UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR.780 - 35

FOURTH AUTUMN COURSE ON MATHEMATICAL ECOLOGY

(24 October - 11 November 1994)

"Ecoscience: Population, Resources, Environment"

Paul R. Ehrlich
Stanford University

Anne H. Ehrlich
Stanford University

John P. Holden
University of California, Berkeley

These are preliminary lecture notes, intended only for distribution to participants.

ECOSCIENCE: POPULATION, RESOURCES, ENVIRONMENT

PAUL R. EHRLICH
STANFORD UNIVERSITY

ANNE H. EHRLICH
STANFORD UNIVERSITY

JOHN P. HOLDREN
UNIVERSITY OF CALIFORNIA, BERKELEY

W. H. FREEMAN AND COMPANY
San Francisco

1. *

The energy that drives the hydrologic cycle is energy from the sun – indeed, this function is the largest single user of the solar energy reaching Earth's surface. The reason so much energy is required is that it takes a great deal of energy to evaporate water – 2250 joules per gram at the boiling point of 100° C and 2440 joules per gram at Earth's average surface temperature of 15° C. (This is the highest heat of vaporization of any known substance.) It takes fifty times as much energy to evaporate a gram of water as it does to lift it to an altitude of 5 kilometers. The energy used to evaporate the water is stored as latent heat of vaporization (see Box 2-1), which is released to the environment as heat whenever and wherever the water vapor condenses into liquid. Thus, energy delivered by the sun at one point on Earth's surface may be released high in the atmosphere over a point 1000 kilometers away. This mechanism of redistributing energy by the transport and condensation of water vapor is a major determinant of Earth's climate.

As noted above, the energy the sun supplies at the time of evaporation reappears as heat at the time of condensation. Similarly, the smaller amount of solar energy that does the work of lifting the water vapor against the force of gravity appears as frictional heat when falling droplets of condensed vapor collide with molecules of air and when rushing mountain streams rub against their rocky beds. That all the energy the sun supplies to terrestrial processes comes back again in one form or another is not coincidence or quirk, but an illustration of the first law of thermodynamics – the law of conservation of energy. Further excursions into the machinery of the physical world – and of human technology – will require some familiarity with this law and with its companion, the second law of thermodynamics, so an introduction to both is provided in Box 2-3.

ATMOSPHERE AND CLIMATE

The blanket of gases that makes up Earth's atmosphere has many functions. Of the four elements required in greatest quantity by living organisms [carbon (C), oxygen (O), hydrogen (H), nitrogen (N)] the atmosphere provides the main reservoir of one (N), the most accessible reservoir of two others (C, O), and an essential link in the continuous recycling of the fourth (H, in the form of H_2O). The atmosphere is substantial enough to protect the organisms on Earth's surface from a variety of harmful particles and radiations that reach the planet from the sun and from space, but it is transparent enough to permit an adequate amount of life-giving sunlight to penetrate to that surface. Acting as a thermal insulator, the atmosphere keeps Earth's surface much warmer, on the average, than it would be if there were no atmosphere. And the stirrings of the atmosphere, transporting energy and moisture from one place to another, are a major part of the patterns of climate so important to the character and distribution of life.

For simplicity, we begin our investigation of the atmosphere by ignoring its internal vertical and horizontal motions and considering its properties as a static body of gas.

Air

The term *air* refers to the particular mixture of gaseous compounds making up the atmosphere. The average composition of this mixture, not including water, is shown in Table 2-11. An important property of gases is that a given number of molecules at a given temperature and pressure will occupy almost exactly the same volume, regardless of the mass or size of the molecules. This property has led to the use of the somewhat confusing terms *percent by volume* or *fraction by volume* to describe the relative abundance of the various constituents of gaseous mixtures. That is, if three-quarters of the molecules in a container of fixed volume are gas A and one-quarter are gas B, one can think of gas A as "occupying" three-fourths of the volume and gas B as "occupying" one-fourth. (What is happening in reality, of course, is that both gases, mixed together, occupy the whole volume, with gas A accounting for three-fourths of the pressure in the volume and gas B accounting for one-fourth of the pressure.) We will use the term *molecular fraction* (number of molecules of a constituent divided by the total number of molecules in the mixture), because it is unambiguous and works for solids and liquids as well as gases. The reader should simply be aware that this term is interchangeable with the term

2.

TABLE 2-11

Average Composition of Clean Dry Air

Constituent	Symbol	Molecular weight	Molecular fraction of air	Mass fraction of air
Nitrogen	N_2	28	0.7809	0.755
Oxygen	O_2	32	0.2095	0.232
Argon	Ar	40	0.0093	0.013
Carbon Dioxide	CO_2	44	320 ppm	486 ppm
Neon	Ne	20	18 ppm	12 ppm
Helium	He	4	5.2 ppm	0.7 ppm
Methane	CH_4	16	2.9 ppm	1.6 ppm
Krypton	Kr	84	1.1 ppm	3.2 ppm
Nitrous Oxide	N_2O	44	0.5 ppm	0.8 ppm
Hydrogen	H_2	2	0.5 ppm	0.03 ppm
Ozone	O_3	48	0.01 ppm	0.02 ppm

Source: Garrels, Mackenzie, and Hunt, *Chemical cycles*.

BOX 2-3 Availability, Entropy, and the Laws of Thermodynamics

Many processes in nature and in technology involve the transformation of energy from one form into others. For example, light from the sun is transformed, upon striking a meadow, into thermal energy in the warmed soil, rocks, and plants; into latent heat of vaporization as water evaporates from the soil and through the surface of the plants; and into chemical energy captured in the plants by photosynthesis. Some of the thermal energy, in turn, is transformed into infrared electromagnetic radiation heading skyward. The imposing science of thermodynamics is just the set of principles governing the bookkeeping by which one keeps track of energy as it moves through such transformations. A grasp of these principles of bookkeeping is essential to an understanding of many problems in environmental sciences and energy technology.

The essence of the accounting is embodied in two concepts known as the first and second laws of thermodynamics. No exception to either one has ever been observed. The first law, also known as the law of conservation of energy, says that energy can neither be created nor destroyed. If energy in one form or one place disappears, the same amount must show up in another form or another place. In other words, although transformations can alter the *distribution* of amounts of energy among its different forms, the *total* amount of energy, when all forms are taken into account, remains the same. The term *energy consumption*, therefore, is a misnomer; energy is used, but it is not really consumed. One can speak of fuel consumption, because fuel, as such, does get used up. But when we burn gasoline, the amounts of energy that appear as mechanical energy, thermal energy, electromagnetic radiation, and other forms are exactly equal all together to the amount of chemical potential energy that disappears. The accounts must always balance; apparent exceptions have invariably turned out to stem from measurement errors or from overlooking categories. The immediate relevance of the first law for human affairs is often stated succinctly as, "You can't get something for nothing."

Yet, if energy is stored work, it might seem that the first law is also saying, "You can't lose!" (by saying that the total amount of stored work in all forms never changes). If the amount of stored work never diminishes, how can we become worse off? One obvious answer is that we can become worse off if energy flows to places where we can no longer get at it—for example, infrared radiation escaping from Earth into space. Then the stored work is no longer accessible to us, although it still exists. A far more fundamental point, however, is that ***different kinds of stored work are not equally convertible into useful, applied work.*** We can therefore become worse off if energy is transformed from a more convertible form to a less convertible one, even though no energy is destroyed and even if the energy has not moved to an inaccessible place. The degree of convertibility of energy—stored work—into applied work is often called ***availability.***

Energy in forms having high availability (that is, in which a relatively large fraction of the

(*Continued*)

stored work can be converted into applied work) is often called high-grade energy. Correspondingly, energy of which only a small fraction can be converted to applied work is called low-grade energy, and energy that moves from the former category to the latter is said to have been degraded. Electricity and the chemical energy stored in gasoline are examples of high-grade energy; the infrared radiation from a light bulb and the thermal energy in an automobile exhaust are corresponding examples of lower-grade energy. The quantitative measure of the availability of thermal energy is temperature. More specifically, the larger the *temperature difference* between a substance and its environment, the more convertible into applied work is the thermal energy the substance contains; in other words, the greater the temperature difference, the greater the availability. A small pan of water boiling at 100° C in surroundings that are at 20° C represents considerable available energy because of the temperature difference; the water in a swimming pool at the same 20° C temperature as the surroundings contains far more total thermal energy than the water in the pan, but the availability of the thermal energy in the swimming pool is zero, because there is no temperature difference between it and its surroundings.

With this background, one can state succinctly the subtle and overwhelmingly important message of the second law of thermodynamics: *all physical processes, natural and technological, proceed in such a way that the availability of the energy involved decreases.* (Idealized processes can be constructed theoretically in which the availability of the energy involved stays constant, rather than decreasing, but in all real processes there is *some* decrease. The second law says that an *increase* is not possible, even in an ideal process.) As with the first law, apparent violations of the second law often stem from leaving something out of the accounting. In many processes, for example, the availability of energy in some *part* of the affected system increases, but the decrease of availability elsewhere in the system is always large enough to result in a net decrease in availability of energy overall. What is consumed when we use energy, then, is not energy itself but its availability for doing useful work.

The statement of the second law given above is deceptively simple; whole books have been written about equivalent formulations of the law and about its implications. Among the most important of these formulations and implications are the following:

1. In any transformation of energy, some of the energy is degraded.
2. No process is possible whose sole result is the conversion of a given quantity of heat (thermal energy) into an equal amount of useful work.
3. No process is possible whose sole result is the flow of heat from a colder body to a hotter one.
4. The availability of a given quantity of energy can only be used once; that is, the property of convertibility into useful work cannot be "recycled."
5. In spontaneous processes, concentrations (of anything) tend to disperse, structure tends to disappear, order becomes disorder.

That Statements 1 through 4 are equivalent to or follow from our original formulation is readily verified. To see that statement 5 is related to the other statements, however, requires establishing a formal connection between order and availability of energy. This connection has been established in thermodynamics through the concept of *entropy,* a well defined measure of disorder that can be shown to be a measure of unavailability of energy, as well. A statement of the second law that contains or is equivalent to all the others is: *all physical processes proceed in such a way that the entropy of the universe increases.* (Not only can't we win—we can't break even, and we can't get out of the game!)

Consider some everyday examples of various aspects of the second law. If a partitioned container is filled with hot water on one side and cold water on the other and is left to itself, the hot water cools and the cold water warms—heat flows from hotter to colder. Note that the opposite process (the hot water getting hotter and the cold getting colder) does not violate the first law, conservation of energy. That it does not occur illustrates the second law. Indeed, many processes can be imagined that satisfy the first law but violate the second and therefore are not expected to occur. As another example, consider adding a drop of dye to a glass of water. Intuition and the second law dictate that the dye will spread, eventually coloring all the water—concentrations disperse, order (the dye/no dye ar-

BOX 2-3 (*Continued*)

rangement) disappears. The opposite process, the spontaneous concentration of dispersed dye, is consistent with conservation of energy but not with the second law.

A more complicated situation is that of the refrigerator, a device that certainly causes heat to flow from cold objects (the contents of the refrigerator – say, beer – which are made colder) to a hot one (the room, which the refrigerator makes warmer). But this heat flow is not the *sole* result of the operation of the refrigerator: energy must be supplied to the refrigeration cycle from an external source, and this energy is converted to heat and discharged to the room, along with the heat removed from the interior of the refrigerator. Overall, availability of energy has decreased, and entropy has increased.

One illustration of the power of the laws of thermodynamics is that in many situations they can be used to predict the maximum efficiency that could be achieved by a perfect machine, without specifying any details of the machine! (Efficiency may be defined, in this situation, as the ratio of useful work to total energy flow.) Thus, one can specify, for example, what *minimum* amount of energy is necessary to separate salt from seawater, to separate metals from their ores, and to separate pollutants from auto exhaust without knowing any details about future inventions that might be devised for these purposes. Similarly, if one is told the temperature of a source of thermal energy – say, the hot rock deep in Earth's crust – one can calculate rather easily the maximum efficiency with which this thermal energy can be converted to applied work, regardless of the cleverness of future inventors. In other words, *there are some fixed limits to technological innovation, placed there by fundamental laws of nature.* (The question of how far from the maximum attainable efficiencies industrial societies operate today is taken up in Chapter 8.)

More generally, the laws of thermodynamics explain why we need a continual input of energy to maintain ourselves, why we must eat much more than a pound of food in order to gain a pound of weight, and why the total energy flow through plants will always be much greater than that through plant-eaters, which in turn will always be much greater than that through flesh-eaters. They also make it clear that *all* the energy used on the face of the Earth, whether of solar or nuclear origin, will ultimately be degraded to heat. Here the laws catch us both coming and going, for they put limits on the efficiency with which we can manipulate this heat. Hence, they pose the danger (discussed further in Chapter 11) that human society may make this planet uncomfortably warm with degraded energy long before it runs out of high-grade energy to consume.

Occasionally it is suggested erroneously that the process of biological evolution represents a violation of the second law of thermodynamics. After all, the development of complicated living organisms from primordial chemical precursors, and the growing structure and complexity of the biosphere over the eons, do appear to be the sort of spontaneous increases in order excluded by the second law. The catch is that Earth is not an isolated system; the process of evolution has been powered by the sun, and the decrease in entropy on Earth represented by the growing structure of the biosphere is more than counterbalanced by the increase in the entropy of the sun. (The process of evolution is discussed in more detail in Chapter 4.)

It is often asked whether a revolutionary development in physics, such as Einstein's theory of relativity, might not open the way to circumvention of the laws of thermodynamics. Perhaps it would be imprudent to declare that in no distant corner of the universe or hitherto-unexplored compartment of subatomic matter will any exception ever turn up, even though our intrepid astrophysicists and particle physicists have not yet found a single one. But to wait for the laws of thermodynamics to be overturned as descriptions of everyday experiences on this planet is, literally, to wait for the day when beer refrigerates itself in hot weather and squashed cats on the freeway spontaneously reassemble themselves and trot away.

5.

fraction by volume, often used elsewhere for gases. In many applications it is also useful to work with the *mass fraction* (grams of constituent/gram of mixture). This is a more precise statement of what is meant by the common term *fraction by weight* or *percent by weight.*[21]

A *mole* of any substance is 6.02×10^{23} molecules, and the mass of a mole is equal to the molecular weight of the substance in grams. For example, the mass of a mole of nitrogen gas (N_2) is 28 grams. It is often convenient to speak of air as if it were a single substance; the term *a mole of air* means 6.02×10^{23} molecules, of which 78.09 percent are nitrogen molecules, 20.95 percent are oxygen molecules, and so on. Such a collection of molecules has a mass of about 29 grams, which is called the *molecular weight of air.* (These definitions will be of importance later in interpreting what is meant by pollution standards expressed in different ways.)

Although nitrogen and oxygen comprise 99 percent of dry air, the trace constituents carbon dioxide (CO_2) and ozone (O_3) play exceedingly important roles because of the special properties of these molecules, as described below. Methane, nitrous oxide, and hydrogen also have roles in atmospheric chemistry and physics, albeit smaller ones. Argon, helium, krypton, and neon, by contrast, are chemically inert, monatomic gases, whose presence in the atmosphere is of interest only as resources for certain applications in technology.[22]

Water Vapor

The water content of the atmosphere varies greatly from place to place and time to time. Three commonly used measures of water content are *absolute humidity, specific humidity,* and *relative humidity.* Absolute humidity is the mass of water vapor per unit volume of air, and it varies from almost zero over the driest deserts to around 25 grams per cubic meter over jungles and tropical seas. Specific humidity is the mass of water vapor per unit mass of air. (A closely related term, the *mixing ratio,* is the mass of water vapor mixed with each unit mass of *dry* air.) Relative humidity, usually expressed as a percentage, is the ratio of the actual molecular fraction of water vapor in air to the molecular fraction corresponding to saturation at the prevailing temperature. (*Saturation* refers to the condition that ensues if air is left for a long time in a sealed container partly filled with pure water; the number of molecules of water vapor per unit volume of air under these circumstances depends only on the temperature.)[23] Relative humidity usually is between 0 and 100 percent, but under special circumstances (supersaturation) it can significantly exceed 100 percent.

Under ordinary circumstances, water vapor in the atmosphere begins to condense into droplets of liquid, forming clouds, as soon as the relative humidity exceeds 100 percent by even a small amount. The process of condensation is greatly facilitated by the virtually universal presence in the atmosphere of small particles that provide surfaces where the condensation can commence. Called *condensation nuclei* when they perform this function, these particles include salt crystals formed by the evaporation of sea spray, dust raised by the wind, ash from volcanoes and forest fires, decomposed organic matter, and, of course, particles produced by various technological activities. Even in "unpolluted" air, particles that might serve as condensation nuclei are seemingly abundant in absolute terms (more than 100 particles per cubic centimeter), but the extent of condensation and precipitation apparently are related to specific physical characteristics of the condensation nuclei as well as to their number.

The molecular fraction of water vapor corresponding to saturation increases as temperature increases—warm air can "hold" more water vapor than cool air. Accordingly, there are two ways in which relative humidity can be raised from less than 100 percent to more, initiating

[21]Weight means the force exerted upon a mass by gravity. Weight and mass are more or less interchangeable (using the relation, weight equals mass multiplied by acceleration of gravity) only if one stays on Earth's surface, where gravity is nearly constant. An astronaut has the same mass on the surface of the moon as on Earth, but a very different weight, because the acceleration of gravity on the moon is much less than on Earth.

[22]For discussions of how the atmosphere came to have the composition it does, the reader should consult Preston Cloud and Aharon Gibor, The oxygen cycle; and Preston Cloud, ed., *Adventures in earth history.*

[23]The often encountered definition of saturation as "the maximum amount of water vapor air can hold at a given temperature" is not quite correct. A good discussion of this and the following points is given by Morris Neiburger, James G. Edinger, William D. Bonner, *Understanding our atmospheric environment,* Chapter 8.

condensation and perhaps precipitation: (1) more water vapor can be added to the air by evaporation from an exposed water surface; (2) the air can be cooled so that the vapor content corresponding to saturation falls. At a given vapor content (a fixed specific humidity), the temperature at which the relative humidity reaches 100 percent is called the *dew point*. Addition of water vapor to the air by evaporation is a slow process, but cooling of the air can be very rapid. Rapid cooling to below the dew point is the mechanism immediately responsible for most condensation phenomena – the appearance of dew and fog at night as air is cooled by radiation of heat to the night sky; formation of clouds and rain in updrafts as the air is cooled by expansion; and formation of beads of water on the outside of a pitcher of ice water on a hot day, as air adjacent to the pitcher is cooled by contact with the cold surface.

Pressure, Temperature, and Vertical Structure

The pressure exerted by the atmosphere on objects at Earth's surface is essentially equal to the weight of the overlying air, which at sea level amounts on the average to 10.3 metric tons per square meter (14.7 lb/in^2). This amount of pressure, defined as 1 atmosphere, is the same as would be exerted at sea level by a column of water about 10 meters high (33 ft) or a column of mercury 760 millimeters (mm) high (29.92 in). This means that the mass of the atmosphere is only equivalent to that of a 10-meter layer of water covering Earth (you can check this in Table 2-2) and that pressure under water increases by the equivalent of 1 atmosphere for every 10 meters of depth. The usual metric unit for the measurement of atmospheric pressure is the millibar; 1 millibar is 100 newtons (N) per square meter (see Box 2-1 for the definition of the newton), and 1 atmosphere is 1013.25 millibars.

Atmospheric pressure is not ordinarily perceived as a force because it acts equally in all directions (up, down, sideways); organisms are not crushed by it because the gases and liquids in tissue are also at atmospheric pressure, so the inward and outward forces balance. Pressure becomes perceptible as a (painful) pressure *difference* if the pressure outside an organism changes more rapidly than the interior pressure can accommodate (an example would be the pain in one's ears associated with a rapid change in altitude).

Unlike water, whose density at the bottom of the deepest ocean trenches at pressures of hundreds of atmospheres is only a few percentage points higher than its density at the surface, the air in the atmosphere is highly compressible – that is, density increases markedly as pressure increases. Indeed, air behaves very much like a "perfect gas," for which pressure (p), density (ρ), and temperature (T) are related by the equation

$$p = \rho R T$$

For this equation to be valid, the temperature T must be measured with respect to *absolute zero*, the temperature at which there is no molecular motion. Temperature measured from this zero point, which is the same for all substances, is called *absolute temperature*, and the corresponding unit of measurement in the metric system is the degree kelvin.[24] The R in the equation is the gas constant, which for dry air equals 287 joules per kilogram per degree kelvin. According to the perfect gas equation, the density of dry air varies in direct proportion to pressure, if temperature is held constant.

Because the atmosphere is compressible, its mass is concentrated in the lower layers. Forty percent of the air in the atmosphere lies below the altitude of the summit of Mount Whitney in California's Sierra Nevada range (4.4 km) and two-thirds lies below the altitude of the summit of Mount Everest (8.9 km). The density of air at an altitude of 12 kilometers, where most subsonic jet airliners fly, is about one-fifth the density at sea level. The average variation of pressure and temperature with altitude above sea level is shown in Figure 2-10.

The atmosphere is subdivided into horizontal layers according to the pattern of temperature variation. The lowest layer, called the *troposphere*, is characterized by a rather uniform average rate of temperature decline with altitude of 6.4° C per kilometer. Almost all the atmo-

[24]Absolute zero, or 0 degrees kelvin (K) equals −273.15° C. An attempt was made recently to standardize the unit of absolute temperature as simply the kelvin, rather than the degree kelvin, but the change has not been generally adopted.

7 .

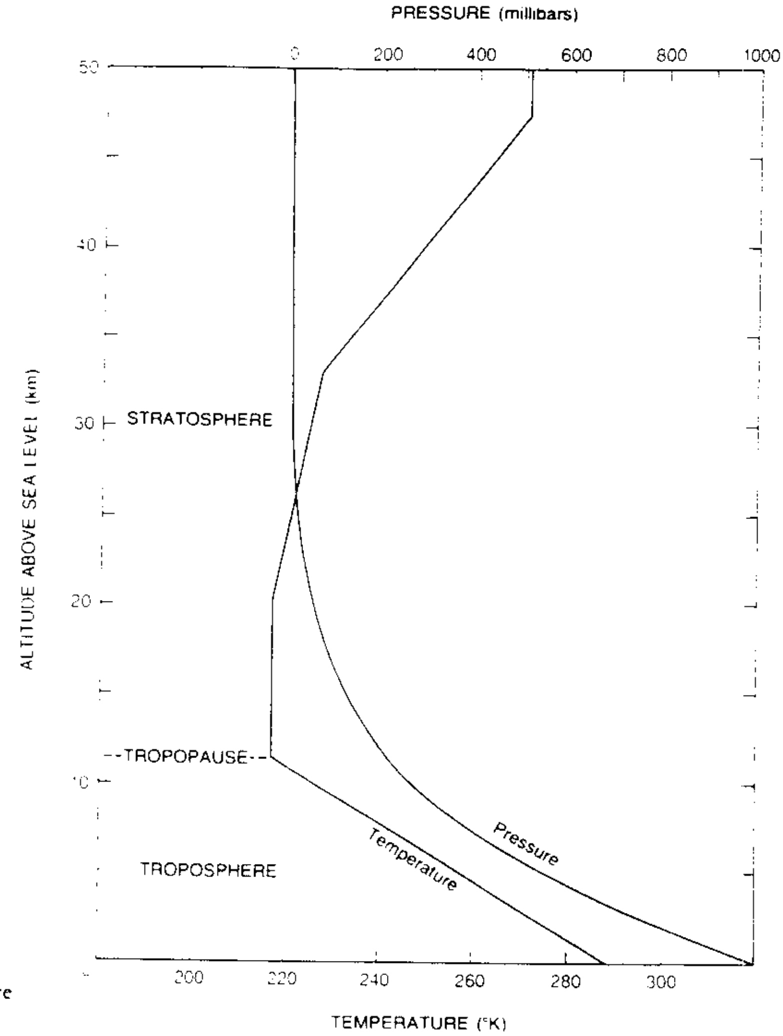


FIGURE 2-10

Variation of atmospheric temperature and pressure with altitude.

spheric phenomena that govern climate take place in the troposphere. The top of the troposphere is called the *tropopause*, where the temperature decline stops and a layer of uniform temperature at about $-55°$ C commences. The tropopause is typically found at an altitude of from 10 to 12 kilometers, but it ranges from a low of 5 or 6 kilometers at the poles to around 18 kilometers at the equator.

The *stratosphere* extends from the tropopause up to the stratopause (about 50 km) and is characterized over much of this interval by temperatures increasing with altitude (reaching almost 0° C at the stratopause). The gaseous composition of the stratosphere is essentially the same as that at sea level, with two significant exceptions. First, there is very little water vapor in the stratosphere; the mixing ratio is typically two or three parts per million (ppm), or 1000 to 10,000 times less than is common near sea level. Second, there is a great deal more ozone in the stratosphere than in the troposphere; the maximum molecular fraction of ozone is 10 ppm near 25 km altitude, or 1000 times more than the average for the whole atmosphere.[25] The air pressure and density at the top of the stratopause are on the order of a thousandth of the values at sea level.

Above the stratosphere lies the *mesosphere* (to about 90 km), wherein the temperature again decreases with altitude. The composition of the mesosphere remains much like that of the lower layers, except for certain trace constituents such as water vapor and ozone. The troposphere, stratosphere, and mesosphere together are called the ***homosphere***, referring to their relatively uniform composition. Above the mesosphere is the ***thermosphere*** (temperature again rising with altitude), which contains the ***heterosphere*** (so named because the molecular constituents are there separated into distinct layers of differing composition) and the *ionosphere* (referring to the presence of free electrons and the positively charged ions from which the electrons have been stripped). The thermosphere has no well defined upper limit; its density at 100 kilometers is around one-millionth of atmospheric density at sea level, and by 10,000 kilometers it has faded off to the density prevailing in interplanetary space.

Radiant Energy Flow in the Atmosphere

What accounts for the complicated way in which temperature changes with altitude in the atmosphere? The answer involves the way in which different atmospheric constituents interact with radiant energy arriving from the sun and with radiant energy trying to escape Earth into space. The same processes, of course, determine how much and what kinds of energy reach Earth's surface, so they are crucial in determining the conditions that govern life. To understand these processes requires at least a modest acquaintance with the character of radiant energy (or electromagnetic radiation), and this is supplied in Box 2-4.

The energy in the electromagnetic radiation reaching the top of Earth's atmosphere from the sun is distributed over a range of wavelengths, as shown in Figure 2-11. One can determine from such a graph that about 9

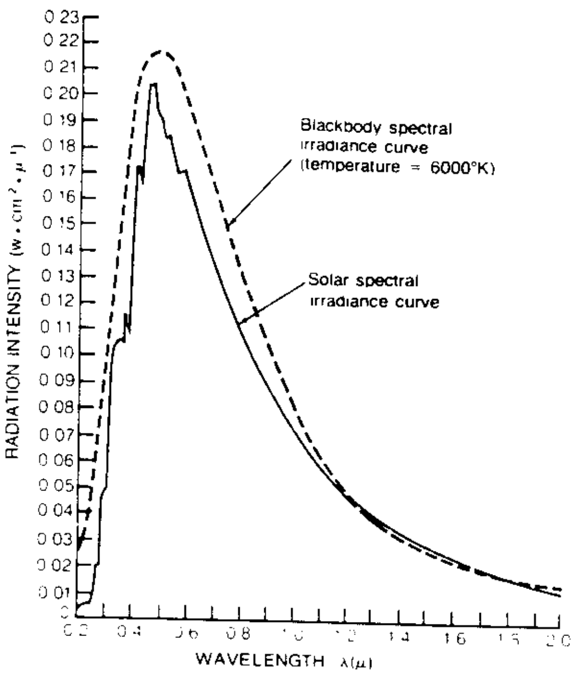


FIGURE 2-11

Solar irradiance spectrum and 6000°K blackbody radiation reduced to mean solar distance. (From Neiburger, Edinger, and Bonner, 1973.)

percent of the total incoming energy is in the ultraviolet part of the electromagnetic spectrum, 41 percent is in the visible part of the spectrum, and 50 percent is in the infrared part. A significant part of this energy is prevented from reaching Earth's surface by gaseous constituents of the atmosphere that are opaque to certain wavelengths. This opacity is due not to reflection but to absorption (the energy in the radiation is absorbed by the gas molecules, warming the atmospheric layers where these processes take place).

The depletion of incoming radiation by absorption in atmospheric gases is summarized in Table 2-12. The main results are that ultraviolet solar radiation with wavelengths less than 0.3 microns (μ) is almost completely absorbed high in the atmosphere, and infrared solar radiation is substantially depleted through absorp-

[25] Study of Critical Environmental Problems (SCEP), *Man's impact on the global environment*, p. 41.

9.

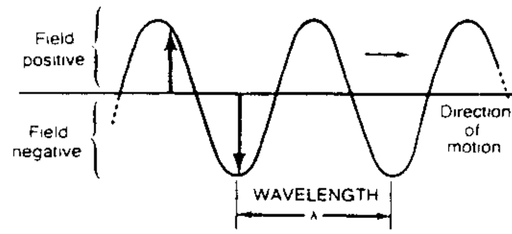
BOX 2-4 Electromagnetic Radiation

Light, X-rays, radio waves, infrared radiation, and radar waves are all variations of the same thing—phenomena with the interchangeable names *electromagnetic radiation, electromagnetic waves,* and *radiant energy.*

Energy in this form travels at the speed of light (c, or 299,792 km/sec in vacuum), and—as the words *in vacuum* imply—requires no material medium to support the energy flow. This property contrasts with the other, much slower forms of energy transport (conduction and convection) which do require a medium. Convection involves the bulk motion of matter; conduction involves molecular motion. What moves in the case of electromagnetic radiation is a combination of electric and magnetic fields of force.

For many purposes, it is useful to visualize this pattern as a traveling wave, as shown in the diagram here. There the curve denotes the spatial pattern of the strength of the electric or magnetic field, and the pattern is moving to the right with speed c. (Actually, there should be separate curves for the electric and magnetic fields, but this detail need not trouble us here.) The *wavelength* (λ) is the distance between successive crests or troughs.* At any fixed point along the path of the wave, the field is seen to oscillate with *frequency* (ν) related to wavelength and the speed of light by the relation $\nu = c/\lambda$.

*Unfortunately, there are more concepts in science than there are Greek letters. Thus, for example, lambda (λ) represents wavelength in physics and finite rate of increase in population biology (Chapter 4).



The different forms of electromagnetic energy are distinguished by their different wavelengths (or, equivalently, their frequencies, since one can be computed from the other using the relation $\lambda\nu = c$). Visible electromagnetic radiation has wavelengths between 0.40 microns (violet light) and 0.71 microns (red light). A micron (μ) is one-millionth of a meter. The entire range of wavelengths that have been observed, from tiny fractions of a micron to tens of kilometers, is called the *electromagnetic spectrum.* Some types of electromagnetic radiation occupying different parts of the spectrum are indicated in the table below.

Type of radiation	Wavelength range
radio	1–10 m
radar (microwaves)	1–30 cm
infrared	0.71–100 μ
visible	0.40–0.71 μ
ultraviolet	0.10–0.40 μ
X-rays	10^{-5}–10^{-2} μ

The way electromagnetic radiation interacts with matter depends in a complicated manner on the wavelength of the radiation, on its intensity (energy flow per unit of area, measured—say—in watts per square meter), and on the properties of the matter. Radiant energy that encounters matter may be transmitted, reflected, or absorbed. To the extent that radiation of a given wavelength is transmitted, the material is said to be transparent to that wavelength; to the extent that the radiation is reflected or absorbed, the material is said to be opaque to that wavelength. Most materials are transparent to some wavelengths and opaque to others. Many gases, for example, are rather opaque to most ultraviolet wavelengths but transparent to radio waves, visible light, and X-rays. Human flesh is opaque to visible light but transparent to X-rays.

Often, transmission, reflection, and absorption all take place at once. Consider a glass window in the morning sunlight: the glint off the window indicates that reflection is taking place; the room behind the window is illuminated and warmed, so there is certainly transmission; and the window itself gets warm, so absorption is happening, too.

Both transmission and reflection can be *direct* or *diffuse*. Direct means that a beam of electromagnetic radiation arriving from a single specific direction is sent on or sent back in a single specific direction. Diffuse means that the incoming beam is split up and sent on or sent back in many different directions. The phenomenon that produces diffuse transmission and reflection is called *scattering*. All light reaching the ground on a completely overcast day is diffuse light that has been scattered by the water droplets in the cloud layer.

All matter that is warmer than absolute zero can *emit* radiation. (The "machinery" by which emission occurs has to do with the behavior of the electrons in matter. We will not dwell on this machinery here, summarizing instead only the main results of its operation.) Some substances emit only radiation of certain wavelengths, and all substances can absorb only the wavelengths they can emit. A body that absorbs all the radiation that hits it in all wavelengths is called a *blackbody*, a useful idealization that is approached in the real world but never quite reached. A blackbody is not only a perfect absorber, but also a perfect emitter; the amount of electromagnetic energy emitted at any given wavelength by a blackbody of a specified temperature is the theoretical maximum that any real body of that temperature can emit at that wavelength. The total amount of radiation emitted by a blackbody is proportional to the fourth power of the body's absolute temperature. The wavelength at which a blackbody emits most intensely decreases in inverse proportion to the absolute temperature—that is, the higher the temperature, the shorter the wavelength.

The characteristics of the sun as a source of electromagnetic radiation are very closely approximated by those of a blackbody with a temperature of 5800° K. The wavelength at which its emission is most intense is 0.5 microns, corresponding to blue-green light (see Figure 2-11).

TABLE 2-12
Absorption of Solar Radiation by Atmospheric Gases

Wavelength range (μ)	Fate of radiation
UNTRAVIOLET	
Less than 0.12	All absorbed by O_2 and N_2 above 100 km
0.12–0.18	All absorbed by O_2 above 50 km
0.18–0.30	All absorbed by O_3 between 25 and 50 km
0.30–0.34	Part absorbed by O_3
0.34–0.40	Transmitted to Earth almost undiminished
VISIBLE	
0.40–0.71	Transmitted to Earth almost undiminished
INFRARED	
0.71–3	Absorbed by CO_2 and H_2O, mostly below 10 km

Source: Neiburger, Edinger, and Bonner, *Understanding our atmospheric environment.*

tion by carbon dioxide and (especially) by water vapor at lower altitudes. The atmosphere's gases are almost completely transparent to the visible wavelengths, where the intensity of solar radiation reaches its peak (Figure 2-11), and to the "near-ultraviolet" radiation, with wavelengths just shorter than the visible. (This near-ultraviolet radiation is an important contributor to sunburn.) Significantly, the only atmospheric gas that is opaque to ultraviolet radiation between 0.18 and 0.30 microns in wavelength is ozone. Without the trace of ozone that exists in the stratosphere, this radiation would reach Earth's surface—where it could be extremely disruptive to the life forms that evolved in the absence of these wavelengths (Chapter 11). The absorption of ultraviolet radiation by ozone in the stratosphere has a second important effect—it produces the stratospheric heating that causes temperature to increase with altitude in this layer of the atmosphere, with consequences discussed in the next section.

Some of the incoming radiation that is not absorbed by atmospheric gases is scattered by them (Box 2-4). Some of the scattered radiation returns to space (diffuse reflection), and some reaches the ground as diffuse solar radiation. The physics of scattering by the molecules in air is such that blue light is scattered much more than is red light. The sky appears blue because what reaches the eye is mostly scattered light and hence—owing to the preferential scattering of blue wavelengths by air molecules—mostly blue. That the sun appears red at sunset is precisely the same phenomenon at work; then one is looking at direct-beam radiation—the unscattered part—from which the blue has been removed (by scattering) and in which mainly the red remains.

Absorption and scattering of solar radiation are done by *aerosols* as well as by gases. An aerosol is a suspension of solid or liquid particles in a gas. Fog, clouds, smoke, dust, volcanic ash, and suspended sea salts are all aerosols. Whether a given aerosol acts mainly as an absorber or mainly as a scatterer depends on the size and composition of the particles, on their altitude, and on the relative humidity of the air they are in. Certainly, the aerosols that interact most strongly with solar radiation are clouds, which at any given time cover about half Earth's surface with a highly reflective layer.

The reflectivity of a surface or a substance is called its *albedo* (formally, albedo equals reflected energy divided by incoming energy). This property depends not only on the characteristics of the surface, but also on the angle at which the incoming radiation strikes the surface (the angle of incidence). The albedo is much higher at shallow angles of incidence than at steep (nearly perpendicular) angles. The albedo of clouds ranges from 0.25 (thin clouds, perpendicular incidence) to more than 0.90.

Averaged around the year and around the globe, the amount of energy that penetrates to Earth's surface is about half of what strikes the top of the atmosphere. Of that which penetrates, somewhat more than half is diffuse and somewhat less than half is direct. Upon

12

TABLE 2-13
Albedo of Various Surfaces

Surface	*Albedo*
Snow	0.50–0.90
Water	0.03–0.80
Sand	0.20–0.30
Grass	0.20–0.25
Soil	0.15–0.25
Forest	0.05–0.25

Sources: SMIC, *Inadvertent climate modification;* Neiburger, Edinger, and Bonner, *Understanding our atmospheric environment.*

reaching the surface, this energy meets several fates. Some is reflected, with the albedo varying from land to water and from one type of vegetation to another—and, of course, depending on the angle of incidence (Table 2-13). Some is absorbed by the melting or sublimation of ice or snow, or the vaporization of water. Some is absorbed to warm the surface and objects on or under it. And a tiny fraction is captured and transformed into chemical energy by the process of photosynthesis in plants.

The fate of solar energy striking the top of Earth's atmosphere is summarized in Figure 2-12. Note that the average albedo of the Earth-atmosphere system is about 0.28, of which two-thirds is accounted for by clouds.

Now from the first law of thermodynamics (Box 2-3), which says energy can neither be created nor destroyed, and from elementary considerations of stocks and flows (Box 2-2), one can draw some important conclusions about energy flow in the Earth-atmosphere system. First, the rate at which Earth's atmosphere and surface are absorbing solar energy must be matched by the rate at which the system loses energy, or else the amount of energy stored on the surface and in the atmosphere would be steadily changing. In other words, if the system is to be in equilibrium, outflow must equal inflow. Lack of equilibrium, if it occurred, would mean a changing stock of energy in the Earth-atmosphere system. This could manifest itself as an upward or downward trend in mean temperature (a changing stock of thermal energy), or in mean absolute humidity or mean volume of ice and snow

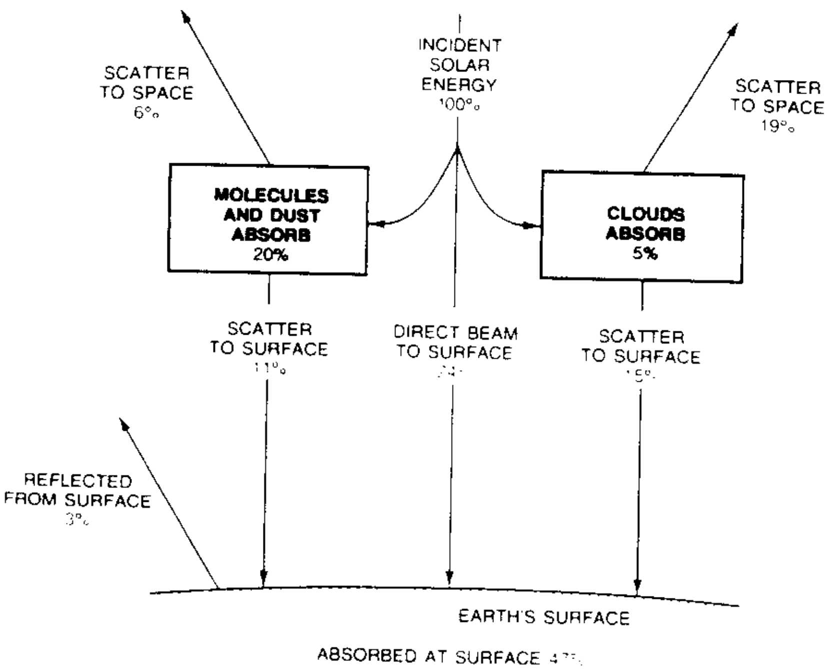


FIGURE 2-12

The fate of incoming solar radiation. Figures represent global annual averages.

(a changing stock of latent energy of vaporization and fusion), or in mean quantity of organic matter (a changing stock of chemical energy). Of course, different parts of Earth's surface and atmosphere at different times of year *are* out of equilibrium—temperature, humidity, snow cover, and quantity of vegetation generally change dramatically with seasons. Inflow and outflow of energy at a given time and place in general do not balance. But they *must* balance on a year-round average for the whole globe or else there will be year-to-year changes in the *mean* values of these indices of stored energy. Such changes have occurred in the past as cooling periods leading into ice ages and warming periods leading out of them. These changes and the potential for further ones, with or without human influence, are discussed below. As a first approximation, however, it is reasonable to assume that inflows and outflows of energy are in balance on the global, annual average.

The second conclusion is that the same considerations of inflow, outflow, and equilibrium that apply to the Earth-atmosphere system as a whole must apply separately to its components. If the atmosphere is to be in equilibrium in a global, time-averaged sense, it must be losing energy at the same rate it is receiving it. The same must hold for Earth's surface.

Clearly, then, the energy flow shown in Figure 2-12 is not the complete picture. The figure shows a net accumulation of energy in the atmosphere (gas molecules, clouds, dust) equal to 22 percent of the solar energy that strikes the top of the atmosphere and a net accumulation of energy at the surface equal to 47 percent of total incoming solar energy.

What is the fate of this absorbed energy, amounting in all to 69 percent of the solar input? The answer is this: after running the machinery of winds, waves, ocean currents, the hydrologic cycle, and photosynthesis, the energy is sent back to space as *terrestrial radiation.*

Terrestrial radiation refers to the electromagnetic radiation emitted by Earth's surface and atmosphere in accordance with the principles summarized in Box 2-4: the amount of energy radiated per unit of area is proportional to the fourth power of the absolute temperature of the radiating substance, and the wavelength of most intense radiation is inversely proportional to the absolute temperature. If the Earth-atmosphere system were a blackbody at temperature, T, the exact relation for the rate of emission of radiation, S, would be:

$$S = \sigma T^4.$$

If S is measured in watts per square meter and T is measured in degrees kelvin, then the proportionality constant σ (the Stefan-Boltzmann constant) has the numerical value 5.67×10^{-8}. On the assumption that the energy returned to space as terrestrial radiation should exactly balance the 69 percent of incoming solar energy that is absorbed, one can compute the effective blackbody temperature of the Earth-atmosphere system to be about 255° K (− 18° C or about 0° F). This is the temperature a perfect radiator would have to have in order to radiate away into space the same amount of energy per unit of area as the Earth-atmosphere system actually does radiate away, on the average.

Actually, the Earth-atmosphere system closely resembles a blackbody radiator, so the above temperature is not unrealistic. Why, then, is that temperature so much lower than the observed mean surface temperature of 288° K (15° C, or 59° F)? The reason is that most of the terrestrial radiation actually escaping from the Earth-atmosphere system is emitted by the atmosphere, not the surface. (You can see from Figure 2-10 that the temperature 255° K corresponds to an altitude of about 5 kilometers, or halfway between sea level and the tropopause.) Most of the terrestrial radiation emitted by the warmer surface does not escape directly to space because the atmosphere is largely opaque to radiation of these wavelengths. The wavelength of radiated energy, you should recall, increases as the temperature of the radiator decreases: Earth, being much cooler than the sun, emits its radiation at longer wavelengths. The peak intensity of terrestrial radiation occurs at a wavelength of around 10 microns, which is in the infrared part of the spectrum, in contrast to solar radiation's peak intensity at a wavelength of about 0.5 microns in the visible part of the spectrum.[26]

[26]There is almost no overlap in the wavelength ranges of solar and terrestrial radiation; at 3 microns, the intensity of solar radiation has fallen to 5 percent of its value at the 0.5-micron peak, while that of terrestrial radiation has only attained about 10 percent of its value at the 10-micron peak. Meteorologists often refer to solar radiation as shortwave radiation and terrestrial radiation as long-wave radiation.

The opaqueness of the atmosphere to outgoing infrared radiation from the surface is due mainly to three atmospheric constituents: carbon dioxide, water vapor, and clouds. (A more modest contribution is made by ozone.) Carbon dioxide absorbs infrared radiation in a narrow band of wavelengths around 3 microns, in another narrow band near 4 microns, and in the wavelengths between 12 and 18 microns. Water vapor absorbs infrared radiation in narrow bands around 1, 1.5, and 2 microns, and in broader ones from 2.5 to 3.5 microns, from 5 to 8 microns, and from 15 microns through the remainder of the infrared. Clouds are very much like blackbodies in the entire infrared part of the electromagnetic spectrum—they absorb most of the infrared radiation that reaches them. The infrared radiation absorbed by carbon dioxide, water vapor, and clouds is subsequently *reradiated* by these substances—much of it being sent back in the direction of Earth's surface and some escaping into space. The atmosphere, therefore, through the properties of clouds, water vapor, and carbon dioxide, acts as a thermal blanket that keeps Earth's surface about 33° C warmer than it would be without these constituents.

Because of this effect of clouds, all else being equal, clear nights are colder than cloudy nights; in the absence of clouds, more infrared radiation leaving the surface escapes directly to space without being intercepted. The additional thermal-blanket effect of water vapor and carbon dioxide is sometimes called the ***greenhouse effect.*** This is because glass, like carbon dioxide and water vapor, is relatively transparent to visible radiation but more opaque to infrared radiation. Light enters a greenhouse more readily than heat can escape, a situation resembling that in the atmosphere. The ***main*** reason a greenhouse is warmer inside than outside, however, is that the glass prevents convection from carrying away sensible heat and latent heat of vaporization.[27] Thus, the term ***greenhouse effect*** is not entirely appropriate for the role of carbon dioxide and water vapor in the atmosphere.

The flows of terrestrial radiation are summarized in Figure 2-13, expressed as a percentage of solar energy striking the top of the atmosphere and computed as a global annual average. Note that the rate at which infrared radiation is emitted from Earth's surface actually exceeds the rate of solar input at the top of the atmosphere and is far larger than the rate at which solar energy is directly received at the surface. This large radiation output and the high temperature responsible

[27] This fact has been demonstrated by building a greenhouse of rock salt (which is as transparent to infrared radiation as to visible radiation) next to a greenhouse of glass. The rock-salt greenhouse, simply by preventing convection, stayed almost as warm inside. (See, for example, Neiburger, et al., *Understanding*.)

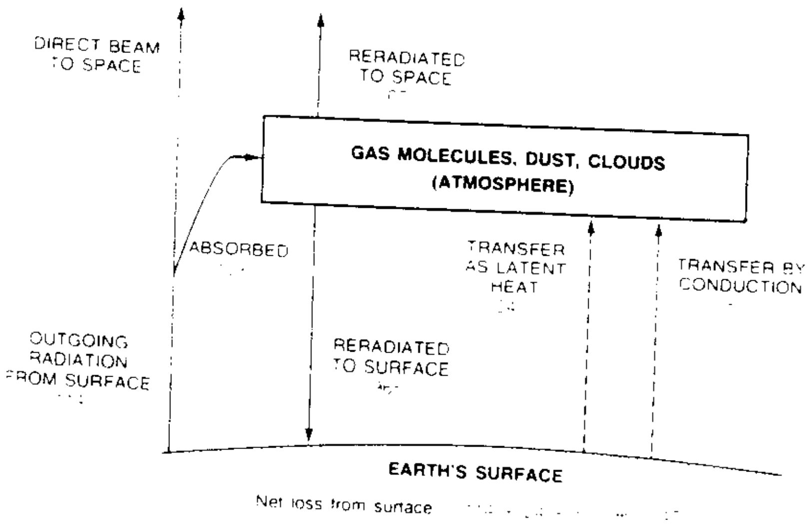


FIGURE 2-13

Flows of terrestrial radiation. Percentage of solar energy incident at the top of the atmosphere (global annual averages).

for it are made possible by the large flow of infrared radiation reradiated to the surface by the atmosphere, as discussed earlier. Any alteration of the composition of the atmosphere by pollution can have an important influence on this downward reradiated energy flow (see Chapter 11).

Two important energy flows other than radiation are represented in Figure 2-13 by broken lines. The larger of these flows is the transfer of energy from the surface to the atmosphere as latent heat of vaporization. That is, energy that has been used to evaporate water at the surface moves into the atmosphere in the form of the latent heat of vaporization that is associated with the water vapor (see Box 2-1). This energy is eventually surrendered to the atmosphere as sensible heat when the water vapor condenses. The magnitude of this surface-to-atmosphere energy flow is equal to almost a quarter of the solar energy flow reaching the top of the atmosphere. The second major nonradiative flow (perhaps a fifth as large as that of latent heat) is the transfer of sensible heat from the surface to the atmosphere by conduction—that is, the warming of the air by contact with the surface. Sensible heat transported across the interface of surface and air in this way then moves upward in the atmosphere by convection. (Of course, in some places and at some times the atmosphere is warmer than the surface, with the result that sensible heat flows from the atmosphere to the surface rather than the reverse. Remember, we have been discussing the global, annual-average situation here.)

Atmospheric Energy Balance and Vertical Motions

With the information in the preceding section, one can construct average energy balances for Earth's surface, for the atmosphere, and for the Earth-atmosphere system as a whole. Such a set of balances is given in Table 2-14, both in terms of the percentage of solar energy flow at the top of the atmosphere and in watts per square meter of Earth's surface area. Energy flow per unit of area is called *flux*. The solar flux through a surface perpendicular to the sun's rays at the top of the atmosphere is called the *solar constant* and is equal to about 1360 watts per square meter [1.95 calories (cal)/cm^2/minute (min), or 1.95 langleys/min, in units often used by meteorologists].[28] As a sphere (to a very good approximation), Earth's total surface area is 4 times the area of the cross section it presents to the sun's rays. Hence, the average amount of solar energy reaching the top of the atmosphere per square meter of Earth's surface is just one-fourth of the solar constant, or 340 watts per square meter, as indicated in Table 2-14. The average solar flux reaching Earth's surface is 173 watts per square meter of horizontal surface. Solar flux measured this way at the surface is often termed *insolation*.

The difference between the amount of energy incident on Earth's surface as radiation (solar and terrestrial) and the amount leaving as radiation is called by meteorologists the *net radiation balance*, or sometimes just net radiation. This amount of energy flow, which as shown in Table 2-14 averages 105 watts per square meter flowing into the surface for the whole globe, is of special meteorological significance because it is the amount of energy available for the climatically crucial processes of evaporation of water and surface-to-air transfer of sensible heat. Many discussions of human impact on climate use the net radiation balance, rather than the incident solar energy, as the yardstick against which civilization's disturbances are measured (Chapter 11).

The pattern of Earth-atmosphere energy flows that has been described here provides the explanation for the observed temperature distribution in the lower atmosphere. Basically, the atmosphere is heated from the bottom—by direct contact with warm land or water surfaces, by the release of latent heat of vaporization when water vapor condenses (virtually entirely in the troposphere and mostly in its lowest third), and by the absorption of terrestrial infrared radiation in water vapor, clouds, and carbon dioxide. Water vapor and clouds are most abundant in the lower troposphere, so that is where a large part of the absorption takes place. These effects, differentially heating the atmosphere near the bottom, produce the general trend of decreasing temperature with altitude in the troposphere. Only above

[28]It is actually not known exactly how constant the solar constant really is (see, for example, S. Schneider and C. Mass, Volcanic dust, sunspots, and temperature trends).

TABLE 2-14
Energy Balances for Earth's Surface, Atmosphere, and Surface-Atmosphere System (global annual averages, accurate to perhaps ± 10%)

	Percentage of solar energy flow at top of atmosphere	*Average energy flow (w/m^2)*
SURFACE-ATMOSPHERE SYSTEM		
Solar radiation reaching top of atmosphere	100	340
Total inflow	100	340
Solar radiation scattered from atmosphere	25	85
Solar radiation scattered from surface	3	10
Terrestrial direct radiation, surface to space	5	17
Terrestrial radiation, atmosphere to space	67	228
Total outflow	100	340
SURFACE		
Solar direct radiation	24	82
Solar diffuse radiation, via clouds	15	51
Solar diffuse radiation, via air, dust	11	37
Terrestrial reradiation from clouds, vapor, CO_2	96	326
Total inflow	146*	496
Reflected solar radiation	3	10
Outgoing terrestrial radiation	114	387
Latent heat to atmosphere	24	82
Sensible heat to atmosphere	5	17
Total outflow	146*	496
ATMOSPHERE		
Solar radiation interacting with clouds, air, dust	76	258
Terrestrial radiation from surface, absorbed	109	371
Latent heat from surface	24	82
Sensible heat from surface	5	17
Total inflow	214*	728
Solar radiation scattered to space	25	85
Solar radiation scattered to surface	26	88
Absorbed radiation reradiated to space	67	228
Absorbed radiation reradiated to surface	96	327
Total outflow	214*	728

*These flows exceed 100 percent of incoming solar radiation because, in addition to the solar throughput, an internal stock of energy is being shifted back and forth between surface and atmosphere.
Source: After Schneider and Dennett, Climatic barriers.

the tropopause does atmospheric heating by absorption of ultraviolet solar radiation (particularly by ozone) reverse the trend and produce temperatures that increase with altitude (refer to Figure 2-10). Ozone also absorbs outgoing terrestrial infrared radiation in a narrow wavelength band at 9.6 microns, to which other atmospheric gases are largely transparent.

The structure of the troposphere, with the warmest air on the bottom, promotes vertical instability (hot air rises, and cold air sinks). Sensible heat and latent heat of vaporization are thus transported upward by convection, and the troposphere tends to be vertically well mixed. The mixing time in the lower half of the troposphere—that is, the time it takes for a molecule at 5 kilometers

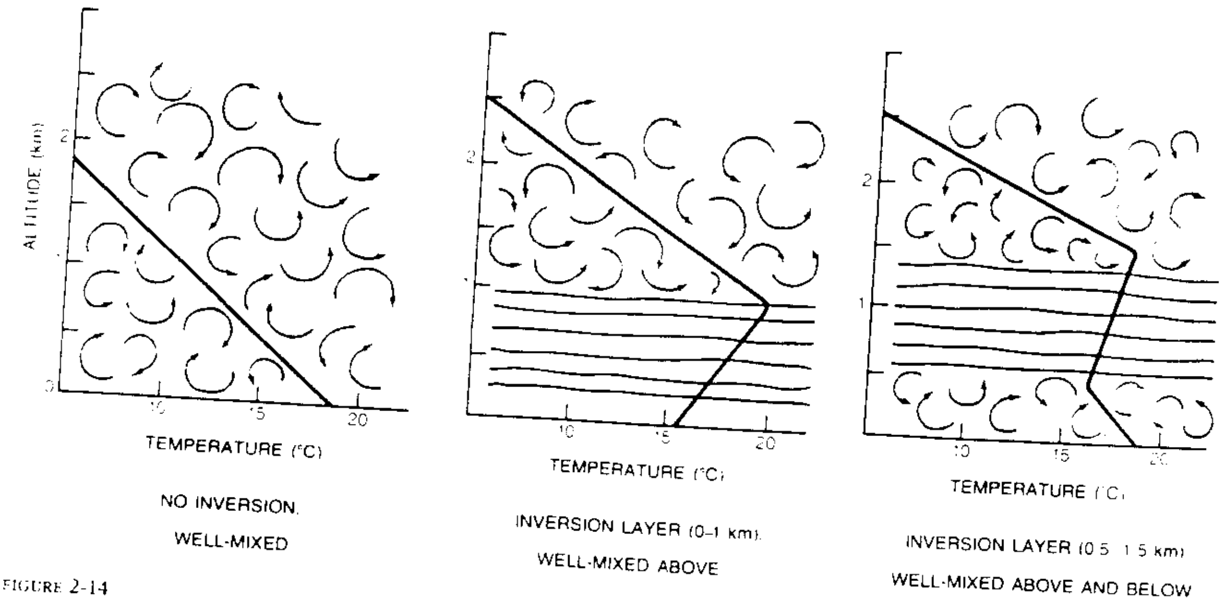


FIGURE 2-14

Inversions in the troposphere. Temperature versus altitude in different circumstances.

altitude to change places with one at the surface—is usually on the order of a few days. (In convective storms it can be hours.) The stratosphere, by contrast, is vertically stratified; like the ocean, its warmest layer is on top, with progressively colder layers below, which tends to suppress vertical motion. The vertical mixing time in the stratosphere is on the order of a year or two.

Under some circumstances, the usual temperature profile of the troposphere near the ground is altered so that temperature increases with altitude. In this situation (called an *inversion*), vertical mixing is suppressed over the altitude range where the temperature is increasing. (Vertical mixing is also suppressed in circumstances where temperature decreases with altitude but not rapidly enough to overcome the stratifying effect of the density variation.) Some of the main possibilities are indicated in Figure 2-14. Inversions are of special importance in environmental science because they inhibit the dilution of pollutants, as is discussed further in Chapter 10.

Variation of Incoming Solar Flux with Place and Season

If the vertical energy flows discussed in the preceding sections were all there were to climate, it would not be so difficult to analyze. So far, however, we have only been working with energy flows averaged over the whole globe and the whole year; the real complexity lies in the tangled patterns of energy flow that arise from differences between day and night, summer and winter, land surfaces and water surfaces, and so on.

Let us first consider the geometrical factors that produce variations from season to season and from latitude to latitude in the amount of solar radiation that strikes the top of the atmosphere. They are:

1. Earth is essentially spherical.
2. Earth travels around the sun in an orbit that is not quite circular.
3. Earth spins on an axis that is tilted 23.5° from perpendicular to the plane of its orbit.

Because Earth is a sphere, the maximum solar flux at any moment is received at the subsolar point (that is, the point that is directly "under" the sun, or equivalently, the point at which the sun appears to be directly overhead) and declining values of flux are received as one moves away from this point on the surface of the sphere. This is so because a flat surface of given area intercepts a maximum of the solar beam when the surface is perpendicular to the beam, and progressively less at angles away from the perpendicular. This situation is illustrated in Figure 2-15.

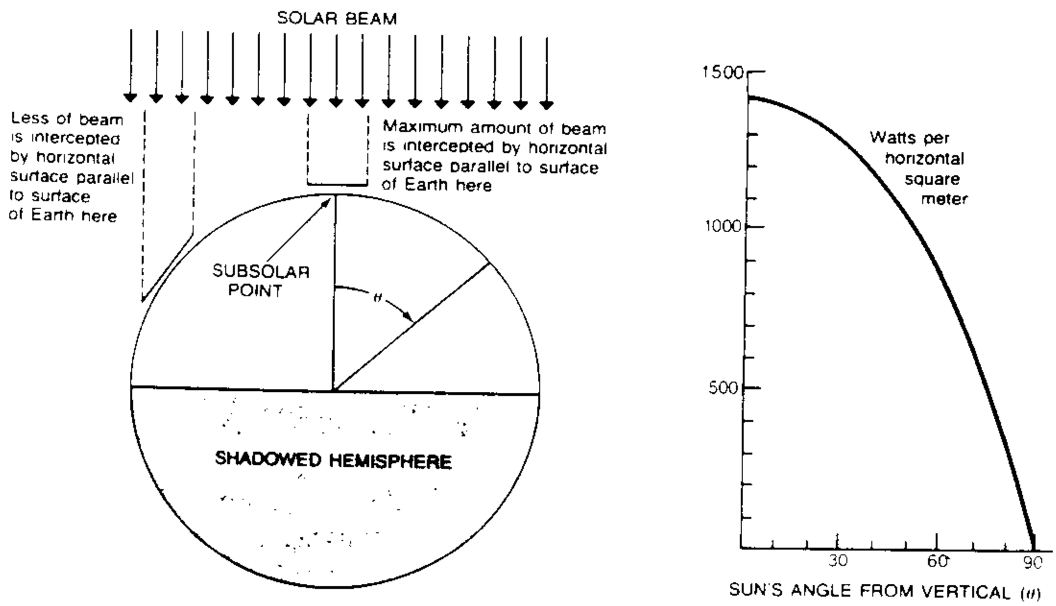


FIGURE 2-15

Insolation at the top of the atmosphere.

Earth's elliptical orbit around the sun deviates from a circle just enough to make the distance between Earth and sun vary by ± 1.7 percent from the mean value of 149.6 million kilometers. This variation in distance produces a difference of ± 3.4 percent in solar flux (that is, the solar energy flow incident on a surface perpendicular to the sun's rays above the atmosphere varies from a maximum of 1406 w/m^2 to a minimum of 1314 w/m^2, the average of 1360 w/m^2 being the "solar constant" given earlier). The minimum distance occurs in the first few days of January, when the Northern Hemisphere is having its winter and the Southern Hemisphere its summer, and the maximum occurs in the first few days of July.

Clearly, the seasons are not produced by the slight ellipticity in Earth's orbit, but rather by the tilt of Earth's axis of rotation. The orientation of this tilt remains fixed as Earth circles the sun, as shown schematically in Figure 2-16. This means that the Northern Hemisphere is tilted directly toward the sun at the June 21 solstice, corresponding to the first day of summer in the Northern Hemisphere and the first day of winter in the Southern Hemisphere, and the Southern Hemisphere is tilted directly toward the sun at the December 21 solstice. On June 21 the subsolar point is on the Tropic of Cancer (23.5° north latitude)[29] and the entire area north of the Arctic Circle (66.5° north latitude) is illuminated by the sun during all twenty-four hours of the day. On December 21 the subsolar point is on the Tropic of Capricorn (23.5° south latitude), and the entire area within the Antarctic Circle is illuminated for all twenty-four hours of the day. At the equinoxes (March 21 and September 23) the subsolar point is on the equator, and day and night are each twelve hours long everywhere on Earth (except in the vicinity of the poles, which experience continuous twilight).

The net effect of these geometrical aspects of the Earth-sun relationship is to produce strong north-south differences, or *gradients*, in the incoming solar flux perpendicular to Earth's surface at the top of the atmosphere. The size and shape of the gradients vary

[29]The subsolar point becomes a subsolar line as Earth performs its twenty-four hour rotation.

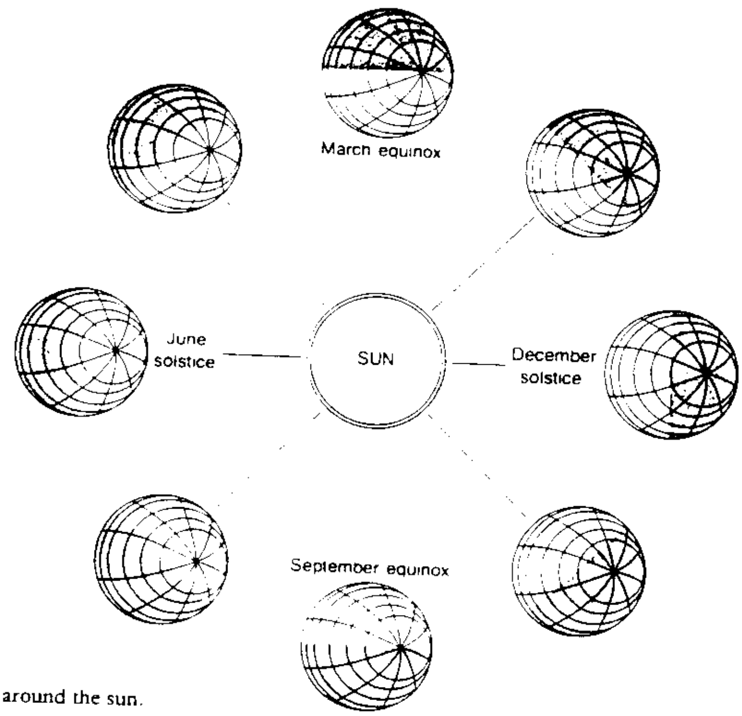


FIGURE 2-16

Earth's orbit around the sun.

with the seasons. Figure 2-17 shows the solar flux at the top of the atmosphere, in calories per square centimeter per day, as it varies throughout the year at three different latitudes in the Northern Hemisphere. Note the strong equator-to-pole contrast in the winter, which is reduced as summer approaches and actually reverses in midsummer. (The summer pole receives more energy per unit of surface area in midsummer than does the equator, because the sun is shining twenty-four hours a day at the summer pole.)

The amount of solar radiation reflected back into space also varies with latitude (more is reflected at the extreme northerly and southerly latitudes, where the radiation strikes the surface at angles far from perpendicular) and with season (ice and snow reflect more than do vegetation and water). Finally, the amount of terrestrial radiation leaving the top of the atmosphere varies with latitude and season, because the emission of such radiation depends on temperature and other characteristics of the atmosphere and the surface below. An accounting of the

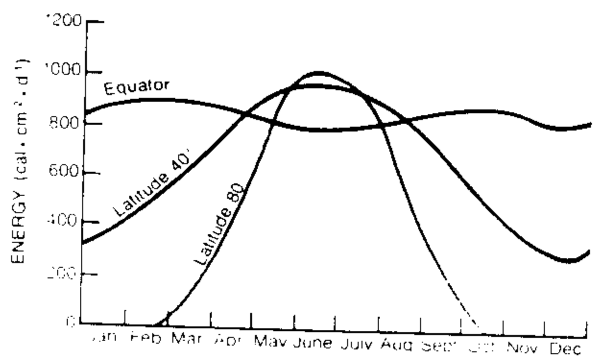


FIGURE 2-17

Seasonal variation of solar flux on a horizontal surface outside the atmosphere. (From Gates, 1971.)

radiation flows across an imaginary surface at the top of the atmosphere, then, must include incoming sunlight, outgoing reflected sunlight, and outgoing terrestrial radiation. Where the inflow exceeds the outflows across such a surface, it is said that the underlying column of atmosphere has a *heating excess*. Where the outflows exceed the inflow, there is said to be a *heating deficit*.

The Machinery of Horizontal Energy Flows

If there were no mechanisms to transfer energy horizontally over Earth's surface from regions of heating excess to regions of heating deficit, the result would be much greater extremes in conditions than actually exist. Energy would accumulate in the areas of heating excess until the additional outgoing infrared radiation produced by higher temperatures restored a balance; similarly, energy would be lost to space from areas of heating deficit until the drop in outgoing radiation associated with lower temperatures restored the incoming-outgoing balance in those areas. To say that the extremes of temperature on Earth's surface would be greater under these circumstances is an understatement. It is plain from Figure 2-17, for example, that the region north of 80° north latitude receives no incoming radiation at all from mid-October to late February (although, of course, the emission of outgoing terrestrial radiation continues). In the absence of energy inflows from warmer latitudes, then, the radiative energy loss in winter at the latitudes having darkness twenty-four hours a day would cool the surface rapidly and continuously toward absolute zero.

But there *are* horizontal energy flows—principally, the transport of sensible heat and latent heat of vaporization by the motions of the atmosphere and the transport of sensible heat by ocean currents. (These *convective* energy transfers are much bigger and faster than conduction. See Box 2-3, to review the difference.) These flows are driven largely by the north-south temperature differences arising from the radiation imbalances just described, and they are largely responsible for the general features of global climate. The overall pattern of the north-south energy flows on an annual average basis is indicated in Table 2-15.

The atmospheric motions that carry energy toward the poles as sensible heat and latent heat must, of course, be balanced by return flows of air toward the equator. Otherwise, air would be piling up at the poles! Two kinds

TABLE 2-15
Average Annual North-South Energy Flows (w/m^2)

Latitude zone	*Net radiation at top of atmosphere (incoming minus outgoing)*	*Net latent heat (precipitation minus evaporation)*	*Net sensible heat transport by atmosphere*	*Net sensible heat transport by ocean*
60–70°N	−65	11	43	11
50–60°N	−40	20	5	15
40–50°N	−16	12	−5	9
30–40°N	5	−17	0	12
20–30°N	19	−41	21	1
10–20°N	31	−15	−3	−13
0–10°N	39	44	−51	−32
0–10°S	41	19	−32	−28
10–20°S	37	−21	−12	−4
20–30°S	27	−43	11	5
30–40°S	12	−25	5	8
40–50°S	−11	11	−8	8
50–60°S	−39	36	−12	15

Note: Sum of entries in any given zone is zero, denoting balance of inflows and outflows averaged for year. (Positive numbers represent net inflow; negative numbers represent net outflow.)

Source: Modified from Budyko, in SMIC, *Inadvertent climate modification*, p. 91.

of circulation that move heat poleward but keep the air distributed are particularly important. The first of these is the thermal circulation illustrated in Figure 2-18A; the poleward flow is high in the troposphere; the equatorward flow, near the surface. The second is the cyclonic circulation shown in Figure 2-18B, which can be thought of as taking place in a horizontal plane. Actually, Earth's major atmospheric circulations are generally combinations of vertical and horizontal components. The term

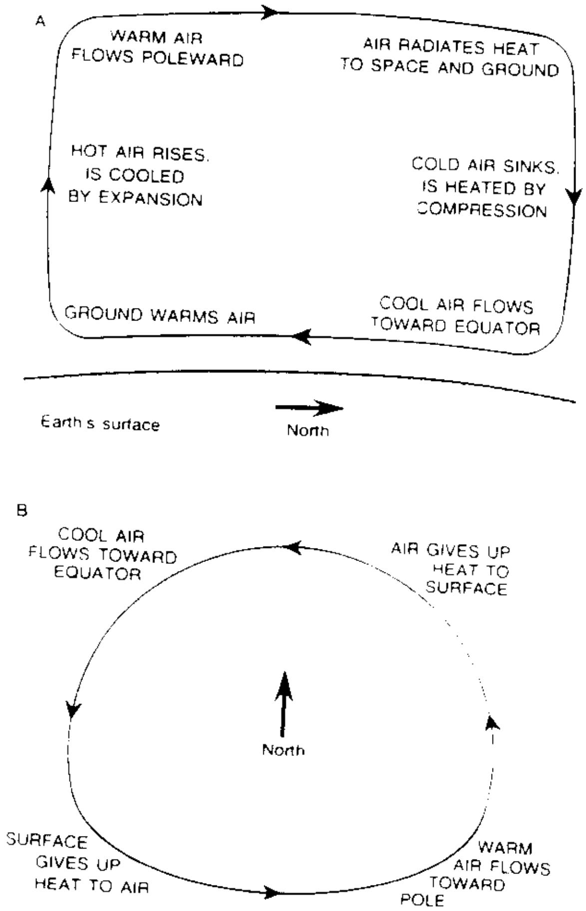


FIGURE 2-18

North-south air flows with poleward transport of heat. A. Thermal circulation, cross-section looking west, parallel to Earth's surface in the Northern Hemisphere. B. Cyclonic circulation, looking down on Earth's surface in the Northern Hemisphere.

wind is usually reserved for the horizontal motion, and the terms *updraft* and *downdraft* describe the vertical. The horizontal motion is almost always much faster than the vertical: typical wind speeds range from 1 to 20 meters per second; typical vertical speeds in large-scale atmospheric circulations are 100 times less, although local updrafts associated with clouds and mountain ranges may be 10 meters per second or more. Typically, then, a "parcel" of air (an arbitrarily defined collection of molecules whose behavior one chooses to trace) might move 100 kilometers horizontally while rising 1 kilometer.

Like solids, gases such as air move in response to the forces exerted upon them in ways described by Newton's laws of motion. One important kind of force in the atmosphere is associated with *pressure gradients*. The pressure-gradient force is a push from regions of high pressure, associated with high temperature and/or density, toward regions of low pressure. Another important force in the atmosphere is gravity, which exerts a downward pull on every molecule of air. (Vertical motions in the atmosphere generally are slow because the vertical pressure-gradient force usually balances the force of gravity almost exactly.) The distribution of regions of high pressure and low pressure in the atmosphere is complex, owing not only to the variations in insolation, reflection, and absorption with latitude and altitude, but also to local differences associated with topography and the distribution of vegetation, land, and water. But the atmospheric circulation patterns are *not* what would be expected from consideration only of the pressure gradients associated with these features, together with the force of gravity, because two other important factors come into play. These are the *Coriolis deflection* (usually inappropriately called the Coriolis force) and *friction*.

The Coriolis deflection is a complication that arises from the rotation of the planet. Specifically, Earth tends to rotate out from under objects that are in motion over its surface (for example, fired artillery shells and moving parcels of atmosphere). That is, such objects do not go quite where they seem to be heading, because the place they were heading is rotating at a different velocity than the point of origin (rotational speed is highest at the equator, where a point must move some 40,000 km/day, and lowest near the poles). Thus, in the Northern

Hemisphere a parcel of air moving north will appear to a terrestrial observer to be deflected to the right (east), the direction of the Earth's rotation, because it will carry the higher velocity of its place of origin. The Coriolis "force" is the *apparent* extra force (besides pressure gradients, gravity, and any other "real" forces that may be acting) needed to explain to an observer on Earth's rotating surface the observed paths of moving objects. We say "apparent" because no work is done to produce the deflection—it is a function of the position of the observer. A person in space observing the trajectory of an artillery shell relative to the solar system (assume, for example, that the shell were visible but the rest of Earth were invisible) would not see any Coriolis deflection. The magnitude of the Coriolis deflection is greatest at the poles and zero on the equator; the magnitude also varies in direct proportion to the speed of an object—a stationary object is not subject to Coriolis deflection.[10] The direction of the Coriolis deflection is always perpendicular to the direction of an object's motion. It changes the *direction* of motion but not the speed. Motions in the Northern Hemisphere are deflected to the right; motions in the Southern Hemisphere, to the left.

Consider the effect of the Coriolis deflection in the Northern Hemisphere on the flow of air into a region of low atmospheric pressure from surrounding regions of higher pressure. The pressure-gradient force tries to drive the flow straight in, but the Coriolis deflection bends it to the right (Figure 2-19). The resulting spiral patterns are actually visible in most satellite photographs of Earth, because the winds carry clouds along with them (Figure 2-20). The spirals associated with low-pressure centers (clockwise in the Southern Hemisphere, counterclockwise in the northern) are called *cyclones*, and the outward spirals associated with high-pressure centers (clockwise in the Northern Hemisphere, counterclockwise in the southern) are called *anticyclones*.

The force of friction adds two features to the wind patterns described thus far. It slows down the wind near Earth's surface—most dramatically in the first few tens of

[10]The magnitude of the Coriolis deflection associated with horizontal motion at velocity v m/sec at latitude ϕ is $2\Omega v \sin\phi$/kg mass, where $\Omega = 7.29 \times 10^{-5}$ radians/sec is the angular velocity of Earth's rotation. (There are 2π radians, or 360 degrees, in one revolution.) See, for example, Neiburger, et al., *Understanding*, pp. 99–104.

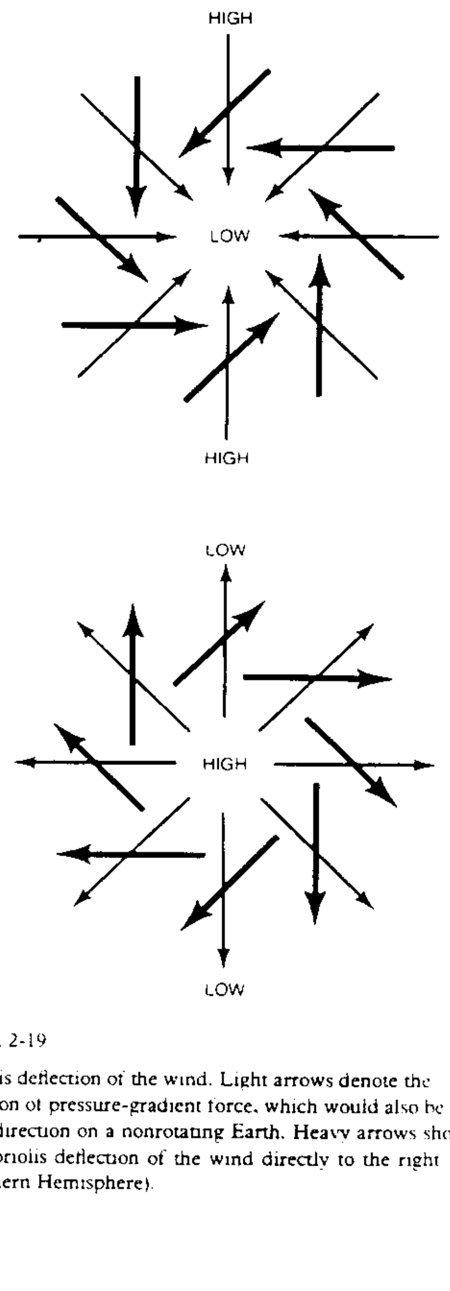


FIGURE 2-19

Coriolis deflection of the wind. Light arrows denote the direction of pressure-gradient force, which would also be wind direction on a nonrotating Earth. Heavy arrows show the Coriolis deflection of the wind directly to the right (Northern Hemisphere).

FIGURE 2-20

Atmospheric circulation patterns as revealed by cloud distributions. (NASA).

meters, but still significantly at altitudes up to a few hundred meters. And, because of the way the friction force interacts with the pressure-gradient force and the Coriolis deflection, it causes the *direction* of the wind to change with altitude for the first several hundred meters. The horizontal pressure gradients themselves may be quite different at one altitude than at another, which also gives rise to significant changes in the wind patterns as one moves away from Earth's surface.[31]

[31]More detailed explanations of the operation of the friction force and the variation of pressure gradients with altitude are given in Neiburger et al., *Understanding*, pp. 109–114.

The General Circulation

The overall pattern of atmospheric motions resulting from the phenomena just described is called the ***general circulation.*** Its main features are illustrated in Figure 2-21. The associated variation of average sea level pressure is indicated in Figure 2-22.

The surface circulation in the tropical regions north and south of the equator is dominated by the trade winds, blowing, respectively, from the northeast and the southeast. These very steady winds are associated with the pressure drop between the subtropical highs and the

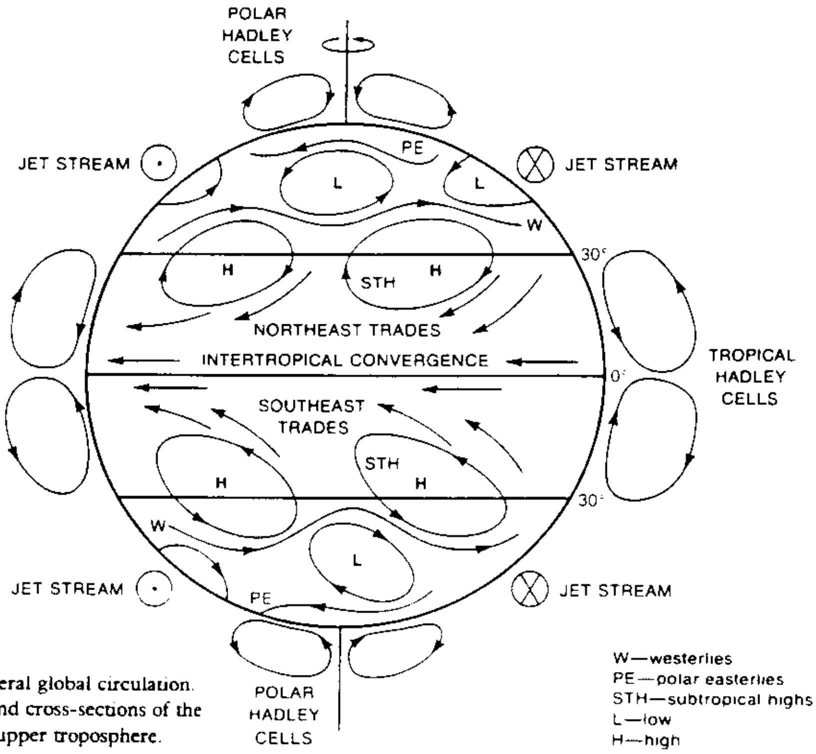


FIGURE 2-21

Main features of the general global circulation. Flows near the surface and cross-sections of the main circulations in the upper troposphere.

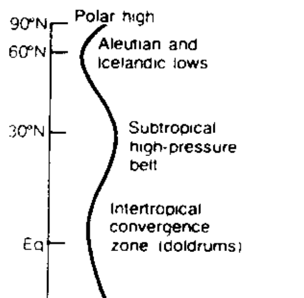


FIGURE 2-22

Variation of average sea-level pressure with latitude. (From Neiburger, Edinger, and Bonner, 1973).

low-pressure *doldrums* on the equator. The trade winds are deflected from the direction of the pressure drop by the Coriolis force—to the right in the Northern Hemisphere, to the left in the southern—as explained above. The region where the trade winds meet to form a belt of easterly winds encircling the globe near the equator is called the *intertropical convergence zone.*

The vertical part of the circulation in the tropics consists of thermal circulations of the form shown in Figure 2-18A, one immediately north of the equator and one immediately south. These are called *Hadley cells,* after the British meteorologist who first postulated their existence. In the Hadley cells, the air rising over the equator is moist as well as warm. As it rises, the air cools, whereupon some of the contained moisture condenses into droplets and falls as rain. The latent heat of

25

vaporization released in this process helps drive the air farther upward, producing more condensation, more release of latent heat, and more rain. Thus, the air rising near the equator in the Hadley cells is largely "wrung out," producing in the process the very rainy climates for which the tropics are known. Having lost its moisture and some of its sensible heat in ascending, the air flowing poleward in the upper part of the Hadley cells continues to lose heat by radiating energy to space more rapidly than it absorbs radiant energy from the warmer atmospheric layers and the surface below. At around 30° north and south latitude, the relatively cold, dry air commences to sink. In sinking into higher pressure it is warmed by compression. This descending flow of warm, dry air in the 30° latitude belts is a major reason these belts are characterized by deserts all around the world. (The Sahara of northern Africa, the Kalahari of southern Africa, the Atacama of Chile, and the Sonoran desert of Mexico and the United States are examples.) Finally, as the dry air moves equatorward on the surface to complete its circuit, it picks up both heat and moisture from the increasingly warm surfaces of the land and water of the tropics.

Thermal circulation patterns similar to the Hadley cells of the tropics are also found in the vicinity of the poles, but they are smaller and weaker than those on either side of the equator. The temperature and pressure differences driving the polar flow are less than those nearer the equator, and the transfers of moisture and latent heat are also much smaller.

Although it was postulated at one time that there must be an indirect cell linking the equatorial and polar cells in each hemisphere – that is, a cell in which air sinks on the side toward the equator, flows poleward on the surface, and rises on the polar side – measurements indicate that this pattern is either extremely weak or entirely missing. Rather than being borne by such circulations, the poleward energy flow in the middle latitudes is accomplished instead by the great, swirling, horizontal flows associated with the subtropical highs (see Figure 2-18B, as well as Figure 2-21) and with the wavy boundary between those highs and the subpolar lows. Along that boundary in both hemispheres, the winds are predominantly westerly (that is, flowing from west to east) both at the surface and high in the troposphere.

Embedded in the westerlies at the upper edge of the troposphere and on the boundary between the subpolar lows and the subtropical highs are the *circumpolar jet streams*, encircling the globe in a meandering path covering latitudes from 40° to 60°. The core of a jet stream is typically 100 kilometers wide and 1 kilometer deep, and is characterized by wind speeds of 50 to 80 meters per second [110 to 180 miles per hour (mph)]. These high speeds are the result of a large pressure change over a relatively short horizontal distance on the boundary between low-pressure and high-pressure circulation systems. In addition to the circumpolar jet streams, there are westerly subtropical jet streams at about 30° north and south latitudes, associated with the poleward edges of the Hadley cells, and some seasonal jet streams of lesser importance.[32]

On the boundary between the subpolar lows and the polar highs, there are weak easterly winds at the surface, giving way to westerlies at higher altitudes. The high-altitude flow, then, is entirely westerly. The part of this flow lying poleward of the subtropical highs is cold air – essentially a great west-to-east spinning cap of it, draped over each pole – and is called the *circumpolar vortex*. The circumpolar vortex in the winter hemisphere is larger and wavier at the edge than the one in the summer hemisphere, because the equator-to-pole temperature difference that is the basic driving force behind these features is much greater in winter. The waviness at the edge of the circumpolar vortex is caused by cold, low-pressure circulation systems being pushed equatorward into the temperate zones, producing the cold fronts and accompanying storms common in winter.

In the summer hemisphere, the characteristics of the general circulation associated with the equator-to-pole temperature difference are much less strongly developed. The intertropical convergence crosses the equator into the summer hemisphere, the summer-hemisphere Hadley cell weakens and nearly vanishes, and the circumpolar vortex subsides. This weakening and relative disorganization of the general circulation in the summer hemisphere permits the pattern to be dominated by asymmetries connected with the distribution of land and bodies of water. Among the most important of these are

[32] Strahler and Strahler, *Environmental geoscience*, pp. 96–98.

the *summer monsoons* of Asia and sub-Saharan Africa, in which moist, cool air sweeps inland over warm landmasses, rises, and drops the moisture as rain. This climatic feature is essential to the food supply of a substantial part of the world's population.

Weather, Climate, and Climate Change

Weather and climate are not the same thing, and the difference between the two terms involves the time span in which one is interested. *Weather* refers to the conditions of temperature, cloudiness, windiness, humidity, and precipitation that prevail at a given moment, or the average of such conditions over time periods ranging from hours to a few days. The weather can change from hour to hour, from day to day, and from week to week.

Climate, on the other hand, means the *average pattern* in which weather varies in time, and the average is determined over longer periods (from a month to decades). Thus, one might speak of the climate of a given region as being characterized by hot, dry summers and severe winters. Within this region, one would still expect some periods (days or weeks) of cool weather in summer and mild weather in winter. If the weather for an entire summer were cooler than usual, one would still speak of an exceptional summer's weather or a short-term climatic fluctuation but not of climate change. But if the weather averaged over ten or twenty consecutive summers were significantly cooler than the average for the previous thirty, then one could begin to call the phenomenon a change in climate.

Local weather and climate are determined by the complicated interaction of regional and global circulation patterns with local topography, vegetation, configuration of lakes, rivers, and bays, and so on.[33] Successful weather prediction requires combining knowledge of these general patterns and known local features with the most detailed available information about the weather conditions of the moment and the past few days—not just at the location whose weather is being predicted, but at other locations, as well. That is, to predict tomorrow's weather, one must know as much as possible about today's. What happens in Los Angeles on Saturday may be largely foreseeable from what was happening in San Francisco on Friday, which could have been a storm that was in Portland on Wednesday, which originated as a disturbance in the Gulf of Alaska on Monday.

As everyone knows, forecasting the weather even one day ahead is not an exact science. This inexactitude is not because there are undiscovered physical processes at work; all the basic physical laws involved are actually known. The imperfections in prediction have two origins: first, the actual system involved—atmosphere, ocean, other water bodies, land—is far too complicated for the known physical laws to be applied exactly, even with the largest computers; second, the initial conditions—the state of the system at the time the analysis begins—can be specified only approximately, owing to fundamental monitoring limitations. The farther in advance one wishes to make a local weather forecast, the more difficult the task becomes, because the larger is the area of the globe whose present conditions can influence subsequent conditions at the place one is interested in. Also, temperature and pressure anomalies small enough to slip through the global network of weather stations today may have grown large enough in a week's time to determine the weather over large regions. Weather satellites have made the meteorologist's task somewhat easier, especially because they provide information about meteorological conditions over the oceans, where surface monitoring stations are relatively scarce; but reliable weather forecasting a week in advance is still not a reality. Because the meteorological system is so complex, reliable forecasting two weeks ahead may never be attainable.

Understanding and predicting the changes in climate that have occurred and will continue to occur over time spans of decades, centuries, and millennia is in some respects even more difficult than forecasting the weather from day to day. The weather forecaster has the disadvantage of having to deal with tremendous detail in terms of variations over short times and short distances, but the advantage of access to a tremendous body of observa-

[33] In this chapter we have emphasized the large-scale processes that govern the patterns of weather and climate over large regions, continents, and hemispheres. Investigating in detail the *microclimates* that influence intimately the human and biological communities in specific small regions would take us deeper into the demanding and technical subject of meteorology than space and the nature of this book permit. The interested reader should consult Neiburger et al., *Understanding our atmospheric environment*, or another basic meteorology text for an introduction to the micrometeorology missing here

27

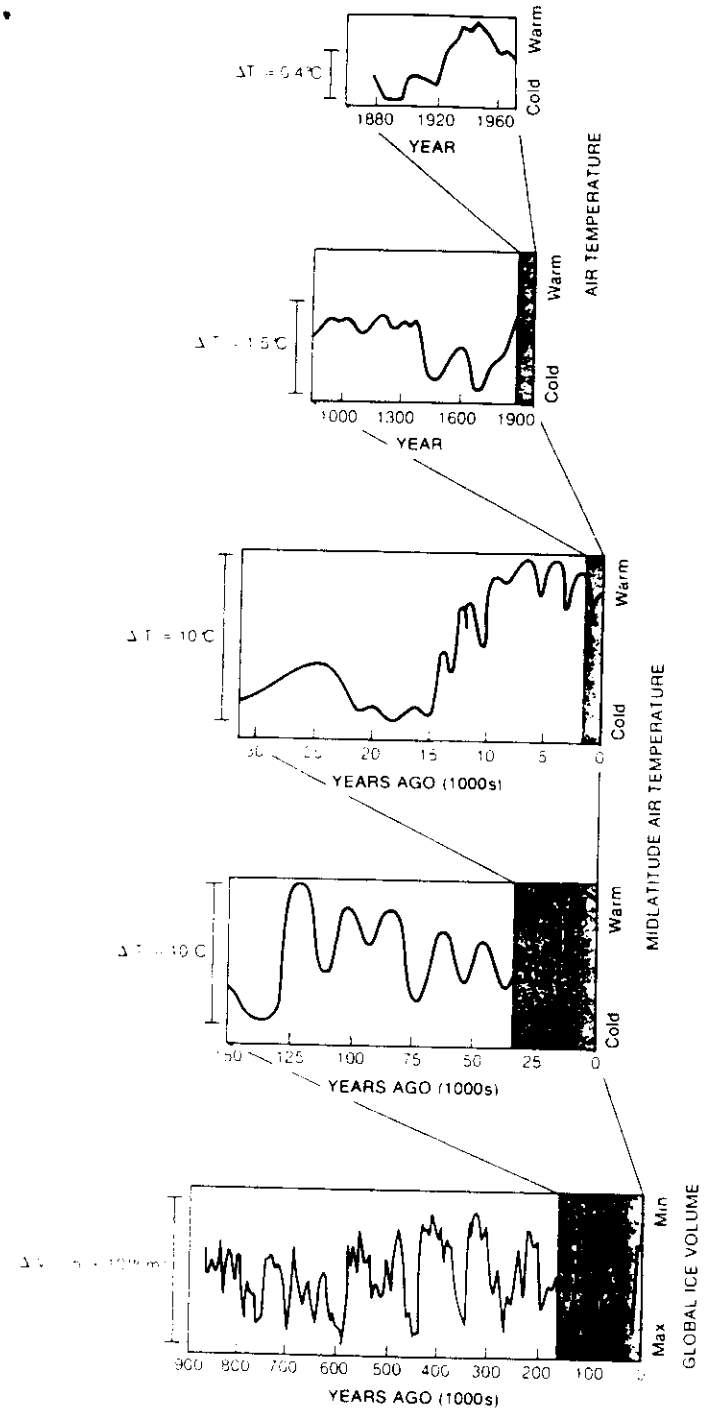


FIGURE 2-23

The Earth's climate—the past million years. (After Bolin, 1974.)

tional data—the weather happens every day, and many skilled observers with good instruments are watching and recording. The climatologist has reasonably good, direct, observational data only for the past few decades, spotty records for the past century or so, and only indirect evidence before that (scattered historical writings dating back several centuries, some archeological evidence going back a few thousand years, and only the fossil record and geological evidence before that). The plight of the climatologist is something like trying to learn all about weather on the basis of good data for the past two days, spotty data for the past two weeks, and only some fuzzy clues as to what might have happened before then.

On the basis of the limited evidence available to them, climatologists have done a remarkable job of reconstructing in a plausible way Earth's climatic history for about the past million years and, more roughly, for the past 60 million years.[34] Some of the main features of the more recent history are illustrated in Figure 2-23. It shows a series of fluctuations in average midlatitude air temperature, with different amplitudes associated with different time scales. (The amplitude of a fluctuation is the difference between the maximum and minimum values associated with it.) The amplitude of the indicated variation in the past hundred years is about 0.5° C (roughly, 1° F); the amplitude of the fluctuations with a time scale of a few hundred to a thousand or so years is from about 1.5° to 2.5° C; and the amplitude of the fluctuations with a time scale of tens of thousands of years is in the range of between 5° and 10° C.

Several points with respect to this climatic history deserve emphasis:

1. Variability has been the hallmark of climate over the millennia. The one statement about future climate that can be made with complete assurance is that it will be variable—a conclusion not without significance for food production (see Chapter 7).

2. Rather small changes in average midlatitude temperatures are likely to be associated with larger changes in the seasonal *extremes* of temperature at those latitudes,

[34] Study of Man's Impact on Climate (SMIC), *Inadvertent climate modification*, pp. 28–45; Bert Bolin, Modelling the climate and its variations.

and with even larger changes in the extremes nearer the poles. These extremes in temperature are likely to have a more crucial influence on flora and fauna than are the averages.

3. Modest changes in average temperature and the accompanying larger changes in temperature extremes are often associated with significant changes in the circulation patterns, humidities, amounts of rainfall, and other features that make up regional climates. These changes, too, can drastically influence the character of the plant and animal communities that exist in different regions.

4. The drop in average midlatitude temperature associated with major ice ages is as little as 4° or 5° C. It is possible that an even smaller drop could *trigger* such an ice age.[35]

5. The onset of significant climatic change in the past has sometimes been quite rapid. There is evidence to suggest, for example, that the advance of the continental ice sheets in a cooling period that commenced about 10,800 years ago destroyed living forests wholesale within the space of a single century or less.[36]

6. The world finds itself in the last part of the twentieth century A.D. in one of the warmest periods in recent climatic history. What people alive today assume is normal – namely, the climate of the past thirty to sixty years – is in reality near one of the extremes of the persistent historical fluctuations. Even without the possibility of inadvertent human influence on climate, it could not be predicted on the basis of present knowledge how much longer this present extreme climate might last. Past evidence suggests, however, that when it ends, it will end (barring human intervention) with a cooling trend.

Several questions present themselves. What has caused the climatic fluctuations of the past? If climate change is a historical fact of life, is there any reason to worry about future changes? Is civilization capable of inadvertently influencing climate – for example, by accelerating natural change or initiating a different trend? Is there any prospect of deliberate intervention to stop a threatening trend in climate? We consider the first two questions in the next few paragraphs; the last two will be touched on only briefly here and then taken up in detail in Chapter 11.

The causes of past changes in climate are not well understood. One possibility is variation in the rate at which the sun emitted energy, but there is no convincing evidence to show that this has occurred, and no convincing theory that predicts it has been proposed.[37] A second set of possibilities involves changes in atmospheric composition, influencing the transmission of incoming solar radiation and/or outgoing terrestrial radiation. For example, periods of intense volcanic activity might have added enough ash to the atmosphere to affect climate significantly, or biological and geophysical processes might have caused the atmospheric stock of carbon dioxide to deviate appreciably from its present value. Evidence to connect these possibilities with the actual onset of ice ages is lacking, however. Still another possibility is variations in Earth's orbit, which would have affected the amount and timing of incident solar energy. Such variations are known to have occurred and to be occurring, but they appear to have been too small *by themselves* to have produced the onset or retreat of ice ages.

A likely contributing factor is that the circulation patterns and other phenomena that produce the gross features of climate are not very stable. If they were stable, the systems governing climate would tend to return after any disturbance to the conditions that prevailed before the disturbance. Many examples are known in physical science of systems that are stable if the disturbances imposed on them are not too large, but unstable – that is, they fall into altogether different patterns of behavior than the initial ones – if a disturbance exceeds some threshold. The idea of stability and the related concept of *feedback mechanisms* are explored by means of some simple examples in Box 2-5.

The number of feedback mechanisms influencing the stability of the global climate, both positively and negatively, is very large.[38] Some of the more important ones are indicated in Figure 2-24. It is not unlikely, under these circumstances, that the ocean-atmosphere

[35] SMIC

[36] H. H. Lamb, *The changing climate*, p. 236

[37] For intriguing speculations, see Schneider and Mass, Volcanic dust.

[38] W. W. Kellogg and S. H. Schneider, Climate stabilization: for better or for worse?

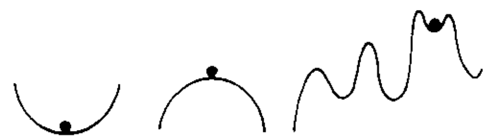
BOX 2-5 Stability and Feedback

In general, a system is said to be *stable* if it tends to return to its initial state after any perturbation, and *unstable* if a perturbation would cause the system to depart permanently from its initial state. Often the character of the perturbation determines whether a system is stable or unstable.

As a simple example, consider a system consisting of a marble in a round bowl, as shown in cross-section in the diagram here. The "state" of this system is the position of the marble, and in the initial state the marble is at rest at the bottom of the bowl (A). The system is stable against small sideways displacements of the marble—that is, if one pushes it to the side, the marble will roll back toward the center and eventually come to rest again in the initial state. The system is *not* stable against displacements so large that the marble is pushed out of the bowl (or against entirely different kinds of perturbations, such as turning the bowl over!). A similar system (B), in which the bowl is inverted and the marble rests on top, is unstable even against small perturbations. Any displacement causes the marble to roll away.

It seems likely that the stability properties of Earth's meteorological system are analogous to those of the third arrangement (C). In this case, there are several states of equilibrium—that is, states in which the system can remain unchanged for an extended period. Each equilibrium state is stable against small perturbations, but a large perturbation can cause the system to shift to a different equilibrium, which in turn is stable until a large enough perturbation happens along to cause another shift. Such systems are often called *metastable*.

When the forces or flows that affect the system are in balance, at least temporarily, then the system will be in equilibrium. A perturbation generally alters the forces or flows, and the *way* it alters them determines whether the system is stable or unstable. In a stable system, a perturbation sets in motion changes in forces or flows that tend to restore the initial state. Processes that work this way are called *negative feedback* (a change in the state of a system induces an effect that reduces the change). *Positive feedback* occurs when a change induces an effect that enlarges the change.



A. STABLE B. UNSTABLE C. METASTABLE

Consider the marble inside the bowl again. A displacement leads to a force that pulls the marble back toward its initial position, reducing the displacement; this is negative feedback. If the marble is on top of an inverted bowl, the force that results from any displacement acts to increase the displacement; this is positive feedback. Cause and effect relations of this kind—cause producing effect that reacts back on cause—are sometimes called *feedback mechanisms* or *feedback loops*, terms that originated in the narrower context of control systems for machinery, aircraft, and so on.

We have already considered some simple feedback mechanisms in this chapter without calling them that. An important one for climate is the negative feedback mechanism connecting surface temperature and outgoing terrestrial radiation. If a perturbation should increase the surface temperature, this would cause the rate of energy outflow in terrestrial radiation to increase; all else being equal, this would cause the surface temperature to fall, reducing the initial perturbation. If a perturbation should reduce the surface temperature, the rate of energy loss via terrestrial radiation would fall, which would tend to raise the temperature again.

Often, one must trace feedback mechanisms through two or more physical processes. Consider, for example, another mechanism involving Earth's surface temperature: an increase in surface temperature causes an increase in evaporation rate, causing increased concentration of water vapor in the atmosphere, causing an enhanced greenhouse effect, causing a further increase in surface temperature. In complicated systems such as those influencing climate, many different feedback mechanisms—some positive, some negative—are operating at the same time, and predicting the net effect of a given perturbation can therefore be very difficult.

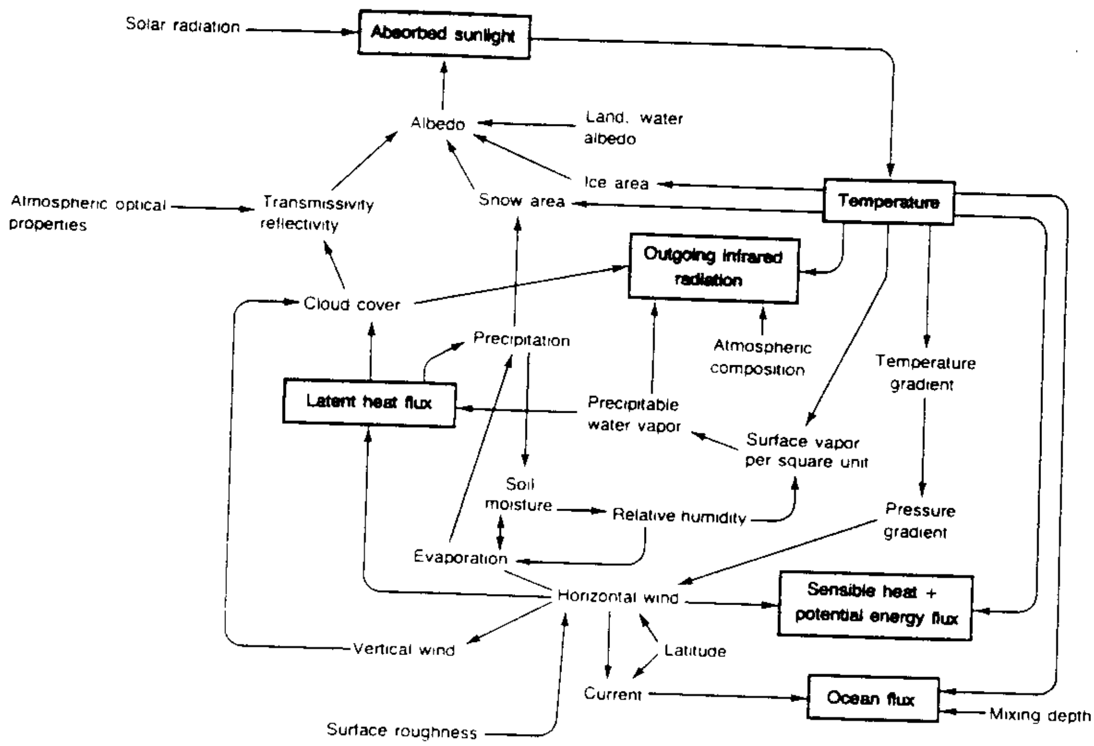


FIGURE 2-24

Some feedback loops governing global climate. (After Kellogg and Schneider, 1974.)

system governing climate belongs to the class of physical systems that are stable against small disturbances but change quite drastically and semipermanently if a larger disturbance happens along. In the case of Earth's climate, a disturbance (such as a change in solar output or in the amount of solar energy reaching Earth) need not involve an energy flow larger than the flows in the main climatic processes in order to start an instability. It would be enough that the energy associated with the disturbance be sufficient to tip the balance between two competing feedbacks, each of which might involve a much larger energy flow than that of the disturbance. Such phenomena—sometimes called *trigger effects*—are encountered frequently in environmental sciences.

There is increasing evidence suggesting that certain large variations in regional weather patterns are caused by a feedback effect in which anomalies in ocean surface temperatures play a major role.[39] The heat capacity of the oceans (that is, the amount of energy stored for each degree the temperature rises) is far greater than those of the atmosphere or the land. As a result, small changes in the temperature of large masses of ocean water absorb or release enormous quantities of heat, and this permits the oceans to serve as thermal buffers, moderating what would otherwise be more extreme changes in seasonal temperatures in the overlying atmosphere and on the adjacent landmasses. That weather and climate are largely the result of the behavior of the ***ocean-atmosphere system***, and not the behavior of the atmosphere alone, has been known for a long time. More specifically, however,

[39] Jerome Namias, Experiments in objectively predicting some atmospheric and oceanic variables for the winter of 1971–72. S. A. Farmer, A note on the long-term effects on the atmosphere of sea surface temperature anomalies in the north Pacific Ocean.

recent studies indicate that weather cycles of hot and cold or wet and dry, which are observed on a time scale of ten to fifteen years in many regions, are connected with changes in ocean surface temperatures that persist over large areas for similar periods. Apparently, such changes can produce a longitudinal (east-west) shift in the wavy pattern of alternating high- and low-pressure zones at the edge of the circumpolar vortex. What causes the changes in ocean temperatures themselves is not completely understood.

A hypothesis believed by many climatologists to explain longer-term climatic change is that the historical changes in Earth's orbit have been large enough to trigger a positive feedback involving the albedo in polar regions: reduced solar input leads to more areas being covered with ice and snow, leading to increased albedo (more reflectivity), leading to less solar energy absorbed at the surface, leading to further expansion of the ice and snow.[40] Eventually, other (negative) feedbacks would come into play to limit the expansion of the ice and snow cover, but the new distribution of surface cover and associated circulation patterns—an ice age—might persist for thousands of years.

Although the causes of ice ages (also called *glaciations*) are not well understood, there is good evidence concerning the actual conditions that prevailed in the Wisconsin glaciation, which was the most recent of several Pleistocene glaciations and ended only about 10,000 years ago. Compared to today's values, mean global temperatures were lowered 5° or 6° C (9° to 11° F), and mean temperatures were lowered by 12° C or more in the vicinity of the ice sheets themselves.[41] Sea level dropped at its lowest to 125 meters below today's level, the water being tied up in the continental ice sheets that covered what is now Canada, the north central United States, Scandinavia, and much of northern Europe (see Figure 2-25).[42] The drop in sea level exposed much of what is now continental shelf, which became a richly vegetated landscape. The snow line in most mountain areas dropped 1000 to 1400 meters, and 700 to 900 meters in

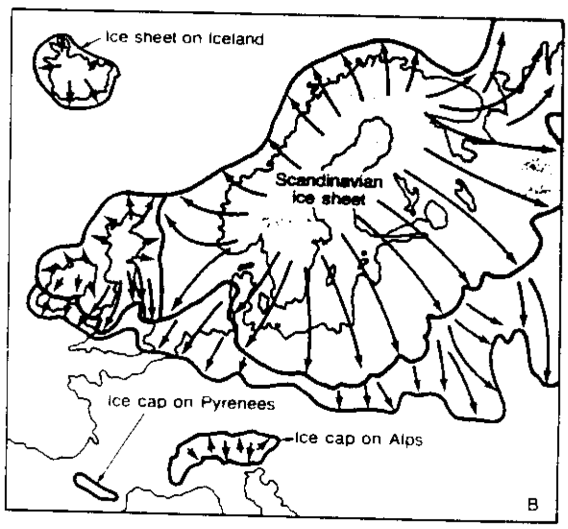


FIGURE 2-25

Maximum extent of Pleistocene glaciation in the Northern Hemisphere. A. North America. B. Europe.

[40] SMIC, pp. 125–130; Kellogg and Schneider, Climate stabilization, p. 1166.
[41] SMIC, pp. 33–34
[42] Strahler and Strahler, *Environmental geoscience*, p. 454.

the tropics, compressing the life zones that exist between the snow line and sea level (see Chapter 4).

It is worth mentioning again that the changes producing these conditions generally may have been—and in some regions certainly were—quite rapid. Although one tends to think of glaciation as a slow process involving the plastic flow of glacial ice (as described earlier), there is a much faster mechanism available for the advance of glaciers during the onset of ice ages: snow falls over a large region, fails to melt, and is compressed under the weight of new snow the following winter. Thus, the area under a semipermanent cover of ice and snow can increase enormously in a single season. This process of rapid glaciation has aptly been termed the *snow blitz.*[43]

The evidence suggests that the departure of the Wisconsin glaciers was even more sudden than the onset. At the peak of the warming period that followed this most recent retreat of the ice sheets, average global temperatures rose to 2° or 3° C warmer than today's, sea level rose to today's level but apparently not above it, and the prevailing circulation patterns produced considerably more rainfall in the Sahara and the eastern Mediterranean lands than occurs today.[44] These conditions, called the postglacial optimum by climatologists, occurred between 5000 and 6000 years ago.

The fact that Earth has a long history of climate change offers small consolation, unfortunately, to today's human population, faced as it is with the prospect of further change in the future. Significant climate change in *any* direction—hotter, colder, drier, wetter—in the world's major food-producing regions would be likely to disrupt food production for years, and even decades, because the animals and crops now relied upon are relatively well adapted to existing climate conditions. The recent historical record and the nearness of present conditions to a temperature maximum, moreover, suggest that the most *likely* major trend to occur next is cooling. This almost certainly would disrupt food production for as long as the lowered temperatures persisted, by reducing the area and growing season available for some of the most important food crops. The dependence of agriculture on climate is explored further in Chapter 7 and Chapter 11. Other ecological effects of climate change could also have serious human consequences, which are treated in Chapter 11.

How and when human activities could themselves cause, accelerate, or prevent climate change is a complicated and imperfectly understood subject, which is also postponed until Chapter 11. What is most relevant to that discussion from the foregoing treatment of the machinery of climate and natural climate change are the complexity and probable instability of the patterns of energy flow in the ocean-atmosphere system, and the speed with which changes, once triggered, may spread and intensify.

[43]Nigel Calder, In the grip of a new ice age?
[44]SMIC, p. 37.

Recommended for Further Reading

Cloud, Preston, ed. 1970. *Adventures in earth history.* W. H. Freeman and Company, San Francisco. Many classic papers in the earth sciences.

Kuenen, P. H. 1963. *Realms of water.* Wiley, New York. Remarkably entertaining and informative treatment of water in the oceans, on land, and in the atmosphere. Splended sections on glaciers and sea ice.

Menard, H. W. 1974. *Geology, resources, and society.* W. H. Freeman and Company, San Francisco. Very readable text, especially good on recent climatic history.

National Academy of Sciences (NAS). 1975. *Understanding climatic change.* NAS, Washington, D.C. Excellent coverage of how climatologists reconstruct climatic history from fragmentary evidence, as well as of many other topics in climate.

Neiburger, Morris; J. G. Edinger; and W. D. Bonner. 1973. *Understanding our atmospheric environment.* W. H. Freeman and Company, San Francisco. Good introductory text, with emphasis on weather.

Press, Frank, and Raymond Siever. 1974. *Earth.* W. H. Freeman and Company, San Francisco. Thorough, magnificently illustrated text on the earth sciences. Emphasis on interaction of experiments and observations with theory.

Strahler, A. N., and A. H. Strahler. 1973. *Environmental geoscience.* Hamilton, Santa Barbara, Calif. Copious illustrations and clear explanations of an extraordinary range of topics in environmental earth science.

Additional References

Bolin, Bert. 1970. The carbon cycle. *Scientific American*, September, pp. 124–132. Includes discussion of carbonates in the sedimentary cycle.

——. 1974. Modelling the climate and its variations. *Ambio*, vol. 3, no. 5, pp. 180–188. Historical record of climate variation, and use of measurements and models to investigate ongoing changes.

Bretherton, F. 1975. Recent developments in dynamical oceanography. *Quarterly Journal of the Royal Meteorological Society*, vol. 101, pp. 705–722. Technical account of the emerging picture of surprisingly vigorous deep-ocean circulation, consisting of large eddies.

Budyko, M. I. 1974. *Climate and life.* Academic Press, New York. Comprehensive technical monograph by one of the world's preeminent climatologists.

Calder, Nigel. 1975. In the grip of a new ice age? *International Wildlife*, July/August, pp. 33–35. Popular treatment of the snow-blitz theory.

Cloud, Preston, and Aharon Gibor. 1970. The oxygen cycle. *Scientific American*, September, pp. 110–123. Quite technical treatment of the evolution of Earth's atmosphere and the role of the sedimentary cycle in that process.

Emery, K. O. 1969. The continental shelves. *Scientific American*, September, pp. 106–122.

Farmer, S. A. 1973. A note on the long-term effects on the atmosphere of sea-surface temperature anomalies in the north Pacific Ocean. *Weather*, vol. 28, no. 3, pp. 102–105.

Frieden, Earl. 1972. The chemical elements of life. *Scientific American*, July, pp. 52–60. Connections between sedimentary cycle and biosphere.

Garrels, R. M.; F. T. Mackenzie; and Cynthia Hunt. 1975. *Chemical cycles and the global environment.* Kaufmann, Los Altos, Calif. Concise and data-rich treatment of sedimentary cycle and chemistry of ocean and atmosphere.

Gates, David. 1971. The flow of energy in the biosphere. *Scientific American*, September, pp. 88–100. Interaction of climate and ecosystems.

Isaacs, J. D. 1969. The nature of oceanic life. *Scientific American*, September, pp. 146–162. Fascinating revelations about the fauna of the deep layers of the ocean, among other topics.

Kellogg, W. W., and S. H. Schneider. 1974. Climate stabilization: For better or for worse? *Science*, vol. 186, pp. 1163–1172 (December 27). Noted here for the useful discussion of feedback mechanisms.

Kukla, G. J., and H. J. Kukla. 1974. Increased surface albedo in the Northern Hemisphere. *Science*, vol. 183, pp. 709–714 (February 22). Changes in snow and ice cover in the Northern Hemisphere were great enough in the early 1970s to cause significant changes in the hemispheric heat balance. Satellite measurements are reported.

Lamb, H. H. 1966. *The changing climate.* Methuen, London. Methods and conclusions of historical climatology.

——. 1972. *Climate: Past, present, and future.* Methuen, London. Excellent introduction to climatology.

Leopold, Luna B. 1974. *Water.* W. H. Freeman and Company, San Francisco. Readable introduction to the hydrologic cycle.

Lorenz, E. N. 1970. Climate change as a mathematical problem. *Journal of Applied Meteorology*, vol. 9, pp. 325–329. The author argues that climate change could be caused by "internal" fluctuations that are characteristic of complicated physical systems.

McDonald, James E. 1952. The coriolis effect. *Scientific American*, May. Reprinted in J. R. Moore, ed., *Oceanography*, pp. 60–63. Readable elaboration and examples.

MacIntyre, Ferren. 1970. Why the sea is salt. *Scientific American*, November, pp. 104–115. Excellent introduction to geochemistry.

Mason, Brian. 1966. *Principles of geochemistry.* 3d ed. Wiley, New York. Good introductory text.

Miller, Albert. 1966. *Meteorology.* Merrill, Columbus, Ohio. Concise and readable paperback.

Moore, J. R., ed. 1971. *Oceanography: Readings from* Scientific American. W. H. Freeman and Company, San Francisco. Collection of articles.

Munk, Walter. 1955. The circulation of the oceans. *Scientific American*, September. Reprinted in J. R. Moore, ed., *Oceanography*, pp. 64–69.

Namias, Jerome. 1969. Seasonal interactions between the north Pacific Ocean and the atmosphere during the 1960s. *Monthly Weather Review*, vol. 97, no. 3, pp. 173–192. Early analysis of a possible controlling role for the ocean in climate fluctuations.

———. 1972. Experiments in objectively predicting some atmospheric and oceanic variables for the winter of 1971–2. *Journal of Applied Meteorology,* vol. 11, no. 8, pp. 1164–74.

Oort, Abraham H. 1970. The energy cycle of the earth. *Scientific American,* September, pp. 54–63. Sizes and functions of major natural energy flows.

Rona, P. A. 1973. Plate tectonics and mineral resources. *Scientific American,* July.

SCEP. *See* Study of Critical Environmental Problems.

Schneider, S. H.; and Roger D. Dennett. 1975. Climatic barriers to long-term energy growth. *Ambio,* vol. 4, no. 2, pp. 65–74. Contains good synopsis of knowledge about the operation of global climatic machinery.

Schneider, S. H., and C. Mass. 1975. Volcanic dust, sunspots, and temperature trends. *Science,* vol. 190, pp. 741–746 (November 21). Interesting speculations and analysis centered around the influence on climate of possible variations in the sun's output.

Sellers, William D. 1965. *Physical climatology.* University of Chicago Press. Excellent text.

Siever, Raymond. 1974. The steady state of the earth's crust, atmosphere, and oceans. *Scientific American,* June, pp. 72–79. An illuminating review of the main ideas in geochemistry.

Skinner, Brian J. 1969. *Earth resources.* Prentice-Hall, Englewood Cliffs, N.J. Capsule introduction to the hydrologic cycle, among other topics.

SMIC. *See* Study of Man's Impact on Climate.

Stewart, R. W. 1969. The atmosphere and the ocean. *Scientific American,* September, pp. 76–86. Useful summary of some ocean-atmosphere interactions, although dated somewhat by recent developments (see Jerome Namias, Seasonal interactions between the north Pacific Ocean and the atmosphere during the 1960s; Namias, Experiments in objectively predicting; and F. Bretherton, Recent developments in dynamical oceanography).

Study of Critical Environmental Problems (SCEP). 1970. *Man's impact on the global environment.* M.I.T. Press, Cambridge, Mass. Contains short summaries of main global climatic processes.

Study of Man's Impact on Climate (SMIC). 1971. *Inadvertent climate modification.* M.I.T. Press, Cambridge, Mass. A wealth of information and references on the study of climate and climate change. Moderately technical.

Wenk, Edward, Jr. 1969. The physical resources of the ocean. *Scientific American,* pp. 166–176.

Wilson, J. Tuzo, ed. 1972. *Continents adrift: Readings from* Scientific American. W. H. Freeman and Company, San Francisco. Articles on plate tectonics and continental drift.

Woodwell, George M. 1970. The energy cycle of the biosphere. *Scientific American,* September, pp. 64–74.

Wyllie, Peter J. 1975. The earth's mantle. *Scientific American,* March, pp. 50–63. Plate tectonics, earthquakes, and the composition of the planet.

35