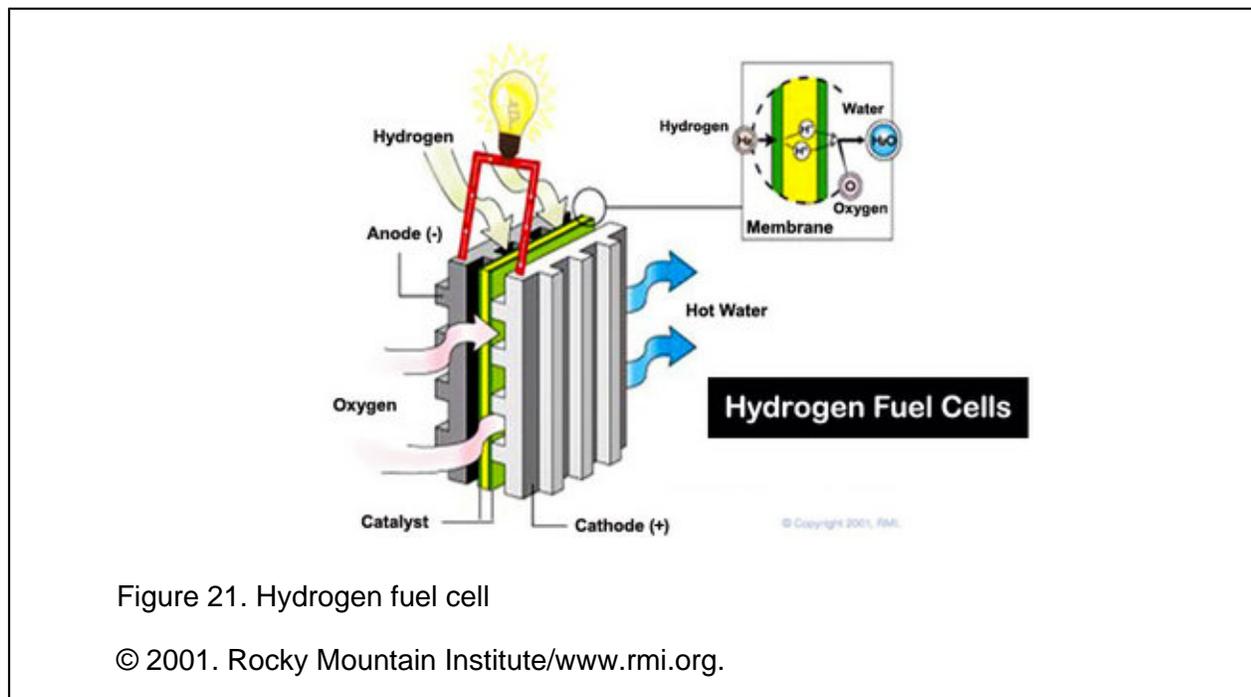


Figure 20. View under the hood of a fuel cell car

© National Renewable Energy Lab.

Today, the oil and chemical industries worldwide use about 50 million tons of hydrogen each year, most of it extracted from natural gas and coal. Deriving hydrogen from fossil fuels emits  $\text{CO}_2$ , so scaling the process up would increase greenhouse gas emissions unless the associated carbon were captured and stored (for more details on carbon capture and sequestration, see Unit 13, "Looking Forward: Our Global Experiment").

Hydrogen can be burned directly to generate energy or used in devices such as [fuel cells](#) that combine hydrogen and oxygen to produce electricity, with water as a byproduct (Fig. 21).



Here is how the basic process works:

- Hydrogen and oxygen flow into opposite sides of the cell, separated by a barrier that only allows positively charged ions to pass through.
- An anode (negatively charged pole) strips electrons from the hydrogen atoms, converting them to positively-charged ions that pass through the barrier. The negatively charged electrons flow around the outside of the cell toward the cathode (positively-charged pole), creating an electrical current. Catalysts speed the reactions at each electrode.
- Oxygen enters the cell near the cathode and combines with the hydrogen ions and electrons to form water, which is removed through an exhaust system.

Existing fuel cell technologies can convert as much as 70 percent of hydrogen's energy content to electricity. None of the basic designs in use today are cheap and technically simple enough yet for mass production, although they have been used for applications such as producing power on manned space missions.

Over the past several years, politicians and scientists have endorsed the idea of converting to a hydrogen economy. This transition poses many challenges. In addition to producing hydrogen

economically and commercializing fuel cells, it takes seven times as much hydrogen on a volume basis to produce the same amount of energy in a gallon of gasoline. Therefore, adopting hydrogen as a fuel will mean building new energy storage and distribution systems nationwide. The devices that convert hydrogen to energy services—cars, heating systems, and consumer goods—will also have to be converted. Most expert assessments of the timing for a hydrogen economy project that such systems will not start to be deployed on a large scale until 2020 or later, and that making a full transition from fossil fuels to hydrogen in the United States would take until approximately 2050 or later.

### 13. Material Resources: Metals

Fuels are not the only resources that we mine from the earth. Modern society could not exist without numerous materials such as metals, rocks, and minerals that serve as raw ingredients for buildings, technologies, and industrial processes. In large part, industrialization is a process of refining and employing these materials in ever more sophisticated ways and of substituting machine power (which relies on metals and other advanced components) for human labor.

Metals are used widely in construction, manufacture, and consumer goods and are important material resources for industrial societies. Steel, which is a blend of iron and other compounds, is king of the metals, accounting for more than 90 per cent of global metal production. Many other metals, such as chromium and nickel, are mined primarily for blending with iron to form steels with desirable characteristics, such as corrosion resistance.

Metals are extracted from **ores**—naturally-occurring rocks and minerals that hold large deposits of high-quality, readily harvested metals. Ores form through many geological processes, including volcanism, underground water flow, and rainfall that alters soils. For example, iron ore deposits generally form when water mobilizes reduced iron ( $\text{Fe}^{2+}$ ) and the iron oxidizes, forming  $\text{Fe}^{3+}$ , which is insoluble in water and precipitates. This process deposited vast quantities of iron ore on ancient seafloors, yielding what are called banded iron formations. Iron ores also form when underground water percolates through porous rocks such as sandstones, forming ironstones.

The first step in making steel is to extract iron from ore in a process called smelting, a chemical reduction process whose basic steps are several thousand years old. Pulverized ore (taconite) is heated by burning coke (a coal derivative) together with limestone to remove waste materials (Fig. 22). The process yields a mixture of iron and carbon, a residue called "slag," and some emissions. Steel, which is stronger than iron and has other superior qualities such as hardness, corrosion resistance, and electrical conductivity, is made by processing the molten iron-carbon mixture to remove excess carbon while adding other elements (for example, nickel, vanadium, or molybdenum) to obtain the desired properties.



Figure 22. Iron smelting, Carrie furnaces, Rankin, Pennsylvania, 1952

© Collection of William J. Gaughan, ais 94:3, Archives Service Center, University of Pittsburgh, Pittsburgh.

Perhaps the second most important metal today is aluminum, which is light, tough, and corrosion-resistant and has high electrical conductivity. Aluminum metal is used today in many manufactured goods, including cars and planes as well as smaller consumer goods.

The primary ore of aluminum is bauxite, which forms when high volumes of rain water move through soils. Typically the water dissolves and removes elements such as sodium, potassium, and calcium, leaving altered soils called laterites that contain significant amounts of highly insoluble metals such as aluminum. Laterites are widespread in tropical environments. To mine aluminum, developers strip off the topsoil and overburden to extract ore, which can require drilling and blasting. Much like coal strip mining, aluminum mining uproots vegetation, displaces wildlife, and pollutes area lakes and rivers.

Aluminum ore is smelted through a complicated process that involves extracting aluminum oxide, then passing high-voltage electricity through it to free the aluminum metal. The process is very energy-intensive: aluminum manufacturers are some of the largest industrial consumers of electricity worldwide and many are located in regions like the Pacific Northwest, where regional electricity prices are relatively low. Aluminum production also generates large quantities of greenhouse

gases, although many major companies have formed partnerships with government to reduce these emissions.

Many other metals are important for specialized industrial and manufacturing purposes. For example, copper is used primarily as a conductor of electricity, while titanium is used as a lightweight metal alloy and for white paint pigments. Mining and smelting operations for many metals are similar to the processes involved in making iron. In some ores, such as copper, the metal is bound with sulfur, so mining and smelting these metals produces sulfuric acid and environmental impacts similar to those of acid rain (for more details, see Unit 11, "Atmospheric Pollution").

Sulfide mines and smelting operations often leave major environmental scars. For example, the Berkeley Pit, a former open-pit copper mine in Butte, Montana, is one of the most polluted water bodies in the United States (Fig. 23). The pit contains over 30 million gallons of water with a pH value of 2.5 (highly acidic) that has drained from mine tunnels and shafts feeding into it, and is laced with arsenic and heavy metals including aluminum, cadmium, copper, iron, lead, zinc, and sulfate. In 1995, 342 migrating snow geese died when they mistook the pit water for a safe resting place. Six companies have agreed to pay for cleaning up the pit at an estimated cost of \$110 million (footnote 8).



Figure 23. Berkeley Pit, Montana

© Steve Peters.

## 14. Other Material Resources

A wide variety of other material resources are extracted from geological deposits, including stone building materials, fertilizers, and industrial chemicals. Building stones, which are chosen for a combination of durability and aesthetic appeal, include marble, granite, limestone, slates, and sandstones. Aggregates such as crushed stone, gravel, and sand are also widely used for construction and to manufacture concrete and other treated rock products. Rock products are used directly as insulators, abrasives, and paint pigments and in many other applications.

Before the industrial era, farmers used manure and bone meal as fertilizers, but manufacturing fertilizer has become a major industry. As discussed in Unit 4, "Ecosystems," and Unit 7, "Agriculture," the primary elements of fertilizers are nitrogen, phosphorus, and potassium. Today potassium and phosphates are mined from ancient marine beds, while nitrogen is extracted from the air by burning natural gas to form ammonia ( $\text{NH}_3$ ).

Rocks and minerals also are used as feed stocks for a wide variety of industrial chemicals such as sulfuric acid, flame retardants, wood preservatives, glass, ceramics, paper coatings, and battery and computer chip components. Ore deposits for the raw materials to manufacture most of these products are generally plentiful, so the economics of extracting, processing, and transporting feed stocks generally dictates which materials are developed.

## 15. Increasing End-Use Efficiency of Energy and Materials

Thirty years of fluctuating world energy prices since the Arab oil embargos of the 1970s, as well as growing concerns about the environmental impacts of producing and using energy, have made it increasingly clear that "more of the same" is not a sufficient basis for national energy planning. Many of the best world oil and gas resources are being fully exploited, and resource depletion is making the remaining supplies more expensive. Some oil experts contend that world oil production will peak in the near future, although others strongly disagree and argue that technology improvements will continue to make new reserves exploitable. We can expand energy supplies by using alternative sources and technologies that usually have lower environmental impacts, but this typically costs more until new sources are fully commercialized. The same is true for some of our important material resources.

When resources become scarce, we can benefit from learning to get more services from a given flow of energy or stock of material. For example, we can design cars that are stronger but weigh less than today's models and get more miles per gallon, buildings that require less energy per square foot to heat and cool than current structures, refrigerators that require less power per cubic foot of interior volume, lights that deliver more illumination for each kilowatt-hour of electricity they consume, and manufacturing processes that use less energy to make a given array of products.



Industrialized and developing countries alike have greatly increased their energy efficiency in recent decades. The United States doubled the overall energy efficiency of its economy between 1970 and 2005. That means that this country was extracting twice as much real gross domestic product from each unit of energy flowing through the economy in 2005 as it did 35 years earlier. There are many areas in which we can continue and accelerate these trends by using energy and materials even more efficiently.

Investments into research and development of end-use efficiency challenges often pay for themselves many times over in resulting savings. Figure 24 shows representative estimated savings from energy efficiency upgrades in a home located in an average U.S. climate and equipped with standard appliances.

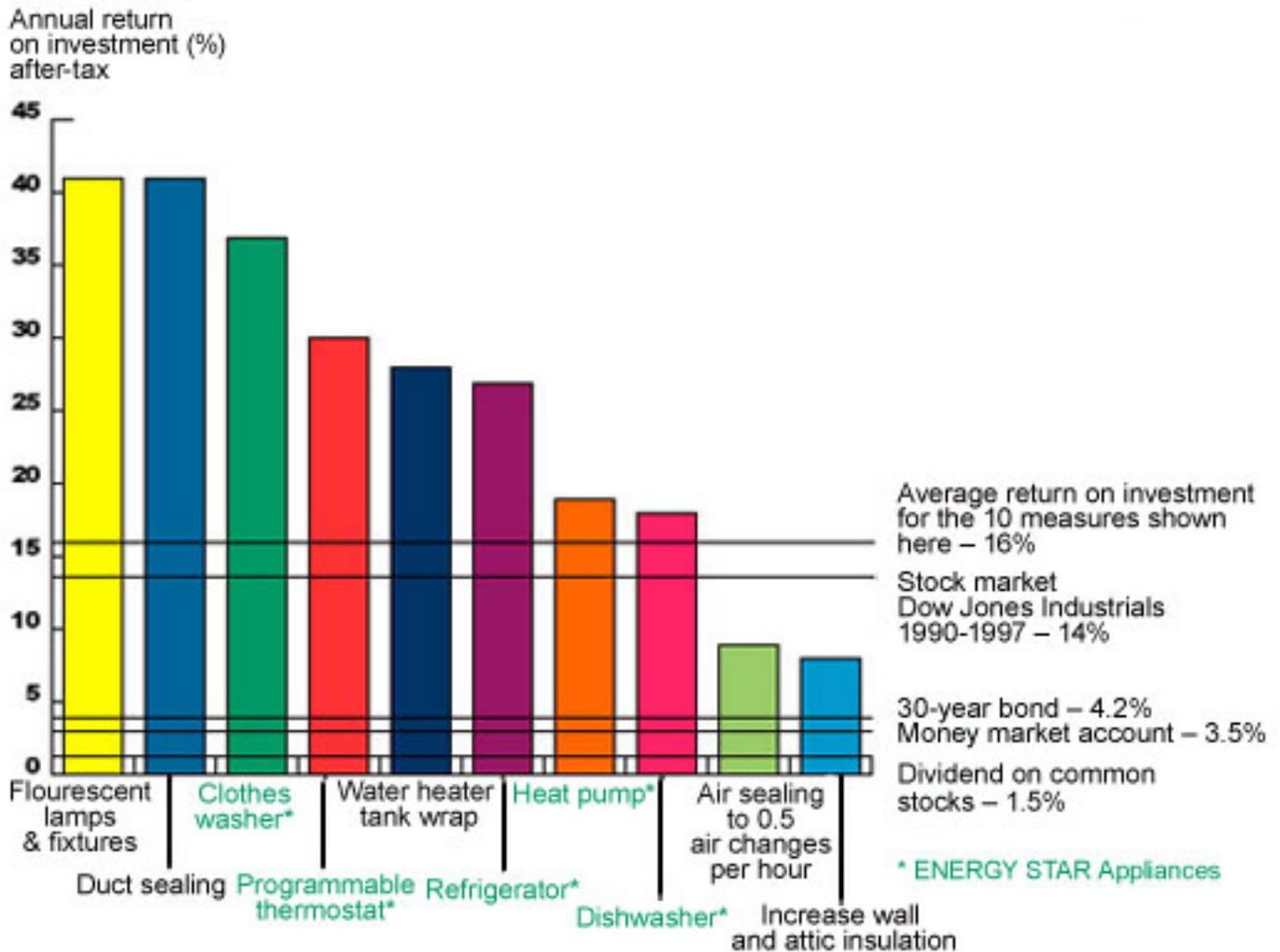


Figure 24. Profitability of energy efficiency upgrades

© Lawrence Berkeley National Laboratory. Environmental Energy Technologies Division.

When a refrigerator saves a kilowatt-hour of electricity or an efficient car saves a liter of fuel, that energy is available for use elsewhere in the economy. This means that improving end-use efficiency is like finding a new supply of energy. It is often cheaper, faster, and cleaner to reap gains from end-use efficiency (sometimes referred to as "negawatts," to connote energy that does not have to be produced) than to expand energy supply through exploration and drilling. Similarly, investing in



recycling programs, better product design, and longer product lifetimes, we can reduce our need for newly-mined minerals.

## 16. Further Reading

"Fuel Cells: Collecting History with the Worldwide Web," Smithsonian Institution, <http://americanhistory.si.edu/fuelcells/intro.htm>. An online project documenting the development of fuel cells, with descriptions and diagrams of many fuel cell technologies.

National Research Council and National Academy of Engineering, **The Hydrogen Economy: Opportunities, Costs, Barriers, and R&D Needs** (Washington, DC: National Academies Press, 2001). A broad assessment of issues involved in making a 50-year transition to a hydrogen economy in the United States.

Michelle Nijhuis, "Selling the Wind," **Audubon**, September/October 2006. Expanding wind power is a balancing act between energy concerns and ecosystem protection.

## Footnotes

1. "The End of the Oil Age," **The Economist**, October 23, 2003.
2. U.S. Energy Information Administration, Emissions of Greenhouse Gases in the United States 2002, table 6-1, <http://www.eia.gov/oiaf/1605/ggrpt/pdf/tab6.1.pdf>.
3. Michael Wang, "The Debate On Energy and Greenhouse Gas Emissions Impacts of Fuel Ethanol," Argonne National Laboratory, August 3, 2005, <http://www.transportation.anl.gov/pdfs/TA/347.pdf>.
4. For details, see the Department of Energy's biomass program web page at [http://www1.eere.energy.gov/biomass/sugar\\_platform.html](http://www1.eere.energy.gov/biomass/sugar_platform.html).
5. U.S. Department of Energy, "Large-Scale Gasification," [http://www1.eere.energy.gov/biomass/large\\_scale\\_gasification.html](http://www1.eere.energy.gov/biomass/large_scale_gasification.html).
6. World Commission on Dams, **Dams and Development: A New Framework for Decision-Making**, chapter 3, pp. 73–85 (London: Earthscan, November 2000), <http://www.dams.org/report/>.
7. For details, see U.S. Department of Energy, [http://www.eere.energy.gov/consumer/renewable\\_energy/ocean/index.cfm/mytopic=50010](http://www.eere.energy.gov/consumer/renewable_energy/ocean/index.cfm/mytopic=50010).
8. U.S. Department of Justice, "United States and Montana Reach Agreement With Mining Companies To Clean Up Berkeley Pit," press release, March 25, 2002, [http://www.justice.gov/opa/pr/2002/March/02\\_enrd\\_180.htm](http://www.justice.gov/opa/pr/2002/March/02_enrd_180.htm).

## Glossary

**biodiesel** : A diesel-equivalent, processed fuel derived from biological sources (such as vegetable oils), that can be used in unmodified diesel-engine vehicles.

**British thermal unit (Btu)** : A unit of heat. One Btu is the energy required to raise one pound of water by one degree Fahrenheit at a constant pressure of one atmosphere.

**clathrates** : A chemical substance consisting of a lattice of one type of molecule trapping and containing a second type of molecule.

**ethanol** : A flammable, colorless, slightly toxic chemical compound with a distinctive perfume-like odor. Also known as ethyl alcohol, drinking alcohol, or grain alcohol, in common usage it is often referred to simply as alcohol.

**fast breeder** : A fast neutron reactor designed to breed fuel by producing more fissile material than it consumes.

**fissile** : Capable of sustaining a chain reaction of nuclear fission.

**flow** : The rate at which new material is added to or removed from the stock.

**fuel cells** : An electrochemical energy conversion device that produces electricity from external supplies of fuel (on the anode side) and oxidant (on the cathode side). Fuel cells differ from batteries in that they consume reactant, which must be replenished, while batteries store electrical energy chemically in a closed system.

**gasification** : A process that converts carbonaceous materials, such as coal, petroleum, petroleum coke or biomass, into carbon monoxide and hydrogen.

**geothermal gradient** : The rate of increase in temperature per unit depth in the Earth.

**hydrates** : Compounds formed by the union of water with some other substance.

**hydrocarbons** : Chemical compounds containing carbon and hydrogen as the principal elements. Oil is composed primarily of hydrocarbons.

**inflow** : General term designating the water or other fluid entering a system.

**methane hydrates** : Natural formations consisting of mounds of icelike material on or just below the sea floor containing large amounts of methane trapped within a lattice of icelike crystals.

**non-renewable resource** : A natural resource such as coal or mineral ores that is not replaceable after its removal.

**oil shale** : A general term applied to a fine-grained sedimentary rock containing enough organic material (called kerogen) to yield oil and combustible gas upon distillation.



**ore** : A mineral or an aggregate of minerals from which a valuable constituent, especially a metal, can be profitably mined or extracted.

**overburden** : The rock and dirt that overlie a mineral deposit and that must be removed before the mineral deposit can be extracted by surface mining.

**photovoltaic** : Producing an electric current as the result of light striking a metal; the direct conversion of radiant energy into electrical energy.

**renewable resources** : Supplies of biological organisms that can be replaced after harvesting by regrowth or reproduction of the removed species, such as seafood or timber.

**source rock** : A rock rich in organic matter which, if heated sufficiently, will generate oil or gas.

**stock** : In ecological cycles and models, the amount of a material in a certain medium or reservoir.

**tar sands** : Sandy deposits containing bitumen, a viscous petroleum-like material that has a high sulfur content.

**thermal maturation** : A process in which as rock layers press down on buried deposits, geothermal energy heats the peat and reduces its oxygen and hydrogen content, converting it to coal.

**uranium enrichment** : A process that results in an increase in the amount of the fissionable isotope of uranium in a given mass of uranium. Used mostly for nuclear weapons, naval propulsion, and smaller quantities for research reactors.

## Unit 11 : Atmospheric Pollution



Air pollution. Source: National Park Service

### Overview

Many forms of atmospheric pollution affect human health and the environment at levels from local to global. These contaminants are emitted from diverse sources, and some of them react together to form new compounds in the air. Industrialized nations have made important progress toward controlling some pollutants in recent decades, but air quality is much worse in many developing countries, and global circulation patterns can transport some types of pollution rapidly around the world. In this unit, discover the basic chemistry of atmospheric pollution and learn which human activities have the greatest impacts on air quality.

### Sections:

1. Introduction
2. Chemicals in Motion
3. Primary Air Pollutants
4. Secondary Air Pollutants
5. Aerosols
6. Smog
7. Acid Deposition
8. Mercury Deposition
9. Controlling Air Pollution
10. Stratospheric Ozone
11. Air Pollution, Greenhouse Gases, and Climate Change
12. Major Laws and Treaties
13. Further Reading

## 1. Introduction

The first week of December 1952 was unusually cold in London, so residents burned large quantities of coal in their fireplaces to keep warm. Early on December 5, moisture in the air began condensing into fog near the ground. The fog mixed with smoke from domestic fires and emissions from factories and diesel-powered buses. Normally the fog would have risen higher in the atmosphere and dispersed, but cold air kept it trapped near the ground. Over the next four days, the smog became so thick and dense that many parts of London were brought to a standstill.

Public officials did not realize that the Great Smog was the most deadly air pollution event on record until mortality figures were published several weeks afterward. Some 4,000 people died in London between December 5-9 of illnesses linked to respiratory problems such as bronchitis and pneumonia, and the smog's effects caused another 8,000 deaths over the next several months. Samples showed that victims' lungs contained high levels of very fine particles, including carbon material and heavy metals such as lead, zinc, tin, and iron.

Air pollution was not news in 1952—London's air had been famously smoky for centuries—but the Great Smog showed that it could be deadly. The event spurred some of the first governmental actions to reduce emissions from fuel combustion, industrial operations, and other manmade sources.

Over the past half-century, scientists have learned much more about the causes and impacts of atmospheric pollution. Many nations have greatly reduced their emissions, but the problem is far from solved. In addition to threatening human health, air pollutants damage ecosystems, weaken Earth's stratospheric ozone shield, and contribute to global climate change (Fig. 1).



Figure 1. Air pollution on the Autobahn in Germany, 2005

Courtesy Zakysant. Wikimedia Commons, GNU Free Documentation License.

Understanding of pollutants is still evolving, but we have learned enough to develop emission control policies that can limit their harmful effects. Some major pollutants contribute to both air pollution and global climate change, so reducing these emissions has the potential to deliver significant benefits. To integrate air pollution and climate change strategies effectively, policy makers need extensive information about key pollutants and their interactions.

This unit describes the most important types of pollutants affecting air quality and the environment. It also summarizes some widely used technical and policy options for controlling atmospheric pollution, and briefly describes important laws and treaties that regulate air emissions. For background on the structure and composition of the atmosphere and atmospheric circulation patterns, see Unit 2, "Atmosphere"; for more on global climate change, especially the role of CO<sub>2</sub>, see Unit 2 and Unit 12, "Earth's Changing Climate."

## 2. Chemicals in Motion

The science of air pollution centers on measuring, tracking, and predicting concentrations of key chemicals in the atmosphere. Four types of processes affect air pollution levels (Fig. 2):



- **Emissions.** Chemicals are emitted to the atmosphere by a range of sources. Anthropogenic emissions come from human activities, such as burning fossil fuel. Biogenic emissions are produced by natural functions of biological organisms, such as microbial breakdown of organic materials. Emissions can also come from nonliving natural sources, most notably volcanic eruptions and desert dust.
- **Chemistry.** Many types of chemical reactions in the atmosphere create, modify, and destroy chemical pollutants. These processes are discussed in the following sections.
- **Transport.** Winds can carry pollutants far from their sources, so that emissions in one region cause environmental impacts far away. Long-range transport complicates efforts to control air pollution because it can be hard to distinguish effects caused by local versus distant sources and to determine who should bear the costs of reducing emissions.
- **Deposition.** Materials in the atmosphere return to Earth, either because they are directly absorbed or taken up in a chemical reaction (such as photosynthesis) or because they are scavenged from the atmosphere and carried to Earth by rain, snow, or fog.

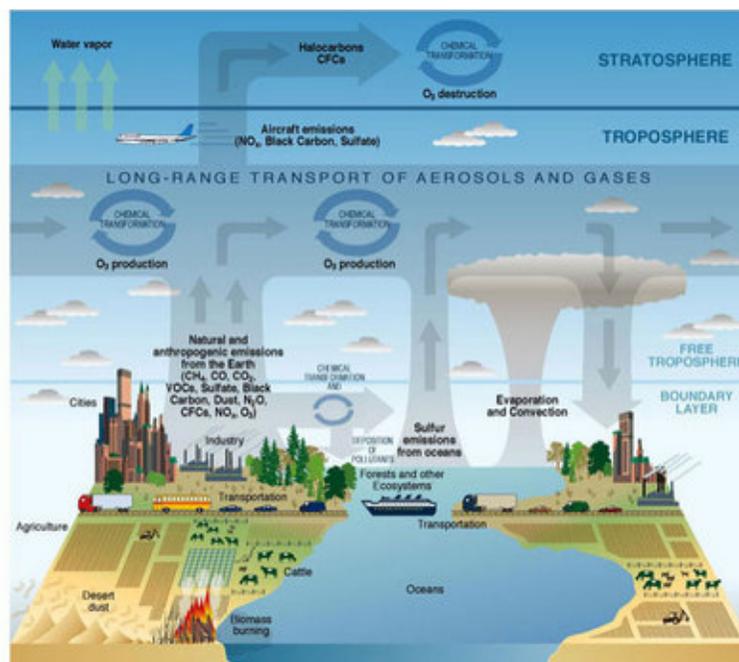


Figure 2. Processes related to atmospheric composition

Courtesy United States Climate Change Science Program (Illustrated by P. Rekacewicz).

Air pollution trends are strongly affected by atmospheric conditions such as temperature, pressure, and humidity, and by global circulation patterns. For example, winds carry some pollutants far from their sources across national boundaries and even across the oceans. Transport is fastest along east-west routes: longitudinal winds can move air around the globe in a few weeks, compared to months or longer for air exchanges from north to south (for more details see Unit 2, "Atmosphere").

Local weather patterns also interact with and affect air pollution. Rain and snow carry atmospheric pollutants to Earth. Temperature inversions, like the conditions that caused London's Great Smog in 1952, occur when air near the Earth's surface is colder than air aloft. Cold air is heavier than warm air, so temperature inversions limit vertical mixing and trap pollutants near Earth's surface. Such conditions are often found at night and during the winter months. Stagnation events characterized by weak winds are frequent during summer and can lead to accumulation of pollutants over several days.

To see the close connections between weather, climate, and air pollution, consider Los Angeles, whose severe air quality problems stem partly from its physical setting and weather patterns. Los

Angeles sits in a bowl, ringed by mountains to the north and east that trap pollutants in the urban basin. In warm weather, cool sea breezes are drawn onshore at ground level, creating temperature inversions that prevent pollutants from rising and dissipating. The region's diverse manufacturing and industrial emitters and millions of cars and trucks produce copious primary air pollutants that mix in its air space to form photochemical smog (Fig. 3).



Figure 3. Smog over Los Angeles

Courtesy United States Environmental Protection Agency.

Scientists can measure air pollutants directly when they are emitted—for example, by placing instruments on factory smokestacks—or as concentrations in the **ambient** outdoor air. To track ambient concentrations, researchers create networks of air-monitoring stations, which can be ground-based or mounted on vehicles, balloons, airplanes, or satellites.

In the laboratory, scientists use tools including laser spectrometers and electron microscopes to identify specific pollutants. They measure chemical reaction rates in clear plastic bags ("smog chambers") that replicate the smog environment under controlled conditions, and observe emission of pollutants from combustion and other sources.

Knowledge of pollutant emissions, chemistry, and transport can be incorporated into computer simulations ("air quality models") to predict how specific actions, such as requiring new vehicle emission controls or cleaner-burning fuels, will benefit ambient air quality. However, air pollutants pass through many complex reactions in the atmosphere and their residence times vary widely, so it

is not always straightforward to estimate how emission reductions from specific sources will impact air quality over time.

### 3. Primary Air Pollutants

**Primary air pollutants** are emitted directly into the air from sources. They can have effects both directly and as precursors of **secondary air pollutants** (chemicals formed through reactions in the atmosphere), which are discussed in the following section.

**Sulfur dioxide** ( $\text{SO}_2$ ) is a gas formed when sulfur is exposed to oxygen at high temperatures during fossil fuel combustion, oil refining, or metal smelting.  $\text{SO}_2$  is toxic at high concentrations, but its principal air pollution effects are associated with the formation of acid rain and aerosols.  $\text{SO}_2$  dissolves in cloud droplets and oxidizes to form sulfuric acid ( $\text{H}_2\text{SO}_4$ ), which can fall to Earth as acid rain or snow or form sulfate aerosol particles in the atmosphere. Associated impacts are discussed below in Section 5, "Aerosols," and Section 7, "Acid Deposition."

**Nitrogen oxides** ( $\text{NO}$  and  $\text{NO}_2$ , referred together as  $\text{NO}_x$ ) are highly reactive gases formed when oxygen and nitrogen react at high temperatures during combustion or lightning strikes. Nitrogen present in fuel can also be emitted as  $\text{NO}_x$  during combustion. Emissions are dominated by fossil fuel combustion at northern mid-latitudes and by biomass burning in the tropics. Figure 4 shows the distribution of  $\text{NO}_x$  emissions to the atmosphere in 2006 as determined by satellite measurements of atmospheric  $\text{NO}_2$  concentrations.

In the atmosphere  $\text{NO}_x$  reacts with volatile organic compounds (VOCs) and carbon monoxide to produce ground-level ozone through a complicated chain reaction mechanism. It is eventually oxidized to nitric acid ( $\text{HNO}_3$ ). Like sulfuric acid, nitric acid contributes to acid deposition and to aerosol formation.

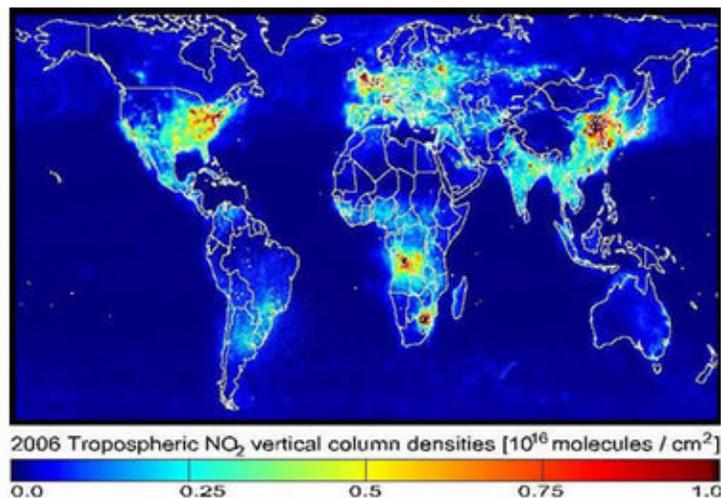


Figure 4. Satellite observations of tropospheric NO<sub>2</sub>, 2006

Courtesy Jim Gleason, USA and Pepijn Veefkind, KNMI, National Aeronautics and Space Administration.

**Carbon monoxide** (CO) is an odorless, colorless gas formed by incomplete combustion of carbon in fuel. The main source is motor vehicle exhaust, along with industrial processes and biomass burning. Carbon monoxide binds to hemoglobin in red blood cells, reducing their ability to transport and release oxygen throughout the body. Low exposures can aggravate cardiac ailments, while high exposures cause central nervous system impairment or death. It also plays a role in the generation of ground-level ozone, discussed below in Section 4.

**Volatile organic compounds** (VOCs), including hydrocarbons (C<sub>x</sub>H<sub>y</sub>) but also other organic chemicals are emitted from a very wide range of sources, including fossil fuel combustion, industrial activities, and natural emissions from vegetation and fires. Some anthropogenic VOCs such as benzene are known carcinogens.

VOCs are also of interest as chemical precursors of ground-level ozone and aerosols, as discussed below in Sections 4 and 5. The importance of VOCs as precursors depends on their chemical structure and atmospheric lifetime, which can vary considerably from compound to compound. Large VOCs oxidize in the atmosphere to produce nonvolatile chemicals that condense to form aerosols. Short-lived VOCs interact with NO<sub>x</sub> to produce high ground-level ozone in polluted environments. Methane (CH<sub>4</sub>), the simplest and most long-lived VOC, is of importance both as a greenhouse gas (Section 11) and as a source of background tropospheric ozone. Major anthropogenic sources of methane include natural gas production and use, coal mining, livestock, and rice paddies.

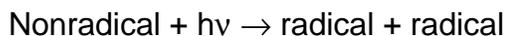
## 4. Secondary Air Pollutants

Secondary pollutants form when primary pollutants react in the atmosphere. Table 1 summarizes common forms of atmospheric reactions.

**Table 1.** Chemical kinetics in the atmosphere.

Type of reaction	Process	Notation
Bimolecular	Two reactants combine to produce two products.	$A + B \rightarrow C + D$
Three-body	Two reactants combine to form one new product. A third, inert molecule (M) stabilizes the end product and removes excess energy.	$A + B + M \rightarrow AB + M$
Photolysis	Solar radiation photon breaks a chemical bond in a molecule	$A + h\nu \rightarrow B + C$
Thermal decomposition	A molecule decomposes by collision with an inert molecule (M)	$A + M \rightarrow B + C$

For reactions to take place, molecules have to collide. However, gases are present in the atmosphere at considerably lower concentrations than are typical for laboratory experiments or industrial processes, so molecules collide fairly infrequently. As a result, most atmospheric reactions that occur at significant rates involve at least one **radical**—a molecule with an odd number of electrons and hence an unpaired electron in its outer shell. The unpaired electron makes the radical unstable and highly reactive with other molecules. Radicals are formed when stable molecules are broken apart, a process that requires large amounts of energy. This can take place in combustion chambers due to high temperatures, and in the atmosphere by **photolysis**:



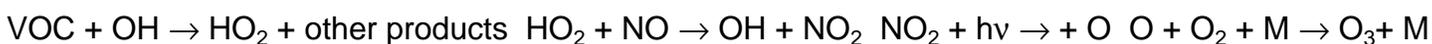
Radical formation initiates reaction chains that continue until radicals combine with other radicals to produce nonradicals (atoms with an even number of electrons). Radical-assisted chain reactions in the atmosphere are often referred to as photochemical mechanisms because sunlight plays a key role in launching them.

One of the most important radicals in atmospheric chemistry is the **hydroxyl radical** (OH), sometimes referred to as the atmospheric cleanser. OH is produced mainly through photolysis reactions that break apart tropospheric ozone, and is very short-lived. It is consumed within about one second by oxidizing a number of trace gases like carbon monoxide, methane, and nonmethane VOCs

(NMVOCs). Some of these reactions eventually regenerate OH in continuous cycles, while others deplete it.

Since OH has a short atmospheric lifetime, its concentration can vary widely. Some anthropogenic emissions, such as carbon monoxide and VOCs, deplete OH, while others such as NO<sub>x</sub> boost OH levels. Measuring atmospheric OH is difficult because its concentration is so low. Long-term trends in OH concentrations are uncertain, although the prevailing view is that trends over the past decades have been weak because of compensating influences from carbon monoxide and VOCs on the one hand and NO<sub>x</sub> on the other hand. Since OH affects the rates at which some pollutants are formed and others are destroyed, changes in OH levels over the long term would have serious implications for air quality.

Ground-level **ozone** (O<sub>3</sub>) is a pernicious secondary air pollutant, toxic to both humans and vegetation (Fig. 5). It is formed in surface air (and more generally in the troposphere) by oxidation of VOCs and carbon monoxide in the presence of NO<sub>x</sub>. The mechanism is complicated, involving hundreds of chemically interactive species to describe the VOC degradation pathways. A simple schematic is:



An important aspect of this mechanism is that NO<sub>x</sub> and OH act as catalysts—that is, they speed up the rate of ozone generation without being consumed themselves. Instead they cycle rapidly between NO and NO<sub>2</sub>, and between OH and HO<sub>2</sub>.



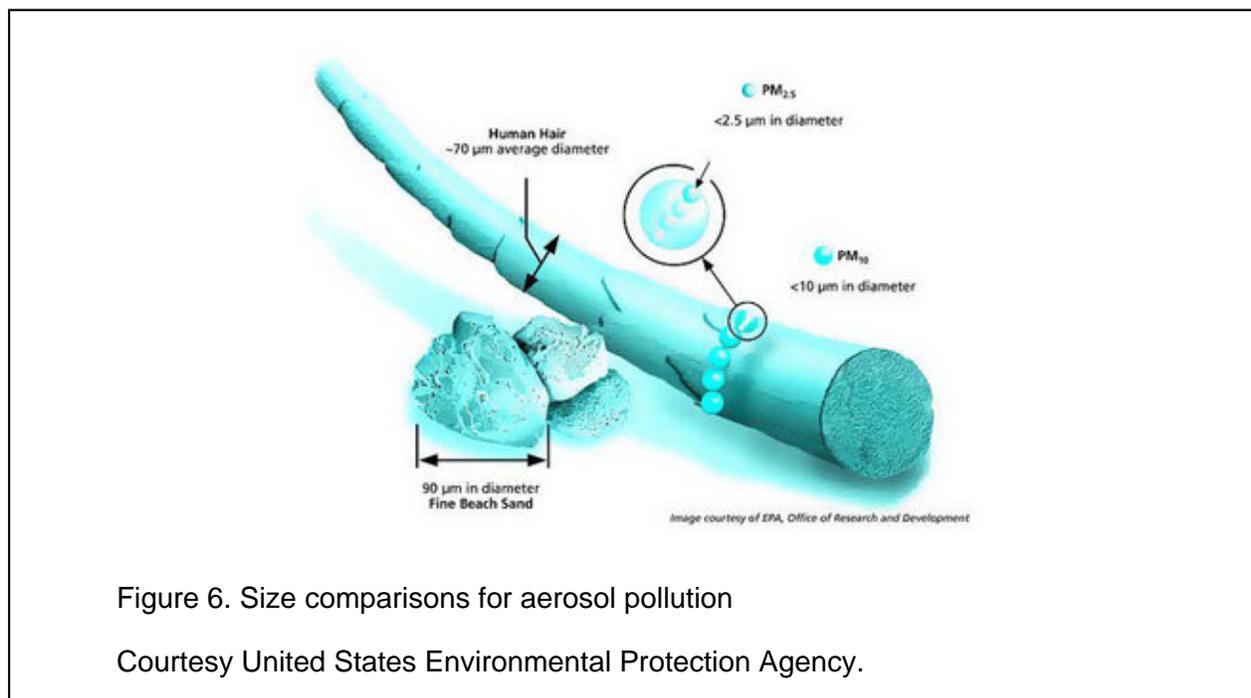
Figure 5. Ozone damage to plant leaves

Courtesy United States Environmental Protection Agency.

This formation mechanism for ozone at ground level is totally different from that for ozone formation in the stratosphere, where 90 percent of total atmospheric ozone resides and plays a critical role in protecting life on Earth by providing a UV shield (for details see Unit 2, "Atmosphere"). In the stratosphere ozone is produced from photolysis of oxygen ( $O_2 + h\nu \rightarrow O + O$ , followed by  $O + O_2 + M \rightarrow O_3 + M$ ). This process does not take place in the troposphere because the strong (< 240 nm) UV photons needed to dissociate molecular oxygen are depleted by the ozone overhead.

## 5. Aerosols

In addition to gases, the atmosphere contains solid and liquid particles that are suspended in the air. These particles are referred to as **aerosols** or **particulate matter (PM)**. Aerosols in the atmosphere typically measure between 0.01 and 10 micrometers in diameter, a fraction of the width of a human hair (Fig. 6). Most aerosols are found in the lower troposphere, where they have a residence time of a few days. They are removed when rain or snow carries them out of the atmosphere or when larger particles settle out of suspension due to gravity.



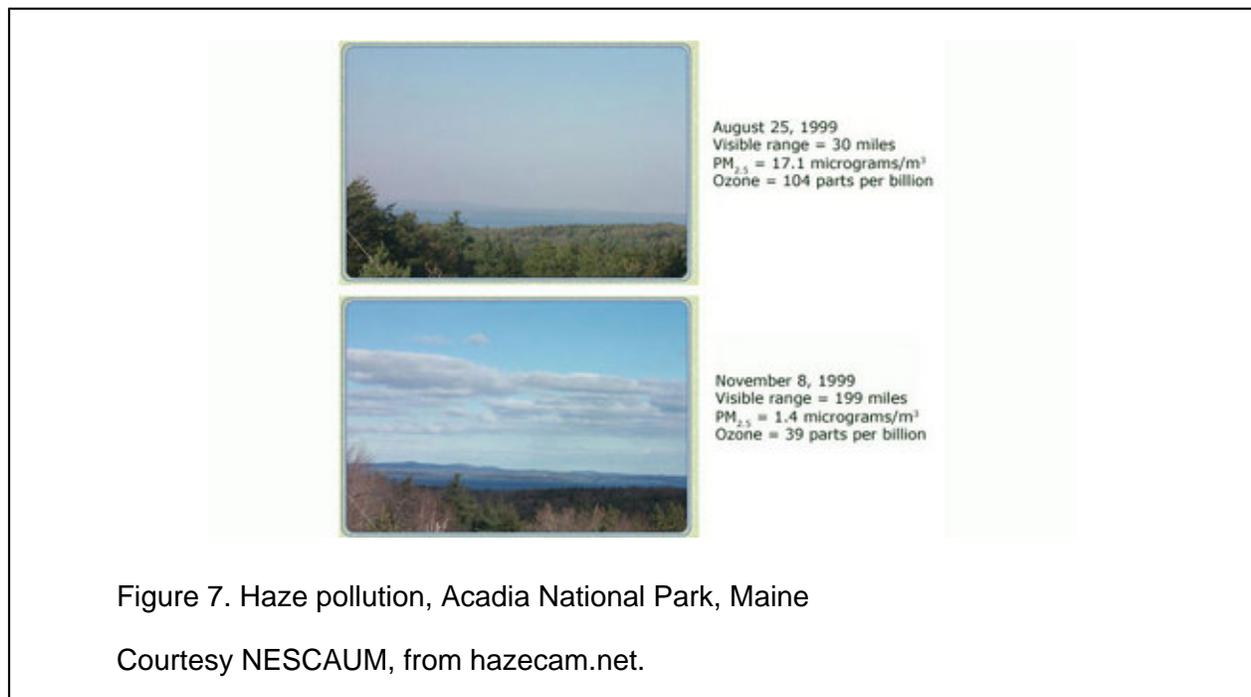
Large aerosol particles (usually 1 to 10 micrometers in diameter) are generated when winds blow sea salt, dust, and other debris into the atmosphere. Fine aerosol particles with diameters less than 1 micrometer are mainly produced when precursor gases condense in the atmosphere. Major components of fine aerosols are sulfate, nitrate, organic carbon, and elemental carbon. Sulfate,



nitrate, and organic carbon particles are produced by atmospheric oxidation of  $\text{SO}_2$ ,  $\text{NO}_x$ , and VOCs as discussed above in Section 3. Elemental carbon particles are emitted by combustion, which is also a major source of organic carbon particles. Light-absorbing carbon particles emitted by combustion are called black carbon or soot; they are important agents for climate change and are also suspected to be particularly hazardous for human health.

High concentrations of aerosols are a major cause of cardiovascular disease and are also suspected to cause cancer. Fine particles are especially serious threats because they are small enough to be absorbed deeply into the lungs, and sometimes even into the bloodstream. Scientific research into the negative health effects of fine particulate air pollution spurred the U.S. Environmental Protection Agency to set limits in 1987 for exposure to particles with a diameter of 10 micrometers or less, and in 1997 for particles with a diameter of 2.5 micrometers or less.

Aerosols also have important radiative effects in the atmosphere. Particles are said to scatter light when they alter the direction of radiation beams without absorbing radiation. This is the principal mechanism limiting visibility in the atmosphere, as it prevents us from distinguishing an object from the background. Air molecules are inefficient scatterers because their sizes are orders of magnitude smaller than the wavelengths of visible radiation (0.4 to 0.7 micrometers). Aerosol particles, by contrast, are efficient scatterers. When relative humidity is high, aerosols absorb water, which causes them to swell and increases their cross-sectional area for scattering, creating haze. Without aerosol pollution our visual range would typically be about 200 miles, but haze can reduce visibility significantly. Figure 7 shows two contrasting views of Acadia National Park in Maine on relatively good and bad air days.



Aerosols have a cooling effect on Earth's climate when they scatter solar radiation because some of the scattered light is reflected back into space. As discussed in Unit 12, "Earth's Changing Climate," major volcanic eruptions that inject large quantities of aerosols into the stratosphere, such as that of Mt. Pinatubo in 1991, can noticeably reduce average global surface temperatures for some time afterward.

In contrast, some aerosol particles such as soot absorb radiation and have a warming effect. This means that estimating the net direct contribution to global climate change from aerosols requires detailed inventories of the types of aerosols in the atmosphere and their distribution around the globe. Aerosol particles also influence Earth's climate indirectly: they serve as condensation nuclei for cloud droplets, increasing the amount of radiation reflected back into space by clouds and modifying the ability of clouds to precipitate. The latter is the idea behind "cloud seeding" in desert areas, where specific kinds of mineral aerosol particles that promote ice formation are injected into a cloud to make it precipitate.

Aerosol concentrations vary widely around the Earth (Fig. 8). Measurements are tricky because the particles are difficult to collect without modifying their composition. Combined optical and mass spectrometry techniques that analyze the composition of single particles directly in an air flow, rather than recovering a bulk composition from filters, have improved scientists' ability to detect and characterize aerosols (footnote 1).

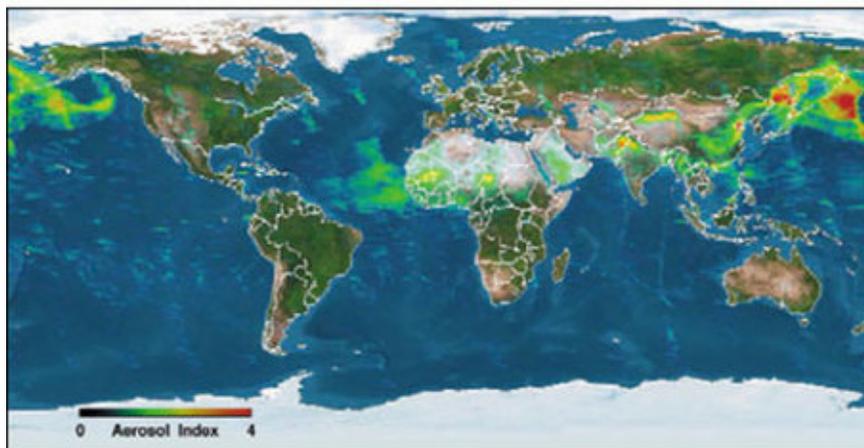


Figure 8. Total ozone mapping spectrometer (TOMS) aerosol index of smoke and dust absorption, 2004

Courtesy Jay Herman, NASA Goddard Space Flight Center.

One important research challenge is learning more about organic aerosols, which typically account for a third to half of total aerosol mass. These include many types of carbon compounds with diverse properties and environmental impacts. Organic aerosols are emitted to the atmosphere directly by inefficient combustion. Automobiles, wood stoves, agricultural fires, and wildfires are major sources in the United States. Atmospheric oxidation of VOCs, both anthropogenic and biogenic, is another major source in summer. The relative importance of these different sources is still highly uncertain, which presently limits our ability to assess anthropogenic influence and develop strategies for reducing concentrations.

## 6. Smog

**Smog** is often used as a generic term for any kind of air pollution that reduces visibility, especially in urban areas. However, it is useful to distinguish two broad types: industrial smog and photochemical smog.

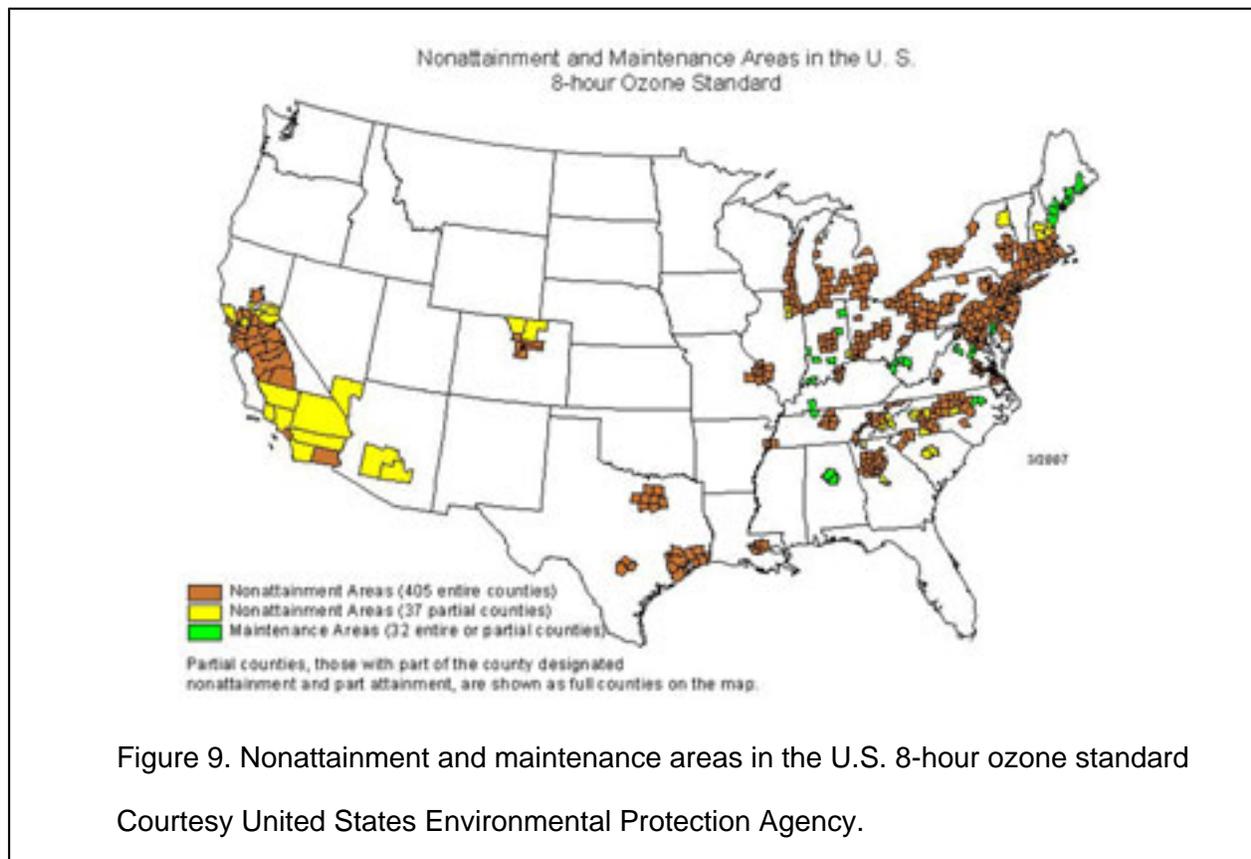
Events like the London smog of 1952 are often referred to as industrial smog because  $\text{SO}_2$  emissions from burning coal play a key role. Typically, industrial smog—also called gray or black smog—develops under cold and humid conditions. Cold temperatures are often associated with inversions that trap the pollution near the surface (see Section 2, "Chemicals in Motion," above). High humidity



allows for rapid oxidation of  $\text{SO}_2$  to form sulfuric acid and sulfate particles. Events similar to the 1952 London smog occurred in the industrial towns of Liege, Belgium, in 1930, killing more than 60 people, and Donora, Pennsylvania, in 1948, killing 20. Today coal combustion is a major contributor to urban air pollution in China, especially from emissions of  $\text{SO}_2$  and aerosols (footnote 2).

Air pollution regulations in developed countries have reduced industrial smog events, but photochemical smog remains a persistent problem, largely driven by vehicle emissions. Photochemical smog forms when  $\text{NO}_x$  and VOCs react in the presence of solar radiation to form ozone. The solar radiation also promotes formation of secondary aerosol particles from oxidation of  $\text{NO}_x$ , VOCs, and  $\text{SO}_2$ . Photochemical smog typically develops in summer (when solar radiation is strongest) in stagnant conditions promoted by temperature inversions and weak winds. Photochemical smog is a ubiquitous urban problem in the developed world and often blankets large populated regions such as the eastern United States and western Europe for extended periods in summer.

Ozone and aerosols are the two main health hazards of photochemical smog. Ozone is invisible, but aerosol particles scatter sunlight as discussed above in Section 5, and are responsible for the whitish haze associated with smog. Because ozone is created in the atmosphere, concentrations are often higher downwind of urban areas than in the urban areas themselves. Figure 9 shows counties in the United States that currently fail to comply with the national standard for ozone levels over an 8-hour period ([nonattainment areas](#)). These cover much of California and the eastern United States on a regional scale.



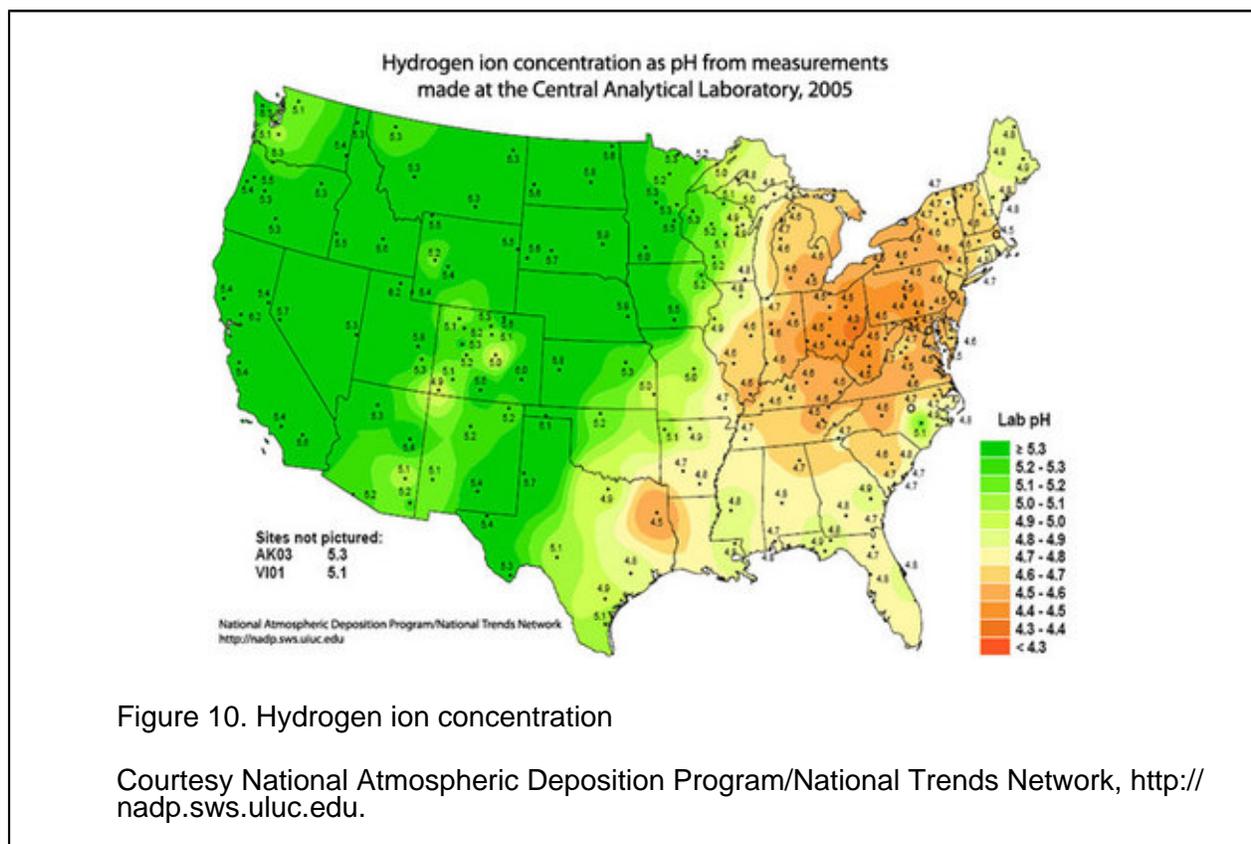
## 7. Acid Deposition

**Acid rain** was first identified in the 19<sup>th</sup> century, when English pharmacist Robert Angus Smith measured high acidity levels in rain falling over industrial regions of England and much lower levels in less-polluted areas near the coast. However, this pattern did not receive sustained attention until biologists began to notice sharp declines in fish populations in lakes in Norway, the northeastern United States, and Canada in the 1950s and 1960s. In each case researchers found that acid precipitation was altering lake chemistry. These findings spurred research into the causes of acid rain.

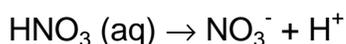
Pure water has a pH value of 7 (neutral), but rainwater falling in the atmosphere always contains impurities. The atmosphere contains natural acids including CO<sub>2</sub> (a weak acid); nitric acid produced naturally from NO<sub>x</sub> emitted by lightning, fires, and soils; and sulfuric acid produced by the oxidation of sulfur gases from volcanoes and the biosphere. It also contains natural bases, including ammonia (NH<sub>3</sub>) emitted by the biosphere and calcium carbonate (CaCO<sub>3</sub>) from suspended soil dust. CO<sub>2</sub>

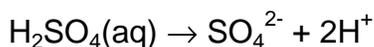
alone at natural levels (280 parts per million volume) would result in a rain pH of 5.7. Taken together, natural contaminants produce natural rain with pH values ranging from about 5 to 7 (recall that the pH scale is logarithmic, so one pH unit represents a factor of 10 difference in acid  $H^+$  concentration).

Acid rain refers to precipitation with pH values below 5, which generally happens only when large amounts of manmade pollution are added to the atmosphere. As Figure 10 shows, acid deposition takes place throughout the eastern United States and is particularly severe in the industrial Midwest due to its concentration of coal-burning power plants. Tall power plant stacks built to protect local air quality inject  $SO_2$  and  $NO_x$  at high altitude where winds are strong, allowing acid rain to extend more than a thousand miles downwind and into Canada.



The main components of acid rain worldwide are sulfuric acid and nitric acid. As discussed above in Section 3, these acids form when  $SO_2$  and  $NO_x$  are oxidized in the atmosphere. Sulfuric and nitric acids dissolve in cloudwater and dissociate to release  $H^+$ :





Human activity also releases large amounts of ammonia to the atmosphere, mainly from agriculture, and this ammonia can act as a base in the atmosphere to neutralize acid rain by converting  $\text{H}^+$  to the ammonium ion ( $\text{NH}_4^+$ ). However, the benefit of this neutralization is illusory because  $\text{NH}_4^+$  releases its  $\text{H}^+$  once it is deposited and consumed by the biosphere. The relatively high pH of precipitation in the western United States is due in part to ammonia from agriculture and in part to suspended calcium carbonate (limestone) dust.

Acid rain has little effect on the environment in most of the world because it is quickly neutralized by naturally present bases after it falls. For example, the ocean contains a large supply of carbonate ions ( $\text{CO}_3^{2-}$ ), and many land regions have alkaline soils and rocks such as limestone. But in areas with little neutralizing capacity acid rain causes serious damage to plants, soils, streams, and lakes. In North America, the northeastern United States and eastern Canada are especially sensitive to acid rain because they have thin soils and granitic bedrock, which cannot neutralize acidity.

High acidity in lakes and rivers corrodes fishes' organic gill material and attacks their calcium carbonate skeletons. Figure 11 shows the acidity levels at which common freshwater organisms can live and reproduce successfully. Acid deposition also dissolves toxic metals such as aluminum in soil sediments, which can poison plants and animals that take the metals up. And acid rain increases leaching of nutrients from forest soils, which weakens plants and reduces their ability to weather other stresses such as droughts, air pollution, or bug infestation.

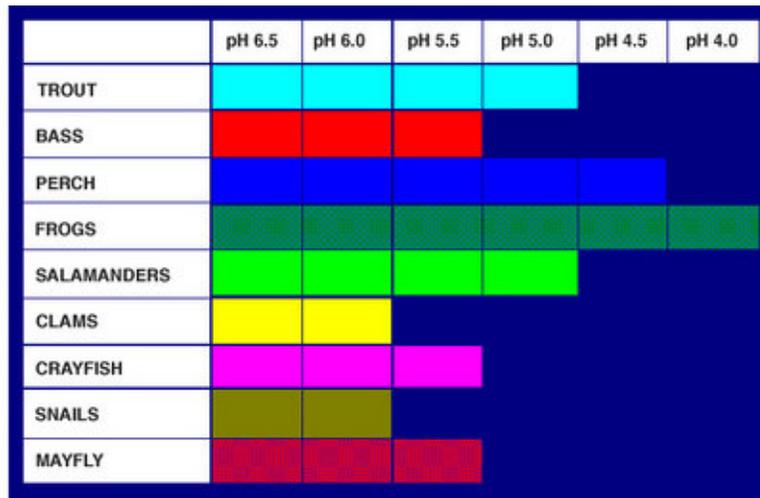


Figure 11. Acid tolerance ranges of common freshwater organisms

Courtesy United States Environmental Protection Agency.

In addition to making ecosystems more acidic, deposition of nitrate and ammonia fertilizes ecosystems by providing nitrogen, which can be directly taken up by living organisms. Nitrogen pollution in rivers and streams is carried to the sea, where it contributes to algal blooms that deplete dissolved oxygen in coastal waters. As discussed in Unit 8, "Water Resources," nutrient overloading has created dead zones in coastal regions around the globe, such as the Gulf of Mexico and the Chesapeake Bay. The main sources of nutrient pollution are agricultural runoff and atmospheric deposition.

Acid rain levels have decreased and acid rain impacts have stabilized in the United States since SO<sub>2</sub> and NO<sub>x</sub> pollution controls were tightened in 1990 (see Section 12, "Major Laws and Treaties," below). However, acid deposition is in large part a cumulative problem, as the acid-neutralizing capacity of soils is gradually eroded in response to acid input, and eventual exhaustion of this acid-neutralizing capacity is a trigger for dramatic ecosystem impacts. Continued decrease in acid input is therefore critical. Figure 12 compares nutrient cycling in a pristine Chilean forest and in a forest impacted by acid deposition.

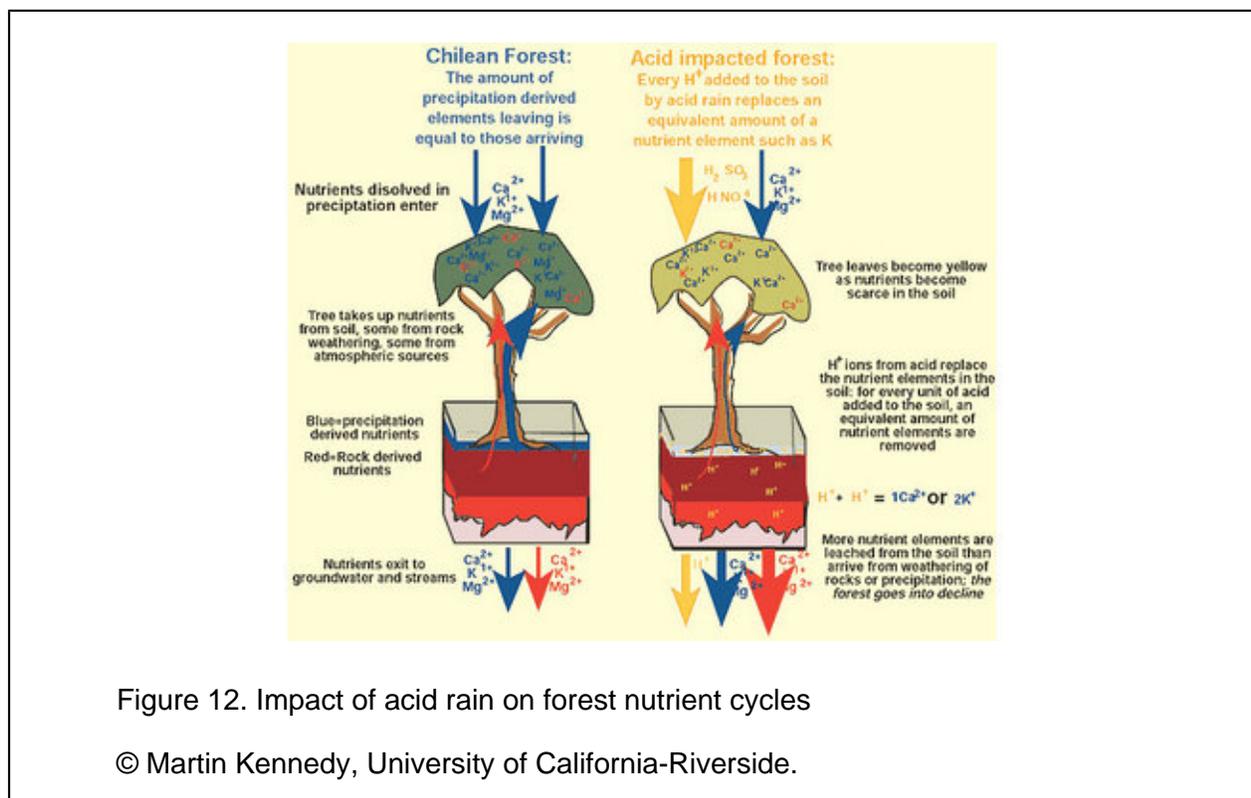


Figure 12. Impact of acid rain on forest nutrient cycles

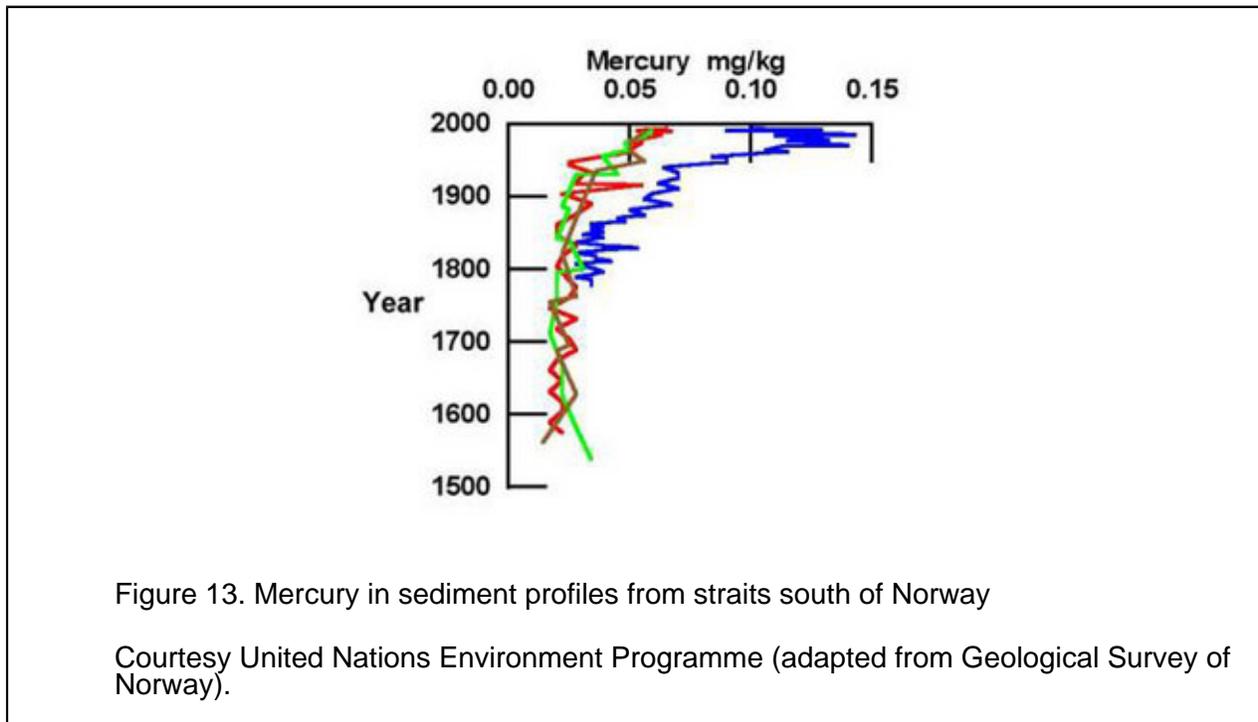
© Martin Kennedy, University of California-Riverside.

## 8. Mercury Deposition

Mercury (Hg) is a toxic pollutant whose input to ecosystems has greatly increased over the past century due to anthropogenic emissions to the atmosphere and subsequent deposition. Mercury is ubiquitous in the environment and is unique among metals in that it is highly volatile. When materials containing mercury are burned, as in coal combustion or waste incineration, mercury is released to the atmosphere as a gas either in elemental form,  $Hg(0)$  or oxidized divalent form,  $Hg^{2+}$ . The oxidized form is present as water-soluble compounds such as  $HgCl_2$  that are readily deposited in the region of their emission. By contrast,  $Hg(0)$  is not water-soluble and must be oxidized to  $Hg^{2+}$  in order to be deposited. This oxidation takes place in the atmosphere on a time scale of one year, sufficiently long that mercury can be readily transported around the world by atmospheric circulation. Mercury thus is a global pollution problem.

Deposition of anthropogenically emitted mercury to land and ocean has considerably raised mercury levels in the biosphere. This accumulation is evident from sediment cores that provide historical records of mercury deposition for the past several centuries (Fig. 13). Ice core samples from Antarctica, Greenland, and the western United States indicate that pre-industrial atmospheric mercury

concentrations ranged from about 1 to 4 nanograms per liter, but that concentrations over the past 150 years have reached as high as 20 ng/L (footnote 3).



Once deposited, oxidized mercury can be converted back to the elemental form  $Hg(0)$  and re-emitted to the atmosphere. This repeated re-emission is called the "grasshopper effect," and can extend the environmental legacy of mercury emissions to several decades. The efficacy of re-emission increases with increasing temperature, which makes  $Hg(0)$  more volatile. As a result mercury tends to accumulate to particularly high levels in cold regions such as the Arctic where re-emission is slow.

Divalent mercury deposited to ecosystems can be converted by bacteria to organic methylmercury, which is absorbed easily during digestion and accumulates in living tissues. It also enters fishes' bodies directly through their skin and gills (Fig. 14). U.S. federal agencies and a number of states have issued warnings against consuming significant quantities of large predatory fish species such as shark and swordfish, especially for sensitive groups such as young children and women of childbearing age (footnote 4).

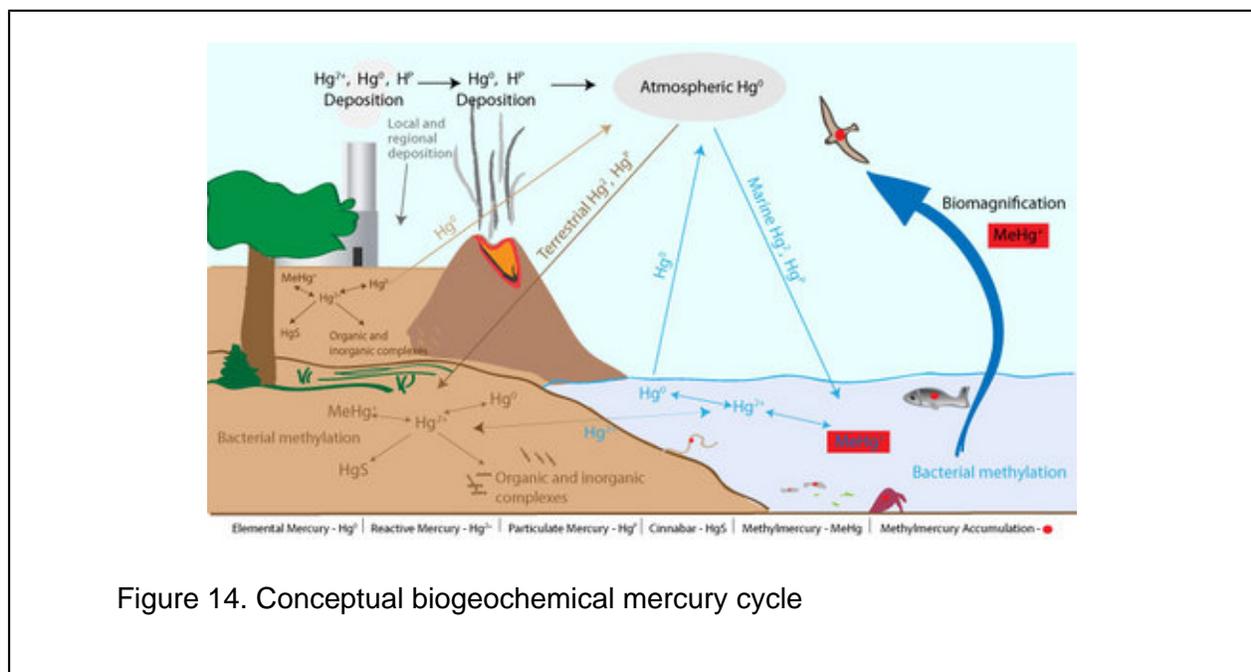


Figure 14. Conceptual biogeochemical mercury cycle

Mercury interferes with the brain and central nervous system. The expression "mad as a hatter" and the Mad Hatter character in Lewis Carroll's **Alice in Wonderland** are based on symptoms common among 19<sup>th</sup>-century English hat makers, who inhaled mercury vapors when they used a mercurous nitrate solution to cure furs. Many hatters developed severe muscle tremors, distorted speech, and hallucinations as a result. Sixty-eight people died and hundreds were made ill or born with neurological defects in Minamata, Japan in the 1950s and 1960s after a chemical company dumped mercury into Minamata Bay and families ate fish from they bay. Recently, doctors have reported symptoms including dizziness and blurred vision in healthy patients who ate significant quantities of high-mercury fish such as tuna (footnote 5).

Developed countries in North America and Europe are largely responsible for the global build-up of mercury in the environment over the past century. They have begun to decrease their emissions over the past two decades in response to the recognized environmental threat. However, emissions in Asia have been rapidly increasing and it is unclear how the global burden of mercury will evolve over the coming decades. Because mercury is transported on a global scale, its control requires a global perspective. In addition, the legacy of past emissions through re-emission and mercury accumulation in ecosystems must be recognized.

## 9. Controlling Air Pollution

Thanks to several decades of increasingly strict controls, emissions of most major air pollutants have declined in the U.S. and other industrialized countries since the 1970s. Figure 15 shows the aggregate decrease in U.S. emissions since 1970. This trend occurred even as economic activity and fuel consumption increased. The reductions came about because governments passed laws limiting allowable pollution levels and required the use of technologies to reduce emissions, such as scrubbers on power plant smokestacks and catalytic converters on vehicles.

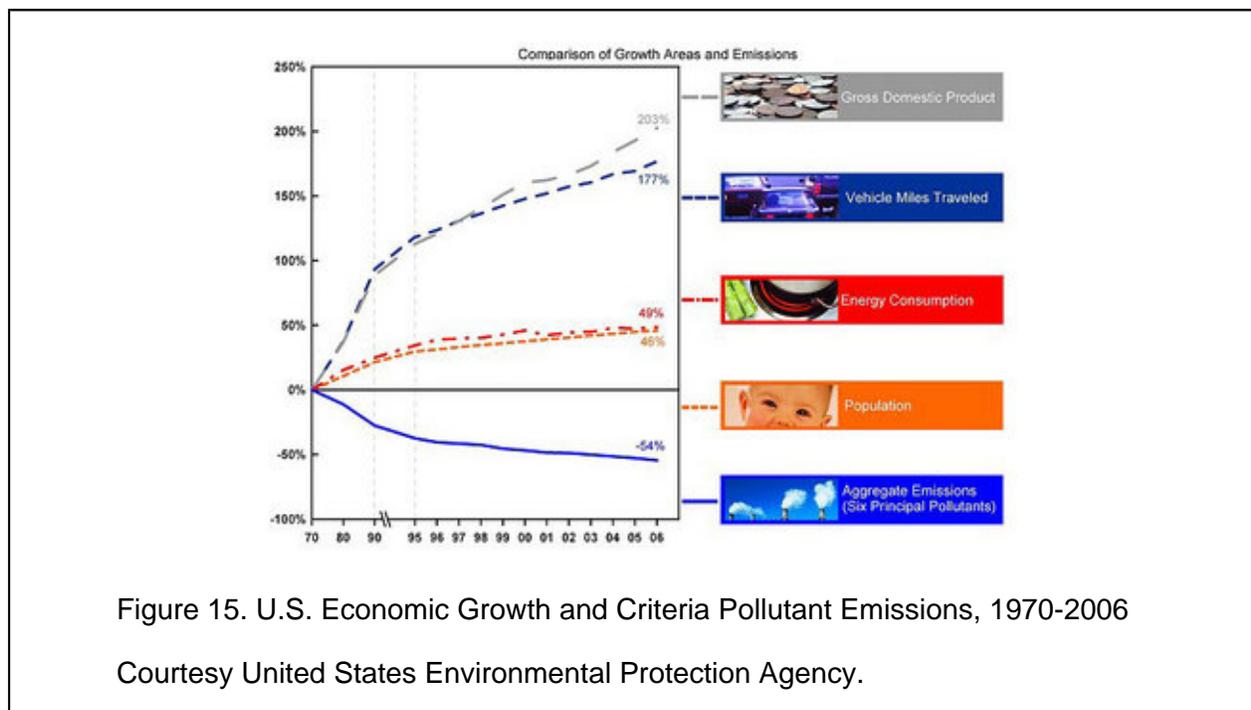


Figure 15. U.S. Economic Growth and Criteria Pollutant Emissions, 1970-2006

Courtesy United States Environmental Protection Agency.

This decrease in emissions has demonstrably reduced levels of the four principal primary pollutants: carbon monoxide, nitrogen dioxide, sulfur dioxide, and lead. Air quality standards for these four pollutants were frequently exceeded in the U.S. twenty years ago but are hardly ever exceeded now.

Progress in reducing the two principal secondary pollutants, ozone and particulate matter, has been much slower. This is largely due to nonlinear chemistry involved in the generation of these pollutants: reducing precursor emissions by a factor of two does not guarantee a corresponding factor of two decrease in the pollutant concentrations (the decrease is often much less, and there can even be an increase). In addition, advances in health-effects research have generated constant pressure for tougher air quality standards for ozone and fine aerosols.



In contrast to improvements in developed countries, air pollution has been worsening in many industrializing nations. Beijing, Mexico City, Cairo, Jakarta, and other megacities in developing countries have some of the dirtiest air in the world (for more on environmental conditions in megacities, see Unit 5, "Human Population Dynamics"). This situation is caused by rapid population growth combined with rising energy demand, weak pollution control standards, dirty fuels, and inefficient technologies. Some governments have started to address this problem—for example, China is tightening motor vehicle emission standards—but much stronger actions will be required to reduce the serious public health impacts of air pollution worldwide.

## 10. Stratospheric Ozone

Earth's stratospheric ozone layer, which contains about 90 percent of the ozone in the atmosphere, makes the planet habitable by absorbing harmful solar ultraviolet (UV) radiation before it reaches the planet's surface. UV radiation damages cells and causes sunburn and premature skin aging in low doses. At higher levels, it can cause skin cancer and immune system suppression. Earth's stratospheric ozone layer absorbs 99 percent of incoming solar UV radiation.

Scientists have worked to understand the chemistry of the ozone layer since its discovery in the 1920s. In 1930 British geophysicist Sydney Chapman described a process in which strong UV photons photolyze oxygen molecules ( $O_2$ ) into highly reactive oxygen atoms. These atoms rapidly combine with  $O_2$  to form ozone ( $O_3$ ) (Fig. 16). This process is still recognized as the only significant source of ozone to the stratosphere. Research and controversy have focused on identifying stratospheric ozone sinks.

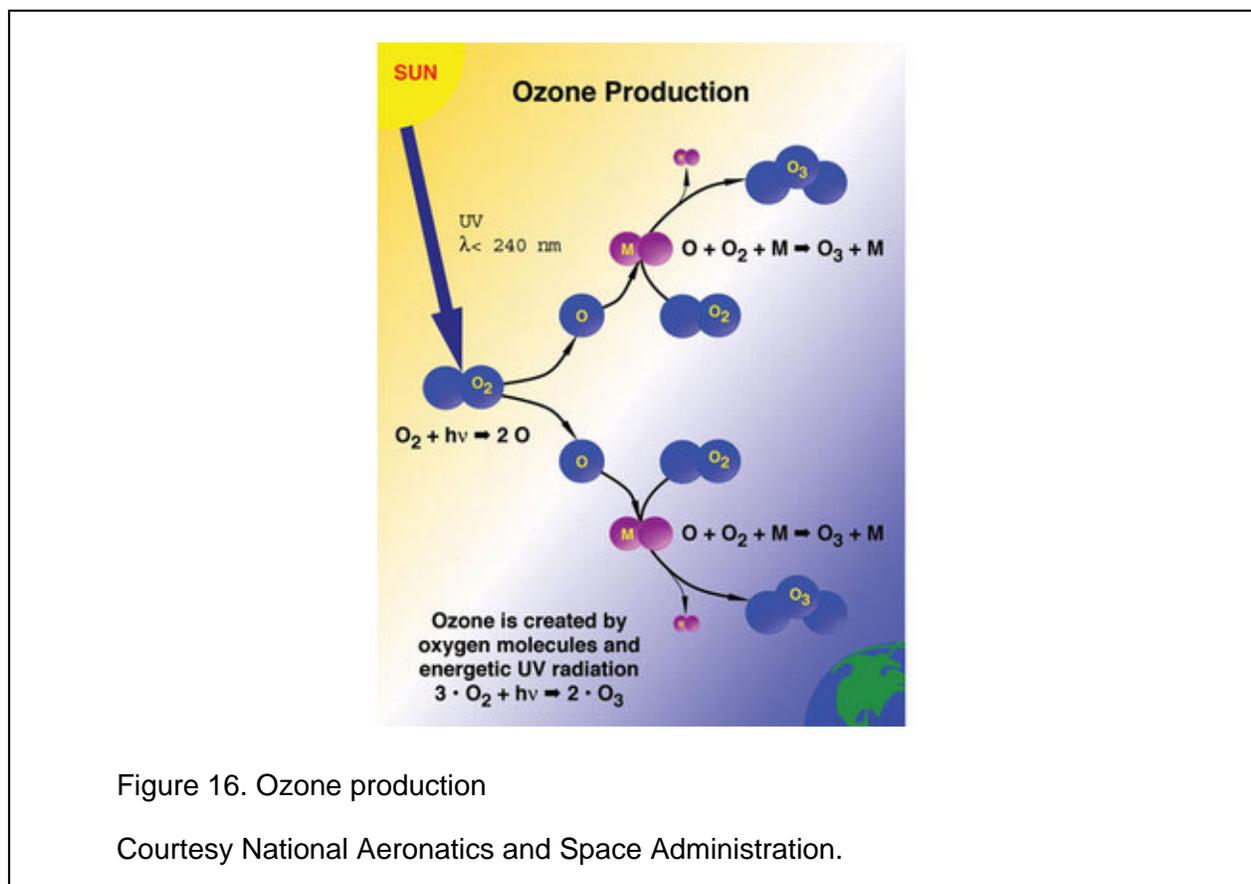


Figure 16. Ozone production

Courtesy National Aeronautics and Space Administration.

Ozone is produced by different processes in the stratosphere, where it is beneficial, and near the Earth's surface in the troposphere, where it is harmful. The mechanism for stratospheric ozone formation, photolysis of  $\text{O}_2$ , does not take place in the troposphere because the strong UV photons needed for this photolysis have been totally absorbed by  $\text{O}_2$  and ozone in the stratosphere. In the troposphere, by contrast, abundance of VOCs promotes ozone formation by the mechanism described above in Section 4. Ozone levels in the stratosphere are 10 to 100 times higher than what one observes at Earth's surface in the worst smog events. Fortunately we are not there to breathe it, though exposure of passengers in jet aircraft to stratospheric ozone has emerged recently as a matter of public health concern.

To explain observed stratospheric ozone concentrations, we need to balance ozone production and loss. Formation of ozone in the stratosphere is simple to understand, but the mechanisms for ozone loss are considerably more complicated. Ozone photolyzes to release  $\text{O}_2$  and O, but this is not an actual sink since  $\text{O}_2$  and O can just recombine to ozone. The main mechanism for ozone loss in the

natural stratosphere is a catalytic cycle involving  $\text{NO}_x$  radicals, which speed up ozone loss by cycling between  $\text{NO}$  and  $\text{NO}_2$  but are not consumed in the process.

The main source of  $\text{NO}_x$  in the troposphere is combustion; in contrast, the main source in the stratosphere is oxidation of nitrous oxide ( $\text{N}_2\text{O}$ ), which is emitted ubiquitously by bacteria at the Earth's surface. Nitrous oxide is inert in the troposphere and can therefore be transported up to the stratosphere, where much stronger UV radiation enables its oxidation. Nitrous oxide emissions have increased over the past century due to agriculture, but the rise has been relatively modest (from 285 to 310 parts per million by volume) and of little consequence for the ozone layer.

In 1974 chemists Sherwood Roland and Mario Molina identified a major threat to the ozone layer: rising atmospheric concentrations of manmade industrial chemicals called **chlorofluorocarbons** (CFCs), which at the time were widely used as refrigerants, in aerosol sprays, and in manufacturing plastic foams. CFC molecules are inert in the troposphere, so they are transported to the stratosphere, where they photolyze and release chlorine (Cl) atoms. Chlorine atoms cause catalytic ozone loss by cycling with ClO (Fig. 17).

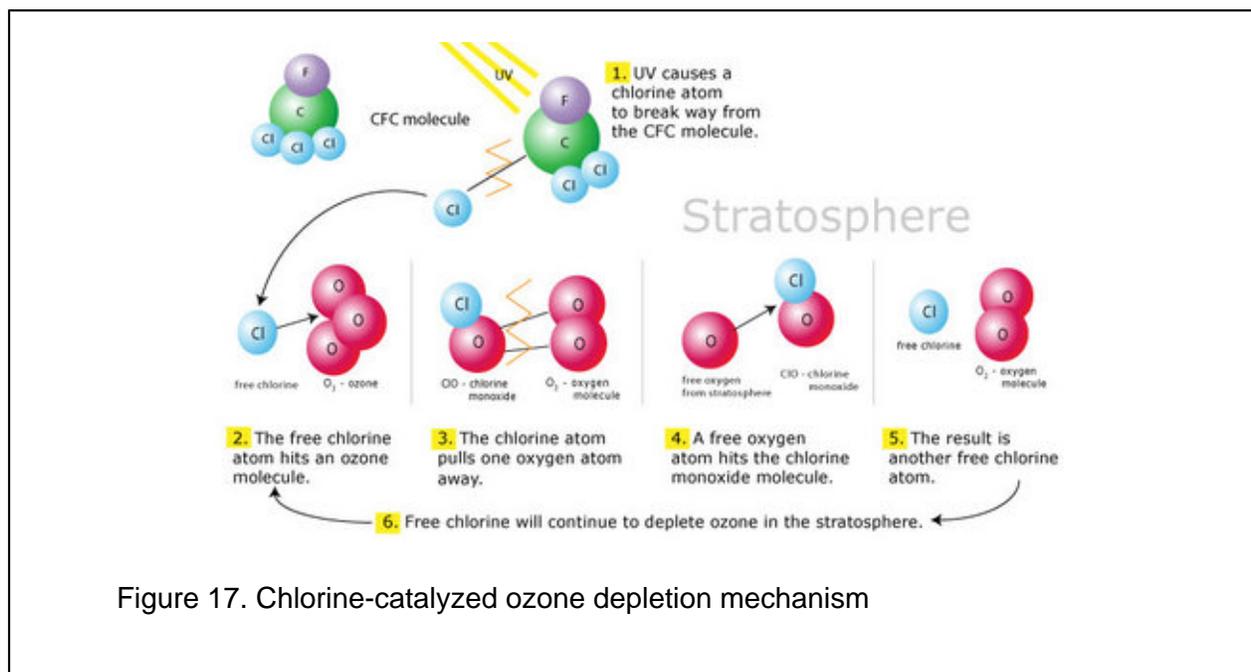


Figure 17. Chlorine-catalyzed ozone depletion mechanism

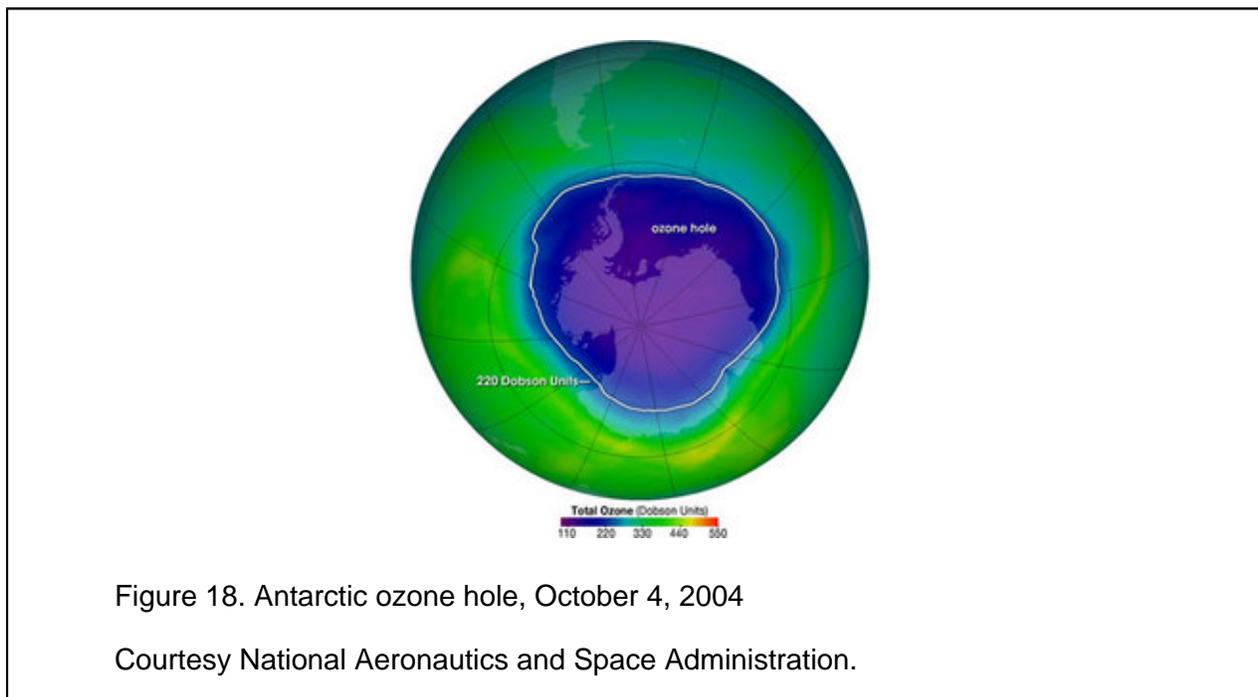
Eventually chlorine radicals (Cl and ClO) are converted to the stable nonradical reservoirs of hydrogen chloride (HCl) and chlorine nitrate ( $\text{ClONO}_2$ ). These reservoirs slowly "leak" by oxidation and photolysis to regenerate chlorine radicals. Chlorine is finally removed when it is transported to the troposphere and washed out through deposition. However, this transport process is slow. Concern

over chlorine-catalyzed ozone loss through the mechanism shown in Figure 17 led in the 1980s to the first measures to regulate production of CFCs.

In 1985 scientists from the British Arctic Survey reported that springtime stratospheric ozone levels over their station at Halley Bay had fallen sharply since the 1970s. Global satellite data soon showed that stratospheric ozone levels were decreasing over most of the southern polar latitudes. This pattern, widely referred to as the "ozone hole" (more accurately, ozone thinning), proved to be caused by high chlorine radical concentrations, as well as by bromine radicals (Br), which also trigger catalytic cycles with chlorine to consume ozone.

The source of the high chlorine radicals was found to be a fast reaction of the chlorine reservoirs HCl and ClNO<sub>3</sub> at the surface of icy particles formed at the very cold temperatures of the Antarctic wintertime stratosphere and called polar stratospheric clouds (PSCs). HCl and ClNO<sub>3</sub> react on PSC surfaces to produce molecular chlorine (Cl<sub>2</sub>) and nitric acid. Cl<sub>2</sub> then rapidly photolyzes in spring to release chlorine atoms and trigger ozone loss.

Ozone depletion has worsened since 1985. Today springtime ozone levels over Antarctica are less than half of levels recorded in the 1960s, and the 2006 Antarctic ozone hole covered 29 million square kilometers, tying the largest value previously recorded in 2000 (Fig. 18). In the 1990s ozone loss by the same mechanism was discovered in the Arctic springtime stratosphere, although Arctic ozone depletion is not as extensive as in Antarctica because temperatures are not as consistently cold.



Rowland and Molina's warnings about CFCs and ozone depletion, followed by the discovery of the ozone hole, spurred the negotiation of several international agreements to protect the ozone layer, leading eventually to a worldwide ban on CFC production in 1996 (for details see Section 12, "Major Laws and Treaties," below). CFCs have lifetimes in the atmosphere of 50-100 years, so it will take that long for past damage to the ozone layer to be undone.

The Antarctic ozone hole is expected to gradually heal over the next several decades, but the effects of climate change pose major uncertainties. Greenhouse gases are well known to cool the stratosphere (although they warm the Earth's surface), and gradual decrease in stratospheric temperatures has been observed over the past decades. Cooling of the polar stratosphere promotes the formation of PSCs and thus the release of chlorine radicals from chlorine reservoirs. The question now is whether the rate of decrease of stratospheric chlorine over the next decades will be sufficiently fast to stay ahead of the cooling caused by increasing greenhouse gases. This situation is being closely watched by atmospheric scientists both in Antarctica and the Arctic.

## 11. Air Pollution, Greenhouse Gases, and Climate Change

Air pollutants are major contributors to climate change. This connection is well known to scientists, although it has not yet permeated environmental policy. Figure 19 shows global climate forcing for the year 2000, relative to 1850, caused by different observed perturbations to the Earth system. Climate forcing from a given perturbation is defined as the mean resulting imbalance between energy input and energy output per unit time and unit area of Earth's surface (watts per square meter or  $W/m^2$ ), with all else remaining constant, including temperature. A positive radiative forcing means a decrease in energy output and hence a warming, shown as red bars in Fig. 19. Negative radiative forcing, shown as blue bars in Fig. 19, means a decrease in energy input and hence a cooling. (For more on climate forcing see Unit 12, "Earth's Changing Climate.")

Figure 19 shows that a large number of perturbation agents have forced the climate since 1850.  $CO_2$  is the single most important agent, but many other agents are also important and together exert greater influence than  $CO_2$ .

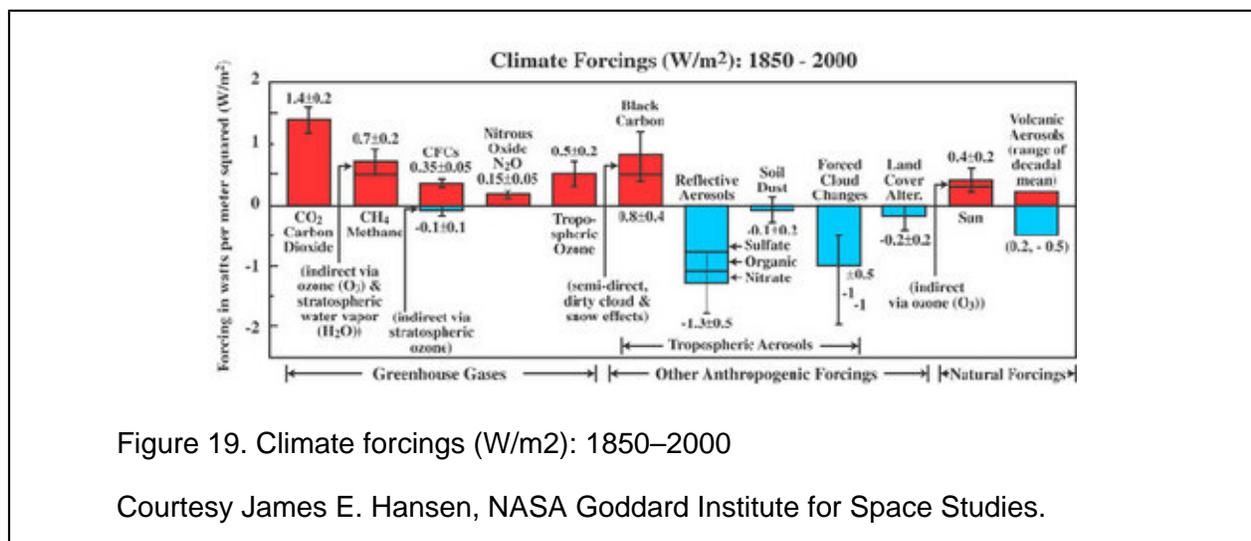


Figure 19. Climate forcings (W/m<sup>2</sup>): 1850–2000

Courtesy James E. Hansen, NASA Goddard Institute for Space Studies.

Among the major greenhouse gases in Figure 19 are methane and tropospheric ozone, which are both of concern for air quality. Light absorption by black carbon aerosol particles also has a significant warming effect. Taken together these three agents produce more radiative forcing than CO<sub>2</sub>. Reductions in these air pollutants thus would reap considerable benefit for climate change.

However, air pollutants can also have a cooling effect that compensates for greenhouse warming. This factor can be seen in Figure 19 from the negative radiative forcings due to non light-absorbing sulfate and organic aerosols originating from fossil fuel combustion. Scattering by these aerosols is estimated by the [Intergovernmental Panel on Climate Change \(IPCC\)](#) to have a direct radiative forcing of -1.3 W/m<sup>2</sup>, although this figure is highly uncertain. Indirect radiative forcing from increased cloud reflectivity due to anthropogenic aerosols is even more uncertain but could be as large as -1 W/m<sup>2</sup>.

Scattering aerosols have thus masked a significant fraction of the warming imposed by increasing concentrations of greenhouse gases over the past two centuries. Aerosol and acid rain control policies, though undeniably urgent to protect public health and ecosystems, will reduce this masking effect and expose us to more greenhouse warming.

Influence also runs the other way. Global climate change has the potential to magnify air pollution problems by raising Earth's temperature (contributing to tropospheric ozone formation) and increasing the frequency of stagnation events. Climate change is also expected to cause more forest fires and dust storms, which can cause severe air quality problems (Fig. 20).



Figure 20. Fire plumes over Southern California, October 26, 2003

Courtesy National Aeronautics and Space Administration.

The link between air pollution and climate change argues for developing environmental policies that will yield benefits in both areas. For example, researchers at Harvard University, Argonne National Laboratory, and the Environmental Protection Agency estimated in 2002 that reducing anthropogenic methane emissions by 50 percent would not only reduce greenhouse warming but also nearly halve the number of high-ozone events in the United States. Moreover, since methane contributes to background ozone levels worldwide, this approach would reduce ozone concentrations globally. In contrast, reducing  $\text{NO}_x$  emissions—the main U.S. strategy for combating ozone—produces more localized reductions to ozone (footnote 6).

Finally, let us draw the distinction between stratospheric ozone depletion and climate change, since these two problems are often confused in the popular press. As summarized in Table 2, the causes, processes, and impacts of these two global perturbations to the Earth system are completely different, but they have some links. On the one hand, colder stratospheric temperatures due to increasing greenhouse gases intensify polar ozone loss by promoting PSC formation, as discussed in Section 9. On the other hand, CFCs are major greenhouse gases, and stratospheric ozone depletion exerts a slight cooling effect on the Earth's surface.

**Table 2.** Comparison of stratospheric ozone depletion and global warming

	Ozone depletion	Global warming
Location	Stratosphere	Troposphere (stratosphere actually cools)
Causative pollutant	Ozone-depleting substances (NO, CFCs)	Greenhouse gases (CO, CH, NO, tropospheric ozone)
Process	Catalytic ozone loss reactions	Trapping of infrared radiation emitted by Earth's surface

## 12. Major Laws and Treaties

The first international treaty to protect the ozone layer was signed in 1987. The [Montreal Protocol on Substances That Deplete the Ozone Layer](#) has been amended several times since then in response to new scientific information, and it has been ratified by 190 countries. Under it industrialized countries phased out production of several classes of ozone-depleting substances by 1996, and developing countries are to follow suit by 2010. An international fund created under the protocol helps developing nations find substitutes for ozone-depleting substances. The World Bank estimates that reductions under the Montreal Protocol through 2003 have avoided up to 20 million cases of cancer, 130 million cases of eye cataracts, and severe damage to ecosystems from increased UV radiation reaching Earth's surface (footnote 7).

When nations began to negotiate agreements on reducing greenhouse gases to slow global climate change, many experts believed that the problem could be addressed through a framework similar to the Montreal Protocol—a treaty that set timetables for phasing out harmful emissions, with technical aid to help developing countries comply. But climate change has proved to be a considerably harder problem for negotiators, for several reasons. These negotiations are discussed in Unit 12, "Earth's Changing Climate."

The Clean Air Act (CAA) sets out a comprehensive set of national standards for controlling air pollutants that are considered harmful to public health and the environment in the United States. Under the CAA the Environmental Protection Agency is directed to establish [National Ambient Air Quality Standards](#) (NAAQS) limiting major air pollutants to levels that will protect public health, including the health of sensitive groups such as children, the elderly, and people with respiratory illnesses. NAAQS are set based on input from scientific advisory committees, and the act specifically directs EPA not to consider costs in setting NAAQS, although states can consider costs when they develop their plans for meeting the standards.

EPA has established NAAQS for six "criteria pollutants": SO<sub>2</sub>, NO<sub>2</sub>, lead, carbon monoxide, particulate matter less than 10 micrometers and less than 2.5 micrometers (PM-10 and PM-2.5), and ozone. States and counties that fail to achieve these standards are required to develop plans

for bringing their air quality into compliance. The CAA also defines major pollution sources, based on their emission levels, and establishes rules governing when new emission sources can be built in polluted areas.

The CAA has been amended several times since its passage in 1970 to tighten standards and institute new controls that reflect advances in scientific understanding of air pollution. The law has achieved some notable successes: for example, it has reduced U.S. automobile emissions considerably from pre-1970 levels, through mechanisms such as phasing out use of leaded gasoline and requiring car manufacturers to install catalytic converters. These devices treat car exhaust in several stages to reduce  $\text{NO}_x$  and oxidize unburned hydrocarbons and carbon monoxide (Fig. 21).



Figure 21. Catalytic converter mounted in a car's exhaust system

Courtesy Wikimedia Commons. Public Domain.

A set of CAA amendments passed in 1990 has produced significant cuts in  $\text{SO}_2$  emissions through what was then a new approach to reducing air pollution: capping the total allowable amount of pollution emitted nationally and then allocating emission rights among major sources (mainly coal-burning electric power plants and industrial facilities). Emitters that reduced pollution below their allowed levels could sell their extra pollution allowances to higher-emitting sources. This approach let sources make reductions where they were cheapest, rather than requiring all emitters to install specific pieces of control equipment or to meet one standard at each location. Some companies cut emissions by installing controls, while others switched to low-sulfur coal or other cleaner fuels.

U.S. SO<sub>2</sub> emissions have fallen by roughly 50 percent since emissions trading was instituted, and the program is widely cited as an example of how this approach can work more effectively than technology mandates. However, some proposals for emissions trading are more controversial—specifically, whether it is a safe approach for cutting toxic pollutants such as mercury. Opponents argue that letting some large sources continue to emit such pollutants could create dangerous "hot spots" that would be hazardous to public health, and that the only safe way to control hazardous pollutants like mercury is to require specific reductions from each individual source.

### 13. Further Reading

C.T. Driscoll et al., **Acid Rain Revisited: Advances in Scientific Understanding Since the Passage of the 1970 and 1990 Clean Air Act Amendments**, Hubbard Brook Research Foundation, Science Links Publication, vol. 1. no. 1 (2001), [http://www.hubbardbrook.org/hbrf/publications/Acid\\_Rain\\_Revisited.pdf](http://www.hubbardbrook.org/hbrf/publications/Acid_Rain_Revisited.pdf).

Daniel Jacob, **Introduction to Atmospheric Chemistry** (Princeton University Press, 1999). An overview of the field that shows how to use basic principles of physics and chemistry to describe a complex system such as the atmosphere.

Peter H. McMurry, Marjorie F. Shepherd, and James S. Vickery, eds., **Particulate Matter Science for Policy Makers: A NARSTO Assessment** (Cambridge University Press, 2004). Scientific information about particulate emissions and their impacts, produced by a North American consortium for atmospheric research in support of air quality management.

National Research Council, **Air Quality Management in the United States** (Washington, DC: National Academies Press, 2004). A Congressionally mandated study of how well major U.S. air quality laws have worked from a scientific and technical perspective, and ways in which they could be strengthened.

### Footnotes

1. Daniel M. Murphy, "Something in the Air," **Science**, March 25, 2005, pp. 1888–1890.
2. National Academies of Science, **Urbanization, Energy, and Air Pollution in China: The Challenges Ahead** (Washington, DC: National Academies Press, 2004), p. 3.
3. U.S. Geological Survey, "Glacial Ice Cores Reveal A Record of Natural and Anthropogenic Atmospheric Mercury Deposition for the Last 270 Years," June 2002, <http://toxics.usgs.gov/pubs/FS-051-02/>.
4. For the latest version, see U.S. Environmental Protection Agency, "Fish Advisories," <http://www.epa.gov/waterscience/fish/>.

5. Jane M. Hightower and Dan Moore, "Mercury Levels in High-End Consumers of Fish," **Environmental Health Perspectives**, Vol. 111., No. 4, April 2003, pp. 604–608; Eric Duhatschek, "Charity Games in Quebec Will Help Kids," **The Globe and Mail**, October 8, 2004, p. R11.
6. Arlene M. Fiore et al., "Linking Ozone Pollution and Climate Change: The Case for Controlling Methane," **Geophysical Research Letters**, Vol. 29, No. 19 (2002), pp. 25–28.
7. World Bank, "The World Bank and the Montreal Protocol," September 2003, <http://siteresources.worldbank.org/INTMP/214578-1110890369636/20489383/WBMontrealProtocolStatusReport2003.pdf>.

## Glossary

**acid rain** : Rainfall with a greater acidity than normal.

**aerosols** : Liquid or solid particles that are suspended in air or a gas. Also referred to as particulate matter.

**ambient** : Surrounding, encircling.

**carbon monoxide** : Odorless, colorless gas that interferes with the delivery of oxygen in the blood to the rest of the body. It is produced as a result of incomplete burning of carbon-containing fuels including coal, wood, charcoal, natural gas, and fuel oil. Depending on the amount inhaled, this gas can impede coordination, worsen cardiovascular conditions, and produce fatigue, headache, weakness, confusion, disorientation, nausea, and dizziness. Very high levels can cause death.

**chlorofluorocarbons** : Any of several organic compounds composed of carbon, fluorine, chlorine, and hydrogen. They were formerly used widely in industry, for example as refrigerants, propellants, and cleaning solvents.

**hydroxyl radical** : The neutral form of the hydroxide ion, often referred to as the "detergent" of the troposphere because it reacts with many pollutants, often acting as the first step to their removal.

**Intergovernmental Panel on Climate Change (IPCC)** : Established in 1988 by two United Nations organizations to assess the risk of human-induced climate change.

**Montreal Protocol on Substances That Deplete the Ozone Layer** : A 1987 international agreement, subsequently amended in 1990, 1992, 1995, and 1997, that establishes in participating countries a schedule for the phaseout of chlorofluorocarbons and other substances with an excessive ozone-depleting potential.

**National Ambient Air Quality Standards** : Standards established by the EPA and required by The Clean Air Act (last amended in 1990) for pollutants considered harmful to public health and the environment.



**nitrogen oxides** : A group of highly reactive gases, all of which contain nitrogen and oxygen in varying amounts. Many of the nitrogen oxides are colorless and odorless. However, one common pollutant, nitrogen dioxide ( $\text{NO}_2$ ), along with particles in the air, can often be seen as a reddish-brown layer over many urban areas.

**nonattainment areas** : Defined by The Clean Air Act as a locality where air pollution levels persistently exceed National Ambient Air Quality Standards, or that contributes to ambient air quality in a nearby area that fails to meet standards.

**ozone** : A triatomic molecule consisting of three oxygen atoms. Ground-level ozone is an air pollutant with harmful effects on the respiratory systems of animals. On the other hand, ozone in the upper atmosphere protects living organisms by preventing damaging ultraviolet light from reaching the Earth's surface.

**particulate matter (PM)** : The sum of all solid and liquid particles suspended in air, many of which are hazardous.

**photolysis** : A chemical process by which molecules are broken down into smaller units through the absorption of light.

**primary air pollutants** : Pollutants that are pumped into our atmosphere and directly pollute the air. Examples include carbon monoxide from car exhausts and sulfur dioxide from the combustion of coal as well as nitrogen oxides, hydrocarbons, and particulate matter (both solid and liquid).

**radical** : Atomic or molecular species with unpaired electrons on an otherwise open shell configuration. These unpaired electrons are usually highly reactive, so radicals are likely to take part in chemical reactions.

**secondary air pollutants** : Pollutant not directly emitted but forms when other pollutants (primary pollutants) react in the atmosphere. Examples include ozone, formed when hydrocarbons (HC) and nitrogen oxides ( $\text{NO}_x$ ) combine in the presence of sunlight;  $\text{NO}_2$ , formed as NO combines with oxygen in the air; and acid rain, formed when sulfur dioxide or nitrogen oxides react with water.

**smog** : A kind of air pollution; the word "smog" is a combination of smoke and fog. Classic smog results from large amounts of coal burning in an area and is caused by a mixture of smoke and sulphur dioxide.

**sulfur dioxide** : A colorless, extremely irritating gas or liquid ( $\text{SO}_2$ ), used in many industrial processes, especially the manufacture of sulfuric acid. In the atmosphere it can combine with water vapor to form sulfuric acid, a major component of acid rain.

**volatile organic compounds** : Organic chemical compounds that have high enough vapour pressures under normal conditions to significantly vaporize and enter the atmosphere.

## Unit 12 : Earth's Changing Climate



Glaciologists. Courtesy of Lonnie Thompson

### Overview

Earth's climate is a sensitive system that is subject to dramatic shifts over varying time scales. Today human activities are altering the climate system by increasing concentrations of heat-trapping greenhouse gases in the atmosphere, which raises global temperatures. In this unit, examine the science behind global climate change and explore its potential impacts on natural ecosystems and human societies.

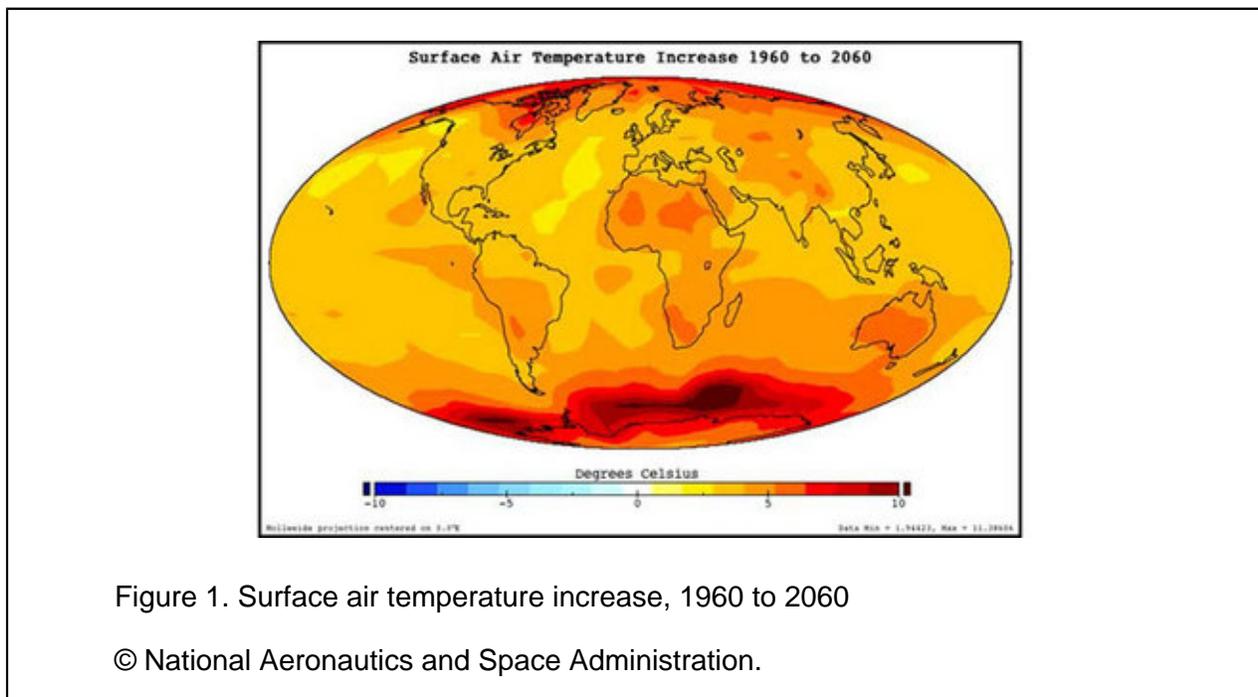
### Sections:

1. Introduction
2. Tipping Earth's Energy Balance
3. Climate Change: What the Past Tells Us
4. Past Warming: The Eocene Epoch
5. Global Cooling: The Pleistocene Epoch
6. Present Warming and the Role of CO<sub>2</sub>
7. Observed Impacts of Climate Change
8. Other Potential Near-Term Impacts
9. Major Laws and Treaties
10. Further Reading

## 1. Introduction

For the past 150 years, humans have been performing an unprecedented experiment on Earth's climate. Human activities, mainly fossil fuel combustion, are increasing concentrations of greenhouse gases (GHGs) in the atmosphere. These gases are trapping infrared radiation emitted from the planet's surface and warming the Earth. Global average surface temperatures have risen about 0.7°C (1.4°F) since the early 20th century.

Earth's climate is a complex system that is constantly changing, but the planet is warmer today than it has been for thousands of years, and current atmospheric carbon dioxide (CO<sub>2</sub>) levels have not been equaled for millions of years. As we will see below, ancient climate records offer some clues about how a warming world may behave. They show that climate shifts may not be slow and steady; rather, temperatures may change by many degrees within a few decades, with drastic impacts on plant and animal life and natural systems. And if CO<sub>2</sub> levels continue to rise at projected rates, history suggests that the world will become drastically hotter than it is today, possibly hot enough to melt much of Earth's existing ice cover. Figure 1 depicts projected surface temperature changes through 2060 as estimated by NASA's Global Climate Model.



Past climate changes were driven by many different types of naturally-occurring events, from variations in Earth's orbit to volcanic eruptions. Since the start of the industrial age, human activities



have become a larger influence on Earth's climate than other natural factors. High CO<sub>2</sub> levels (whether caused by natural phenomena or human activities) are a common factor between many past climate shifts and the warming we see today.

Many aspects of climate change, such as exactly how quickly and steadily it will progress, remain uncertain. However, there is strong scientific consensus that current trends in GHG emissions will cause substantial warming by the year 2100, and that this warming will have widespread impacts on human life and natural ecosystems. Many impacts have already been observed, including higher global average temperatures, rising sea levels (water expands as it warms), and changes in snow cover and growing seasons in many areas.

A significant level of warming is inevitable due to GHG emissions that have already been released, but we have options to limit the scope of future climate change—most importantly, by reducing fossil fuel consumption (for more details, see Unit 10, "Energy Challenges"). Other important steps to mitigate global warming include reducing the rate of global [deforestation](#) to preserve forest carbon [sinks](#) and finding ways to capture and sequester carbon dioxide emissions instead of releasing them to the atmosphere. (These responses are discussed in Unit 13, "Looking Forward: Our Global Experiment.")

## 2. Tipping Earth's Energy Balance

Earth's climate is a dynamic system that is driven by energy from the sun and constantly impacted by physical, biological, and chemical interactions between the atmosphere, global water supplies, and ecosystems (Fig. 2).



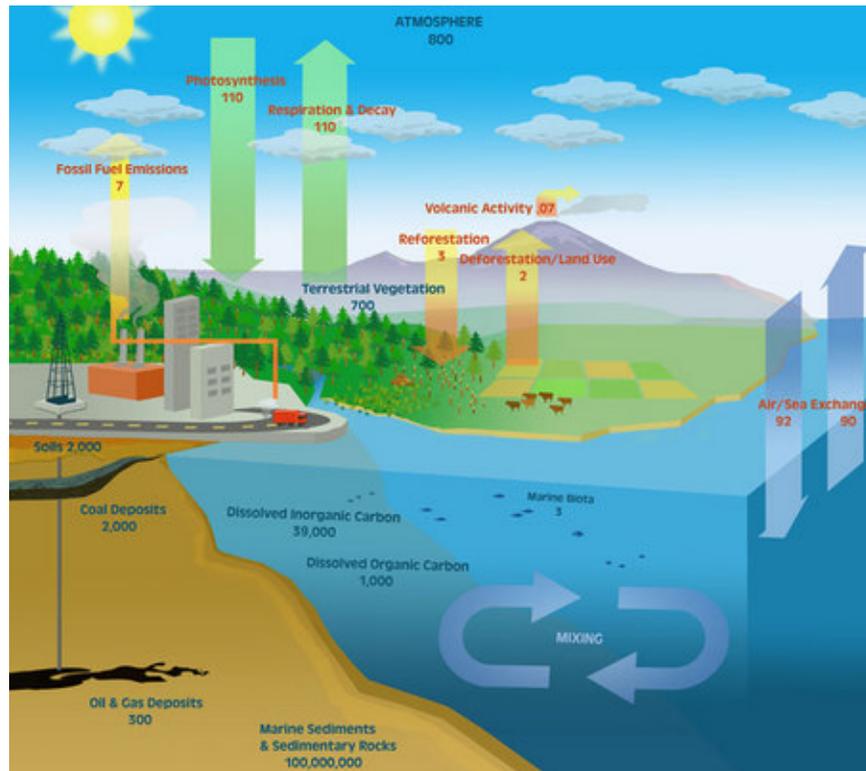


Figure 3. Earth's energy balance

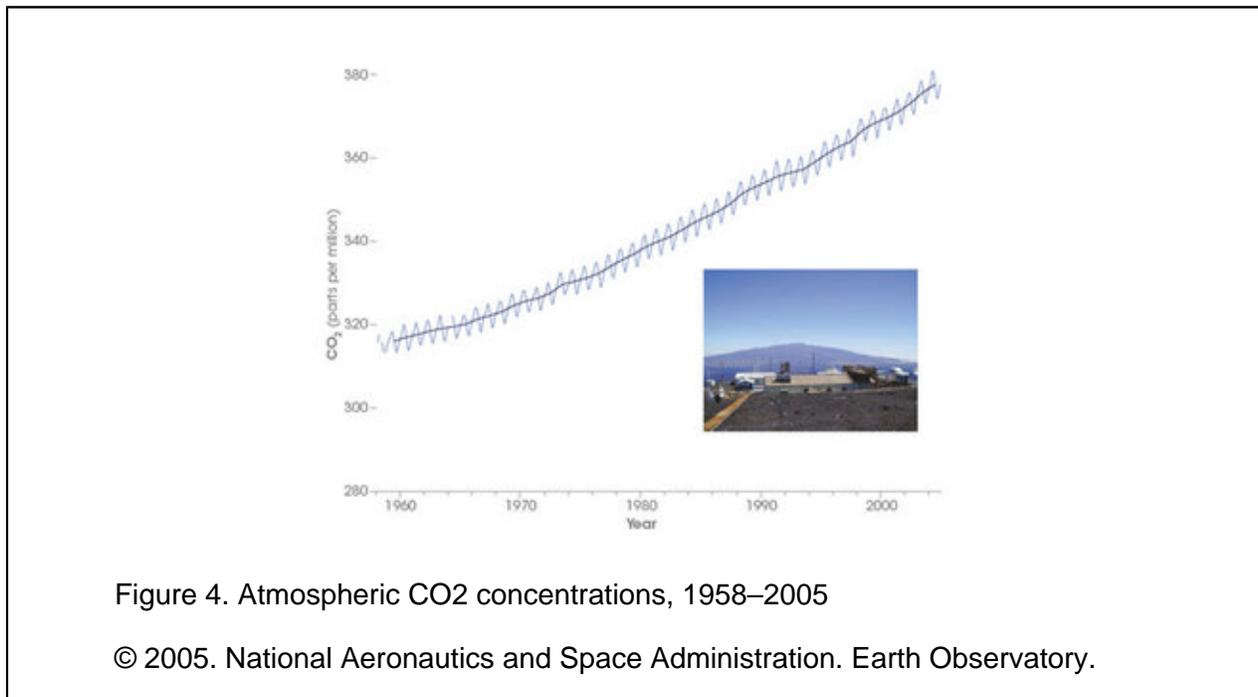
Courtesy Jared T. Williams. © Dan Schrag, Harvard University.

Many GHGs, including water vapor, ozone, CO<sub>2</sub>, methane (CH<sub>4</sub>), and nitrous oxide (N<sub>2</sub>O), are present naturally. Others are synthetic chemicals that are emitted only as a result of human activity, such as chlorofluorocarbons (CFCs), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), and sulfur hexafluoride (SF<sub>6</sub>). Important human activities that are raising atmospheric GHG concentrations include:

- fossil fuel combustion (CO and small quantities of methane and NO);
- deforestation (CO releases from forest burning, plus lower forest carbon uptake);
- landfills (methane) and wastewater treatment (methane, NO);

- livestock production (methane, NO);
- rice cultivation (methane);
- fertilizer use (NO); and
- industrial processes (HFCs, PFCs, SF).

Measuring CO<sub>2</sub> levels at Mauna Loa, Hawaii, and other pristine air locations, climate scientist Charles David Keeling traced a steady rise in CO<sub>2</sub> concentrations from less than 320 parts per million (ppm) in the late 1950s to 380 ppm in 2005 (Fig. 4). Yearly oscillations in the curve reflect seasonal cycles in the northern hemisphere, which contains most of Earth's land area. Plants take up CO<sub>2</sub> during the growing season in spring and summer and then release it as they decay in fall and winter.



Global CO<sub>2</sub> concentrations have increased by one-third from their pre-industrial levels, rising from 280 parts per million before the year 1750 to 377 ppm today. Levels of methane and N<sub>2</sub>O, the most influential GHGs after CO<sub>2</sub>, also increased sharply in the same time period (see Table 1 below).

If there are so many GHGs, why does CO<sub>2</sub> get most of the attention? The answer is a combination of CO<sub>2</sub>'s abundance and its **residence time** in the atmosphere. CO<sub>2</sub> accounts for about 0.1 percent of

the atmosphere, substantially more than all other GHGs except for water vapor, which may comprise up to 7 percent depending on local conditions. However, water vapor levels vary constantly because so much of the Earth's surface is covered by water and water vapor cycles into and out of the atmosphere very quickly—usually in less than 10 days. Therefore, water vapor can be considered a feedback that responds to the levels of other greenhouse gases, rather than an independent climate forcing (footnote 1).

Other GHGs contribute more to global climate change than CO<sub>2</sub> on a per-unit basis, although their relative impacts vary with time. The **global warming potential** (GWP) of a given GHG expresses its estimated climate impact over a specific period of time compared to an equivalent amount by weight of carbon dioxide. For example, the current 100-year GWP for N<sub>2</sub>O is 296, which indicates that one ton of N<sub>2</sub>O will have the same global warming effect over 100 years as 296 tons of CO<sub>2</sub>. Internationally-agreed GWP values are periodically adjusted to reflect current research on GHGs' behavior and impacts in the atmosphere.

However, CO<sub>2</sub> is still the most important greenhouse gas because it is emitted in far larger quantities than other GHGs. Atmospheric concentrations of CO<sub>2</sub> are measured in parts per million, compared to parts per billion or per trillion of other gases, and CO<sub>2</sub>'s atmospheric lifetime is 50 to 200 years, significantly longer than most GHGs. As illustrated in Table 1, the total extent to which CO<sub>2</sub> has raised global temperature (referred to as radiative forcing and measured in watts per square meter) since 1750 is significantly larger than forcing from other gases.

**Table 1.** Current greenhouse gas concentrations.

Gas	Pre-1750 concentration	Current tropospheric concentration	100-year GWP	Atmospheric lifetime (years)	Increased radiative forcing (watts/meter)
<b>Carbon dioxide</b>	280 parts per million	377.3 parts per million	1	Variable (up to 200 years)	1.66
<b>Methane</b>	688-730 parts per billion	1,730-1,847 parts per billion	23	12	0.5
<b>Nitrous oxide</b>	270 parts per billion	318-319 parts per billion	296	114	0.16
<b>Tropospheric ozone</b>	25	34	Not applicable due to short residence time	Hours to days	0.35

Gas	Pre-1750 concentration	Current tropospheric concentration	100-year GWP	Atmospheric lifetime (years)	Increased radiative forcing (watts/meter)
<b>Industrial gases (HFCs, PFCs, halons)</b>	0	Up to 545 parts per trillion	Ranges from 140 to 12,000	Primarily between 5 and 260 years	0.34 for all halocarbons collectively
<b>Sulfur hexafluoride</b>	0	5.22 parts per trillion	22,200	3,200	0.002

A look at current emissions underlines the importance of CO<sub>2</sub>. In 2003 developed countries emitted 11.6 billion metric tons of CO<sub>2</sub>, nearly 83 percent of their total GHG emissions. Developing countries' reported emissions were smaller in absolute terms, but CO<sub>2</sub> accounted for a similarly large share of their total GHG output (footnote 2). In 2004, CO<sub>2</sub> accounted for 85 percent of total U.S. GHG emissions, compared to 7.8 percent from methane, 5.4 percent from N<sub>2</sub>O, and 2 percent from industrial GHGs (footnote 3).

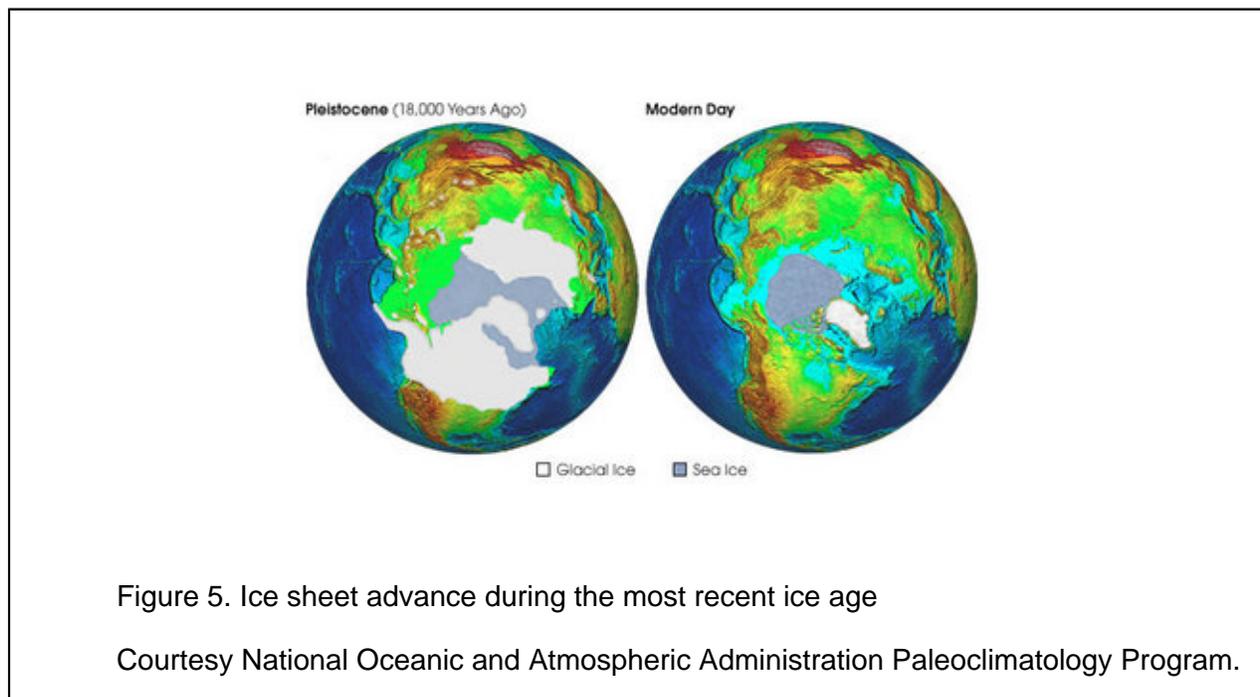
These emissions from human activities may reshape the global carbon cycle. As discussed in Units 2 ("Atmosphere") and 3 ("Oceans"), roughly 60 percent of CO<sub>2</sub> emissions from fossil fuel burning remain in the atmosphere, with about half of the remaining 40 percent absorbed by the oceans and half by terrestrial ecosystems. However, there are limits to the amount of anthropogenic carbon that these sinks can take up. Oceans are constrained by the rate of mixing between upper and lower layers, and there are physical bounds on plants' ability to increase their photosynthesis rates as atmospheric CO<sub>2</sub> levels rise and the world warms.

Scientists are still trying to estimate how much carbon these sinks can absorb, but it appears clear that oceans and land sinks cannot be relied on to absorb all of the extra CO<sub>2</sub> emissions that are projected in the coming century. This issue is central to projecting future impacts of climate change because emissions that end up in the atmosphere, rather than being absorbed by land or ocean sinks, warm the earth.

### 3. Climate Change: What the Past Tells Us

Throughout much of its 4.5 billion year history, Earth's climate has alternated between periods of warmth and relative cold, each lasting for tens to hundreds of millions of years. During the warmest periods, the polar regions of the world were completely free of ice. Earth also has experienced repeated ice ages—periods lasting for millions of years, during which ice sheets advanced and retreated many times over portions of the globe. During the most extreme cold phases, snow and ice covered the entire globe (for more details, see Unit 1, "Many Planets, One Earth").

From the perspective of geological time our planet is currently passing through a relatively cold phase in its history and has been cooling for the past 35 million years, a trend that is only one of many swings between hot and cold states over the last 500 million years. During cold phases glaciers and snow cover have covered much of the mid-latitudes; in warm phases, forests extended all the way to the poles (Fig. 5).



Scientists have analyzed [paleoclimate](#) records from many regions of the world to document Earth's climate history. Important sources of information about past climate shifts include:

- **Mineral deposits in deep sea beds.** Over time, dissolved shells of microscopic marine organisms create layers of chalk and limestone on sea beds. Analyzing the ratio of oxygen-18 (a rare isotope) to oxygen-16 (the common form) indicates whether the shells were formed during glacial periods, when more of the light isotope evaporated and rained down, or during warm periods.
- **Pollen grains trapped in terrestrial soils.** Scientists use [radiocarbon dating](#) to determine what types of plants lived in the sampled region at the time each layer was formed. Changes in vegetation reflect surface temperature changes.

- **Chemical variations in coral reefs.** Coral reefs grow very slowly over hundreds or thousands of years. Analyzing their chemical composition and determining the time at which variations in corals' makeup occurred allows scientists to create records of past ocean temperatures and climate cycles.
- **Core samples from polar ice fields and high-altitude glaciers.** The layers created in ice cores by individual years of snowfall, which alternate with dry-season deposits of pollen and dust, provide physical timelines of glacial cycles. Air bubbles in the ice can be analyzed to measure atmospheric CO levels at the time the ice was laid down.

Understanding the geological past is key to today's climate change research for several reasons. First, as the next sections will show, Earth's climate history illustrates how changing GHG levels and temperatures in the past shaped climate systems and affected conditions for life. Second, researchers use past records to tune climate models and see whether they are accurately estimating dynamics like temperature increase and climate feedbacks. The more closely a model can replicate past climate conditions, the more accurate its future predictions are likely to be.

#### 4. Past Warming: The Eocene Epoch

Scientists have looked far back in time to find a period when atmospheric GHG concentrations were as high as they could rise in coming decades if current emission trends continue. The Eocene epoch, which lasted from 55 million to 38 million years ago, was the most recent time when scientists think that CO<sub>2</sub> was higher than 500 parts per million.

Fossil evidence shows that Earth was far warmer during the Eocene than it is now. Tropical trees grew over much larger ranges to the north and south than they occupy today. Palm trees grew as far north as Wyoming and crocodiles swam in warm ocean water off Greenland. Early forms of modern mammals appeared, including small creatures such as cat-sized horses whose size made them well adapted to a warm climate (Fig. 6). Without ice cover at the poles, sea levels were nearly 100 meters higher than today. The deep ocean, which today is near freezing, warmed to over 12°C.

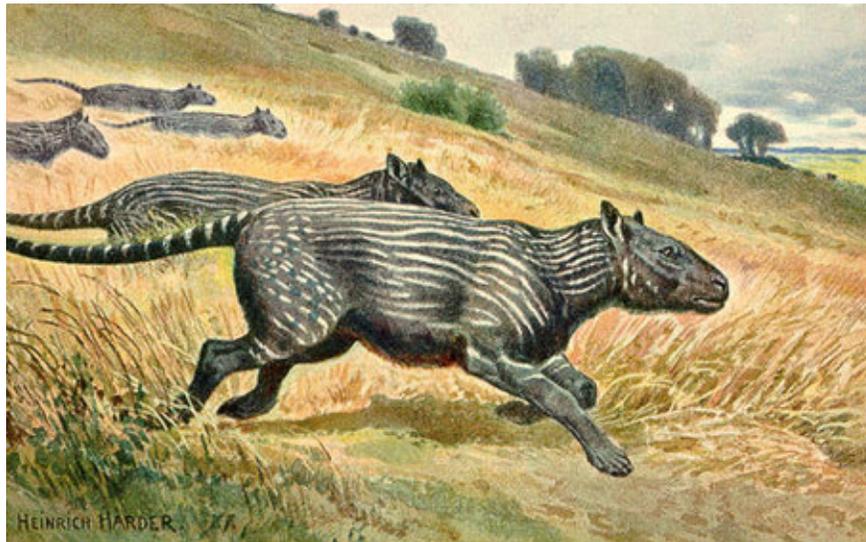


Figure 6. Phenacodus, a sheep-sized herbivore found in the Eocene era  
Courtesy Wikimedia Commons. Public Domain.

Scientists cannot measure CO<sub>2</sub> levels during the Eocene—there are no ice cores because there is no ice this old—but from indirect measurements of ocean chemistry they estimate that atmospheric CO<sub>2</sub> levels were three to ten times higher than pre-industrial levels (280 ppm). These concentrations were probably related to a sustained increase in CO<sub>2</sub> released from volcanoes over tens of millions of years. Because this climate persisted for tens of millions of years, living species and the climate system had time to adapt to warm, moist conditions. If humans release enough GHGs into the atmosphere to create Eocene-like conditions in the next several centuries, the transition will be much more abrupt, and many living organisms—especially those that thrive in cold conditions—will have trouble surviving the shift.

A troubling lesson from the Eocene is that scientists are unable to simulate Eocene climate conditions using climate models designed for the modern climate. When CO<sub>2</sub> levels are raised in the computer models to levels appropriate for what scientists think existed during the Eocene, global temperatures rise but high latitude temperatures do not warm as much as what scientists measure, particularly in winter. Some scientists believe that this is because there are unrecognized feedbacks in the climate system involving types of clouds that only form when CO<sub>2</sub> levels are very high. If this theory is correct, future climate could warm even more in response to anthropogenic release of CO<sub>2</sub> than most models predict.

The beginning of the Eocene also hosted a shorter event that may be the best natural analogue for what humans are doing to the climate system today. Fifty-five million years ago a rapid warming episode called the Paleocene-Eocene Thermal Maximum (PETM) occurred, in which Earth's temperature rose by 5 to 6°C on average within 10,000 to 30,000 years. Several explanations have been proposed for this large, abrupt warming, all of which involve a massive infusion of GHGs into the atmosphere, resulting in a trebling or perhaps a quadrupling of CO<sub>2</sub> concentrations, not unlike what is predicted for CO<sub>2</sub> levels by 2100 (footnote 4).

## 5. Global Cooling: The Pleistocene Epoch

During the Pleistocene epoch, which began about 2 million years ago, Earth's average temperature has always been cold enough to maintain ice at high latitudes. But Pleistocene climate has not been constant: ice coverage has fluctuated dramatically, with continental ice sheets advancing and retreating over large parts of North America and Europe. These peak glacial periods are often referred to as "Ice Ages" or "Glacial Maxima." During the Pleistocene, Earth has experienced more than 30 swings between prolonged glacial periods and brief warmer interglacial phases like the one we live in today.

Glacial advances and retreats shaped Earth's topography, soils, flora, and fauna (Fig. 7). During glaciation events, huge volumes of water were trapped in continental ice sheets, lowering sea levels as much as 130 meters and exposing land between islands and across continents. These swings often changed ocean circulation patterns. During the most extreme cold phases, ice covered up to 30 percent of the Earth's surface.

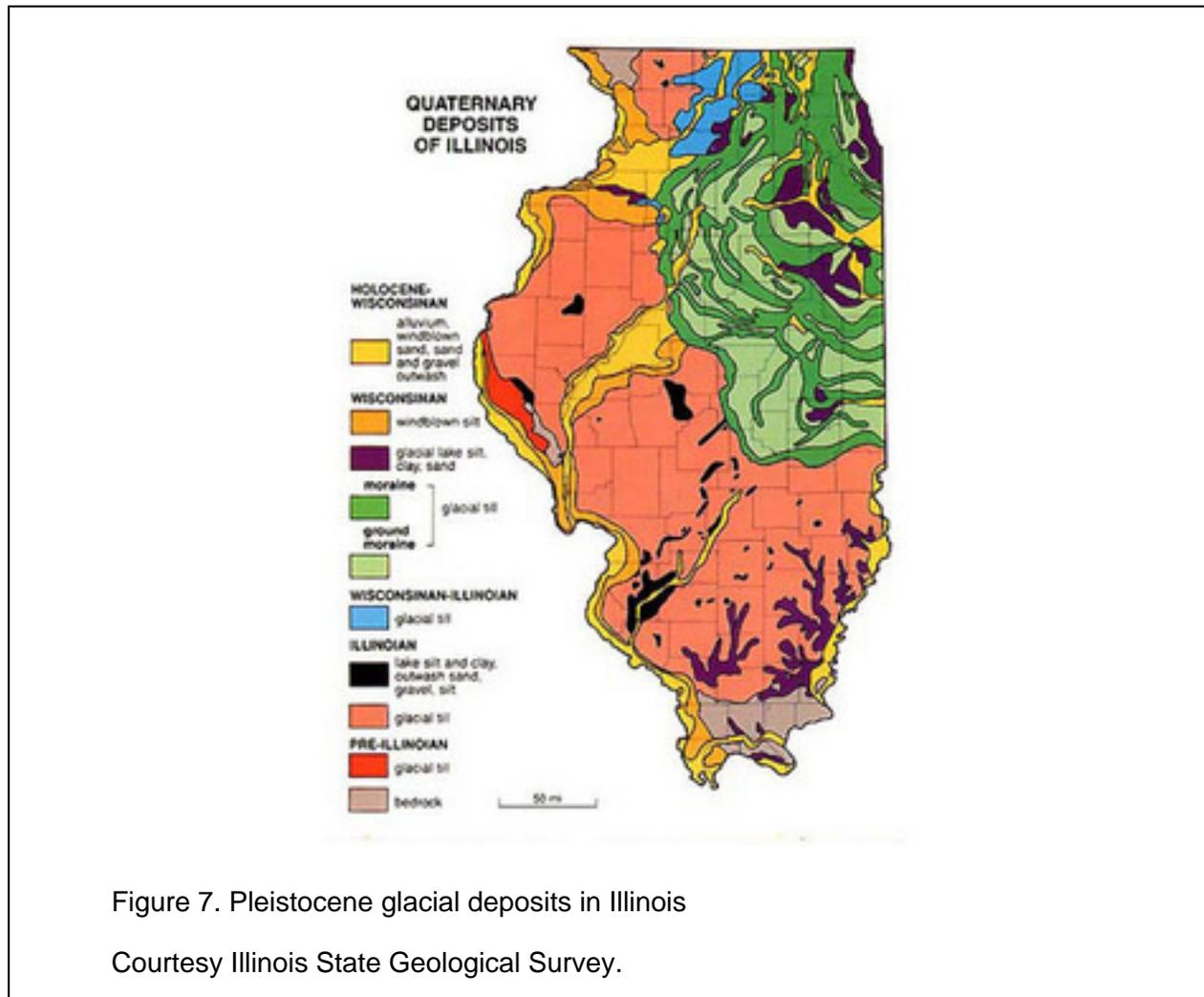


Figure 7. Pleistocene glacial deposits in Illinois

Courtesy Illinois State Geological Survey.

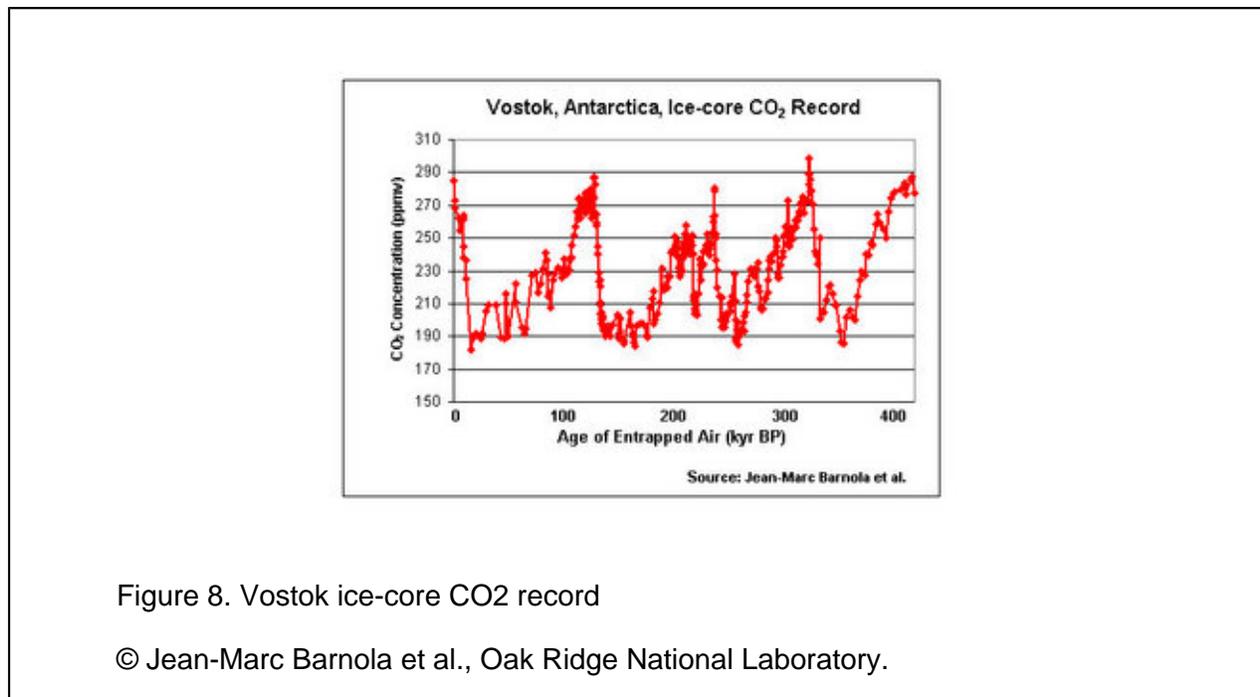
As glaciers advanced and retreated at high latitudes, ecosystems at lower latitudes evolved to adapt to prevailing climate conditions. In North America, just south of the advancing glaciers, a unique type of grass steppe supported distinctive cold-adapted fauna dominated by large mammals such as the mammoth, woolly rhinoceros, and dire wolf.

Why did Pleistocene temperatures swing back and forth so dramatically? Scientists point to a combination of factors. One main cause is variations in Earth's orbit around the sun. These variations, which involve the tilt of the Earth's pole of rotation and the ellipticity of the Earth's orbit, have regular timescales of 23,000, 41,000, and 100,000 years and cause small changes in the distribution of solar radiation received on the Earth (footnote 5). The possibility that these subtle variations could drive changes in climate was first proposed by Scottish scientist James Croll in the 1860s. In the 1930s,

Serbian astronomer Milutin Milankovitch developed this idea further. Milankovitch theorized that variations in summer temperature at high latitudes were what drove ice ages—specifically, that cool summers kept snow from melting and allowed glaciers to grow.

However, changes in summer temperature due to orbital variations are too small to cause large climate changes by themselves. Positive feedbacks are required to amplify the small changes in solar radiation. The two principal feedbacks are changes in Earth's **albedo** (the amount of light reflected from the Earth's surface) from snow and ice buildup, and in the amount of CO<sub>2</sub> in the atmosphere.

Ice core samples from the Vostok station and the European Project for Ice Coring in Antarctica (EPICA) document that CO<sub>2</sub> levels have varied over glacial cycles. From bubbles trapped in the ice, scientists can measure past concentrations of atmospheric CO<sub>2</sub>. The ice's chemical composition can also be used to measure past surface temperatures. Taken together, these records show that temperature fluctuations through glacial cycles over the past 650,000 years have been accompanied by shifts in atmospheric CO<sub>2</sub>. GHG concentrations are high during warm interglacial periods and are low during glacial maxima. The ice cores also show that atmospheric CO<sub>2</sub> concentrations never exceeded 300 parts per million—and therefore that today's concentration is far higher than what has existed for the last 650,000 years (Fig. 8).



One important lesson from ice cores is that climate change is not always slow or steady. Records from Greenland show that throughout the last glacial period, from about 60,000 to 20,000 years ago,

abrupt warming and cooling swings called Dansgaard-Oeschger, or D-O, events took place in the North Atlantic. In each cycle temperatures on ice sheets gradually cooled, then abruptly warmed by as much as 20°C, sometimes within less than a decade. Temperatures would then decline gradually over a few hundred to a few thousand years before abruptly cooling back to full glacial conditions. Similar climate fluctuations have been identified in paleoclimate records from as far away as China.

These sharp flips in the climate system have yet to be explained. Possible causes include changes in solar output or in sea ice levels around Greenland. But they are powerful evidence that when the climate system reaches certain thresholds, it can jump very quickly from one state to another. At the end of the Younger Dryas—a near-glacial phase that started about 12,800 years ago and lasted for about 1,200 years—annual mean temperatures increased by as much as 10°C in ten years (footnote 6).

## 6. Present Warming and the Role of CO<sub>2</sub>

There is clear evidence from many sources that the planet is heating up today and that the pace of warming may be increasing. Earth has been in a relatively warm interglacial phase, called the Holocene Period, since the last ice age ended roughly 10,000 years ago. Over the past thousand years average global temperatures have varied by less than one degree—even during the so-called "Little Ice Age," a cool phase from the mid-fourteenth through the mid-nineteenth centuries, during which Europe and North America experienced bitterly cold winters and widespread crop failures.

Over the past 150 years, however, global average surface temperatures have risen, increasing by 0.6°C +/- 0.2°C during the 20th century. This increase is unusual because of its magnitude and the rate at which it has taken place. Nearly every region of the globe has experienced some degree of warming in recent decades, with the largest effects at high latitudes in the Northern Hemisphere. In Alaska, for example, temperatures have risen three times faster than the global average over the past 30 years. The 1990s were the warmest decade of the 20th century, with 1998 the hottest year since instrumental record-keeping began a century ago, and the ten warmest years on record have all occurred since 1990 (Fig. 9).

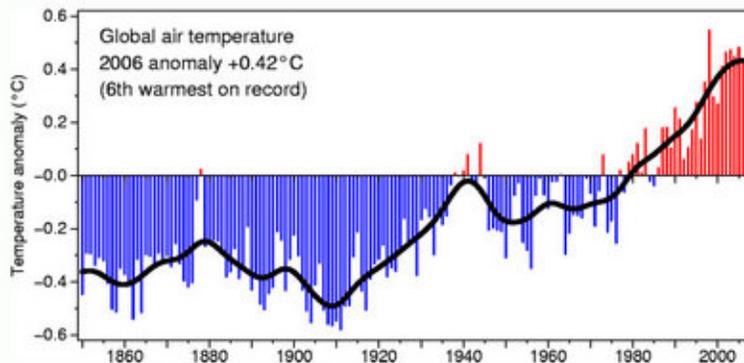


Figure 9. Global temperature record

Courtesy Phil Jones. © Climactic Research Unit, University of East Anglia and the U.K. Met Office Hadley Centre.

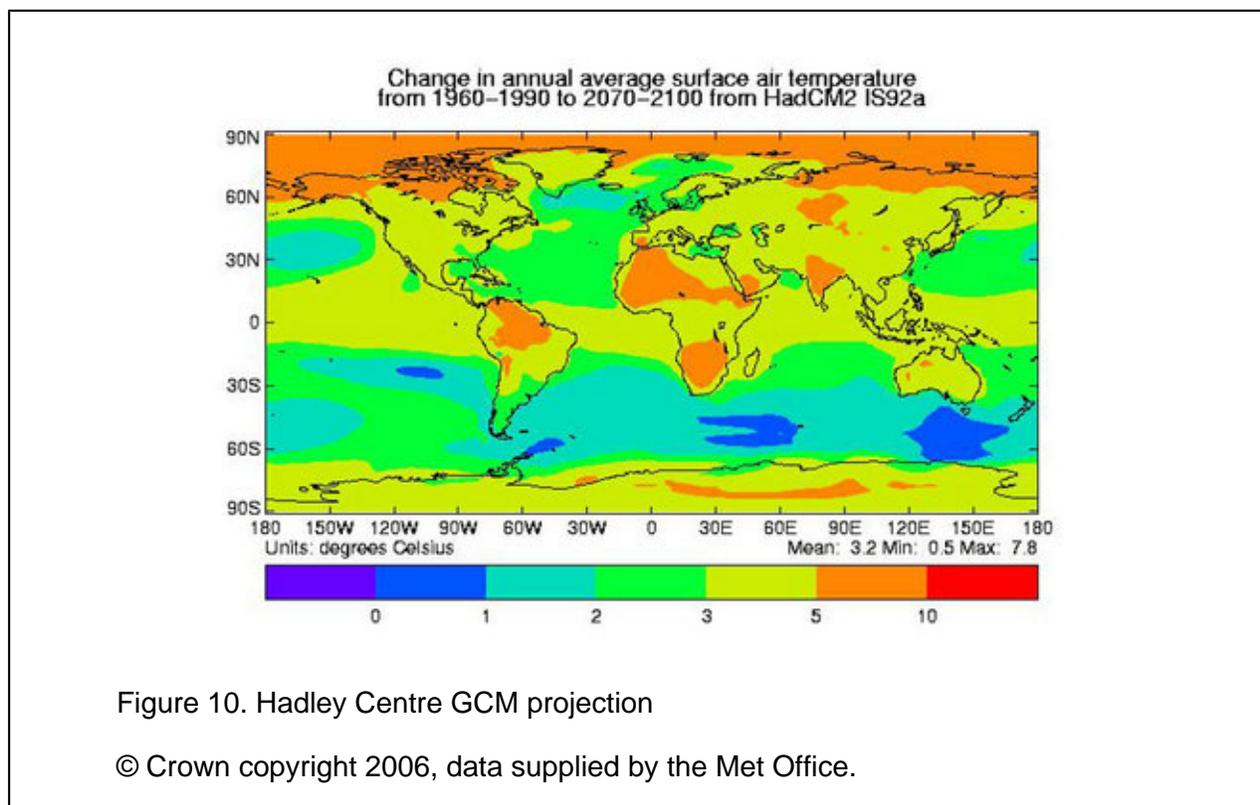
As temperatures rise, snow cover, sea ice, and mountain glaciers are melting. One piece of evidence for a warming world is the fact that tropical glaciers are melting around the globe. Temperatures at high altitudes near the equator are very stable and do not usually fluctuate much between summer and winter, so the fact that glaciers are retreating in areas like Tanzania, Peru, Bolivia, and Tibet indicates that temperatures are rising worldwide. Ice core samples from these glaciers show that this level of melting has not occurred for thousands of years and therefore is not part of any natural cycle of climate variability. Paleoclimatologist Lonnie Thompson of Ohio State University, who has studied tropical glaciers in South America, Asia, and Africa, predicts that glaciers will disappear from Kilimanjaro in Tanzania and Quelccaya in Peru by 2020.

*“The fact that every tropical glacier is retreating is our warning that the system is changing.”*

Lonnie Thompson, Ohio State University

Rising global temperatures are raising sea levels due to melting ice and thermal expansion of warming ocean waters. Global average sea levels rose between 0.12 and 0.22 meters during the 20th century, and global ocean heat content increased. Scientists also believe that rising temperatures are altering precipitation patterns in many parts of the Northern Hemisphere (footnote 7).

Because the climate system involves complex interactions between oceans, ecosystems, and the atmosphere, scientists have been working for several decades to develop and refine General Circulation Models (also known as Global Climate Models), or GCMs, highly detailed models typically run on supercomputers that simulate how changes in specific parameters alter larger climate patterns. The largest and most complex type of GCMs are coupled atmosphere-ocean models, which link together three-dimensional models of the atmosphere and the ocean to study how these systems impact each other. Organizations operating GCMs include the National Aeronautic and Space Administration (NASA)'s Goddard Institute for Space Studies and the United Kingdom's Hadley Centre for Climate Prediction and Research (Fig. 10).



Researchers constantly refine GCMs as they learn more about specific components that feed into the models, such as conditions under which clouds form or how various types of [aerosols](#) scatter light. However, predictions of future climate change by existing models have a high degree of uncertainty because no scientists have ever observed atmospheric CO<sub>2</sub> concentrations at today's levels.

Modeling climate trends is complicated because the climate system contains numerous feedbacks that can either magnify or constrain trends. For example, frozen tundra contains ancient carbon and methane deposits; warmer temperatures may create a positive feedback by melting frozen ground



and releasing CO<sub>2</sub> and methane, which cause further warming. Conversely, rising temperatures that increase cloud formation and thereby reduce the amount of incoming solar radiation represent a negative feedback. One source of uncertainty in climate modeling is the possibility that the climate system may contain feedbacks that have not yet been observed and therefore are not represented in existing GCMs.

Scientific evidence, including modeling results, indicates that rising atmospheric concentrations of CO<sub>2</sub> and other GHGs from human activity are driving the current warming trend. As the previous sections showed, prior to the industrial era atmospheric CO<sub>2</sub> concentrations had not risen above 300 parts per million for several hundred thousand years. But since the mid-18th century CO<sub>2</sub> levels have risen steadily.

In 2007 the [Intergovernmental Panel on Climate Change](#) (IPCC), an international organization of climate experts created in 1988 to assess evidence of climate change and make recommendations to national governments, reported that CO<sub>2</sub> levels had increased from about 280 ppm before the industrial era to 379 ppm in 2005. The present CO<sub>2</sub> concentration is higher than any levels over at least the past 420,000 years and is likely the highest level in the past 20 million years. During the same time span, atmospheric methane concentrations rose from 715 parts per billion (ppb) to 1,774 ppb and N<sub>2</sub>O concentrations increased from 270 ppb to 319 ppb (footnote 8).

Do these rising GHG concentrations explain the unprecedented warming that has taken place over the past century? To answer this question scientists have used climate models to simulate climate responses to natural and anthropogenic forcings. The best matches between predicted and observed temperature trends occur when these studies simulate both natural forcings (such as variations in solar radiation levels and volcanic eruptions) and anthropogenic forcings (GHG and aerosol emissions) (Fig. 11). Taking these findings and the strength of various forcings into account, the IPCC stated in 2007 that Earth's climate was unequivocally warming and that most of the warming observed since the mid-20th century was "very likely" (meaning a probability of more than 90 percent) due to the observed increase in anthropogenic GHG emissions (footnote 9).

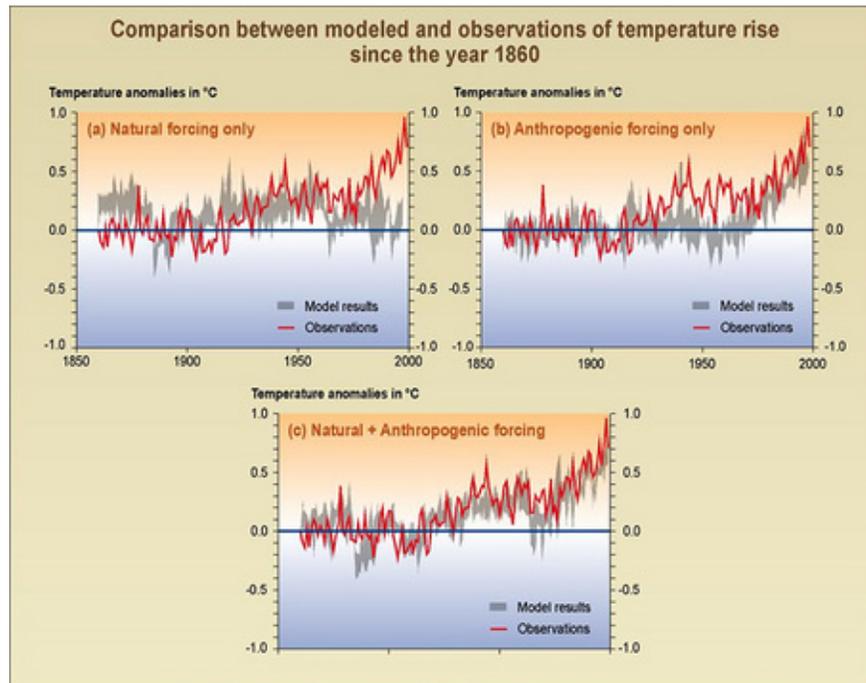


Figure 11. Comparison between modeled and observations of temperature rise since the year 1860

© Intergovernmental Panel on Climate Change, Third Assessment Report, 2001. Working Group 1: The Scientific Basis, Figure 1.1.

Aerosol pollutants complicate climate analyses because they make both positive and negative contributions to climate forcing. As discussed in Unit 11, "Atmospheric Pollution," some aerosols such as sulfates and organic carbon reflect solar energy back from the atmosphere into space, causing negative forcing. Others, like black carbon, absorb energy and warm the atmosphere. Aerosols also impact climate indirectly by changing the properties of clouds—for example, serving as nuclei for condensation of cloud particles or making clouds more reflective.

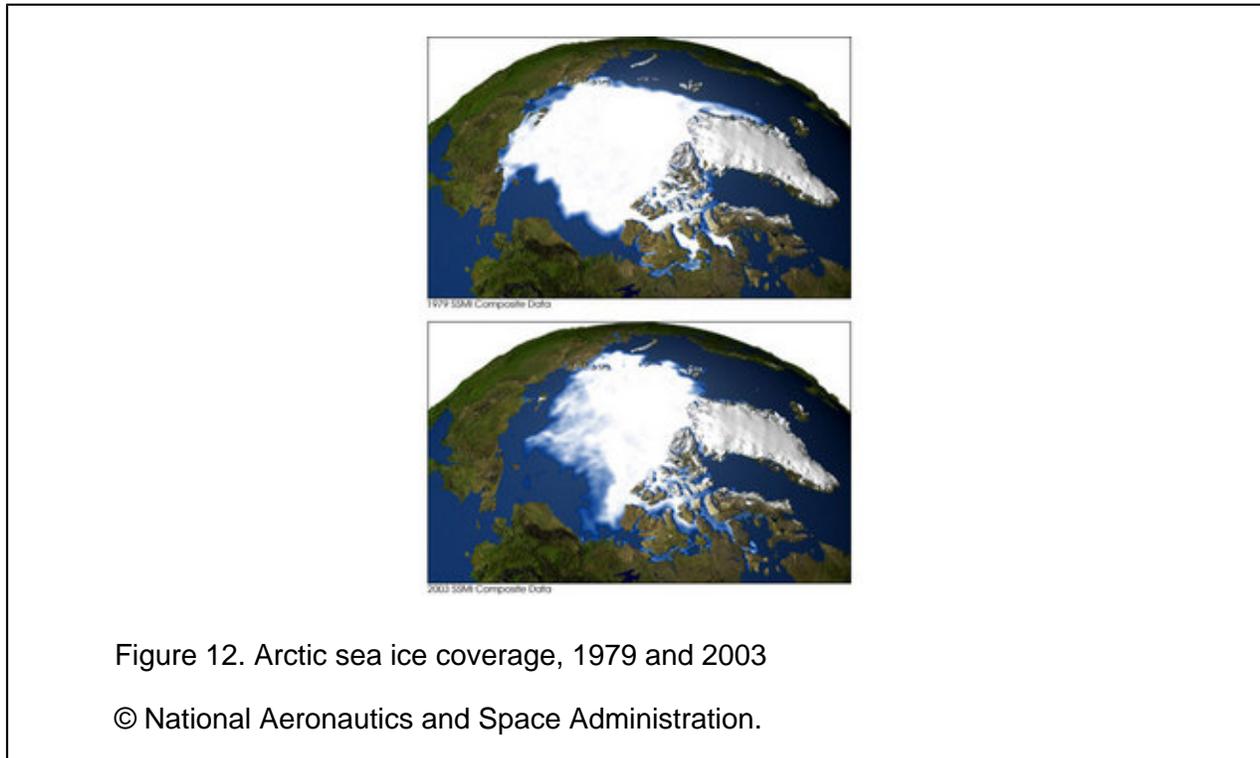
Researchers had trouble explaining why global temperatures cooled for several decades in the mid-20th century until positive and negative forcings from aerosols were integrated into climate models. These calculations and observation of natural events showed that aerosols do offset some fraction of GHG emissions. For example, the 1991 eruption of Mount Pinatubo in the Philippines, which injected 20 million tons of  $\text{SO}_2$  into the stratosphere, reduced Earth's average surface temperature by up to 1.3°F annually for the following three years (footnote 10).

But cooling from aerosols is temporary because they have short atmospheric residence times. Moreover, aerosol concentrations vary widely by region and sulfate emissions are being reduced in most industrialized countries to address air pollution. Although many questions remain to be answered about how various aerosols are formed and contribute to radiative forcing, they cannot be relied on to offset CO<sub>2</sub> emissions in the future.

## 7. Observed Impacts of Climate Change

Human-induced climate change has already had many impacts. As noted above, global average surface temperatures rose by 0.6°C +/- 0.2°C and sea levels rose by 0.12 to 0.22 meters during the 20th century. Other observed changes in Earth systems that are consistent with anthropogenic climate change include:

- Decreases by about two weeks in the duration of ice cover on rivers and lakes in the mid- and high latitudes of the Northern Hemisphere over the 20th century;
- Decreases by 10 percent in the area of snow cover since satellite images became available in the 1960s;
- Thinning by 40 percent of Arctic sea ice in late summer to early autumn in recent decades, and decrease by 10 to 15 percent in extent in spring and summer since the 1950s (Fig. 12);
- Widespread retreat of non-polar glaciers;
- Increases by about 1 to 4 days per decade in growing seasons in the Northern Hemisphere, especially at higher latitudes, during the last 40 years; and
- Thawing, warming, and degrading of [permafrost](#) in some regions .



The Earth is not warming uniformly. Notably, climate change is expected to affect the polar regions more severely. Melting snow and ice expose darker land and ocean surfaces to the sun, and retreating sea ice increases the release of solar heat from oceans to the atmosphere in winter. Trends have been mixed in Antarctica, but the Arctic is warming nearly twice as rapidly as the rest of the world; winter temperatures in Alaska and western Canada have risen by up to 3–4°C in the past 50 years, and Arctic precipitation has increased by about 8 percent over the past century (mostly as rain) (footnote 12).

Observed climate change impacts are already affecting Earth's physical and biological systems. Many natural ecosystems are vulnerable to climate change impacts, especially systems that grow and adapt slowly. For example, coral reefs are under serious stress from rapid ocean warming. Recent **coral bleaching** events in the Caribbean and Pacific oceans have been correlated with rising sea surface temperatures over the past century (footnote 13). Some natural systems are more mobile. For example, tree species in New England such as hemlock, white pine, maple, beech, and hickory have migrated hundreds of meters per year in response to warming and cooling phases over the past 8,000 years (footnote 14). But species may not survive simply by changing their ranges if other important factors such as soil conditions are unsuitable in their new locations.



Insects, plants, and animals may respond to climate change in many ways, including shifts in range, alterations of their hibernation, migrating, or breeding cycles, and changes in physical structure and behavior as temperature and moisture conditions alter their immediate environments. A recent review of more than 40 studies that assessed the impacts of climate change on U.S. ecosystems found broad impacts on plants, animals, and natural ecosystem processes. Important trends included:

- Earlier spring events (emergence from hibernation, plant blooming, and onset of bird and amphibian breeding cycles);
- Insect, bird, and mammal range shifts northward and to higher elevations; and
- Changes in the composition of local plant and animal communities favoring species that are better adapted to warming conditions (higher temperatures, more available water, and higher CO levels).

Because many natural ecosystems are smaller, more isolated, and less genetically diverse today than in the past, it may be increasingly difficult for them to adapt to climate change by migrating or evolving, the review's authors concluded (footnote 15). This is especially true if climate shifts happen abruptly so that species have less response time, or if species are adapted to unique environments (Fig. 13).



Figure 13. Polar bear hunting on Arctic sea ice

© Greenpeace/Beltra.

## 8. Other Potential Near-Term Impacts

In its 2007 assessment report the IPCC projected that global average surface temperatures for the years 2090 to 2099 will rise by 1.1 to 6.4°C over values in 2001 to 2010. The greatest temperature increases will occur over land and at high northern latitudes, with less warming over the southern oceans and the North Atlantic (footnote 16).

This rate of warming, driven primarily by fossil fuel consumption, would be much higher than the changes that were observed in the 20th century and probably unprecedented over at least the past 10,000 years. Based on projections like this, along with field studies of current impacts, scientists forecast many significant effects from global climate change in the next several decades, although much uncertainty remains about where these impacts will be felt worldwide and how severe they will be.

Climate change is likely to alter hydrologic cycles and weather patterns in many ways, such as shifting storm tracks, increasing or reducing annual rainfall from region to region, and producing more extreme weather events such as storms and droughts (Fig. 14). While precipitation trends vary widely over time and area, total precipitation increased during the 20th century over land in high-latitude regions of the Northern Hemisphere and decreased in tropical and subtropical regions (footnote 17).



Figure 14. Flooding in New Orleans after Hurricane Katrina, 2005

© National Oceanic and Atmospheric Administration.

Rising temperatures and changing hydrological cycles are likely to have many impacts, although it is hard to predict changes in specific regions—some areas will become wetter and some dryer. Storm tracks may shift, causing accustomed weather patterns to change. These changes may upset natural ecosystems, potentially leading to species losses. They also could reduce agricultural productivity if new temperature and precipitation patterns are less than optimal for major farmed crops (for example, if rainfall drops in the U.S. corn belt). Some plant species may migrate north to more suitable ecosystems—for example, a growing fraction of the sugar maple industry in the northeastern United States is already moving into Canada—but soils and other conditions may not be as appropriate in these new zones.

Some natural systems could benefit from climate change at the same time that others are harmed. Crop yields could increase in mid-latitude regions where temperatures rise moderately, and winter conditions may become more moderate in middle and high latitudes. A few observers argue that rising CO<sub>2</sub> levels will produce a beneficial global "greening," but climate change is unlikely to increase overall global productivity. Research by Stanford University ecologist Chris Field indicates that

elevated CO<sub>2</sub> prevents plants from increasing their growth rates, perhaps by limiting their ability to utilize other components that are essential for growth such as nutrients. This finding suggests that terrestrial ecosystems may take up less carbon in a warming world than they do today, not more.

Undesirable species may also benefit from climate change. Rising temperatures promote the spread of mosquitoes and other infectious disease carriers that flourish in warmer environments or are typically limited by cold winters (Fig. 15). Extreme weather events can create conditions that are favorable for disease outbreaks, such as loss of clean drinking water and sanitation systems. Some vectors are likely to threaten human health, while others can damage forests and agricultural crops.

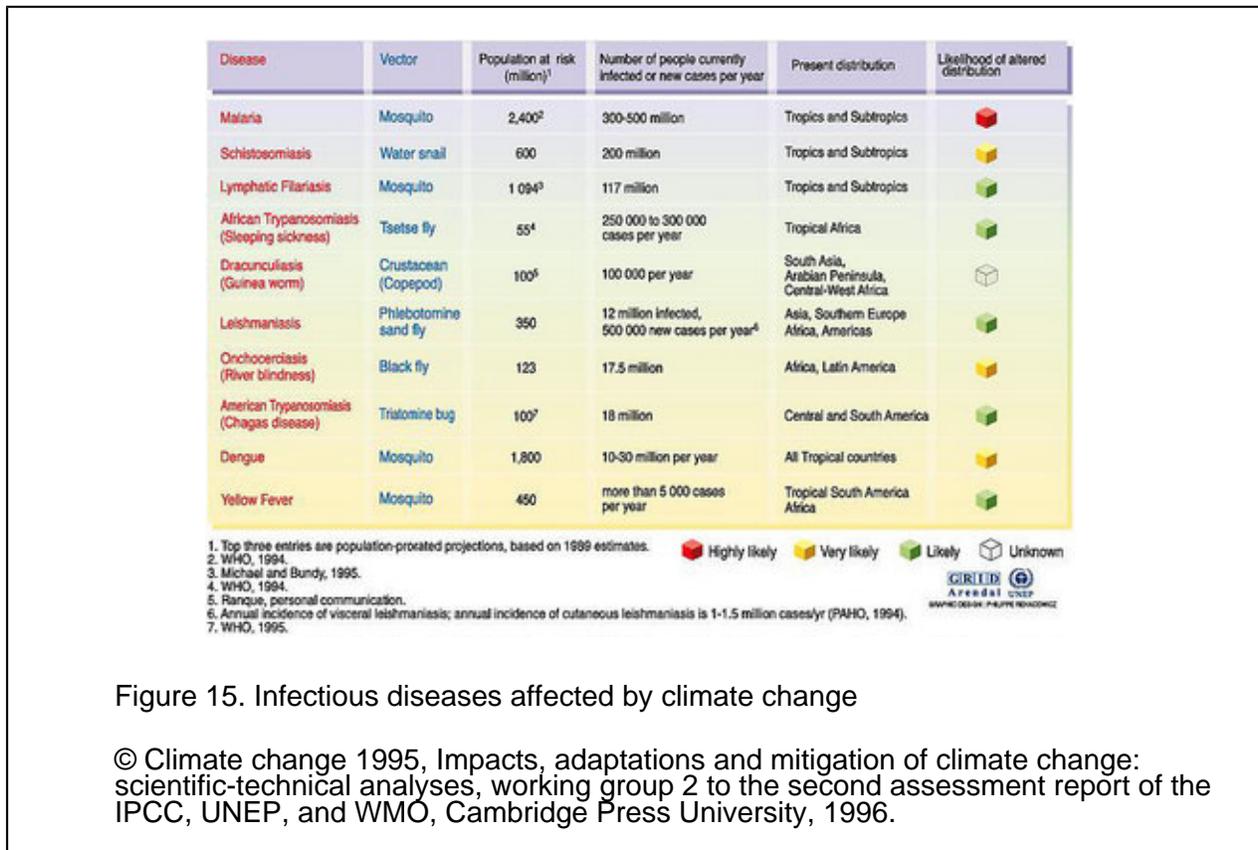


Figure 15. Infectious diseases affected by climate change

© Climate change 1995, Impacts, adaptations and mitigation of climate change: scientific-technical analyses, working group 2 to the second assessment report of the IPCC, UNEP, and WMO, Cambridge Press University, 1996.

Melting of polar ice caps and glaciers is already widespread and is expected to continue throughout this century. Since the late 1970s Arctic sea ice has decreased by about 20 percent; in the past several years, this ice cover has begun to melt in winter as well as in summer, and some experts predict that the Arctic could be ice-free by 2100. Ice caps and glaciers contain some 30 million cubic kilometers of water, equal to about 2 percent of the volume of the oceans. Further melting of sea ice will drive continued sea-level rise and increase flooding and storm surge levels in coastal regions.

Warmer tropical sea surface temperatures are already increasing the intensity of hurricanes, and this trend may accelerate as ocean temperatures rise (footnote 18). Stronger storms coupled with rising sea levels are expected to increase flooding damage in coastal areas worldwide. Some scientists predict that extreme weather events, such as storms and droughts, may become more pronounced, although this view is controversial. In general, however, shifting atmospheric circulation patterns may deliver "surprises" as weather patterns migrate and people experience types of weather that fall outside their range of experience, such as flooding at a level formerly experienced only every 50 or 100 years.

Human societies may already be suffering harmful impacts from global climate change, although it is important to distinguish climate influences from other socioeconomic factors. For example, financial damages from storms in the United States have risen sharply over the past several decades, a trend that reflects both intensive development in coastal areas and the impact of severe tropical storms in those densely populated regions. Human communities clearly are vulnerable to climate change, especially societies that are heavily dependent on natural resources such as forests, agriculture, and fishing; low-lying regions subject to flooding; water-scarce areas in the subtropics; and communities in areas that are subject to extreme events such as heat episodes and droughts. In general, developed nations have more adaptive capacity than developing countries because wealthier countries have greater economic and technical resources and are less dependent on natural resources for income.

And more drastic changes may lie in store. As discussed above, climate records show that the climate can swing suddenly from one state to another within periods as short as a decade. A 2002 report by the National Research Council warned that as atmospheric GHG concentrations rise, the climate system could reach thresholds that trigger sudden drastic shifts, such as changes in ocean currents or a major increase in floods or hurricanes (footnote 19).

*“Just as the slowly increasing pressure of a finger eventually flips a switch and turns on a light, the slow effects of drifting continents or wobbling orbits or changing atmospheric composition may ‘switch’ the climate to a new state.”*

Richard B. Alley, Chair, Committee on Abrupt Climate Change, National Research Council

How much the planet will warm in the next century, and what kind of impacts will result, depends on how high CO<sub>2</sub> concentrations rise. In turn, this depends largely on human choices about fossil fuel consumption. Because fossil fuel accounts for 80 percent of global energy use, CO<sub>2</sub> levels will continue to rise for at least the next 30 or 40 years, so additional impacts are certain to be felt. This means that it is essential both to mitigate global climate change by reducing CO<sub>2</sub> emissions and to adapt to the changes that have already been set in motion. (For more on options for mitigating and adapting to climate change, see Unit 13, "Looking Forward: Our Global Experiment.")



## 9. Major Laws and Treaties

Science plays a central role in international negotiations to address global climate change. In 1988, the World Meteorological Organization and the United Nations Environment Programme established the Intergovernmental Panel on Climate Change (IPCC), an organization composed of official government representatives that is charged with assessing scientific, technical, and socio-economic information relevant to understanding climate change risks, potential impacts, and mitigation and adaptation options (footnote 20). The IPCC meets regularly to review and assess current scientific literature and issues "assessment reports" at approximately five-year intervals (most recently in 2007). IPCC reports are adopted by consensus and represent a broad cross-section of opinion from many nations and disciplines regarding current understanding of global climate change science. The panel's recommendations are not binding on governments, but its models and estimates are important starting points for international climate change negotiations.

The most broadly-supported international agreement on climate change, the [United Nations Framework Convention on Climate Change](#) (FCCC), was opened for signature in 1992 and entered into force in 1994 (footnote 21). To date it has been ratified by 189 countries, including the United States. FCCC signatories pledge to work toward stabilizing atmospheric GHG concentrations "at a level that would prevent dangerous anthropogenic interference with the climate system," but the Convention does not define that level. As a result, it has not been a significant curb on GHG emissions, although it creates a system for nations to report emissions and share other relevant information and for developed countries to provide financial and technical support for climate change initiatives to developing countries.

Recognizing that the FCCC commitments were not sufficient to prevent serious climate change, governments negotiated the [Kyoto Protocol](#), which commits industrialized countries to binding GHG emission reductions of at least 5 percent below their 1990 levels by the period of 2008–2012 (footnote 22). The Protocol focuses on developed countries in reflection of the fact that they are the source of most GHGs emitted to date, although it allows developed countries to fulfill their reduction commitments partially through projects to reduce or avoid GHG reductions in developing countries.

The Kyoto Protocol entered into force in 2005 and has been ratified to date by 163 countries, representing 61.6 percent of developed countries' GHG emissions. The United States signed the Protocol but has not ratified it. President George W. Bush argued that the economic impact of its assigned reductions (7 percent below 1990 levels) would be too severe and instead emphasized voluntary domestic reduction commitments.

For all of the controversy that it has generated, the Kyoto Protocol alone will not reduce the threat of major climate change because it covers only 40 percent of global GHG emissions without U.S. participation, does not require emission reductions from rapidly developing countries, like India and China, that are major fossil fuel consumers, and only covers emission through the year 2012. No single option has emerged yet as a follow-on, but analysts widely agree that the next phase of global



action against climate change will have to take a longer-term approach, address the costs of reducing GHG emissions, and find ways to help developing countries reap the benefits of economic growth on a lower-carbon pathway than that which industrialized countries followed over the past 150 years. Continually improving our scientific understanding of climate change and its impacts will help nations to identify options for action.

## 10. Further Reading

Center for Health and the Global Environment, Harvard Medical School, **Climate Change Futures: Health, Ecological, and Economic Dimensions** (Boston, MA: Harvard Medical School, November 2005).

Climate Timeline (<http://www.ngdc.noaa.gov/paleo/ctl/index.html>). This web site, maintained by the National Geophysical Data Center at the National Oceanographic and Atmospheric Administration, links changes in weather and climate through history to specific historic events.

"The Discovery of Global Warming," (<http://www.aip.org/history/climate>). Created by Spencer Weart, author of the book of the same title, this site includes detailed essays on the history of climate change science, case studies, and links to relevant scientific and historical publications.

James E. Hansen, "Can We Still Avoid Dangerous Human-Made Climate Change?" February 10, 2006, [http://www.columbia.edu/~jeh1/newschool\\_text\\_and\\_slides.pdf](http://www.columbia.edu/~jeh1/newschool_text_and_slides.pdf). In this speech and accompanying slides, a leading U.S. climate scientist makes the case for action to slow global climate change.

## Footnotes

1. Water vapor contributes to climate change through an important positive feedback loop: as the atmosphere warms, evaporation from Earth's surface increases and the atmosphere becomes able to hold more water vapor, which in turn traps more thermal energy and warms the atmosphere further. It also can cause a negative feedback when water in the atmosphere condenses into clouds that reflect solar radiation back into space, reducing the total amount of energy that reaches Earth. For more details, see National Oceanic and Atmospheric Administration, "Greenhouse Gases: Frequently Asked Questions," <http://lwf.ncdc.noaa.gov/oa/climate/gases.html>.

2. **Key GHG Data: Greenhouse Gas Emissions Data for 1990–2003 Submitted To the United Nations Framework Convention on Climate Change** (Bonn: United Nations Framework Convention on Climate Change, November 2005), pp. 16, 28.

3. U.S. Environmental Protection Agency, "The U.S. Inventory of Greenhouse Gas Emissions and Sinks: Fast Facts," April 2006, <http://yosemite.epa.gov/oar/globalwarming.nsf/content/ResourceCenterPublicationsGHGEmissions.html>.



4. John A. Higgins and Daniel P. Schrag, "Beyond Methane: Towards a Theory for the Paleocene-Eocene Thermal Maximum," **Earth and Planetary Science Letters**, vol. 245 (2006), pp. 523–537.
5. National Oceanographic and Atmospheric Administration, Paleoclimatology Branch, "Astronomical Theory of Climate Change," <http://www.ncdc.noaa.gov/paleo/milankovitch.html>; Spencer R. Weart, **The Discovery of Global Warming** (Cambridge, MA: Harvard University Press, 2003), pp. 74–77.
6. "Abrupt Climate Change," Lamont-Doherty Earth Observatory, Columbia University, <http://www.ldeo.columbia.edu/res/pi/arch/examples.shtml>.
7. Intergovernmental Panel on Climate Change, **Climate Change 2007: The Scientific Basis**, Summary for Policymakers (Cambridge, UK: Cambridge University Press, 2007), pp. 4–6.
8. Ibid., pp. 2–3.
9. Ibid., p. 8.
10. U.S. Geological Survey, "Impacts of Volcanic Gases on Climate, The Environment, and People," May 1997, <http://pubs.usgs.gov/of/1997/of97-262/of97-262.html>.
11. IPCC, **Climate Change 2001: Synthesis Reports**, Summary for Policymakers (Cambridge, UK: Cambridge University Press, 2001), p. 6.
12. ACIA, **Impacts of a Warming Arctic: Arctic Climate Impact Assessment** (Cambridge, UK: Cambridge University Press, 2004), p. 12.
13. J.E. Weddell, ed., **The State of Coral Reef Ecosystems of the United States and Pacific Freely Associated States, 2005**, NOAA Technical memorandum NOS NCCOS 11 (Silver Spring, MD: NOAA/NCCOS Center for Coastal Monitoring and Assessment's Biogeography Team, 2005), pp. 13–15, [http://ccma.nos.noaa.gov/ecosystems/coralreef/coral\\_report\\_2005/](http://ccma.nos.noaa.gov/ecosystems/coralreef/coral_report_2005/).
14. David R. Foster and John D. Aber, eds., **Forests in Time: The Environmental Consequences of 1,000 Years of Change in New England** (New Haven: Yale University Press, 2004), pp. 45–46.
15. Camille Parmesan and Hector Galbraith, **Observed Impacts of Global Climate Change in the U.S.** (Arlington, VA: Pew Center on Global Climate Change, 2004), [http://www.pewclimate.org/global-warming-in-depth/all\\_reports/observedimpacts/index.cfm](http://www.pewclimate.org/global-warming-in-depth/all_reports/observedimpacts/index.cfm).
16. IPCC, **Climate Change 2007: The Scientific Basis**, p. 749.
17. United Nations Environment Programme, "Observed Climate Trends," <http://www.grida.no/climate/vital/trends.htm>.
18. Kerry Emanuel, "Increasing Destructiveness of Tropical Cyclones Over the Past 30 Years," **Nature**, vol. 436, August 4, 2005, pp. 686–88, and "Anthropogenic Effects on Tropical Cyclone Activity," <http://wind.mit.edu/~emanuel/anthro2.htm>.
19. National Research Council, **Abrupt Climate Change: Inevitable Surprises** (Washington, DC: National Academy Press, 2002).

20. <http://www.ipcc.ch/>.

21. [http://unfccc.int/essential\\_background/convention/items/2627.php](http://unfccc.int/essential_background/convention/items/2627.php).

22. [http://unfccc.int/essential\\_background/kyoto\\_protocol/items/2830.php](http://unfccc.int/essential_background/kyoto_protocol/items/2830.php).

## Glossary

**aerosols** : Liquid or solid particles that are suspended in air or a gas. Also referred to as particulate matter.

**albedo** : The fraction of electromagnetic radiation reflected after striking a surface.

**anthropogenic** : Describing effects or processes that are derived from human activities, as opposed to effects or processes that occur in the natural environment without human influences.

**coral bleaching** : Refers to the loss of color of corals due to stress-induced expulsion of symbiotic, unicellular algae called zooxanthellae that live within their tissues. Stress can be induced by: increased water temperatures (often attributed to global warming), starvation caused by a decline in zooplankton levels as a result of overfishing, solar irradiance (photosynthetically active radiation and ultraviolet band light), changes in water chemistry, silt runoff, or pathogen infections.

**deforestation** : Removal of trees and other vegetation on a large scale, usually to expand agricultural or grazing lands.

**global warming potential** : A measure of how much a given mass of greenhouse gas is estimated to contribute to global warming. Compares the gas in question to that of the same mass of carbon dioxide.

**Intergovernmental Panel on Climate Change (IPCC)** : Established in 1988 by two United Nations organizations to assess the risk of human-induced climate change.

**Kyoto Protocol** : An amendment to the international treaty on climate change, assigning mandatory targets for the reduction of greenhouse gas emissions to signatory nations.

**paleoclimate** : Referring to past climates of the Earth.

**permafrost** : Soil that stays in a frozen state for more than two years in a row.

**radiocarbon dating** : A radiometric dating method that uses the naturally occurring isotope carbon-14 to determine the age of carbonaceous materials up to about 60,000 years.

**residence time** : A broadly useful concept that expresses how fast something moves through a system in equilibrium; the average time a substance spends within a specified region of space, such as a reservoir. For example, the residence time of water stored in deep groundwater, as part of the water cycle, is about 10,000 years.



**sinks** : Habitats that serve to trap or otherwise remove chemicals such as plant nutrients, organic pollutants, or metal ions through natural processes.

**United Nations Framework Convention on Climate Change** : A treaty signed by nations at the Earth Summit in 1992 to stabilize and reduce greenhouse gas emissions. In 1997 the Kyoto Protocol, an agreement among 150 nations, was added, setting specific reduction levels.

## Unit 13 : Looking Forward: Our Global Experiment



View of Earth from the Moon. Source: NASA

### Overview

Emerging technologies offer potential solutions to environmental problems. Over the long-term, human ingenuity may ensure the survival not only of our own species but of the complex ecosystems that enhance the quality of human life. In this unit, examine the wide range of efforts now underway to mitigate the worst effects of man-made environmental change, looking toward those that will have a positive impact on the future of our habitable planet.

### Sections:

1. Introduction
2. Measuring (and Reducing) the Human Footprint
3. Multiple Stresses on Interconnected Systems
4. Confronting the Climate-Energy Challenge
5. Further Reading



## 1. Introduction

The preceding units have described how humans have both affected and been affected by the Earth system. Over the next century, human society will have to confront these changes as the scale of environmental degradation reaches a planetary scale. As described throughout this course, human appropriation of natural resources—land, water, fish, minerals, and fossil fuels—has profoundly altered the natural environment. Many scientists fear that human activities may soon push the natural world past any number of tipping points—critical points of instability in the natural Earth system that lead to an irreversible (and undesirable) outcome. This chapter discusses how environmental science can provide solutions to some of our environmental challenges. Solving these challenges does not mean avoiding environmental degradation altogether, but rather containing the damage to allow human societies and natural ecosystems to coexist, avoiding some of the worst consequences of environmental destruction.

It would be impossible to address the question of how human society will deal with environmental challenges in the future without realizing that people make decisions not just based on science, but more often based on economic and political considerations. This is not the focus of this course, and so it will not be discussed here. Instead, this chapter will examine some of the scientific constraints on our environmental challenges over the next century that will guide decision making into the future. In addition, some of the strategies discussed here depend on technological developments that cannot be anticipated. Environmental science cannot predict the future, as the future depends on technological and economic choices that will be made over the next century. However, environmental science can help us make better choices, using everything we know about the Earth system to anticipate how different choices will lead to different outcomes. A discussion of some of those outcomes is presented here.

## 2. Measuring (and Reducing) the Human Footprint

Population growth and economic development over this century present many different environmental challenges. As described in previous units, some are local in scale, such as certain types of water pollution. Some are regional in scale, such as [acid rain](#). And some are global, such as climate change. In all cases, one simple strategy to minimize harmful impacts on the natural environment is to reduce the human "footprint" on the environment—although this simple concept is sometimes quite complicated to apply to a specific environmental problem.

At some basic level, most of our environmental challenges are related to the rapid increase in human population. As discussed in Unit 5, "Human Population Dynamics," demographers estimate that global population will increase through the middle of this century to approximately 9 billion and may stabilize or even decrease after that. A simple way of thinking about how to solve environmental problems—from atmospheric pollution to climate change—is in the context of how an individual appropriates some of the natural environment for his or her own needs. Addressing climate change

can be discussed in terms of how much carbon dioxide and other **greenhouse gases** are emitted by each person; habitat loss can be discussed in terms of how much land each person requires to extract food and other services; air pollution can be discussed in terms of the amount of pollutants each person emits, etc. In this framework, population growth can be seen as a primary driver of environmental degradation, as the footprint of human society will increase in direct proportion to the number of people. In the purest sense, one's ecological footprint refers to how much land is required to support one's various activities (Fig. 1). However, the concept of a footprint is often used in a more general sense, applied not only to the amount of land, but also to water use, pollution emitted, etc.

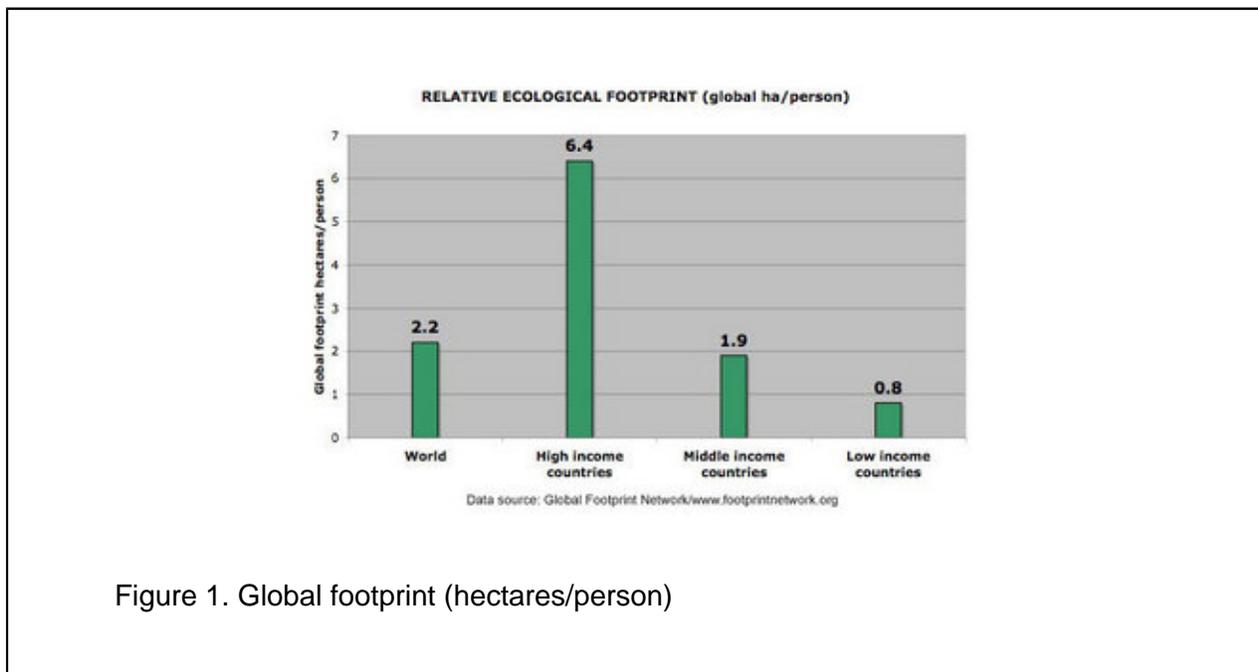


Figure 1. Global footprint (hectares/person)

For some issues, like water resources, calculating one's footprint is relatively straightforward. Per capita water use depends on dietary choices as discussed in Unit 8, "Water Resources," as well as on water use for sanitation, drinking, and other purposes, so calculating the average water needs for a particular society is quite feasible.

Because there is more than one cause for biodiversity loss, quantifying one's footprint is more complicated. As discussed in Unit 9, "Biodiversity Decline," many fish species are declining in number because of human fishing (or overfishing). For other species, such as those that live in the tropical rainforest, a major threat is the destruction of habitat. Still other species are threatened by toxic pollution. Quantifying exactly how one person affects the decline in biodiversity is therefore a much more complicated affair.

Calculating the impact of the human footprint on climate change brings other complications. At a basic level, one can calculate how much fossil fuel an individual uses and therefore how much carbon dioxide is emitted. However, greenhouse gases are produced not only when we use energy directly but also when we buy products that require energy to make them, from a new house or car to fresh produce that require energy for transportation. This is also an issue at a national scale. For example, the carbon dioxide emission footprint of a country like the United States only includes the fossil fuel that is actually used in the United States, but excludes the energy that is used to make products in other countries that are then shipped to American consumers.

Considering one's environmental footprint—however it is calculated—leads to a fundamental tension between economic development and environmental impacts. As discussed above, population growth is at the root of many environmental problems. But population growth is not the only driver of environmental degradation, and perhaps not even the primary one. It is true that many environmental problems would be much easier to solve if the population were much smaller, but over the next 50 years, demographers predict that the world population will increase another only 50 percent or so and then will start to decline. In comparison, human consumption of goods and services—sometimes measured by economists as gross domestic product (GDP) per capita—is predicted to grow by a factor of ten or even more through this century (Fig. 2). What this means is that the footprint of human society is getting larger, partially because the human population is growing (i.e., more individual footprints), but mostly because humans are getting richer, appropriating more and more of the natural environment for their needs, impacting almost every environmental challenge discussed in this course.

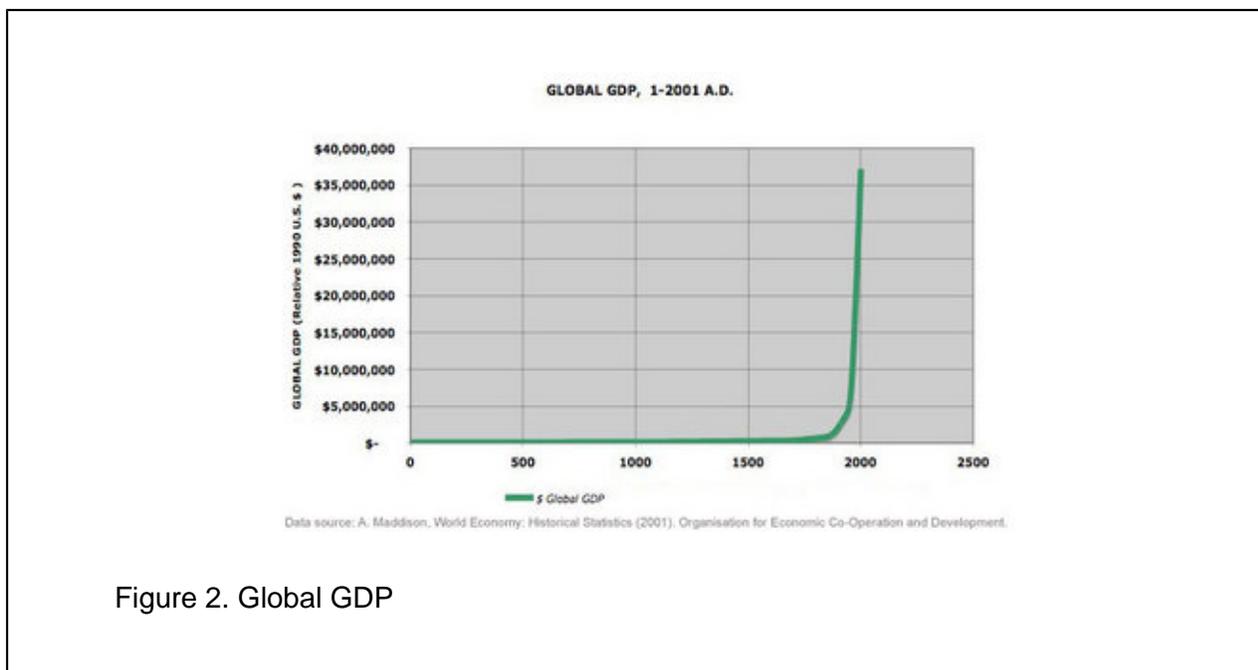


Figure 2. Global GDP

Some people believe a strategy for fixing environmental problems involves restraining economic growth, reducing the human footprint on the environment by using less of the natural world. In many cases, this can be accomplished without reducing the quality of human life. For example, there are many ways to conserve water or electricity that do not sacrifice quality of life. However, preserving the environment is unlikely to happen simply at the expense of economic development. Economic development leads to better quality of life for people all over the world; it raises people up from desperate poverty and gives our societies the capacity to fix many of the environmental challenges. So how can we increase the quality of human life, encourage economic development, and still protect the environment as human appropriation of the natural world becomes greater and greater? The answer may involve new technology. In some cases, new technologies allow us to reduce our environmental footprint while still providing the goods and services we need, allowing our economic well-being to flourish. A good example is the catalytic converter on automobiles that reduced air pollution, improving human health but still allowing us to drive our cars. New technology may not be a panacea for all environmental problems, but it can help societies balance their needs for economic development with their goals for protecting the environment.

### 3. Multiple Stresses on Interconnected Systems

As we discuss solutions to environmental challenges, we must acknowledge that many environmental problems are interconnected, so we cannot protect one part of the natural environment without considering all the different stresses on the system as a whole. A good example of this is the multiple threats to coral reefs.

Coral reefs occur in a variety of forms throughout tropical and subtropical seas and account for half of the Earth's present production of calcium carbonate (Wood, 1999). The [scleractinian coral](#), the most important Order of reef building corals (also called "stony corals"), contains symbiotic algae within its tissues known as [zooxanthellae](#). While the coral animal is capable of ingesting nutritious particulate material from seawater, an important component of its nutrition is derived from the photosynthetic products of the algae. The algal requirement for light limits the depth to which reef-building corals can live and form new reef complexes.

Coral reef communities are among the most diverse assemblages in the marine environment. Coral reefs occupy less than 0.25 percent of the marine environment but contain more than 25 percent of marine fishes, and in terms of total species diversity, coral reefs are the marine analogue of tropical rainforests. Because coralline skeletal material provides the physical substrate for the reef, the loss of the coral makes the balance of the reef community vulnerable.

Many studies have documented the decline in coral reefs around the world. However, identifying the primary cause of coral reef decline is very difficult because there are so many different environmental stresses on coral reef ecosystems. First, there is the direct destruction of coral reefs by anchoring. Most species of corals grow very slowly—usually around 1 cm of extension per year. This means that when fragments of coral are destroyed by ship anchors, even when the destruction is relatively minor,

the cumulative effect of many different ships over several years can ultimately destroy a reef because the reef requires so much time to repair the damage.

A more extreme way of destroying the reef is through dynamite fishing. Dynamite fishing involves setting off charges in the water that stun or kill fish, making them easy to gather with basic skin-diving equipment. Although it is banned in most tropical countries, it is still quite common, especially in poor regions with limited access to deep-water fisheries. A by-product of the blast that kills the fish is the total demolition of coral within many meters of the blast. Left behind is a pile of coral rubble, unsuitable for supporting the diverse communities of organisms that live in a healthy coral reef ecosystem.

Another threat to coral reefs is overfishing. Some of this fishing is not even for food, but for live tropical fish for aquaria in homes. In addition, even [pelagic](#) fishing can cause problems for coral reefs because marine food webs are very complicated. Depletion of one species—large, predatory fish, for example—can lead to unforeseen consequences resulting in other species collapsing, ultimately affecting the coral reef ecosystem.

Coral reefs are also threatened by human land use. In many tropical regions, development near the coast has led to an increase in erosion of soil, which can kill coral reefs either by the direct effect of terrestrial soil material on the coral or by adding excess nutrients, which stimulate algal populations that can outcompete the stationary corals for light. Other forms of pollution associated with development are also a problem for coral reefs.

The final threat to coral reefs comes from climate change, although there are two separate impacts. Changes in ocean chemistry resulting from higher CO<sub>2</sub> levels are likely to be more serious threats to the health of coral reef communities. Aragonite—the mineral used by corals for their skeleton—is supersaturated in seawater today by about 400 percent. This will decline as CO<sub>2</sub> concentration in the atmosphere rises, with corresponding reduction in pH. Calculations suggest that aragonite saturation will decline by approximately 30 percent at atmospheric CO<sub>2</sub> concentrations twice the pre-industrial level, and this will lead to lower calcification rates for corals. It is possible that reduced rates of calcification will make reef corals more susceptible to storm damage.

An additional threat to coral reefs from climate change is a condition known as “[coral bleaching](#),” which occurs when the corals lose their symbiotic algae in response to environmental stress (Fig. 3). Some corals do recover following brief periods of bleaching, although the means by which the algae become reestablished is highly speculative. If this fails to happen, the coral tissue dies, leaving the calcareous reef substratum exposed to physical damage and dissolution. Experiments have shown that this condition can be caused by elevated temperatures, reduced salinity, and excessive suspended fine particulate matter, and one or more of these factors has been associated with numerous observed bleaching events. There is also evidence that at elevated temperatures virulence of bacterial [pathogens](#) of corals may increase and that these may be involved in the bleaching process.



Figure 3. Coral reef after a bleaching event

© 2003. Reef Futures. Courtesy Ray Berkelmans, Australian Institute of Marine Science.

Corals in today's tropical and subtropical oceans are very near their upper limits for temperature (some within 2°C) during the warm seasons of the year. The response of different species of coral to warmer temperatures is probably sensitive to both the magnitude of the increment of temperature and the rate at which this increase is experienced. Bleaching that isn't necessarily fatal can occur in response to an increase as small as 1°C above normal seasonal maxima. There is evidence that thermal anomalies greater than 3°C are fatal to several coral species. With the death of coral tissue, the reef substrate is subject to erosion from physical and dissolution processes and colonization by other organisms, especially seaweeds.

With so many different threats to coral reefs, how can they be protected? The challenge is to solve many of the different threats simultaneously. Protecting coastal marine ecosystems by setting aside marine preserves will not solve the issue by itself. This is because under predicted climate change conditions, raised carbon dioxide levels and warmer temperatures will destroy reefs even in protected areas.

This basic problem can be generalized to many types of environmental issues—in particular, the relationship between biodiversity loss and climate change. The primary strategy for protecting endangered species has been to set aside natural habitat, either as national parklands or wilderness areas. However, climate change threatens to undo much of the good work accomplished by conservation efforts, as the same barriers to that keep people and development out of these natural habitats also serve to prevent many species of plants and animals from migrating to preferred climate

zones as the climate changes. Isolating natural ecosystems into specific protected areas bounded by agricultural or urban areas means that migration of these ecosystems in response to climate change becomes impossible. Thus, if we cannot avoid the most extreme climate change scenarios, many of the conservation efforts will fail. This does not mean that preservation of habitat is not important. Human appropriation of land continues to be the major threat to biodiversity, particularly in tropical forests. However, conservation is not enough when faced with the grand challenge of global climate change.

#### 4. Confronting the Climate-Energy Challenge

The realization that humans are changing the Earth's climate is profound, and yet it is only one of many ways in which humans are changing the physical and biological environment. As discussed in Unit 1, "Many Planets, One Earth," the arrival of humans to the Americas at the end of the last ice age, approximately 14,000 years ago, was accompanied by the extinction of most large mammals, including mammoths and mastodons, presumably from excessive hunting. As discussed in Unit 9, "Biodiversity Decline," human land-use has caused an enormous reduction in biodiversity, as sensitive ecosystems such as wetlands, tropical rainforests, and coral reefs are encroached or destroyed by human activities. Confronted with these and other impacts of human activities, why is **anthropogenic** climate change so troubling? The answer is that climate change has the potential to make many of the other environmental challenges much more difficult to solve because of the global scale of the impacts and the huge magnitude of the change relative to what the Earth system has experienced over the last tens of millions of years.

Given the dramatic changes we are observing today in the climate system, coupled with our view from Earth history, what can be done? The first challenge we must confront in working toward a solution to future climate change is that any "solution" will be incomplete. Some amount, perhaps even a substantial amount, of climate change is unavoidable. Reducing CO<sub>2</sub> emissions so that they are below the level of CO<sub>2</sub> uptake by the oceans and biosphere will not happen in a decade or two, but only through prolonged actions over 50 years and perhaps longer. In addition, the oceans would continue to warm for decades even if emissions were halted. Ecological changes due to climate change that has already occurred will continue to unfold for decades. CO<sub>2</sub> resides in the atmosphere and surface ocean for centuries and is only slowly taken up by the deep ocean. If we were to reduce our emissions to zero immediately, it would take more than 200 years for terrestrial and oceanic uptake of carbon to restore the atmosphere to its pre-industrial condition. Thus, there is great momentum in the climate system, in the heat capacity of the oceans, in ice sheets, and in the residence time of carbon dioxide in the atmosphere, and this fact makes a certain amount of climate change inevitable. Future impacts discussed in Unit 12, "Earth's Changing Climate," include sea level rise, changes in rainfall patterns, early melting of mountain snow pack and glaciers (which serve as the primary water supply for billions of people), changes in storms and tropical cyclones, and other ecological changes that affect ecosystems crucial to human society.

One source of confusion in discussions on how to reduce CO<sub>2</sub> emissions is that our energy system is really more than one system. As discussed in Unit 10, "Energy Challenges," we use energy for transportation, for electricity to power our lights and electronics, to heat or cool our homes, for manufacturing, and for agriculture. Our energy choices within each of these sectors come with different technological constraints that require different types of solutions if reductions in CO<sub>2</sub> emissions are to be achieved. For example, the internal combustion engine (along with the gas turbine in airplanes) currently dominates the transportation sector and is fueled almost exclusively by petroleum, making transportation responsible for approximately 40 percent of global CO<sub>2</sub> emissions. The electricity industry has a much broader set of energy sources, including coal, natural gas, nuclear, hydroelectric, wind, solar, biomass, and geothermal—although coal, natural gas, and nuclear are currently dominant. Thus, the discussion of strategies for mitigating climate change must address not just sources of energy, but sources in relationship to different societal needs (Fig. 4).

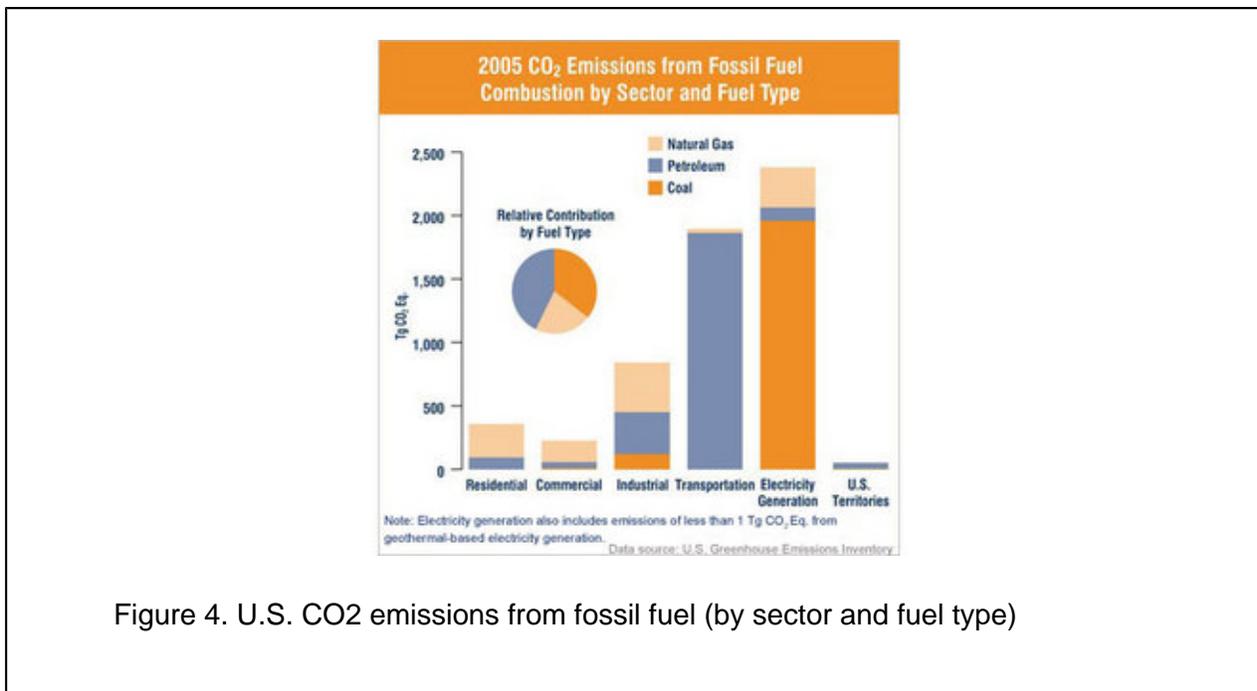
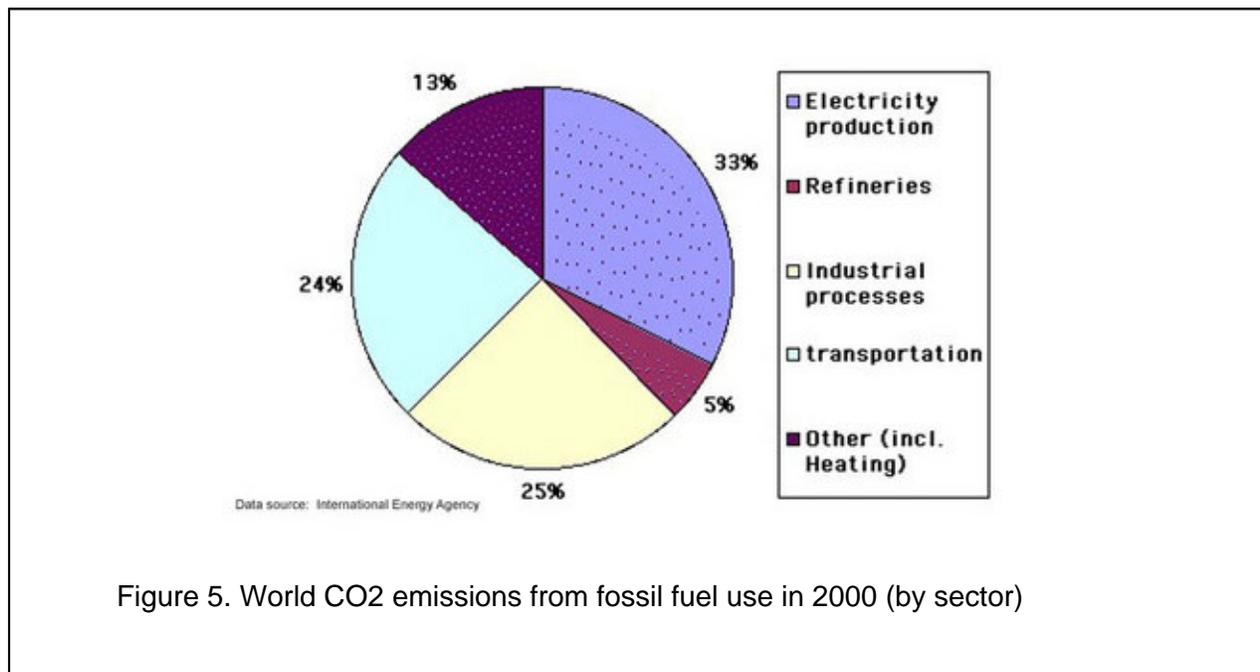


Figure 4. U.S. CO<sub>2</sub> emissions from fossil fuel (by sector and fuel type)

Another consideration is differences in energy technologies among countries. Some countries, such as Saudi Arabia and Russia, are rich in hydrocarbon resources, and this guides their energy decisions. Other countries, such as Japan, have almost no domestic energy resources and turn to technological solutions such as solar and nuclear power. The end use of energy also varies among countries. For example, both China and the United States have large coal reserves, but China consumes almost twice as much, in part due to the large manufacturing industry in China versus the service economy of the U.S. (Fig. 5). This means there can be no single strategy for how the world will address climate change but rather we need a portfolio of strategies. Rapidly developing countries

will have different solutions from those of developed countries. Even countries with similar levels of economic development will employ different solutions because of geography, political and cultural attitudes, and political systems.

This does not mean that individual solutions must be created for all countries. CO<sub>2</sub> emissions are not distributed evenly; a handful of countries contribute most of the emissions and will be responsible for bringing about most of the reductions. If the United States, China, India, the European Union, Russia, Japan, Australia, Canada, and perhaps Indonesia and Brazil each take significant steps to reduce emissions, it is likely that such efforts will be successful in reducing the impacts of climate change on the rest of the world (Fig. 5).



Another constraint is the timescale over which it is possible to build new energy systems. Eliminating carbon emissions from electricity generation by using nuclear power, for example, would require building two large nuclear plants each week for the next 100 years. This rate of change is simply not possible given current constraints on steel production, construction capacity, the education of operators, and many other practical considerations. Taken together with the diverse uses of energy and the different needs of different nations, this means that there is no silver bullet solution for the climate–energy challenge. Myriad approaches are required. One can group these approaches into three broad categories, each of which will play an essential part in any serious climate mitigation effort.

## Reduction of Energy Demand

The first category involves reducing CO<sub>2</sub> emissions by reducing energy consumption, as discussed in section 2, "Measuring (and Reducing) the Human Footprint." This does not necessarily require reducing economic activity, i.e. consuming less (although this can be part of the solution); rather, it means restructuring society, either by investing in low-energy adaptations such as efficient public transportation systems or by adopting energy-efficient technologies in buildings, in automobiles, and throughout the economy.

Huge discrepancies in energy efficiency exist today among developed countries. In general, countries with higher historical energy prices, such as most of Western Europe, are more efficient than countries with inexpensive energy including petroleum, although the differences can also be explained by historical investments in cities and suburbs and in highways and public transportation systems, as well as by a variety of other factors. But whatever the cause of the current differences among countries, there is great potential across the developed and the developing world to dramatically lower energy use through smarter and better energy systems.

Much of the efficiency gains can be accomplished with existing technologies, such as compact-fluorescent lighting or more efficient building designs; these are often referred to as the "low hanging fruit," as they are often economically advantageous because they are simple and inexpensive. In addition, there are technological improvements in end use that would contribute greatly to any emissions reductions effort by making large jumps in energy efficiency. For example, if we can develop batteries for electric automobiles that are economical and reliable, and if their use is broadly adopted, we will be able to replace the low-efficiency internal combustion engine with the high-efficiency electric motor. Moreover, electric cars would break the monopoly that petroleum currently has as the source of energy for transportation, thus addressing security concerns involving the geopolitics of oil and allowing transportation fuel to come from carbon-free sources. Whether better batteries are technically possible remains a question.

## Non-Fossil Energy Systems

The second category of solutions to the climate–energy challenge involves expansion of non-fossil energy systems, including wind, solar, biomass, geothermal, and nuclear power. Again, there is no silver bullet. Wind is currently the most economical of these energy systems for electricity generation. However, wind requires huge excess capacity because of problems with intermittency, and so it cannot become a source for base load power unless storage technologies improve. Solar-generated electricity has similar problems with energy storage and is also expensive compared with wind or nuclear power. Nuclear power can be used for **base load power**, unlike wind or solar, but issues of safety, storage, and handling of nuclear waste and security concerns about nuclear weapons proliferation will have to be addressed before widespread expansion is likely, at least in the United States and Western Europe (aside from France, which already has made a significant commitment to nuclear energy).

This category is one with great hopes for technological "breakthroughs"—such as fusion, inexpensive solar, and inexpensive fuel cells—that may revolutionize our energy systems. Thus, basic research and development must be a part of any climate mitigation strategy. However, no responsible strategy should rely exclusively on breakthrough technologies; they may not exist for decades, if ever.

Outside of the electric realm, biomass converted to biofuel may play a major role in reducing CO<sub>2</sub> emissions in the transportation sector, at least until powerful, inexpensive, and reliable battery technologies or some alternative transportation technologies are developed. For example, Brazil currently obtains most of its transportation fuel from fermentation of sugar cane into ethanol, and similar programs are being implemented around the world.

A more efficient technology may be the conversion of biomass into synthetic diesel fuel via the [Fischer-Tropsch process](#), which was used by the Germans in World War II to transform coal into liquid fuel. This process has the advantage of creating a more diverse range of fuel products, including jet fuel for air transport, and of being more efficient through use of all types of biomass, not just sugar (or cellulose for a cellulosic conversion process). The Fischer-Tropsch process, which involves gasification of the biomass by heating it in the presence of oxygen, produces carbon monoxide and hydrogen. This "syngas" is then converted to liquid fuel by passing the gas over a cobalt or iron catalyst.

## Carbon Sequestration

The third category of solutions involves CO<sub>2</sub> capture from emissions sources and storage in geologic repositories, a process often referred to as carbon sequestration. This is an essential component of any climate mitigation portfolio because of the abundance of inexpensive coal in the largest economies of the world. Even with huge improvements in efficiency and increases in nuclear, solar, wind, and biomass power, the world is likely to depend heavily on coal, especially the five countries that hold 75 percent of the world's reserves: the United States, Russia, China, India, and Australia. However, as a technological strategy, carbon capture and storage (CCS) need not apply only to coal; any point source of CO<sub>2</sub> can be sequestered, including biomass gasification, which would result in negative emissions.

The scientific questions about CCS deal with the reliability of storage of vast quantities of CO<sub>2</sub> in underground repositories—and the quantities are indeed vast. Reservoir capacity required over the next century is conservatively estimated at one trillion tons of CO<sub>2</sub>, and it may exceed twice this quantity. This amount far exceeds the capacity of old oil and gas fields, which will be among the first targets for sequestration projects because of additional revenues earned from enhanced oil recovery. However, there is more than enough capacity in deep saline [aquifers](#) to store centuries of emissions, and also in deep-sea sediments, which may provide leakproof storage in coastal sites. In general, the storage issues do not involve large technological innovations, but rather improved understanding of the behavior of CO<sub>2</sub> at high pressure in natural geologic formations that contain fractures and faults. Geologic storage does not have to last forever—only long enough to allow the natural carbon

cycle to reduce the atmospheric CO<sub>2</sub> to near pre-industrial levels. This means that storage for 2000 years is long enough if deep-ocean mixing is not impeded significantly by stratification. It seems likely that many geological settings will provide adequate storage, but the data to demonstrate this over millennia do not yet exist. A more expansive program aimed at monitoring underground CO<sub>2</sub> injections in a wide variety of geologic settings is essential if CCS is to be adopted before the middle of the century.

The technological advances in CCS necessitate improving the efficiency of the capture of CO<sub>2</sub> from a coal-fired power plant. Capture can take place either by postcombustion **adsorption**, or through design of a power plant (either oxy-combustion or gasification) that produces a pure stream of CO<sub>2</sub> as an **effluent**. Either way, the capture of CO<sub>2</sub> is expensive, both financially and energetically. It has been suggested that capture and storage combined would use roughly 30 percent of the energy from the coal combustion in the first place and may raise the cost of generating electricity from coal by 50 percent, with two-thirds of this increase coming from capture. Even though these estimates are uncertain, given that carbon sequestration is not yet practiced at any coal plant, it is clear that technological innovation in the capture of CO<sub>2</sub> from a mixed gas stream is important.

Carbon sequestration also occurs through enhanced biological uptake such as reforestation or fertilization of marine phytoplankton. These approaches could be considered a separate category, as, for example, planting trees is quite different from injecting vast quantities of CO<sub>2</sub> underground. If pursued aggressively, such strategies might offset CO<sub>2</sub> emissions by as much as 7 Gigatons (Gt) of CO<sub>2</sub> (2 Gt of carbon) per year by the end of the century, out of total emissions of more than 80 Gt per year of CO<sub>2</sub> (22 Gt of carbon) as forecast in most business-as-usual scenarios. These approaches might be an important piece of a solution, but they will not replace the need for improved energy efficiency, non-fossil energy sources, and carbon sequestration.

The nature of the climate experiment means that no one truly knows what a safe level of CO<sub>2</sub> really is, apart from the impossible goal of the pre-industrial level of 280 parts per million (ppm). It is possible that nations will implement many of the approaches outlined above over the next few decades, which would stabilize atmospheric CO<sub>2</sub> below 600 ppm; it is difficult to imagine that a much lower stabilization level will be realized given the current state of the world energy systems. It is possible that this effort will be enough, that the world will warm another 2 or 3°C, that ice sheets will slowly melt, and that most of the severe consequences will be gradual, allowing adaptation by humans and natural ecosystems. On the other hand, it is also possible that even with concerted effort and cooperation among the large nations of the world, the climate system will respond too quickly for humans to adapt, that the Greenland and West Antarctic ice sheets will decay more quickly than expected, and that the impacts of a warmer world on humans and on natural ecosystems will be worse than we now predict.

It is very difficult to know which scenario is correct. The magnitude of the consequences depends in part on how we deal with them. Because of the potential for catastrophe, it seems prudent to ask



what societies might do if the rate of climate change were to accelerate over the next few decades and if the consequences were to be much worse than anticipated. One approach that has long been discussed is the engineering of our climate system by adjusting the incoming solar radiation by means of reflectors in space or in the upper atmosphere; indeed, there may be some ways to accomplish a reduction in solar radiation at very low cost relative to other strategies of mitigation. Recently, such ideas have gained more prominence, not as a substitute for serious emissions reductions, but in the sober realization that efforts to reduce emissions may not be sufficient to avoid dangerous consequences. The power to engineer the climate comes with an awesome responsibility. How could we engineer such a system to be failsafe? Which countries would control this effort? Who would decide how much to use, or when? And what would happen if something went wrong, if we discovered some unforeseen consequences that required shutting the effort down once human societies and natural ecosystems depended on it?

Ultimately, our path in dealing with climate change, as with many other environmental challenges, will depend on the choices we make—not just we as individuals, but nations and human society as a whole. The good news is that there are strategies that can solve these problems. Tropical rain forests can be protected from deforestation. Marine ecosystems can be protected from overfishing. And although we are already committed to substantial climate change, we can choose to rebuild our energy infrastructure to avoid the worst impacts. Some of these choices involve new technologies that require spending money; others simply involve a change in behavior, perhaps enforced by laws and regulations. Environmental science helps clarify what the consequences of our choices are likely to be and hopefully guides society to make better choices in caring for our habitable planet.

## 5. Further Reading

Cohen, **How Many People Can the Earth Support?** New York: W. W. Norton. 532 pp. 1995.

Pacala and Socolow, "Stabilization Wedges: Solving the Climate Problem for the Next 50 Years with Current Technologies," **Science** 305, 968–972, 2004.

Pandolfi et al., "Global Trajectories of the Long-Term Decline of Coral Reef Ecosystems," **Science** 301, 955–958, 2003.

Schrag, "Confronting the Climate-Energy Challenge," **Elements**, 3, 171–178, 2007. (Text is reproduced from this source with permission.)

Schrag and McCarthy, "Biological-physical interactions and global climate change: Some lessons from Earth history," in **The Sea**, vol. 12, edited by Allan R. Robinson, James J. McCarthy, and Brian J. Rothschild, John Wiley & Sons, Inc., New York, 605–619, 2002.

Wilson, **Consilience: The Unity of Knowledge**, Knopf, 1998.

## Glossary

**acid rain** : Rainfall with a greater acidity than normal.

**adsorption** : Process that occurs when a gas or liquid solute accumulates on the surface of a solid or, more rarely, a liquid (adsorbent), forming a molecular or atomic film (the adsorbate). It is different from absorption, in which a substance diffuses into a liquid or solid to form a solution.

**anthropogenic** : Describing effects or processes that are derived from human activities, as opposed to effects or processes that occur in the natural environment without human influences.

**aquifers** : Underground formations, usually composed of sand, gravel, or permeable rock, capable of storing and yielding significant quantities of water.

**aragonite** : A carbonate mineral that forms naturally in almost all mollusk shells, as well as the calcareous endoskeleton of warm- and cold-water corals

**base load power** : The average amount of electricity consumed at any given time. Base load power stations are designed to operate continuously, unlike peaking power stations that generally run only when there is a high demand.

**coral bleaching** : Refers to the loss of color of corals due to stress-induced expulsion of symbiotic, unicellular algae called zooxanthellae that live within their tissues. Stress can be induced by: increased water temperatures (often attributed to global warming), starvation caused by a decline in zooplankton levels as a result of overfishing, solar irradiance (photosynthetically active radiation and ultraviolet band light), changes in water chemistry, silt runoff, or pathogen infections.

**effluent** : An outflowing of water from a natural body of water, or from a man-made structure, generally considered to be pollution, such as the outflow from a sewage treatment facility or the wastewater discharge from industrial facilities.

**Fischer-Tropsch process** : A catalyzed chemical reaction in which carbon monoxide and hydrogen are converted into liquid hydrocarbons of various forms. The principal purpose of this process is to produce a synthetic petroleum substitute, typically from coal or natural gas, for use as synthetic lubrication oil or as synthetic fuel.

**greenhouse gases** : Atmospheric gases or vapors that absorb outgoing infrared energy emitted from the Earth naturally or as a result of human activities. Greenhouse gases are components of the atmosphere that contribute to the Greenhouse effect.

**pathogen** : A biological agent that causes disease or illness to its host.

**pelagic** : Water coming from the part of the open sea or ocean that is not near the coast

**scleractinian corals** : Stony or hard corals responsible for the very existence of the reef. As living animals, they provide habitats for many other organisms. The breakdown of their skeletons during



calcium-carbonate accretion and especially after death provides material for redistribution and consolidation into the reef framework.

**zooxanthellae** : Unicellular yellow-brown (dinoflagellate) algae which live symbiotically in the gastrodermis of reef-building coral.