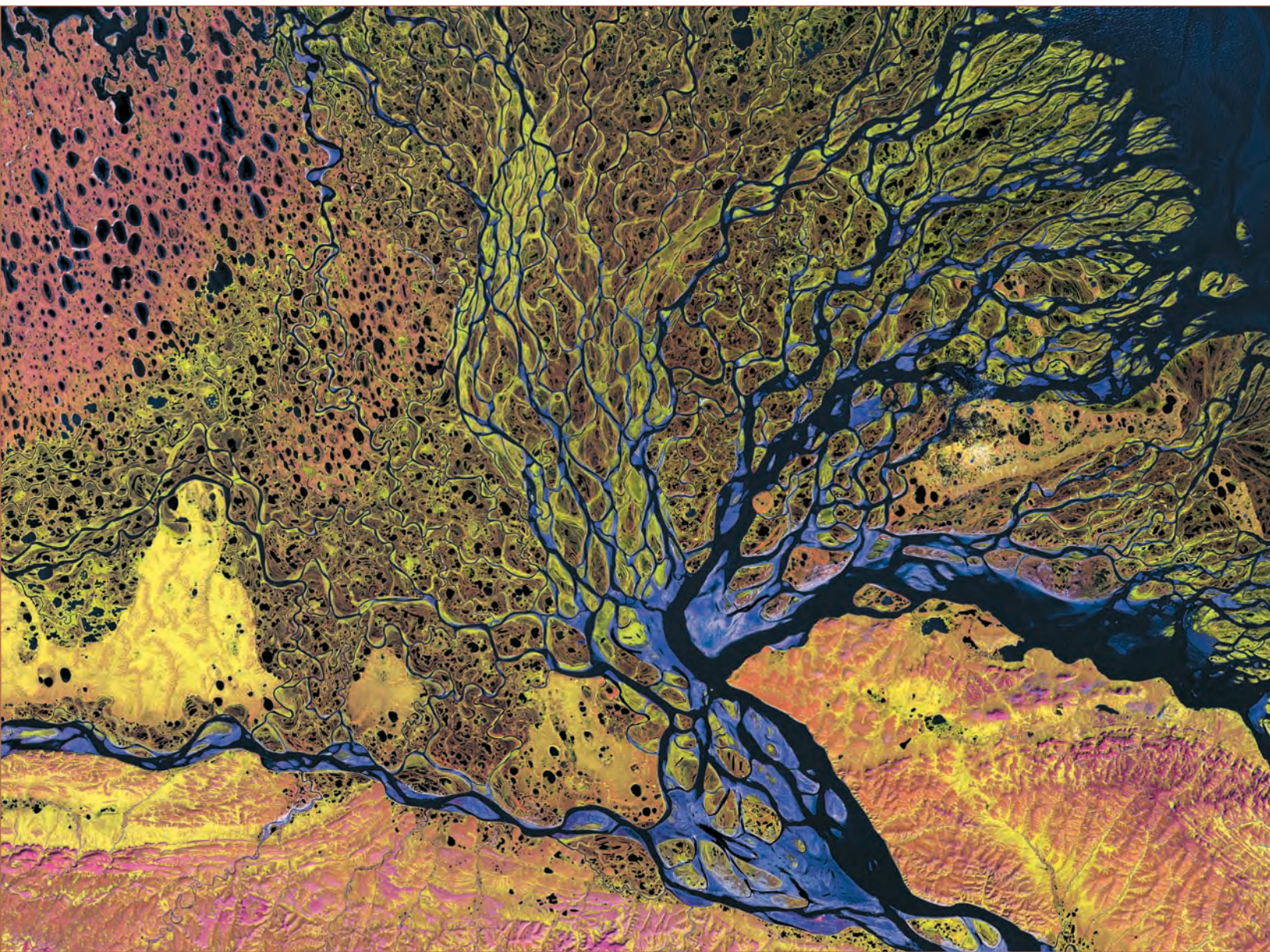


UNCORRECTED PAGE PROOFS • © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

UNIT TWO



THE GEOGRAPHIC AND ECOLOGICAL FOUNDATIONS OF BIOGEOGRAPHY

UNCORRECTED PAGE PROOFS • © 2010 Sinauer Associates, Inc. This material cannot be copied, reproduced, manufactured, or disseminated in any form without express written permission from the publisher.

Previous Page: The Lena River Delta in Russia. Courtesy of USGS National Center for EROS and NASA Landsat Project Science Office.



3

THE GEOGRAPHIC TEMPLATE: Visualization and Analysis of Biogeographic Patterns

Definition and Components of the Geographic Template 47

The geographic template 47
Climate 49
Soils 58
Aquatic environments 63
Time 68

Two-Dimensional Renderings of the Geographic Template 69

Early maps and cartography 69
Flattening the globe: Projections and geographic coordinate systems 70

Visualization of Biogeographic Patterns 71

History and exemplars of visualization in biogeography 71
The GIS revolution 76
Cartograms and strategic "distortions" 77

Obtaining Geo-Referenced Data 77

Humboldt's legacy: a global system of observatories 77
Remote sensing and satellite imagery 79
Interpolation over space and time 79

Analysis of Biogeographic Patterns 80

Definition and Components of the Geographic Template

Organisms can be found almost everywhere on Earth: from the cold, rocky peaks of high mountains to the hot, windswept sand dunes of lowland deserts; from the dark, near-freezing depths of the ocean floor to the steaming waters of hot springs. Some organisms even live around hydrothermal vents in the deep ocean, where temperatures exceed 100° C (but the water does not boil because of the extreme pressure). Yet no single kind of organism lives in all of these places. Each species has a restricted geographic range in which it encounters a limited range of environmental conditions. Polar bears and caribou are confined to the Arctic, whereas palms and corals are rare outside the tropics. There are a few species, such as *Homo sapiens* and the peregrine falcon, that we call cosmopolitan because they are distributed over all continents and over a wide range of latitudes, elevations, climates, and habitats. These species, however, are not only exceptional but also much more limited in distribution than they appear at first glance. Humans and peregrines, for example, are absent from the three-fourths of the Earth that is covered with water; indeed, they are little more than rare visitors to large expanses of the terrestrial realm with extremely harsh environments.

The geographic template

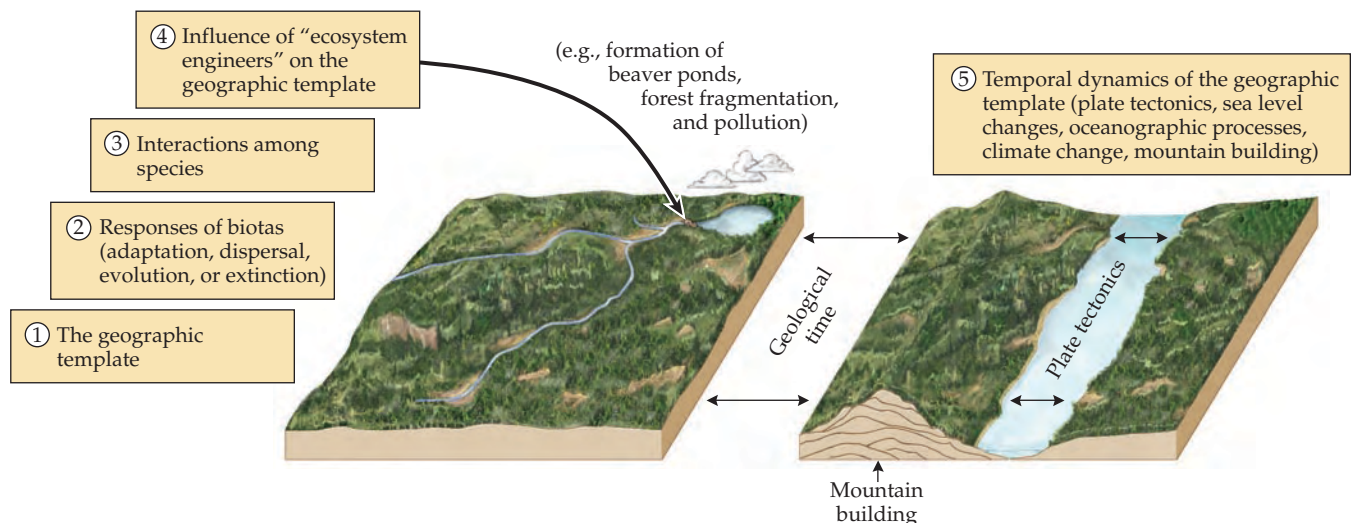
As we will point out in later chapters, we may need to invoke unique historical events or ecological interactions with other organisms to account for the limited geographic ranges of some species, but the most obvious patterns in the distributions of organisms occur in response to variations in the physical environment. In terrestrial habitats, these patterns are largely determined by climate (primarily temperature and precipitation) and soil type. The distributions of aquatic organisms are limited largely by water temperature, salinity, light, and pressure.

As Edward Forbes and other early biogeographers observed (see Chapter 2), climate, soil type, water chemistry, and a long list of other environmental conditions vary in a highly nonrandom manner across geographic gradients of latitude, elevation, depth, and proximity to major landforms such as coastlines and mountain ranges. In addition, regardless of whether we consider the aquatic or terrestrial realms, the environmental variables tend to exhibit strong spatial autocorrelation, or what is sometimes referred to as **distance-decay**, which simply means that similarity of environmental conditions between sites decreases as we compare more distant sites.

Taken together, these nonrandom patterns of spatial variation in environmental conditions constitute a multifactor **geographic template**, which forms the foundations of all biogeographic patterns. That is, most biogeographic patterns ultimately derive from this very regular, spatial variation in environmental conditions. For example, as we move from low to high latitudes, from the Equator to the poles, or from the ocean through estuaries and then upstream, environmental temperatures tend to cool, and this either directly or indirectly influences biotic communities along each of these geographic and environmental gradients (see Chapter 15). Diversity, species composition, and vital processes (e.g., productivity and decomposition) of biotic communities along these gradients change in a highly predictable manner, with similarity among communities being much higher for those located in close proximity along the geographic template.

The conceptual model illustrated in **Figure 3.1** may prove useful in understanding a great diversity of biogeographic patterns. Again, the fundamental layer or foundation for all patterns of variation across space is the geographic template. Organisms can then respond to this highly nonrandom spatial variation through adaptations (both behavioral and physiological), dispersal, or speciation, or if unsuccessful at any of these responses, populations may suffer local extinction. All of these responses, taken together, ultimately determine the geographic distributions and patterns of variation among populations, species, and communities across geographic gradients. To more fully account for the diversity of patterns biogeographers study, this conceptual model must include at least two additional layers of complexity—both of which may be viewed as feedback or interactions among system components. Not only are species distributions influenced by environmental conditions, but those distributions and related patterns are also influenced by interactions among the species themselves (e.g., mutualism, parasitism, and

FIGURE 3.1 All biogeographic patterns are ultimately influenced by the geographic template. Distributions and patterns of geographic variation of life-forms and communities result from (1) the highly nonrandom patterns of spatial variation in environmental characteristics across the Earth (i.e., the geographic template); (2) responses of biotas (including adaptation, dispersal, evolution, or extinction) to this variation; (3) interactions among species (e.g., competitive exclusion and mutualistic interactions); (4) impacts of particular species—“ecosystem engineers” such as beavers, prairie dogs, and humans—on the geographic template; and (5) the temporal dynamics of the geographic template (including the so-called TECO events—plate Tectonics [including orogeny], Eustatic changes in sea level, Climate change, and Oceanographic processes).



competitive exclusion; see Chapter 4). Furthermore, some species—often referred to as ecosystem engineers (e.g., beavers, prairie dogs, and humans)—can modify the geographic template itself. Even before our own species rose to become the world's dominant **ecosystem engineer**, microscopic and otherwise “primitive” life-forms fundamentally altered Earth's atmosphere, thereby increasing its oxygen content and its abilities to store heat, modifying its climate, and eventually molding and recasting the geographic template across both the terrestrial and aquatic realms. We still need to introduce one final but nonetheless fascinating layer of complexity to this conceptual model. The geographic template itself is dynamic—not just across space but across time as well. Over the 3.5 billion year history of life on Earth, and indeed even over much shorter timescales, climates have swung dramatically from glacial to interglacial episodes of the Pleistocene (including most of the past 2 million years) and from the much earlier periods of the so-called snowball Earth to the global sauna of the Mid Eocene (roughly 50 million years ago; see Chapter 8). Just as fundamental, and sometimes driving these major climatic shifts, Earth's crust itself is highly dynamic, emerging from the mantle below to drift, split, and collide in a kaleidoscopic jigsaw puzzle known as plate tectonics and continental drift (see Chapter 8).

All of these dynamics in space and time are themselves driven by two great engines, which are powered by two different sources of energy. The energy stored in the Earth's core at the time the solar system was formed is also supplemented by compressional heating due to gravity. A portion of this energy is gradually but continuously being dissipated through the Earth's mantle and crust and ultimately out into space. This transfer of heat energy moves and shapes the Earth's crust, shifting the positions of the crustal plates containing the continents, thrusting up mountains, creating or consuming ocean basins, and causing earthquakes and volcanic eruptions.

The other great engine is driven by the energy of the sun. Radiant energy emitted by the sun strikes the Earth's surface, where it is absorbed and converted into heat, warming the surface of the land and water and the atmosphere above them. The resulting differences in the temperature and density of air and water cause them to move over the Earth's surface, both horizontally and vertically, creating the Earth's major wind patterns and ocean currents. The heating of surface water also causes evaporation, and the resulting water vapor is carried by the air and redeposited as rain or snow. These processes, which are responsible for the Earth's climate and for many physical characteristics of its oceans and fresh waters, are the subject of the following sections of this chapter.

Climate

Solar energy and temperature regimes

SOLAR RADIATION AND LATITUDE. Sunlight sustains life on Earth. Not only does solar energy warm the Earth's surface and makes it habitable, but it also is captured by green plants and converted into chemical forms of energy that power the growth, maintenance, and reproduction of most living things.

According to the principles of thermodynamics, heat is transferred from objects of higher temperature to those of lower temperature by one of three mechanisms: (1) conduction, a direct molecular transfer (especially through solid matter); (2) convection, the mass movement of liquid or gaseous matter; or (3) radiation, the passage of waves through space or matter. Heat flows as radiant energy from the hot sun across the intervening space to the cooler Earth. When incoming solar radiation strikes matter such as water or soil, some of it is absorbed, and the matter is heated. Some solar radiation is ini-

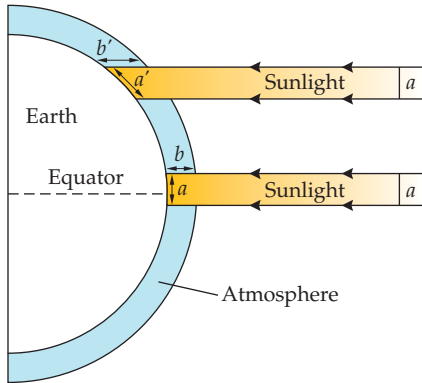


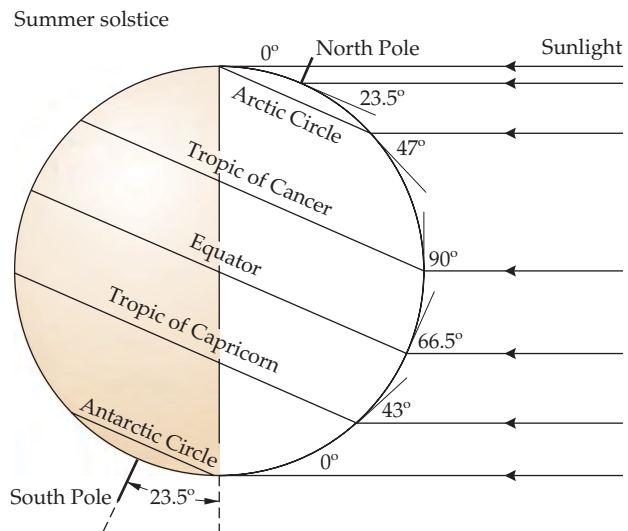
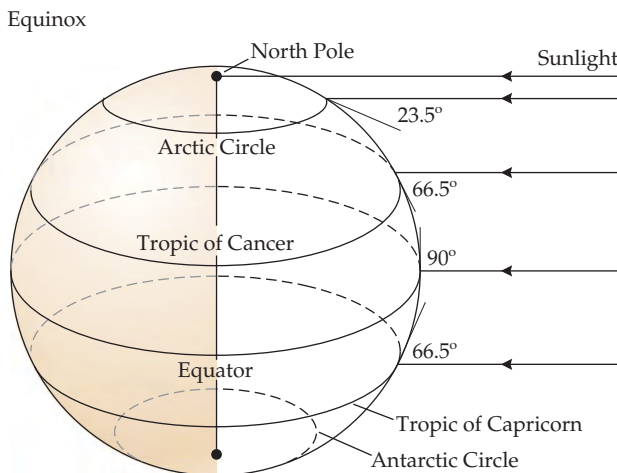
FIGURE 3.2 Average input of solar radiation to the Earth's surface as a function of latitude. Heating is most intense when the sun is directly overhead, when incoming solar radiation strikes perpendicular to the Earth's surface. The higher latitudes are cooler than the tropics because the same quantity of solar radiation is dispersed over a greater surface area (a' as opposed to a) and passes through a thicker layer of filtering atmosphere (b' as opposed to b).

tially absorbed by the air, particularly if it contains suspended particles of water or dust (e.g., clouds), but most passes through the sparse matter of the atmosphere and is absorbed by the denser matter of the Earth's surface. This surface is not heated uniformly. Soil, rocks, and plants absorb much of the radiation and may be heated intensely. Although air is heated to some extent by absorption of incoming solar radiation, most of the heating of air occurs at the Earth's surface, where it is warmed by direct contact with warm land and water, by latent heat released by the condensation of water, and by long-wave infrared radiation emitted from the surfaces of warm objects such as leaves and bare soil. In contrast, much of the solar radiation striking the surface of the oceans is reflected back toward the atmosphere, while the rest penetrates and warms layers of the water column below.

The angle of incoming radiant energy relative to the Earth's surface affects the quantity of heat absorbed. The most intense heating occurs when the surface is perpendicular to incident solar radiation, for two reasons: (1) the greatest quantity of energy is delivered to the smallest surface area; and (2) a minimal amount of radiation is absorbed or reflected back into space during passage through the atmosphere, because the distance it travels through air is minimized (**Figure 3.2**). This differential heating of surfaces at different angles to the sun explains why it is usually hotter at midday than at dawn or dusk, why average temperatures in the tropics are higher than at the poles, and why south-facing hillsides are warmer than north-facing ones in the Northern Hemisphere (and the reverse in the Southern Hemisphere).

Because the Earth is tilted 23.5° from vertical on its axis with respect to the sun, solar radiation falls directly, that is perpendicularly, on different parts of the Earth during an annual cycle. This differential heating produces the seasons. The seasons are also characterized by different lengths of day and night. Only at the Equator are there exactly 12 hours of daylight and darkness every 24 hours throughout the year (**Figure 3.3**). At the spring and fall equinoxes

FIGURE 3.3 Seasonal variation in day length with latitude is due to the inclination of the Earth on its axis. At either equinox (A), the sun is directly overhead at the Equator, and all parts of the Earth experience 12 hours of light and 12 hours of darkness each day. At the summer solstice (B) in the Northern Hemisphere, however, the 23.5° angle of inclination causes the sun to be directly over the Tropic of Cancer, while the Arctic Circle and areas farther north experience 24 hours of continuous daylight; at the same time, all regions in the Southern Hemisphere experience less than 12 hours of sunlight per day, and the sun never rises in areas south of the Antarctic Circle.



(March 21 and September 22, respectively) the sun's rays fall perpendicularly on the Equator, equatorial latitudes are heated most intensely, and every place on Earth experiences the same day length. At the summer solstice (June 22), sunlight falls directly on the Tropic of Cancer (23.5° N latitude). The Northern Hemisphere is heated most intensely and begins to experience longer days than nights and enjoys summer, while the Southern Hemisphere has winter. At the winter solstice (December 22), the sun shines directly on the Tropic of Capricorn (23.5° S latitude), and the Southern Hemisphere enters its summer, while the Northern Hemisphere experiences winter, cold temperatures, and long nights. The result of all this is that the seasonality of climate increases with increasing latitude, with the Tropics of Cancer and Capricorn marking the northern- and southernmost latitudes, respectively, that receive direct sunlight at least one day each year (i.e., on the north or south summer solstice). At the Arctic and Antarctic Circles (66.5° latitude), there is one day each year of continuous daylight when the sun never sets and one day of continuous darkness—each at a solstice. Although every location on the Earth theoretically experiences the same amount of daylight and darkness over an annual cycle, the sun is never directly overhead at high latitudes; however, considerable solar radiation is absorbed during the long summer days. Temperatures in excess of 30° C are commonly recorded in Alaska. The warmest days typically are in July (after the summer solstice) because of the time lag required to heat the Earth's surface.

THE COOLING EFFECT OF ELEVATION. The processes just described account for seasonal and latitudinal variation in temperature, but it remains to be explained why it gets colder as we ascend to higher altitudes. The fact that Mount Chimborazo and Mount Kilimanjaro (nearly on the Equator in tropical South America and East Africa, respectively) are capped with permanent ice and snow seems to be in conflict with our intuitive expectation and with our previous discussion on the intensity of solar radiation in the tropics. Mountain peaks are nearer the sun, so why are they cooler than nearby lowlands? The answer lies in the physical and thermal properties of air. As a climber moves up a mountain, the length (and the pressure) of the column of air that lies above the climber decreases. Thus, the density and pressure of air decrease with increasing elevation. When air is blown across the Earth's surface and forced upward over mountains, it expands in response to the reduced pressure. Expanding gases undergo what is called **adiabatic cooling**, a process where they lose heat energy as their molecules move farther apart (temperature essentially is a measure of activity and the frequency of collisions of molecules). The same process occurs in a refrigerator as freon gas expands after leaving the compressor. The rate of adiabatic cooling of dry air is about 10° C per km elevation, as long as no condensation of water vapor and cloud formation occurs.

Higher elevations are also colder because the less dense air allows a higher rate of heat loss by radiation back through the atmosphere. Water vapor and carbon dioxide in the atmosphere would typically absorb radiant heat—the so-called **greenhouse effect**. As the name implies, these gases act like the glass in a greenhouse: They allow the short wavelengths of incoming solar radiation to pass through, but they trap the longer wavelength radiation (infrared, or heat) emitted by surfaces that have been warmed by the sun. The resulting warming effect of greenhouse gases is most pronounced in moist lowland areas, where air is laden with water vapor and carbon dioxide. In contrast, mountains and deserts typically experience extreme daily temperature fluctuations, because the local atmosphere is thinner or drier, respectively.

Winds and rainfall

WIND PATTERNS. Differential heating of the Earth's surface also causes the winds that circulate heat and moisture. As we have already seen, the most intense heating is at the Equator, especially during the equinoxes, when the sun is directly overhead. As this tropical air is heated, it expands, becomes less dense than the surrounding air, and rises. This rising air produces an area of reduced atmospheric pressure over the Equator. Denser air from north and south of the Equator flows into the area of reduced pressure, resulting in surface winds that blow toward the Equator (**Figure 3.4**). Meanwhile, the rising equatorial air cools adiabatically, becomes denser, is pushed away from the Equator by newly warmed rising air, and eventually descends again at about 30°N and S latitude (the Horse Latitudes). This vertical circulation of the atmosphere results in three convective cells (Hadley, Ferrel, and Polar) in each hemisphere, with warm air ascending at the Equator and at about 60°N and S latitude, and cool air descending at about 30°N and S and at the poles. These circulating air masses produce surface winds that typically blow toward the Equator between 0° and 30° and toward the poles between 30° and 60°. In the upper atmosphere between the convective cells are the jet streams—high-speed winds blowing approximately parallel to the Equator.

The surface winds do not blow exactly in a north–south direction; instead, they appear to be deflected toward the east or west by the **Coriolis effect**. Although the Coriolis effect is often called the *Coriolis force*, it is not a force but a straightforward consequence of the law of conservation of angular momentum. Every point on the Earth's surface makes one revolution every 24 hours. Because the circumference of the Earth is about 40,000 km, a point at the Equator moves from west to east at a rate of about 1700 km h⁻¹. But the parallel lines (really circles) of latitude become increasingly shorter as we move from the Equator to the poles. Therefore, points north or south of the Equator travel a shorter distance each 24-hour rotation of the Earth;

that is, they move at a slower rate than points closer to the Equator. Consider what happens at the Equator if you shoot a rocket straight upward. Where does it come down? Right where it was launched; the rocket travels not only up and down but also eastward at a rate of 1700 km h⁻¹, the same rate as the Earth moving beneath it. Now suppose the rocket is propelled northward away from the Equator. It continues to travel eastward at 1700 km h⁻¹, but the Earth underneath it moves ever more slowly as the rocket travels farther north, and consequently its path appears to be deflected toward the right. The Coriolis effect describes this tendency of moving objects to veer to the right in the Northern Hemisphere and to the left in the

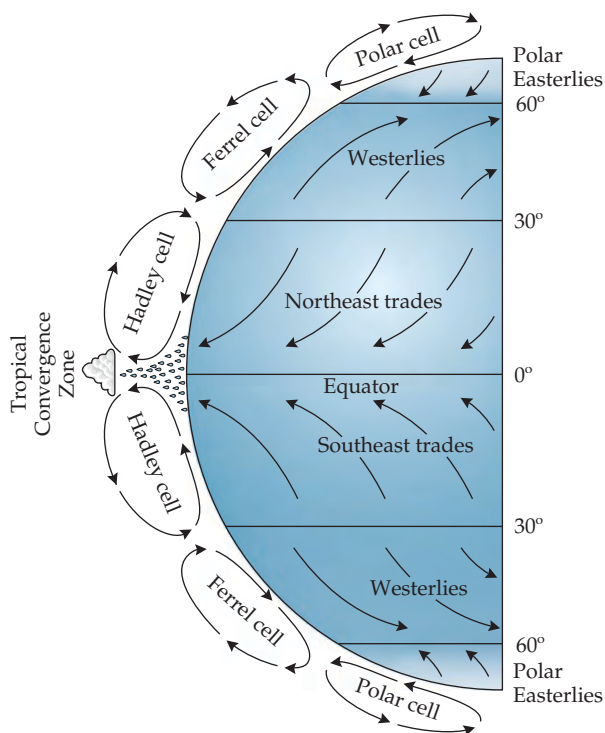


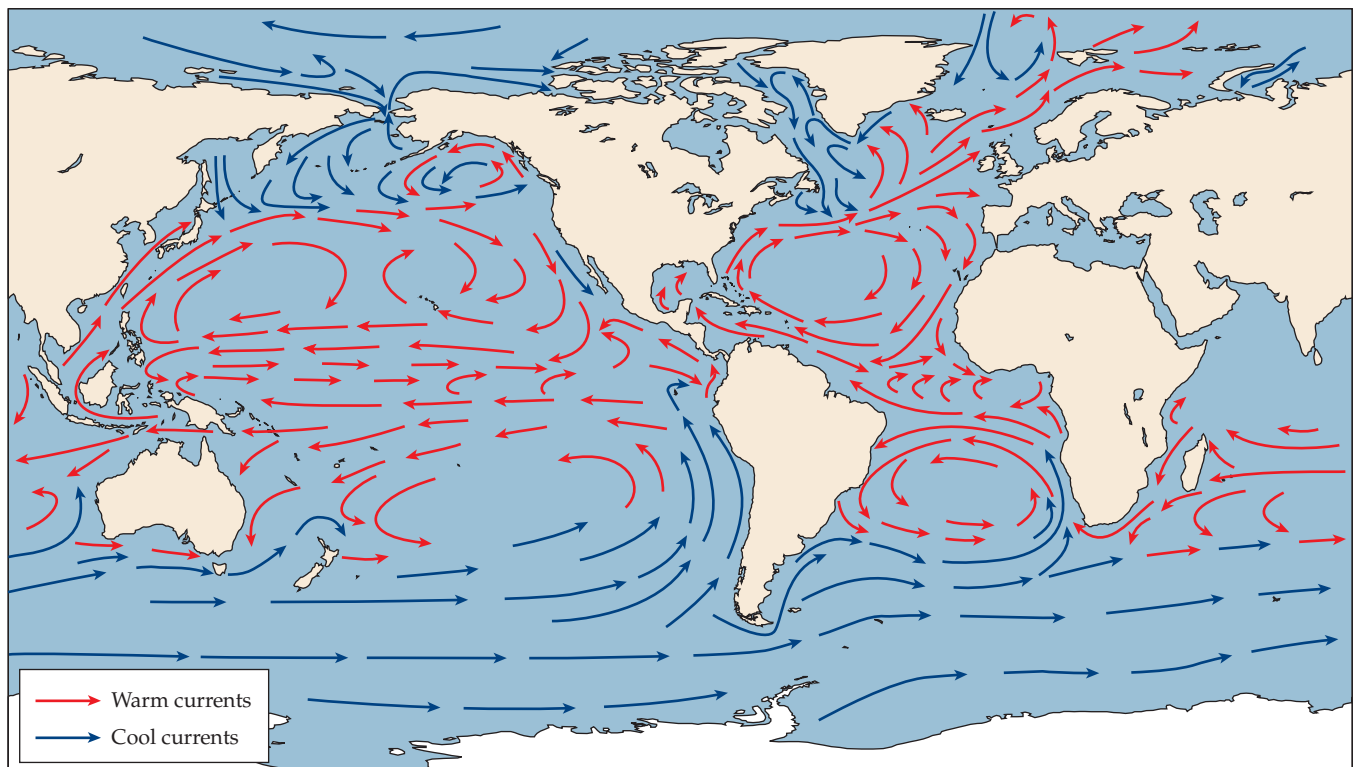
FIGURE 3.4 Relationship between vertical circulation of the atmosphere and wind patterns on the Earth's surface. There are three convective cells (Hadley, Ferrel, and Polar) of ascending and descending air in each hemisphere. As the winds move across the Earth's surface in response to this vertical circulation, they are deflected by the Coriolis effect, which produces easterly trade winds in the tropics, and Westerlies at temperate latitudes. The latitudinal locations of these cells shift with the seasons as the latitude of the most direct sunlight and most intense heating (i.e., the Tropical Convergence Zone) shifts between 23.5°N and 23.5°S (the Tropics of Cancer and Capricorn, respectively).

Southern Hemisphere. The winds approaching the Equator from the Horse Latitudes appear to be deflected to the west and are therefore called northeast or southeast **trade winds**. (Winds are described based on the direction of their *sources*.) Winds blowing toward the poles between about 30° and 60°N and S latitude are called **Westerlies** and are deflected to the east (see Figure 3.4). These winds naturally were very important to commerce in the days of sailing ships, when both the Westerlies and the trade winds (or *trades*) got their names. Ships coming to the New World from Europe traveled south to the Canary Islands and Azores in tropical latitudes to intercept the trades before heading westward, but they returned to Europe at higher latitudes with the Westerlies behind them.

The surface winds, influenced by the Coriolis effect, initiate the major ocean currents. The trade winds push surface water westward at the Equator, whereas the Westerlies produce eastward-moving currents at higher latitudes. Responding to the Coriolis effect, these water masses are deflected toward the east or west, and the net result is that the ocean currents move in great circular gyres—clockwise in the Northern Hemisphere and counterclockwise in the Southern Hemisphere (**Figure 3.5**). Warm currents flow from the tropics along eastern continental margins; as these water masses reach high latitudes, they are cooled, producing cold currents that flow down the western margins.

PRECIPITATION PATTERNS. By superimposing these patterns of temperature, winds, and ocean currents, we can begin to understand the global distribution of rainfall. We will also need some additional background in physics. As air warms, it can absorb increasing amounts of water vapor evaporated from the land and water. As air cools, it eventually reaches the **dew point**, at which it is saturated with water vapor. Further cooling then results in condensation and the formation of clouds. When the particles of water or ice in clouds become too heavy to remain airborne, rain or snow falls. In the tropics, the cool-

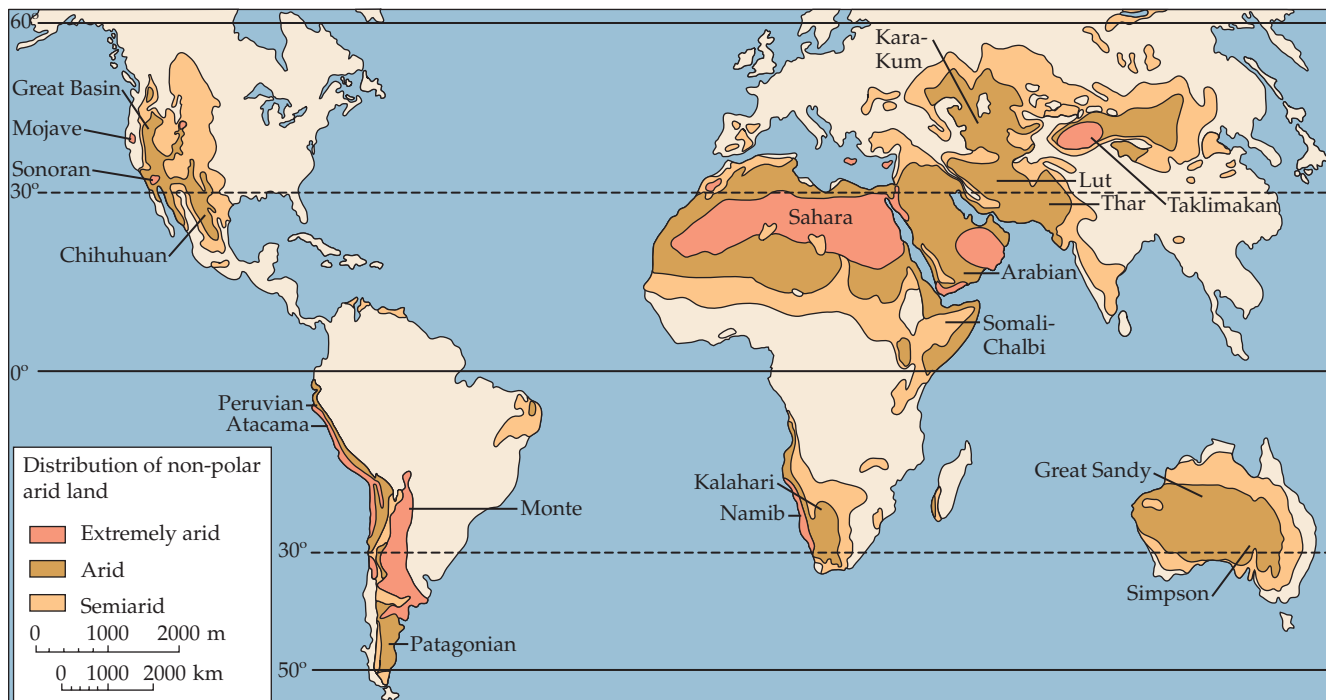
FIGURE 3.5 Main patterns of circulation of the surface currents of the oceans. In each ocean, water moves in great circular gyres, which move clockwise in the Northern Hemisphere and counterclockwise in the Southern Hemisphere. These patterns result in warm currents along the eastern coasts of continents and cold currents along the western coasts. Note the Pacific equatorial countercurrent: the small current along the Equator that flows from west to east, opposite to the gyres, and which strengthens in some years to cause the El Niño phenomenon.



ing of ascending warm air laden with water vapor produces heavy rainfall at low and middle elevations, where rain forests and cloud forests occur. Rainy seasons in the tropics tend to occur when the sun is directly overhead and the most intense heating occurs. The tropical grasslands of Kenya and Tanzania in East Africa, which lie virtually on the Equator but at higher elevations than rain forests, experience two rainy seasons each year, corresponding approximately to the equinoxes (when the **Tropical Convergence Zone** shifts overhead; see Figure 3.4), and two dry seasons, which correspond to the solstices. In contrast, the area around the Tropic of Cancer in central Mexico has only one principal rainy season—in the summer. Most tropical regions have at least one dry season.

At the Horse Latitudes, where cool air descends from the upper atmosphere, two belts of relatively dry climate encircle the globe. Descending air warms and can therefore absorb more moisture, drying the land. In these belts lie most of the Earth's great deserts (including the Mojave, Sonoran, and Chihuahuan in southwestern North America; the Sahara in North Africa; and the Arid Zone in central Australia; **Figure 3.6**), and adjacent to these deserts are regions of semiarid climates and grassy or shrubby vegetation. Here the seasonality of climate is very marked on the western sides of continents, which experience **Mediterranean climates**. Parts of coastal California, Chile, the Mediterranean region in Europe, southwestern Australia, and southernmost Africa have dry, usually hot summers and mild, rainy winters. In winter, when the land tends to be cooler than the ocean water, the westerly winds bring ashore warm, moisture-laden air; cooling and condensation occur, and fog and rain result. In summer, when the land is warmer than the ocean, the *Westerlies* blowing inland from the cold offshore currents are warmed; this increases their capacity to hold more water vapor and creates the relatively dry climates on land. The effects of cold currents are even more pronounced in localized regions of western South America and southwestern Africa, where they contribute to the formation of coastal deserts that are the driest areas on Earth (Amiran and Wilson 1973).

FIGURE 3.6 Major deserts of the world are not randomly distributed but tend to occur near 30° N or S latitude or along the leeward slopes of mountains, where descending air masses undergo adiabatic warming, increasing their capacity to hold water and drying local environments. (Meigs, 1953.)



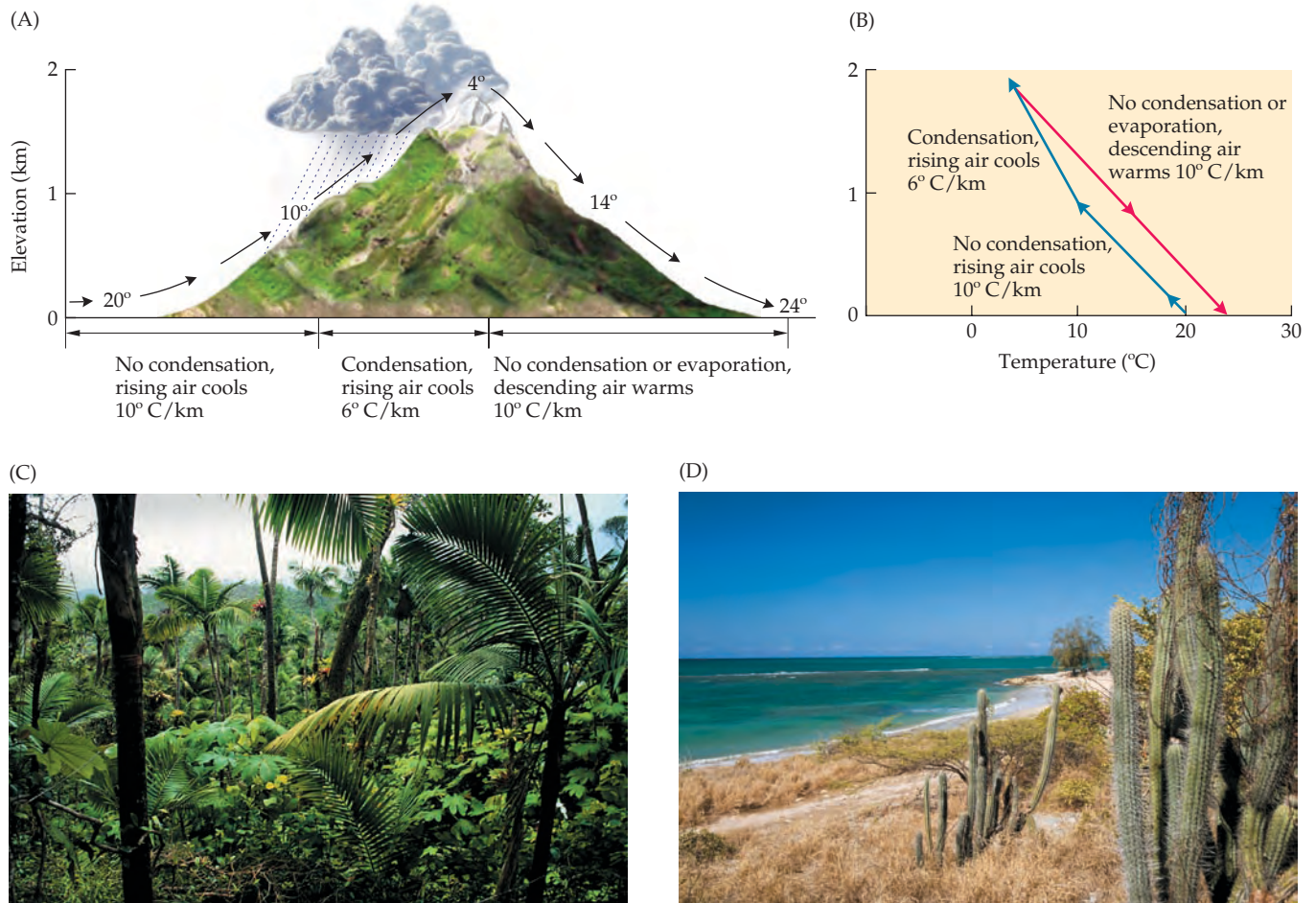


FIGURE 3.7 Factors causing rain shadow deserts. (A) Air blowing over a mountain cools as it rises, water vapor condenses, and the air loses much of its moisture as rain on the windward side, so the leeward side experiences warm, dry winds. (B) The rate of change in air temperature with elevation is greater for drier air, resulting in warmer, drier conditions on the leeward side than at the same elevation on the windward side. Comparison of vegetation on opposite sides of the central mountain range on the tropical island of Puerto Rico. (C) On the northeastern side, which receives the moisture-laden trade winds, lush rain forests occur. (D) In marked contrast, the southwestern side lies in a rain shadow, has a hot and dry climate, and has cacti and other plants typical of desert regions. (A,B after Flohn 1969; C © Lawrence Sawyer/istock; D © Michele Falzone/AGE Fotostock.)

Several of the deserts between 30° and 40° N and S latitude are located not only on the western sides of continents, but also on the eastern sides of major mountain ranges. As westerly winds blow over the mountains, they are cooled until eventually the dew point is reached and clouds begin to form. Condensation releases the latent heat of evaporation, so wet air cools adiabatically at a slower rate than dry air—6° C per km of elevation, as opposed to 10° C per km for dry air. As the air continues to rise and cool, most of its moisture falls as precipitation on the western side of the mountain range. When the air passes over the crest and begins to descend, the remaining clouds quickly evaporate, and the dry air warms at the faster rate. This **rain shadow** effect causes the warm, dry climates found on the leeward sides of temperate mountains (Figure 3.7). Thus, for example, the Sierra Nevada in California has lush, wet forests of giant sequoias and other conifers on its western slopes but arid woodlands of piñons and junipers on its eastern (leeward) slopes; a bit farther east, with an elevation below sea level, lies Death Valley—the driest place on the North American continent. Similarly, the Monte Desert of South America is in the rain shadow on the eastern side of the Andes in Argentina.

These global patterns of temperature and precipitation frequently are summarized in climatic maps like the one in Figure 3.8. Such maps are useful, but they can be misleading because they fail to show the local-scale patterns of spatial and temporal variation that influence the abundance and distribution of organisms.

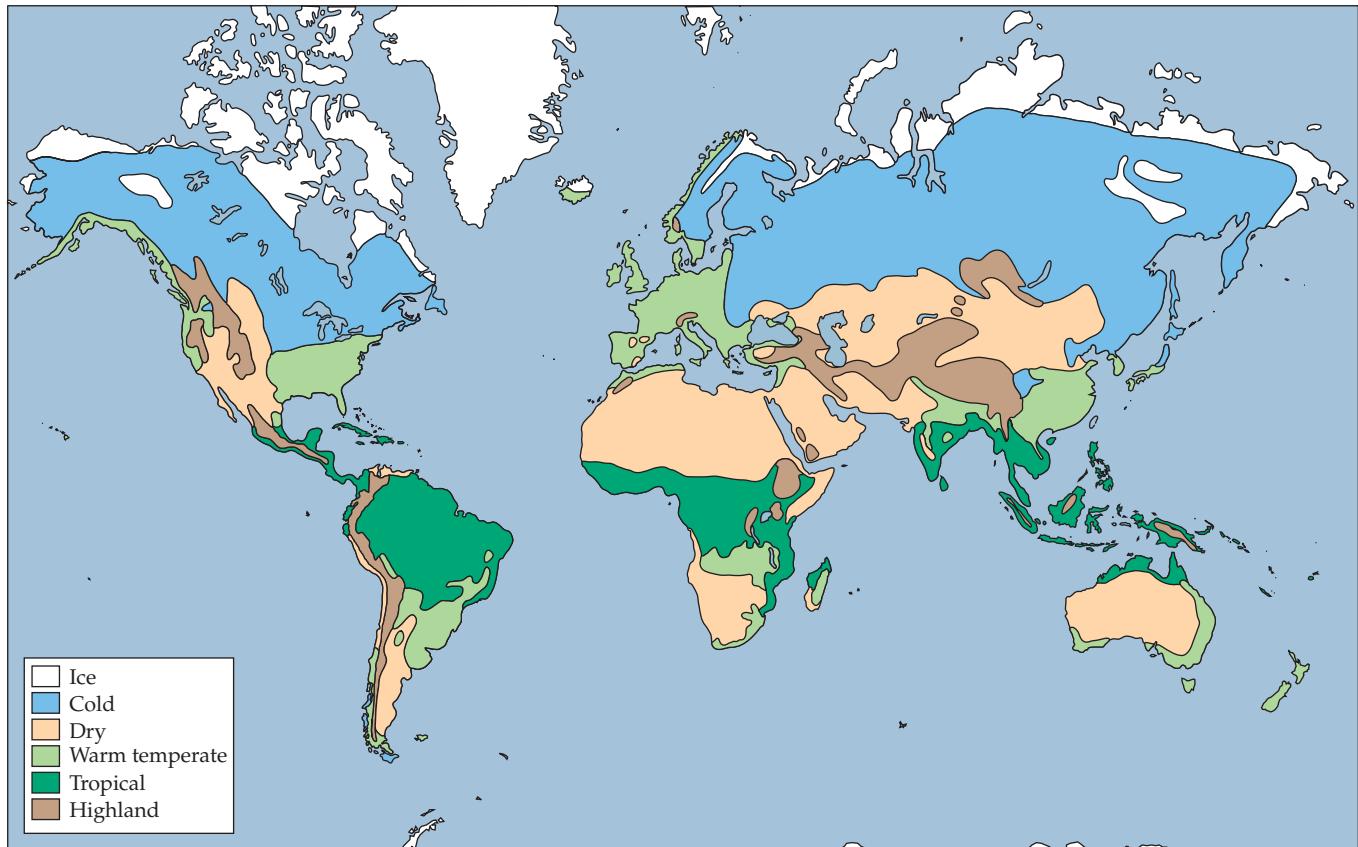


FIGURE 3.8 Major climatic regions of the world. Note that these regions occur in distinct patterns with respect to latitude and the positions of continents, oceans, and mountain ranges. (After Strahler and Strahler 1973.)

SMALL-SCALE SPATIAL AND TEMPORAL VARIATION. The processes that we have just described on a global scale can also produce great climatic variation on a local scale. The effect of mountains is particularly great, as we can illustrate with several examples. From Tucson, Arizona, it is only 25 km by a paved road to the top of Mount Lemmon (2800 m elevation) in the Santa Catalina Mountains. But the climate and the plants at the summit are far more similar to those in northern California and Oregon—1500 km to the north—than to those in the desert just below (**Table 3.1**). Similarly, the spruce–fir forests on the summit of the Great Smoky Mountains in Tennessee are more similar to the boreal forests of northern Canada than to the deciduous forests in the valleys below. Puerto Rico, which lies in the Caribbean Sea at 18°N latitude, is about 150 km long and 50 km wide and has a central mountainous backbone

TABLE 3.1 *The Influence of Elevation on Climate*

Site	Elevation (m)	Temperature (°C)				Mean annual precipitation (cm)
		Mean January	Mean July	Lowest	Highest	
Tucson, Arizona	745	10.8	30.7	−9.4	46.1	27.3
Mt. Lemmon, Arizona	2791	2.3	17.8	−21.7	32.8	70.0
Salem, Oregon	60	3.2	19.2	−24.4	40.0	104.3

Source: Data from U.S. Weather Bureau.

Note: Two of the sites are near one another in Arizona; the third site is in Oregon. Note that the climate of the high-elevation site in Arizona—Mt. Lemmon—is much more similar to that of Salem, Oregon, 1700 km to the north, than to that of Tucson, only 25 km away but 2000 m lower in elevation.

rising to about 1000 m. The lowlands on the northern and eastern sides are lush and tropical, but much more rain falls at higher elevations on the northeastern slopes, and this is where the best-developed rain forests are found. So much moisture is lost as the northeast trade winds traverse the mountains that the southwestern corner of Puerto Rico is extremely dry; the cacti and shrubby vegetation that occur there remind a visitor of the deserts and tropical thorn forests of western Mexico (see Figure 3.7C, D). Even more dramatic are the combined effects of the cold Humboldt Current and the rain shadow cast by the westward-flowing southeast trades over subtropical regions of the Andes in Peru and Chile. Here, up to 10 m of precipitation per year drenches the tropical rain forests on the eastern slope, while there may be several years in succession with no rain at all in the Atacama Desert on the western slope. Note that because they are located in regions where the prevailing winds come from different directions, the Atacama Desert (10° – 15° S, with westward-flowing southeast trades) and the Monte Desert (around 30° S, with eastward-flowing Westerlies) are located on opposite sides of the Andes.

There are also year-to-year and longer-term temporal variations in climate. The entire global system of moving air masses, ocean currents, and patterns of precipitation fluctuates on a five-to-seven-year cycle. The fluctuations appear to be initiated by events in the vast tropical Pacific Ocean (although similar events occur in the tropical Atlantic). This pattern is called the **El Niño Southern Oscillation**, or **ENSO** for short. We are uncertain about its initial cause—perhaps a variation in the output of solar radiation or intrinsic fluctuations in the atmosphere–ocean system. Whatever the ultimate cause, the pattern of tropical ocean circulation changes. While the primary ocean currents are the great hemispheric gyres mentioned above, close inspection of Figure 3.5 will show a small current running west to east right along the Equator. It is called the **equatorial countercurrent** because it runs in the opposite direction of the gyres. It is usually small, as the figure suggests, but in some years it becomes much stronger and pushes warm water away from the Equator up the coasts of North and South America. As the westerly winds pass over this warm water, they pick up moisture and carry it onto the adjacent continents, causing heavy precipitation in the winter when the land is colder than the offshore waters. This phenomenon is called **El Niño** (literally, “little boy” in Spanish, the predominant common language in western South America) because the resulting rains tend to fall around Christmas, the celebration of the birth of the Christ child. El Niño years are the only times that it rains in the extremely arid coastal deserts of South America. The seemingly lifeless Atacama Desert bursts into bloom, as plants that have survived as seeds (dormant in the soil) germinate, grow, and reproduce. In contrast, ENSO events also are characterized by reduced coastal upwelling, which leads to dramatic reductions in nutrients and affects the entire food chain of oceanic communities. Seabirds and other marine organisms along the Pacific coast and in the Galápagos Islands suffer wholesale reproductive failure and mortality due to the unusual rain and reduced upwelling.

Other kinds of temporal variation can also have important biogeographic consequences. For example, a hurricane may pass over a Caribbean island only once in a century on average, yet these rare, unpredictable storms wreak incredible devastation. Hurricanes probably are one of the primary causes of disturbance on Caribbean islands. Such large, infrequent storms may increase or decrease biodiversity; while they can inundate tiny islets, causing extinction of some terrestrial animals and plants, they also clear space in forests and coral reefs, which facilitates the continued existence of competitively inferior species (Spiller et al. 1998). The more general lesson for biogeographers cannot be overstated: Seemingly unpredictable and extremely rare events such as

hurricanes, cataclysmic volcanic eruptions, long-distance chance dispersal to isolated oceanic islands, or collision with a wayward asteroid can fundamentally transform the development and distributions of life on Earth.

Soils

Primary succession

Except for the polar ice caps and the perpetually frozen peaks of the tallest mountains, almost all terrestrial environments on Earth can and do support life. Areas of bare rock and other sterile substrates created by volcanic eruptions or other geological events are gradually transformed into habitats capable of supporting living ecological communities by a process called **primary succession**. This process involves the formation of soil, the development of vegetation, and the assembly of a complement of microbial, plant, and animal species.

We cannot understand the distribution of soils without knowledge of the role of climate and organisms in successional processes. The type of vegetation covering a region depends primarily on three ingredients: climate, type of soil, and history of disturbance. For example, three distinct vegetation types (temperate deciduous forest, pine barrens, and salt marsh) occur in northern New Jersey in close proximity to one another but on different soil types (Forman 1979). Moreover, if a mature stand of deciduous forest is destroyed, such as at the hands of humans or by natural fire, it is not reestablished immediately. Instead, certain plant species colonize the area and are in turn replaced by later colonists, beginning with weedy pioneer species and continuing until the mature or **climax** vegetation is reestablished. This process of community development on existing soils (as opposed to that on volcanic ash or bare rock) is called **secondary succession**. Throughout this process, both the microclimate and the soil of the site also change—becoming more favorable for some species and less favorable for others.

Soil formation is both a chemical and a biological process resulting from weathering of rock and the accumulation of organic material from dead and decaying organisms. The process by which new soil is formed from mineral substrates is usually long and complicated. Physical processes such as freezing and thawing, and water and wind erosion, break down the parent rock material. Organisms also play key roles: Lichens hasten the weathering of rock; decaying corpses of plants, animals, and microbes add organic material; the activities of roots and microbes alter the chemical composition of the soil; and burrowing animals mix and aerate it.

Totally organic soils (or **histosols**), such as peat, form in certain unusual environments where cold, acidic, or other conditions inhibit decomposition of accumulating plant and animal debris. In fact, the rate of soil formation varies widely depending largely on the nature of the parent material and the climatic setting. The formation of shallow soils may take thousands of years in Arctic and desert regions, where temperature and moisture regimes are extreme (e.g., McAuliffe 1994). For example, soils only a few centimeters deep cover much of eastern Canada where the retreat of the last Pleistocene ice sheets left bare rock only about 10,000 years ago. In other cases, especially when soils are formed from sand, lava, or **alluvial** materials in regions with warm, moist climates, primary succession can be amazingly rapid. In 1883, the small tropical island of Krakatau in Indonesia experienced an explosive volcanic eruption that exterminated all living things and left only sterile volcanic rock and ash. Organisms rapidly recolonized Krakatau from the large neighboring islands of Java and Sumatra, and by 1934—only 50 years after

the eruption—35 cm of soil had been formed and a lush tropical rain forest containing almost 300 plant species was rapidly developing (Docters van Leeuwen 1936; Thornton 1996; Whittaker 1998).

Formation of major soil types

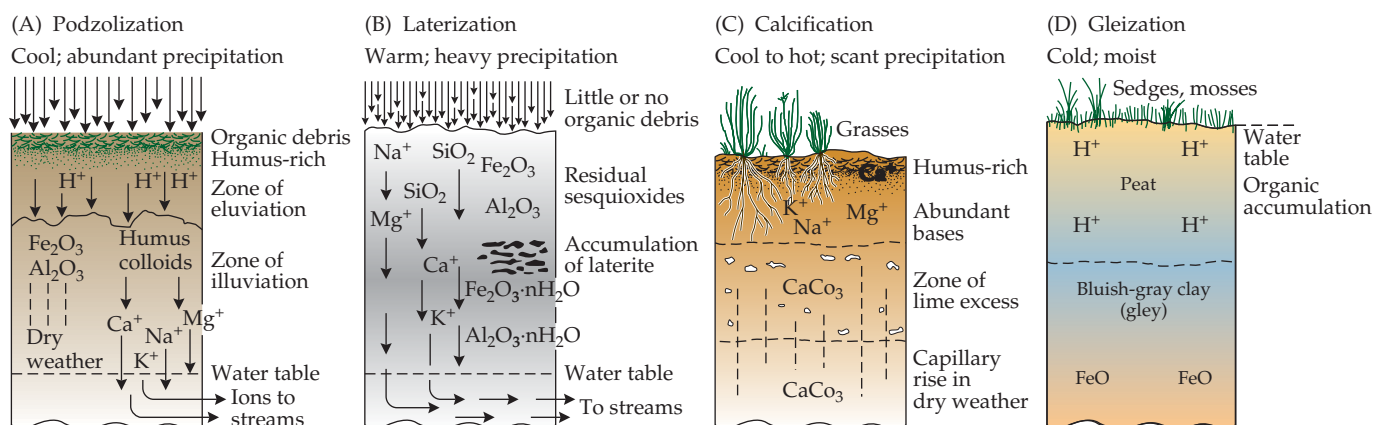
Anything we write about soils must be a gross oversimplification because both the classification and the distributions of soils are very complex, even controversial. Visit the vast flat plains of the United States or the Ukraine and you will find just one or a few soil types distributed as far as the eye can see, but in other geographic regions—especially mountainous areas—soil and geological maps are mosaics that look like complicated abstract paintings. For example, Great Britain has a series of unusual organic soil types formed in cold, wet environments, as well as soils overlaid onto a complex geological foundation and greatly modified by centuries of human activities.

We can begin to appreciate the diversity and distribution of soils by studying the four major processes that produce the primary (or *zonal*) soil types. These so-called **pedogenic regimes** are those that typically occur in habitats characterized by temperate deciduous and coniferous forests (**podzolization**), tropical forests (**laterization**), arid grasslands and shrublands (**calcification**), and waterlogged tundra (**gleization**).

Podzolization occurs at temperate and subarctic latitudes and at high elevations where temperatures are cool and precipitation is abundant. In such climates plant growth may be substantial, but the low temperatures inhibit microbial activity, so organic matter, called **humus**, accumulates. As the humus decays, organic acids are released and carried downward (**leached**) through the soil profile by percolating water. The hydrogen ions of these acids tend to replace cations that are important for plant growth, such as calcium, potassium, magnesium, and sodium, which are removed by leaching from the soil (**Figure 3.9A**). This process leaves behind a silica-rich upper soil containing oxidized iron and aluminum compounds, but few cations. Coniferous forests, which thrive in such acidic conditions, are a characteristic vegetation type on podzolic soils.

In the humid tropics, which experience high temperatures and heavy rainfall, little humus can accumulate, because microbes and other organisms rapidly break down dead organic material. In the absence of organic acids, oxides of iron and aluminum precipitate to form red clay or a bricklike layer (laterite). The heavy rainfall causes silica and many cations such as potassium, so-

FIGURE 3.9 Schematic representations of the four major pedogenic regimes showing the resulting soil profiles: (A) podzolization, (B) laterization, (C) calcification, and (D) gleization. (After Strahler 1975.)



dium, and calcium to be leached out of the soil (**Figure 3.9B**), leaving behind a firm and porous soil with very low fertility. In some areas, if the tropical forest cover is removed, the organic material and its bound nutrients are easily lost and the intense equatorial sun bakes the exposed lateritic soils hard, retarding secondary succession and making the area unsuitable for agriculture.

Calcareous soils typically occur in arid and semiarid environments, particularly in regions where thick layers of calcium carbonate were deposited beneath ancient shallow tropical seas. Where rainfall is relatively low, such that evaporation and transpiration exceed precipitation, cations are generally not leached out. Instead they are carried downward through the soil profile to the depth of greatest water penetration, where they precipitate and form a layer rich in calcium carbonate (**Figure 3.9C**). In desert soils, the scanty rainfall penetrates only a short distance below the surface, where it leaves behind a rocklike layer of calcium carbonate called **caliche** or **petrocalcic horizons**. In regions where precipitation is higher, water and roots penetrate deeper into the soil profile, leading to the formation of deep, fertile soils rich in organic material and essential nutrients such as potassium, nitrogen, and calcium. Such soils are typical of tallgrass and shortgrass prairie habitats, although little of the former remains because these soils are so highly prized for agriculture.

In cold and wet polar regions, gleization is the typical process of soil formation. At the permanently wet (or frozen) surface, where the low temperatures and waterlogged conditions prevent decomposition, acidic organic matter builds up, sometimes forming a layer of peat that can be several meters thick (**Figure 3.9D**). Below this organic upper layer an inorganic layer of grayish clay, containing iron in a partially reduced form, typically accumulates. While few nutrients are lost through leaching, the highly acidic conditions cause nutrients to be bound up in chemical compounds that cannot be used by plants. Thus, gley soils typically support a sparse vegetation of acid-tolerant species.

The above descriptions represent four idealized cases of soil formation processes. Given the complex variation in parent material and climate over the Earth's surface, pedogenic regimes vary in complex but predictable ways. The processes of soil formation and soil types described above occur where the chemical composition of the parent material is typical of the common rock types: sandstone, shale, granite, gneiss, and slate. The soils that are derived from these "typical" rocks are called **zonal soils**. A simplified summary of the relationship between climate and zonal soil type is given in **Figure 3.10**.

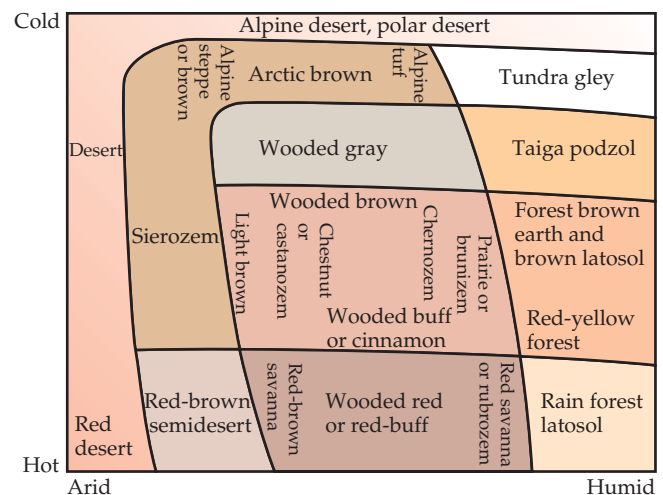


FIGURE 3.10 Schematic diagram depicting the relationships between major soil types and climate, showing that different combinations of temperature and precipitation cause the formation of distinctive soil types. (After Whittaker 1975.)

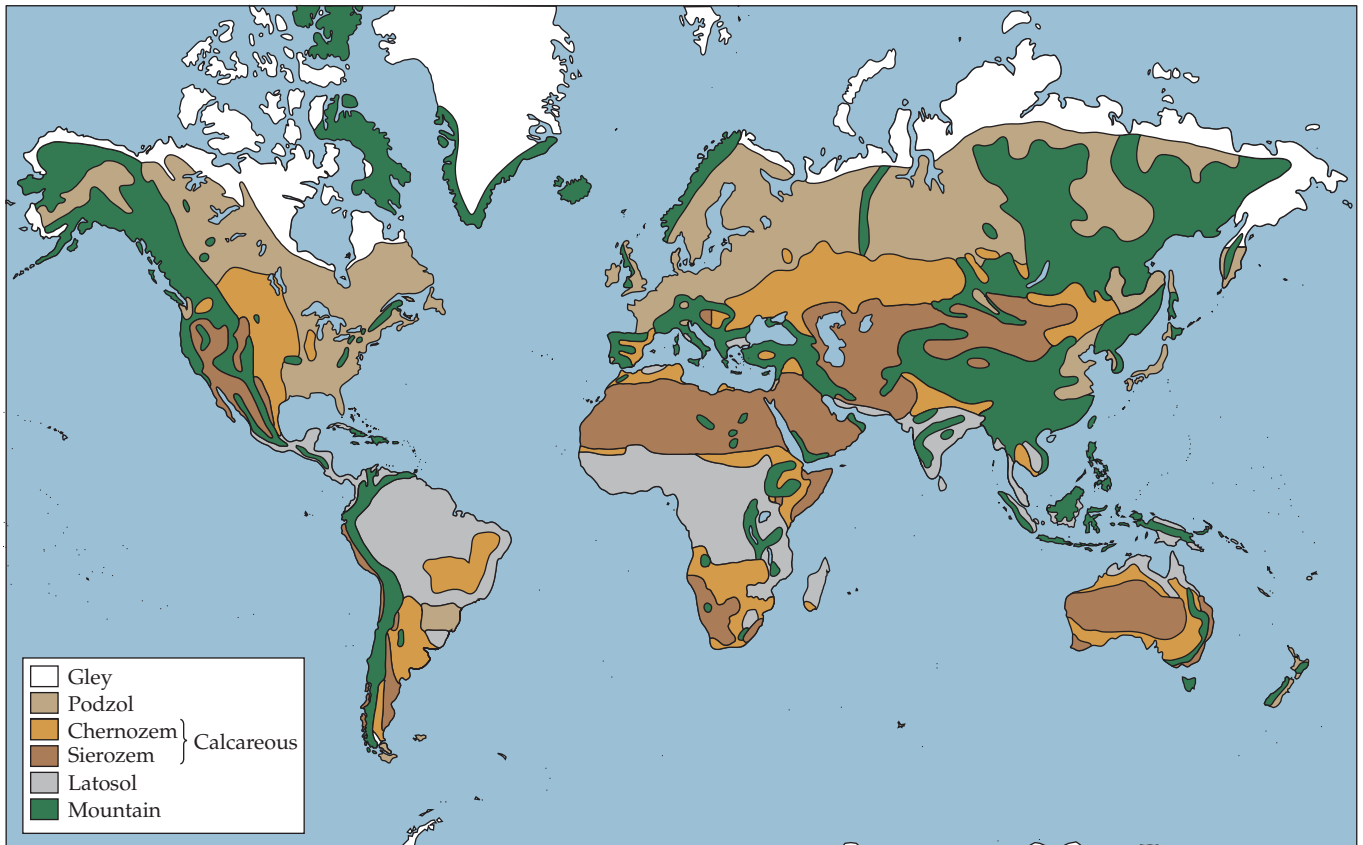


FIGURE 3.11 World distribution of major soil types. Note the close correlation of these soil types with the climatic zones shown in Figure 3.7, which reflects the influence of temperature and precipitation on soil formation.

The global distribution of zonal soil types (**Figure 3.11**) can also be compared with the global climate map (see Figure 3.8) to demonstrate the close relationship between soils and climate.

Unusual soil types requiring special adaptations

In addition to such zonal soils, there are unusual soil types derived from parent material of unusual chemical composition. Certain rock types such as gypsum, serpentine, and limestone contain unusually high amounts of some compounds and little of others. Serpentine, for example, is particularly deficient in calcium, and gypsum contains an excess of sulfate. Few plant species can tolerate such azonal soils, and the low-diversity plant communities that do grow on such soils have special physiological adaptations for dealing with their unusual chemical composition.

One example of a soil type that requires special adaptations by plants is **halomorphic soil**, which contains very high concentrations of sodium, chlorides, and sulfates. Halomorphic soil typically occurs near the ocean in estuaries and salt marshes, and in arid inland basins where shallow water accumulates and evaporates, leaving behind high concentrations of salts. A small number of specialized **halophytic** (salt-loving) plant species grow in such areas. They include a variety of taxonomic and functional groups, each of which has special adaptations for dealing with the problem of maintaining osmotic and ionic balance in these environments. Some species of mangroves and grasses excrete salts from specialized cells in their leaves, whereas pickleweeds and ice plants store salts in special cells in their succulent leaves.

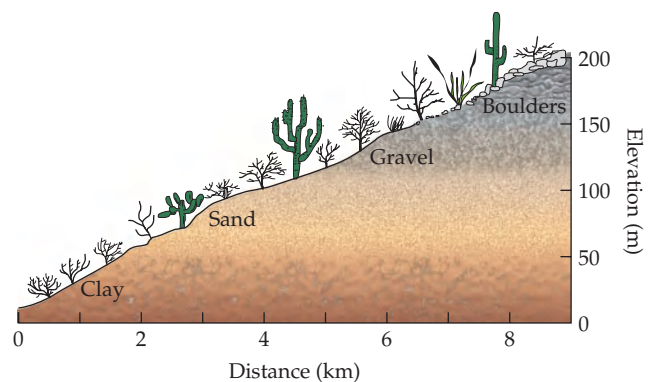
As mentioned above, highly acidic soil conditions cause essential nutrients—especially nitrogen and phosphorus—to be bound in compounds that

plants cannot use. Pitcher plants, sundews, Venus's flytraps, and other insectivorous plants can grow in highly acidic soils or other environments where nutrients are severely limited. These plants obtain their nitrogen and phosphorus by capturing living insects, digesting them, and assimilating the nutrients. A less spectacular adaptation to acidic and other nutrient-poor soils is evergreen vegetation (Beadle 1966). Because nutrients are lost when leaves are dropped, and because more minerals must then be taken up by the roots to produce new leaves, plants can use limited nutrients more efficiently by retaining their leaves for longer periods. In mesic (moist) temperate climates, where the predominant vegetation is usually deciduous forest, it is common to find evergreens growing on acidic and nutrient-poor soils. Examples are the pine barrens of the eastern United States and the eucalyptus forests of Australia (Daubenmire 1978; Beadle 1981).

In addition to their chemical composition, the physical structure of soils can influence the distribution of plant species and the nature of vegetation. In arid regions, for example, the size and porosity of soil particles affect the availability of the limited moisture to plants by affecting the runoff, infiltration, penetration, and binding of water. Thus, even within a small region of uniform climate, differences in soil texture can cause large differences in vegetation. A striking example is provided by the *bajadas* (or extended alluvial fans) of desert regions (Figure 3.12). These interesting geological formations are made up of sediments carried out of mountains by infrequent but heavy flooding of the canyons. As the floodwater gradually loses energy, it deposits sediments in a gradient—dropping large, heavy rocks at the mouths of the canyons and small sand- and clay-sized particles at the bottom of the fan. The resulting *bajada* shows a corresponding gradient in water availability and vegetation (Bowers and Lowe 1986). Cacti predominate on the coarse, rocky, well-drained soils high on the *bajada*, where water is available only for short periods during and after rains. These succulents can take up water rapidly through their extensive shallow roots and store it in their expandable tissues. Shrubs and grasses are much more common farther down the *bajada*, where their roots can extract the water held on and among the smaller soil particles.

A somewhat similar situation occurs along the coast of the Gulf of Mexico in the southeastern United States. The uplands have coarse, sandy, well-drained soils that support drought-tolerant, coniferous woodland/savanna vegetation. In contrast, the lowlands have accumulated fine, water-retaining soils, and this—as well as their proximity to the water table—allows them to support much more mesic vegetation. Thus, a person interested in the factors influencing plant distributions and community composition at this local to

FIGURE 3.12 Schematic representation of the local elevational distribution of soil particle size and vegetation on a desert *bajada* on the Sonoran coast of the Gulf of California (Sea of Cortez). At the upper end of the alluvial fan, where large boulders have been deposited, the vegetation is dominated by cacti and other succulents that can take up water rapidly before it percolates below the root zone. At the lower end, where water infiltration is poor and the existing water is tightly bound by fine clay particles, the vegetation consists of sparse, shallowly rooted shrubs. The greatest water availability, productivity, and species diversity occur at intermediate elevations, where the soils are sandy, infiltration is high, and water is not tightly bound by soil particles.



regional scale must pay particular attention to how subtle characteristics of soil structure affect the runoff, infiltration, and retention of rainwater.

Although we have concentrated here on the relationship between soils and vegetation, soils also affect the distributions of animals—both indirectly by controlling which plant species are present, and directly through the effects of the chemical and physical environment on their life cycles. Many kinds of mammals, reptiles, and invertebrates are restricted to particular types of soils that meet their specialized requirements for burrowing and locomotion. For example, in North American deserts, lizards of the genus *Uma*, the kangaroo rat *Dipodomys deserti*, and the kangaroo mouse *Microdipodops pallidus* are all restricted to dunes and similar patches of deep, sandy soil. Another set of species, including chuckwalla (*Sauromalus obesus*), collared lizards (*Crotophytus collaris*), and rock pocket mice (*Chaetodipus intermedius*) show just the opposite habitat requirement, that of being restricted to rocky hillsides and boulder fields.

Aquatic environments

As anyone who has ever tried to keep tropical fish knows, warm and relatively stable temperatures are essential for their survival and reproduction. Salinity, light, inorganic nutrients, pH, and pressure also play key roles in the distributions of aquatic organisms. Like terrestrial climates, the physical characteristics of water often exhibit predictable patterns along geographic gradients, which can be understood with a basic background in physics.

Stratification

THERMAL STRATIFICATION. When solar radiation strikes water, some is reflected but most penetrates the surface and is ultimately absorbed. Although water may be transparent, it is much denser than air, and its absorption of radiation is rapid. Even in exceptionally clear water, 99 percent of the incident solar radiation is absorbed in the upper 50 to 100 m, and this absorption occurs even more rapidly if many organisms or colloidal substances are suspended in the water column. Longer wavelengths of light are absorbed first; the shorter wavelengths—which have more energy—penetrate farther, giving the depths their characteristic blue color.

This rapid absorption of sunlight by water has two important consequences. First, it means that photosynthesis can occur only in surface waters where the light intensity is sufficiently high (the **photic zone**). Virtually all of the primary production that supports the rich life of oceans and lakes comes from plants living in the upper 10 to 30 m of water. Along shores and in very shallow bodies of water, some species such as kelp are rooted in the substrate. These plants may attain considerable size and structural complexity, and may support diverse communities of organisms. In the open waters that cover much of the globe, however, the primary producers are tiny, often unicellular algae (called **phytoplankton**), which are suspended in the water column. **Zooplankton**, tiny crustaceans and other invertebrates that feed on phytoplankton, migrate vertically on a daily cycle: up into the surface waters at night to feed, and down into the dark, deeper waters during the day to escape predatory fish that rely on light to detect prey.

Second, the rapid absorption of solar radiation by water means that only surface water is heated. Any heat that reaches deeper water must be transferred by conduction or convection by vertical currents. Consequently, deep waters are characteristically cold, even in the tropics. The density of pure water is greatest at 4° C and declines as its temperature rises above or falls below this point. This unusual property of water is significant for the sur-

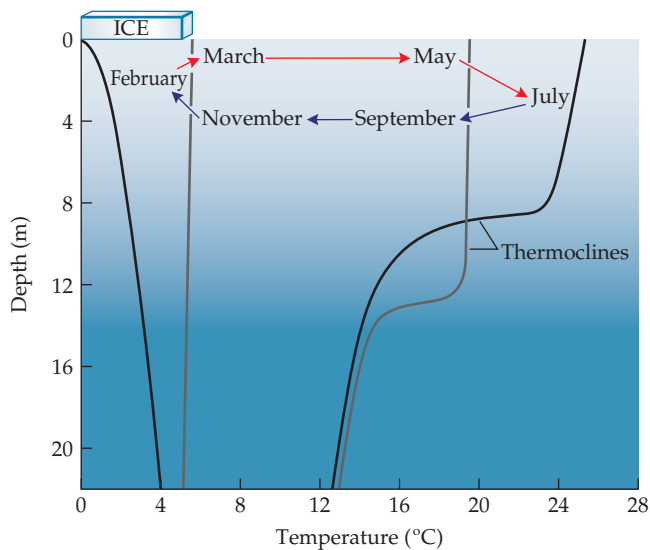


FIGURE 3.13 Vertical temperature profiles of Lake Mendota, Wisconsin, at different dates from summer through winter showing the loss of thermal stratification as the lake cools. In July the thermocline is pronounced and shallow; in September it is less pronounced and deeper; and by November it has disappeared, allowing surface and deep waters to mix during the fall overturn. Because the density of water increases as it cools down to 4° C, but then decreases as water molecules begin to form ice crystals, thermal stratification is typically reversed under the ice cover in winter (February). (After Birge and Juday 1911.)

ture between the surface layer and deeper water is called a **thermocline** (Figure 3.13). Mixing of the surface water by wave action determines the depth of the thermocline and maintains relatively constant temperatures in the water above it. In small temperate ponds and lakes that do not experience high winds and heavy waves, the thermocline is often so abrupt and shallow that swimmers can feel it by letting their feet dangle a short distance. In large lakes and oceans, where there is more mixing of surface waters, the thermocline is usually deeper and less abrupt.

Tropical lakes and oceans show pronounced permanent stratification of their physical properties, with warm, well-oxygenated, and lighted surface water giving way to frigid, nearly anaerobic, and dark (aphotic) deep water. Oxygen cannot be replenished at great depths where there are no photosynthetic organisms to produce it, and the stable thermal stratification prevents mixing and reoxygenation by surface water. Only a relatively small but fascinating menagerie of organisms can exist in these extreme conditions. The feces and dead bodies of organisms living in the surface waters sink to the depths, taking their mineral nutrients with them. The lack of vertical circulation thus limits the supply of nutrients to the phytoplankton in the photic zone. Consequently, deep tropical lakes are often relatively unproductive and depend on continued input from streams for the nutrients required to support life.

OVERTURN IN TEMPERATE LAKES. The situation is somewhat different in temperate and polar waters. Deep temperate lakes, in particular, undergo dramatic seasonal changes: They develop warm surface temperatures and a pronounced thermocline in summer, but freeze over in winter. Twice each year, in spring and fall, the entire water column attains equal temperature and equal density (see Figure 3.13), the temperature/density stratification is eliminated, and moderate winds may generate waves that then mix deep and shallow water, producing what is called **overturn**. This semiannual mixing carries oxygen downward and returns inorganic nutrients to the surface. Phosphorus and other mineral nutrients may be depleted during the summer, when warm temperatures allow algae to grow and reproduce at high rates; overturn replenishes these nutrients by stimulating the growth of phytoplankton. Temperate lakes, such as the Great Lakes of North America, are often quite productive and support abundant plant and animal life, including valuable commercial fisheries. However, abnormally high nutrient inputs—

vival of many temperate and polar organisms because it means that ice floats. Ice provides an insulating layer on the surface that prevents many bodies of water from freezing solid. The presence of salts in water lowers its freezing point, and some organisms are therefore able to exist in unfrozen water below 0° C (de Vries 1971).

A more general consequence of the relationship between water density and temperature is that water tends to acquire stable thermal stratification. When solar radiation heats the water surface above 4° C, the warm surface water becomes lighter than the cool deeper water, so it tends to remain on the surface where it may be heated further and become even less dense. In tropical areas and in temperate climates during the summer, the surfaces of oceans and lakes are usually covered by a thin layer of warm water. Unless these bodies of water are shallow, the deep water below this layer is much colder (sometimes near 4° C). The change in temperature

often due to runoff from agricultural fields and discharges of inadequately treated sewage—can cause excessive production, rapid algal growth, depletion of oxygen, fish kills, and other environmental problems.

Oceanic circulation

The vertical and horizontal circulation of oceans is more complicated than that of lakes, in part because oceans are so vast, extending through many climatic zones, and in part because salinity affects the density of water. Salts are dissolved solids carried into the oceans by streams and concentrated by evaporation over millions of years. The presence of salts in water increases its density, causing swimmers to experience greater buoyancy in the ocean than in freshwater. Varying salinity and density have important effects on ocean circulation. Rivers and precipitation continually supply freshwater to the surface of the ocean, and this lighter water tends to remain at the surface. If you have ever flown over the mouth of a large, muddy river such as the Mississippi, Thames, or Nile, you may have noticed that its water remains relatively intact, flowing over the denser ocean water for many kilometers out to sea. In polar regions, the input of freshwater to the ocean from rivers and precipitation generally exceeds losses from evaporation, but the reverse is true in the tropics. This pattern creates a somewhat confusing situation because warm tropical surface water tends to become concentrated by evaporation and to increase in density, counteracting to some extent stratification owing to temperature. Conversely, cold polar water—which would be expected to show little stratification—may become somewhat stabilized as low-density freshwater accumulates on the surface.

Vertical circulation occurs in oceans, but the rates of water movement are so slow that a water mass may take hundreds or even thousands of years to travel from the surface to the bottom and back again. Areas of descending water tend to occur at the convergence of warm and cold currents in polar regions, where the colder, denser water sinks under the warmer, lighter water. Areas of rising water, called **upwellings**, are found where ocean currents pass along the steep margins of continents. This happens, for example, along the western coast of North and South America, where there is little continental shelf and the land drops sharply offshore. As the Pacific gyres sweep toward the Equator along these shores, the Coriolis effect and, in tropical latitudes, the easterly trade winds tend to deflect the surface water offshore, and water wells up from the depths to replace it. Because upwelling, like the overturn in lakes, returns nutrients to the surface, productivity tends to be high in areas of upwelling (see the global productivity map of Figure 5.30A). Probably the greatest commercial fishery in the world is located in the zone of upwelling off the coasts of Chile and Peru, making the episodic ENSO events that were discussed earlier in this chapter particularly devastating to local economies.

Surface currents, such as the great hemispheric gyres (see Figure 3.5), are relatively shallow and rapidly moving, so they tend to form discrete water masses, each of which has a characteristic salinity and temperature profile distinct from those of neighboring water masses. Some organisms with limited capacity for locomotion may drift in currents for long distances without leaving a single uniform water mass. Organisms that can move actively to overcome the currents must also be able to tolerate the contrasting physical environments in different water masses.

Although oceanographers have recognized the existence of distinct water masses within the oceans for many years, modern technology has revealed the extent of spatial heterogeneity in shallow ocean waters. For example, investigators from the Woods Hole Oceanographic Institution have studied the

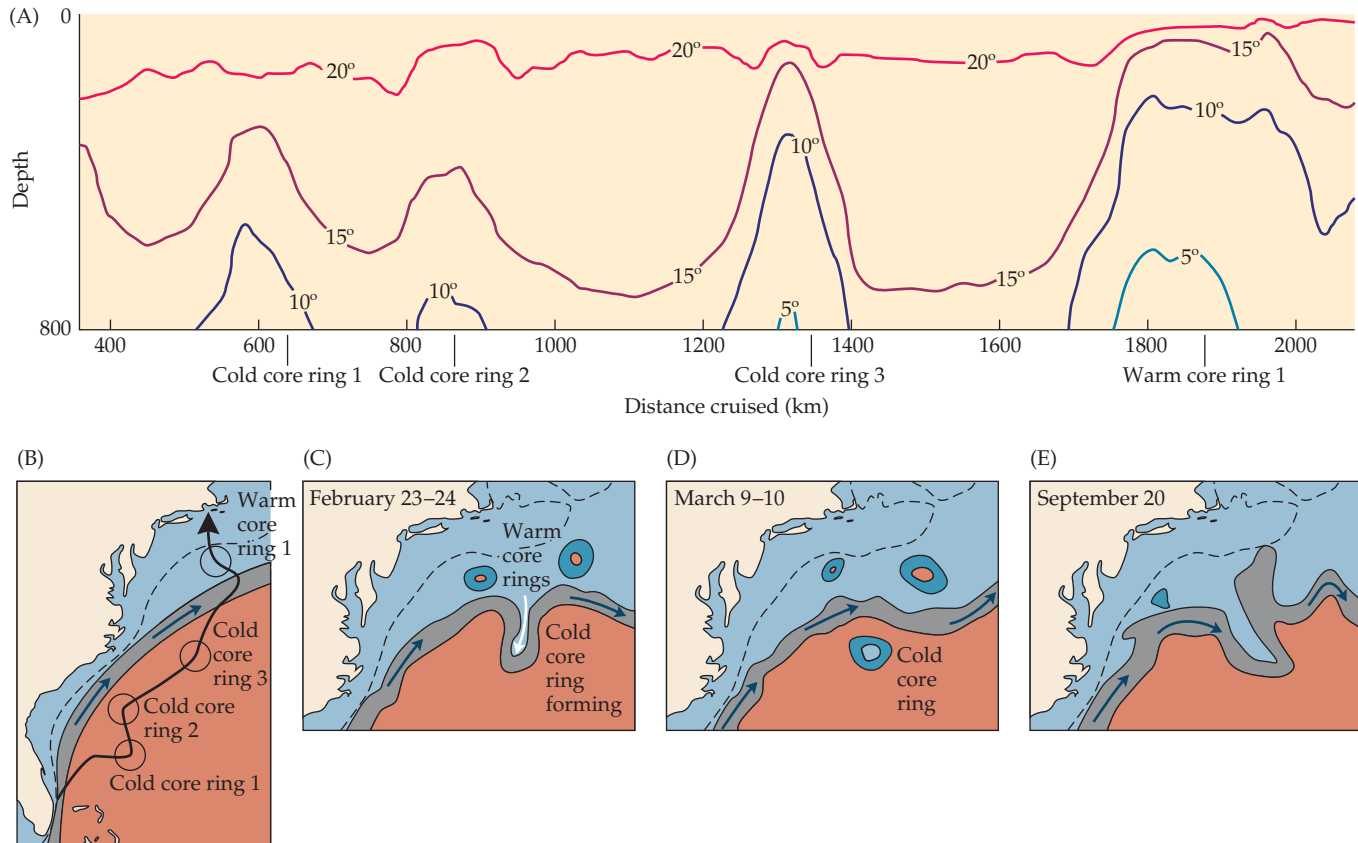


FIGURE 3.14 Local- to regional-scale spatial and temporal heterogeneity of surface waters in the North Atlantic Ocean is caused by meanders of the Gulf Stream (dark gray shading) that create great rings (some as much as 300 km in diameter) of Gulf Stream water which encircle water of different origin. Rings to the north of the Gulf stream encircle a core of water derived from the Sargasso Sea, and thus tend to be relatively warm (around 18°C), while the core of those to the south of the stream tend to be cold (< 10°C), drawing their waters from northern regions (see cold core ring forming in part C). (A) Temperature/depth profile recorded by an oceanographic vessel that traveled through several rings, as indicated by the dashed line on map (B). (C–E) Changes in water surface temperatures as mapped by infrared satellite imagery showing the formation, movement, and disappearance of rings. (After Wiebe 1982.)

physical environment and the biota of Gulf Stream “rings” (Wiebe 1976, 1982; Lai and Richardson 1977; Katsman et al. 2003). These rings are small masses of cold or warm water that have broken away from the southern or northern edges of the Gulf Stream to drift through water of contrasting temperature in the North Atlantic. They can be readily seen on infrared satellite images that show sea surface temperatures (Figure 3.14). These rings not only have physical environments that are strikingly different from their surroundings, but also contain a unique biota that can persist in these special conditions far from its normal distribution in the Gulf Stream. The possible roles of these floating warm- or cold-water eddies in trans-Atlantic dispersal—both now and in the past—are intriguing subjects for future biogeographic research.

Pressure and salinity

Pressure and salinity vary greatly among aquatic habitats. These variations have major effects on the distributions of organisms, because special physiological adaptations are necessary to tolerate the extremes. As every scuba diver knows, water pressure increases rapidly with depth. It becomes a major problem for organisms in the ocean, where the deepest areas are up to 6 km below the surface. Pressure increases at a rate of about 1 atmosphere (about 1.5 mega Pascals) for every 10 m of depth. In the abyssal depths, pressures are more than 200 times greater than at the surface. Organisms adapted to living in surface waters cannot withstand the pressures of the deep sea, and vice versa.

Variation in salinity is relatively discontinuous. The vast majority of the Earth’s water is in the oceans and is therefore highly saline (greater than 34 parts per thousand). In contrast, freshwater lakes, marshes, and rivers, which account for less than 1 percent of the Earth’s waters, contain very few dissolved salts. Habitats of intermediate or fluctuating salinity, such as salt marshes and

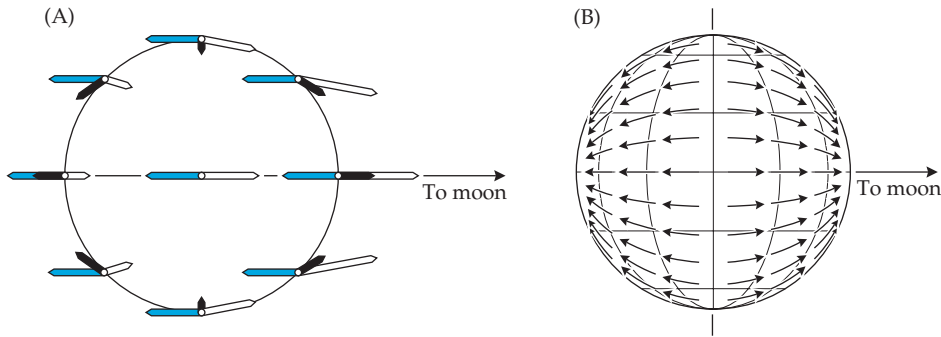


FIGURE 3.15 (A) Schematic representation of how the centrifugal force of the spinning Earth and the gravitational force of the moon cause the tides. On the side of the Earth closest to the moon, the gravitational force (white) is stronger than the centrifugal force (blue), and the net tidal force (black) tends to draw surface water toward the moon. On the side opposite the moon, the gravitational force is weaker than the centrifugal force, and the net tidal force tends to draw water away from the moon. In between these extremes, the gravitational and centrifugal forces are balanced, and there is essentially no net tidal force. (B) The movement of surface waters in response to these tidal forces.

estuaries, constitute only a tiny fraction of the Earth's aquatic habitats. Consequently, most aquatic organisms are physiologically adapted and geographically restricted either to freshwater, where the physiological problem is obtaining sufficient salts to maintain osmotic balance, or to salt water, where the problem can be eliminating excess salt. Only a few widely tolerant (**euryhaline**) organisms have the special physiological mechanisms required to survive in the widely fluctuating salinities of estuaries and salt marshes.

Tides and the intertidal zone

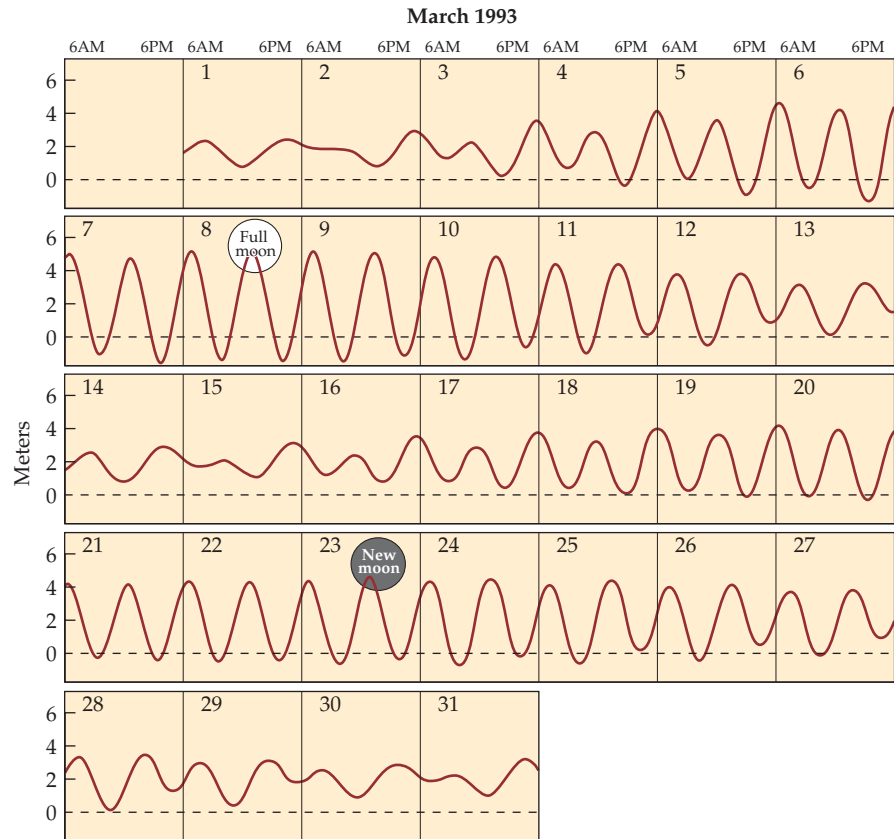
We can learn a great deal about the factors determining the distributions of organisms by studying environmental gradients: both gradual changes such as variation in light and pressure with depth in lakes and oceans, and rapid changes such as the variation in temperature in the cooling outflow of a hot spring. One of the steepest, best-studied, and most interesting environmental gradients occurs where the ocean meets the land. Along the shore is a narrow region that is alternately covered and uncovered by seawater. It is called the **intertidal zone** because it experiences a regular pattern of inundation and exposure caused by tides.

Sir Isaac Newton explained how the gravitational influences of the moon and sun interact to cause the global fluctuations in sea level that we call tides. The entire story is complicated, but the main pattern and its mechanism are simple. The tides are flows of surface waters. They occur in response to a net tidal force, which reflects a balance between the centrifugal force of the Earth and moon revolving around their common center of mass, and the gravitational forces of the moon and sun (**Figure 3.15**). Because the gravitational force exerted by one object on another is equal to its mass divided by the square of the distance between those objects, the smaller but nearer moon has a greater effect than the sun.

Most shores typically experience both a daily and a monthly tidal cycle: There are two high and two low tides every 24 hours, and there are two periods of extreme tides each month, corresponding to the new and full moons (**Figure 3.16**). During these periods, the moon and sun are in the same plane as the Earth, and their gravitational effects are additive, causing high-amplitude tides, or **spring tides**, with the highs occurring at dawn and dusk and the lows near noon and midnight. During the quarter moons, the sun and moon are at right angles to each other (from the perspective of Earth), and their gravitational effects tend to cancel each other, resulting in low-amplitude tides, or **neap tides**.

A distinct community of plant and animal species lives in the intertidal zone. Nearly all aspects of the lives of these organisms are dictated by the cyclical pattern of inundation by seawater at high tide and exposure to desiccating conditions at low tide. Most species are confined to a very narrow zone of tidal exposure, so their distributions form thin bands running hori-

FIGURE 3.16 A tide calendar for the northern Gulf of California (Sea of Cortez) showing the typical pattern of tides due to the gravitational influences of the moon and sun. Note that there are two high and two low tides each day. There are also two periods of low-amplitude (neap) and high-amplitude (spring) tides each month; the latter correspond to the times of the new and full moons when the gravitational forces of moon and sun are aligned. (Courtesy of D. A. Thomson.)



zontally along the shore. As we shall see in Chapter 4, the narrow ranges of species in the intertidal zone (typically only a few centimeters or meters) and the ease with which critical environmental conditions can be manipulated experimentally have produced a wealth of information about the factors limiting the distributions and regulating the diversity of species.

Time

Before discussing important methodological issues associated with describing and then exploring the response of biotas to variation in the geographic template, we first should emphasize one dimension that, although fundamental to all patterns and underlying processes in biogeography, ecology, and evolution, is often overlooked—time. There are at least two reasons why time should be explicitly considered in all attempts to understand the geography of nature. First, and perhaps most obvious, is that all features of the geographic template discussed above—climate, soils, and water chemistry—are temporally dynamic, varying in some complex but important manner over various scales of time from that of days and generations to millennia and geological periods. Second, but perhaps less intuitive, is that time itself influences the abilities of individuals, species, and entire biotas to respond to features of the geographic template even during periods of relative environmental stasis. That is, the fundamental biogeographic processes of dispersal, range expansion, and subsequent extinction or evolution in new regions takes time, such that biogeographic responses often lag far behind the actual dynamics of local to global environments. For example, geographic ranges of many species of plants and animals are still expanding poleward in response to retreat of the glaciers that occurred some 15,000 years ago, while the bio-

geographic imprint of plate tectonics, in terms of distinctiveness of biotas among the continents, remains indelible hundreds of millions of years since the supercontinent of Pangaea began to split apart (see Chapters 8 and 9).

Two-Dimensional Renderings of the Geographic Template

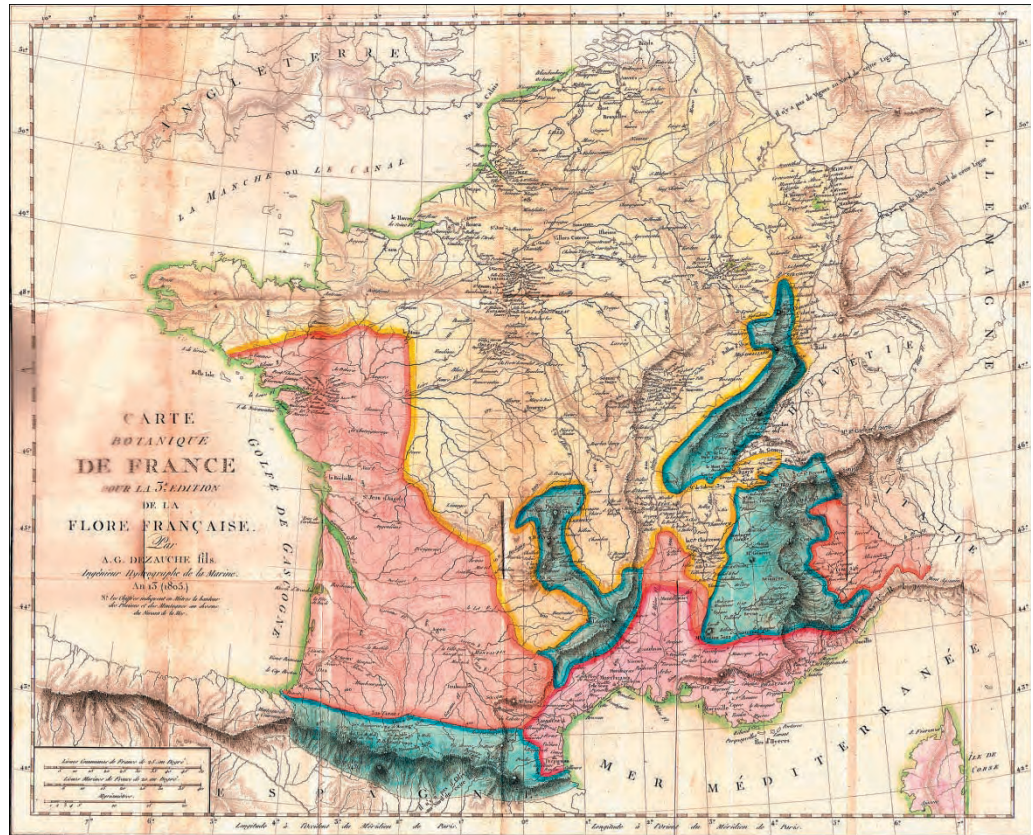
Early maps and cartography

As we observed in the previous chapter, setting the precise date of any major discovery or innovation in science is bound to be an elusive if not futile exercise. So it is with maps and cartography, whose developmental history is intricately linked with advances in perhaps all physical sciences from geology to astronomy. Maps, at least in their most rudimentary forms, are probably ancient and may well date back to the earliest renderings of game species and elements of their surroundings (e.g., a local river, mountain, or shoreline) that early humans etched into the sands at their feet or onto the walls of their dwellings. The earliest formal biogeographic maps, however, are not known to have existed until the early nineteenth century, perhaps the first of these commissioned by two luminaries in the history of biogeography, evolution, and ecology—Augustin-Pyramus de Candolle (1779–1841) and Jean-Baptiste Lamarck (1744–1829) (see Chapter 2).

In their discussion of the history of biogeographic maps, Ebach and Goujet (2006) make the distinction between maps that serve two fundamental objectives in biogeography: (1) maps of “**chorology**,” a term coined by Ernst Haeckel in 1866 to describe the science of the geographic spread of organisms, and (2) **systematic biogeographic maps**, which describe the distinctiveness and/or similarities among biotas from local (provincial) to global scales. We have already discussed examples of **chorological maps** in Chapter 2, including Buffon’s description of the spread of life-forms from his hypothesized Northern Origins (see Figure 2.2) and the extensionists’ maps of transoceanic land bridges across which biotas were hypothesized to have migrated among the continents (see Figure 2.8). Not surprisingly, one of the earliest and most comprehensive chorological maps was that created by Haeckel himself (see Figure 2.9), which vividly illustrates his hypothesis on the monophyletic origins and migration routes of early humans. Haeckel’s map also nicely captures the interdependence between these two types of maps and objectives in biogeography. That is, systematic maps describe the differences and similarities among regions in terms of their biotas (geographic races, species, etc.), while chorological maps provide hypothetical explanations for those patterns (migrations and subsequent evolution of species in isolated and environmentally disparate regions, and the dynamics of land, sea, and climates that affect those migrations). In fact, even the most descriptive biogeographic map, such as Lamarck and Candolle’s 1805 map of France’s botanical provinces (**Figure 3.17**) or Wallace’s world map of biogeographic regions (frontispiece) are both systematic and chorological, illustrating differences among regional biotas and simultaneously, at least implicitly, proposing explanations for those differences based on hypothetical migration paths or barriers to dispersal.

Next, we will briefly introduce the basic methods for creating maps, which are essentially two-dimensional representations (and distortions) of the three-dimensional curved surface of the Earth, and we will describe geographic coordinate systems for identifying particular locations across this surface. We will then explain, again in very basic terms, how scientists have delineated the fundamental unit of biogeography—the geographic range—before showcasing some advances in **visualization** and analysis of spatial pat-

FIGURE 3.17 Perhaps the first systematic, biogeographic map was the *Carte Botanique de France*, which was commissioned by Jean-Baptiste de Lamarck and Augustin-Pyramus de Candolle for the third edition of *Flore Française* (Lamarck and Candolle 1805). The five regions are distinguished by color: green = maritime plants; blue = mountain plants; crimson red = Mediterranean plants; yellow = plants widely dispersed throughout much of France and surrounding areas (lacking endemics); vermillion = vegetation that is intermediate or transitional between those of northern plains and southern provinces. (For additional descriptions of this map, see Ebach and Goujet 2006.)



terns—technological advances that provided the tools and new perspectives integral to the revitalization of biogeography that took place during the latter decades of the twentieth century.

Flattening the globe: Projections and geographic coordinate systems

Before we can describe and analyze distributional patterns and search for explanations for those patterns, it seems appropriate that we review the basis for visualizing these patterns (i.e., maps, which are simplifications and, quite often, distortions of the true geographic template). The distortion arises from two challenges: (1) describing and somehow transforming three-dimensional patterns across the curved surface of the Earth onto the two-dimensional plane of a map and (2) locating particular points or areas of interest on that plane. Fortunately, cartographers have been tackling these challenges for centuries, providing us with a wide array of **projections** and **geographic coordinate systems** to address each of these challenges, respectively. As **Figure 3.18** illustrates, developing a map projection is relatively simple in theory, but in practice it turns out to be impossible to accurately represent a three-dimensional entity on a two-dimensional surface without distortion. For relatively small areas, across which the curvature of the Earth is negligible, map distortions are so minor that they can be ignored for most applications. At larger scales, however, map projections can cause substantial distortions of shape, area, distance, and direction. Fortunately, cartographers have developed a diversity of projections with known properties such that we can choose the one most appropriate to our question. For example, conformal, equal-area, equidistant, and true directional projections preserve the shape, area, distances between points, and directions, respectively.

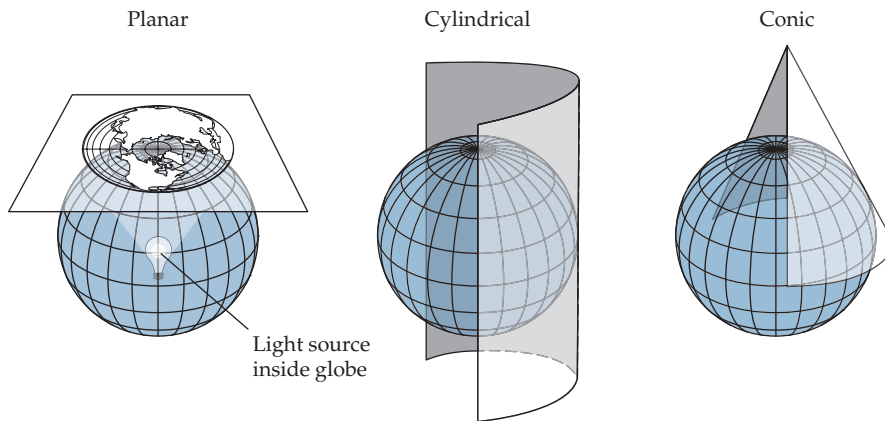


FIGURE 3.18 Although map projections are created using various mathematical formulae to transform spatial data from a curved three-dimensional surface to a flat two-dimensional map, the basic approach is not difficult to understand. If the features of interest (e.g., continental outlines or geographic ranges) are drawn onto a transparent model of the globe, those features can be transformed (“projected”) by a light source that passes through the globe and then onto a piece of paper located either inside or outside the globe. By folding the paper (e.g., to form a plane, cylinder, or cone) and by moving it to different positions around the globe, projections can be adjusted to better preserve one or more of the key geographic features (i.e., shape, area, distance, or direction).

Once we have decided on an appropriate projection, we still cannot locate particular points or areas of interest without a geographic coordinate system. The two most commonly used types of coordinate systems are (1) latitude and longitude and (2) the Universal Transverse Mercator (UTM) system. Locations in the UTM system are expressed as number of meters east of any of 60 arbitrary reference meridians (“false eastings”) and, in the Northern Hemisphere, number of meters north of the Equator. In the Southern Hemisphere, “eastings” are expressed in the same manner, while “northings” are expressed as number of meters north of an imaginary “false origin” located 10,000 m south of the Equator. With the addition of elevation, each of these systems can better approximate true complexities of the Earth’s surface.

Visualization of Biogeographic Patterns

History and exemplars of visualization in biogeography

Visualization is much more than just a means of displaying a particular phenomenon such as a geographic range; it is a creative and insightful process that ideally serves to fundamentally transform our abilities to view and conceptualize the world. As with Galileo’s telescope and van Leeuwenhoek’s microscope, biogeographic maps and graphical models provide new perspectives on these phenomena and, thus, serve as invaluable tools for exploring the interplay of processes that may be causally related to patterns in the geographic variation of nature. The most familiar visualizations in biogeography are, of course, maps, but the most informative and insightful biogeographic maps are much more than just accurate portrayals of landscapes and seascapes. Indeed, there exists throughout the history of the field a great wealth of transformative visualizations—maps and otherwise—whose diverse forms were limited only by the scientific and artistic ingenuity of their creators. Success in terms of the ability of any visualization to fundamentally alter the views of contemporary scientists and advance the field depends largely on striking a delicate balance between two competing objectives that constitute a fundamental tension in visualizations: achieving accuracy and precision (in describing geographic, environmental, and biotic features) on the one hand, versus simplification and abstraction (to develop a conceptual synthesis of the interplay and interdependence of underlying processes) on the other.

Here, we showcase three of what were arguably the most influential visualizations in the history of biogeography: William Smith’s “map that changed the world,” Alexander von Humboldt’s *Tableau Physique* (the keystone illustration in his *Essay on the Geography of Plants*; see Chapter 2), and Robert H. MacArthur and Edward O. Wilson’s graphical model of island biogeogra-



FIGURE 3.19 William Smith (1769–1839), although suffering great tragedy throughout his life, fundamentally transformed the way scientists of the eighteenth and nineteenth centuries viewed the world and the evolutionary dynamics of its rocks and life-forms. (Portrait by Hugues Fourau.)

phy. These three exemplars span a range of approaches in scientific visualization, from largely descriptive (but still fundamentally insightful and aesthetically compelling) to predominately abstract (with little geographic context but great versatility for understanding a broad diversity of patterns in biodiversity). Later, we provide a brief and unavoidably incomplete introduction to the various visualization and analytical tools that became available and, in no small way, contributed to the revitalization of biogeography that took place during the latter decades of the twentieth century. In Unit 4, we will introduce a new generation of mapping that projects three-dimensional renderings of phylogenetic trees onto geographic maps.

“The map that changed the world”

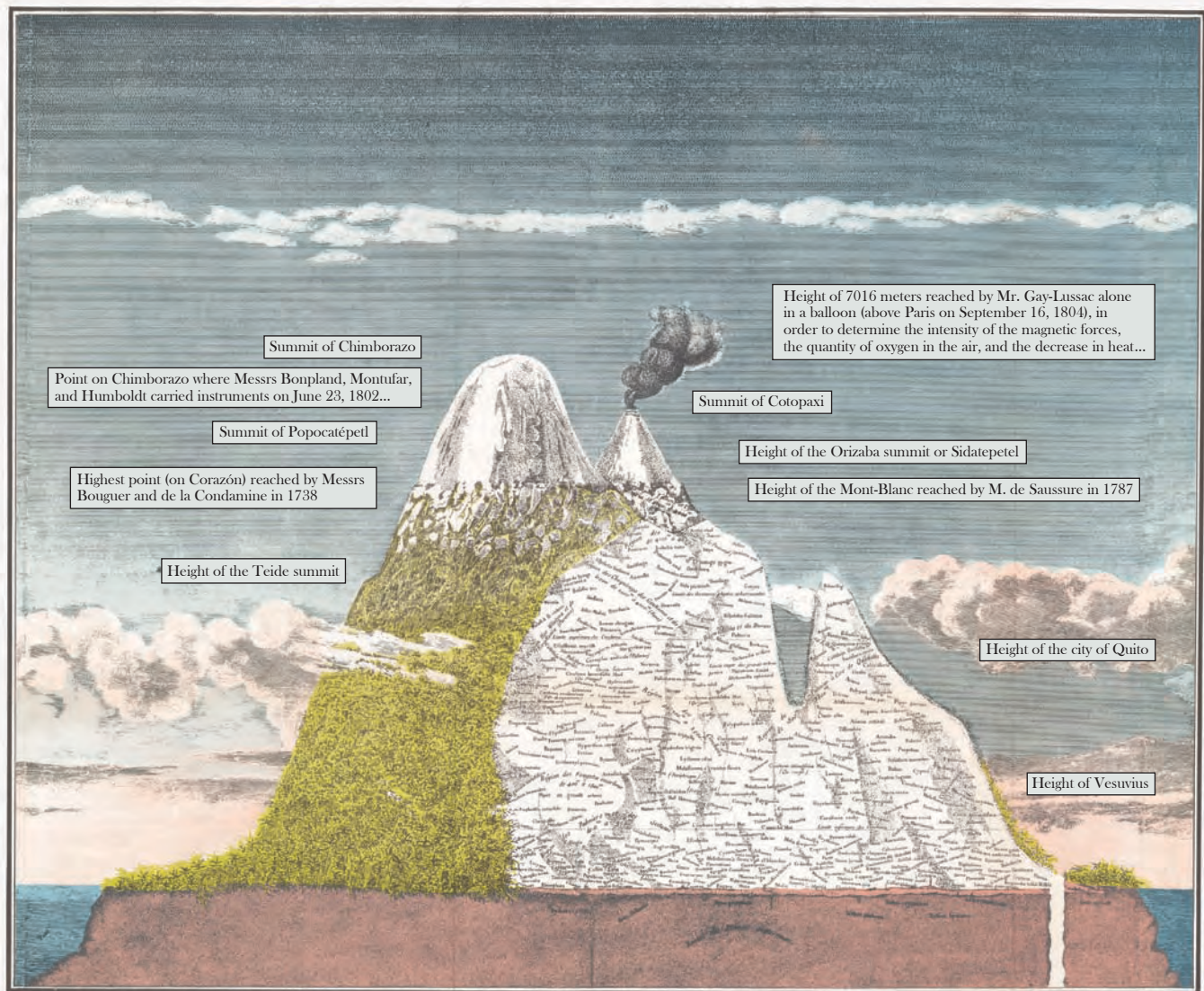
The story of William Smith (1769–1839; **Figure 3.19**) and his “map that changed the world” is a heartrending saga of tragedy and triumph (Winchester 2001). Unlike most of the distinguished scientists of that era—including Linnaeus, Buffon, Lamarck, and Humboldt, who were blessed with the privileges and wealth of aristocracy—Smith was of humble origins. His father, the village blacksmith, died at a young age and left him an orphan to be raised by his uncle. Indefatigable throughout his challenged life, Smith was a voracious, albeit by necessity self-taught, student, taking up the trade of surveying at 18 and eventually becoming an accomplished surveyor in his own right—working for Somerset Coal Canal Company for some eight years until he was dismissed in 1799. By then, at the age of 29, he had amassed a great wealth of information, meticulously recording the changes in rocks and associated fossils both horizontally across the landscapes of England and vertically along the slopes of pit mines. Most importantly, he realized that the series of those rocks and ancient, fossilized life-forms was invariant across England and that they formed what eventually would be recognized as the “geological record” of succession in ages of the Earth and its species.

He drafted the first, rough sketch of his map of England’s geological strata in 1801. Unfortunately, he was unable to find consistent employment for most of the next decade, which freed him to travel across the countryside to further develop his map, but also left him unable to fund its substantial production and publication costs. Eventually, he found a benefactor who fully appreciated the map’s great promise, and it was finally published in 1815. Rather than a time of heralded triumph, this marked the beginning of some 15 years of professional and personal suffering for Smith. His map was plagiarized and sold for less than the cost of the original. Smith became bankrupt and was sentenced to debtors’ prison; he became homeless and was forced to leave London, and his wife fell into insanity. It wasn’t until he was in his early 60s and working as an itinerant surveyor that he was finally recognized for his contributions, which truly did transform the way scientists viewed the world (**Figure 3.20**) to that of an ancient but dynamic system where evolution of its biotas is intricately linked to that of the Earth itself.

FIGURE 3.20 William Smith’s seminal map, *A Delineation of the Strata of England and Wales with Parts of Scotland*, first published in 1815, is often referred to as “the map that changed the world” (see Winchester 2001). This marked the first time geologists firmly established the connection between geographic variation in rocks and the temporal development of those rocks into distinctive strata—each identified by their unique associations (and evolutionary series) of fossils. In this region, uplift and erosion have exposed the strata along a roughly northwest to southeast gradient of decreasing antiquity (i.e., oldest strata in the upper

left-hand corner of this map; we have enlarged the cross-sectional key of rock strata and oriented it in the approximate latitudinal context; see Figure 8.1 for an enlarged image of the key to this map). This colossal visualization (the original comprised 15 separate copperplate engravings and spanned 74 × 105 inches) finally made it possible to fully appreciate the antiquity and dynamics of the Earth and its life-forms, and was thus foundational to the fields of geology, biogeography, ecology, and evolution. (From University of New Hampshire, Durham, NH, www.unh.edu/esci/wmsmith.html.)





GÉOGRAPHIE DES PLANTES ÉQUINOXIALES.

Tableau physique des Andes et Pays voisins

Dressé d'après des Observations & des Mesures prises Sur les lieux depuis le 10. degré de latitude boréale jusqu'à au 10. de latitude australe en 1799, 1800, 1801, 1802 et 1803.

PAR
ALEXANDRE DE HUMBOLDT ET AÏME BONPLAND.

Esquisse et crayon par M. de Humboldt dessinée par Adrien Berger et Jacques de Paris gravée par Bonquet, la Lettre par Bonaldi, imprimée par Langlois.

FIGURE 3.21 Alexander von Humboldt's *Tableau physique des Andes et pays voisins*, published first in 1807, was far more than just a vividly descriptive portrait but a landmark in visualization and conceptualization of the interrelationships among geography, climate and soils, and vegetation. This was the centerpiece illustration of his seminal book, *Essai sur la Géographie des Plantes*. Humboldt's *Essai* demonstrated to generations of scientists, including Charles Darwin and Alfred Russel Wallace, how biogeography, ecology, and evolution could be advanced through meticulous observations complemented by integrative, conceptual explanations based on the interplay of causal processes. Labels within the central plate delineate distributions of particular plant species (along the central and right-hand slopes of the mountain), and labels against the atmospheric background reference key elevations, including those of the world's tallest mountains and the highest elevations reached by climbers and balloonists at that time (translations provide here for five of these). Humboldt's *Tableau* also included detailed side panels with columns in tables describing elevational gradients in meters; light refraction; distance at which mountains are visible from the sea; elevations measured in various parts of the Earth; electrical phenomena; cultivation of the soil; decrease in gravity; azure color of the sky; decrease in humidity; atmospheric air pressure; scale in toises (a measure approximately equal to 6 feet); air temperature; chemical composition of the atmosphere; elevation of the lowest limit of perpetual snow at various latitudes; scale of animals according to where they live; temperature at which water boils at various altitudes; geological aspects; intensity of light at various elevations. (For a full translation of the *Tableau* and Humboldt's entire *Essai*, see Jackson and Romanowski 2009.)

Humboldt's *Tableau Physique*

Unlike William Smith, Alexander von Humboldt (1769–1859) was born into wealth and, to his credit, used it well to unselfishly support the scientific and intellectual endeavors of many of his colleagues as well as his own explorations and creations. As we observed in the previous chapter, his masterful works—including his multivolumed *Cosmos*, the narrative of his travels through the equinoctial (tropical) regions of America, and his *Essai sur la Géographie des Plantes*—inspired the generation of scientists that were to establish the fields of biogeography, ecology, and evolution. The centerpiece illustration in Humboldt's 1807 *Essai* (the *Tableau physique des Andes et pays voisins*) is deservedly recognized as one of the hallmarks of visualization and conceptualization in science.

Like William Smith's map, Humboldt's *Tableau* is a colorful illustration of the geographic variation of nature. The *Tableau*, however, is much more conceptualized and abstract—purposely distorting, exaggerating, or altering the juxtaposition of landforms and associated soil types, climates, plants, and animals in order to emphasize the interdependence among these components of the natural world. As Stephen Jackson observes, the *Tableau* is a comprehensive and compelling visualization that illustrates “the first mature, integrated statement of Humboldt's view of a unified nature” (see Jackson and Romanowski 2009). The *Tableau*'s center illustration along with its tabular side panels (Figure 3.21) describe the regular and interdependent variation in environmental characteristics and dependent life-forms as one ascends from sea level to the summit of what was then recognized as the world's highest peak—Mount Chimborazo, in the equatorial Andes of South America.

Humboldt's *Essai*, in its entirety, was finally translated into English in 2009, making this masterful work once again accessible to a modern audience who should benefit, not just from its historical perspective, but from its ability to serve as an exemplar for advancing today's frontiers of science. The lesson has come full circle, such that modern biogeographers are now adopting a very holistic approach in many ways similar to that utilized by the founding figures of biogeography, from Humboldt and Smith through Darwin and Wallace: meticulous observation and unbiased interpretation of all relevant phenomena, complemented by integrative, conceptual explanations premised on the interplay of causal processes.

MacArthur and Wilson's graphical model

The final in the three exemplars in the history of visualizations in biogeography is both the most recent and also the most abstract, devoid of color and unquestionably the most limited in its artistic aesthetics. Robert H. MacArthur and Edward O. Wilson's graphical model of their equilibrium theory of island biogeography (MacArthur and Wilson 1963, 1967) is not even, at least explicitly, geographic (Figure 3.22). It is not a map, and indeed, their seminal 1967 monograph—which revolutionized the fields of ecology and island biogeography—is notable for its surprising dearth of maps (save for a scale-less map illustrating the process of fragmentation, and a map depicting latitudinal gradients in avian diversity [MacArthur and Wilson 1967: their figures 1 and 37, respectively]).

The abstraction and de-emphasis of geographic context was, of course, purposeful and a central reason for the success of their theory (Losos and Ricklefs 2009). Following the dictum that “beauty is found in simplicity,” MacArthur and Wilson's graphical model became one of the most compelling and versatile visualizations and conceptual tools in biogeography and biodiversity research, in general: one that could easily be adapted to explain a variety of phenomena, including the diversity of biotas from microscopic to

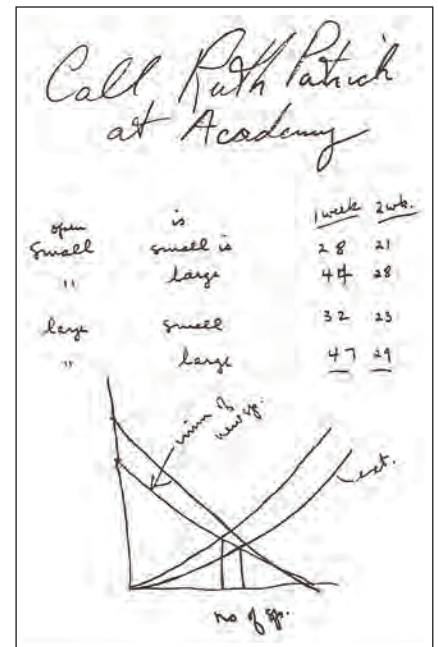


FIGURE 3.22 The graphical equilibrium model, first sketched in a letter from Robert H. MacArthur to Edward O. Wilson in April of 1962, became one of the most compelling features of their theory of island biogeography and was integrally responsible for its rapid acceptance as the new paradigm of the field in the late 1960s. The theory itself and its influence on the development of the field are discussed in Chapter 13 (see also various contributions to the volume on MacArthur and Wilson's theory edited by Losos and Ricklefs 2009). (From the Robert Helmer MacArthur [1930–1972] Papers, Department of Ecology and Evolutionary Biology, Princeton University.)

global scales and from ancient to recent time periods, the extinction dynamics of imperiled species in the face of fragmentation and climate change, and the patterns of variation in linguistic diversity among regional populations of our own species (see Chapter 13).

The spartan nature of MacArthur and Wilson's approach to advancing science was, we believe, a sign of the times: when simplification and reductionism influenced many fields of science and their practitioners' quests for truly transformative approaches to understanding the universal laws of nature. Interestingly, as we remark elsewhere throughout this book, it appears that biogeographers are once again adopting more holistic approaches that may—like those of Humboldt, Darwin, and the founding fathers of the field—strike a balance between empirically realistic and theoretically insightful approaches for understanding the geographic and temporal dynamics of life. Contributing in no small way to the enhanced “vision” and integrative powers of these modern biogeographers are the recent waves of technical advances in our abilities to record geographically explicit information, to visualize it in truly innovative ways, and to develop conceptual and predictive models to advance both basic and applied research.

The GIS revolution

One of the most important innovations that contributed to the revitalization of biogeography during the latter decades of the twentieth century was the advent and eventual wide availability of **geographic information systems (GIS)** (Figure 3.23). While they can be used to create some aesthetically compelling maps, they are much more than this. Put simply, a GIS is a system of technologically sophisticated but wonderfully accessible tools for visualizing, modifying, and analyzing patterns among spatially referenced observations. In its applications to biogeography, that information is typically in the form of locations of particular biological features—such as the locations of individuals, populations, species, or entire biotas—and the characteristics of those features and the local (geographically referenced) environmental factors that may underlie biogeographic patterns of interest.

There now exists an impressive variety of GIS software, but all of these utilize one or a combination of two georeferencing platforms. In **raster-based GIS**, the georeference platform is a system of cells (typically rectangular, but sometimes hexagonal) that tessellate to form a virtual grid work of geographic units representing the area to be mapped and analyzed. Because each cell has a unique identity, it can be attributed with particular biological,

geographic, and environmental features (e.g., a list of species occurring within that location, its elevation, and its climate and soils). The alternative georeferencing platform is a **vector-based GIS**, where each feature or attribute is assigned to a precise and exact coordinate location (most commonly, latitude and longitude, or the Northing and Easting of the Universal Transverse Mercator system described above), or it is assigned to vectors and polygons created by interconnecting lines

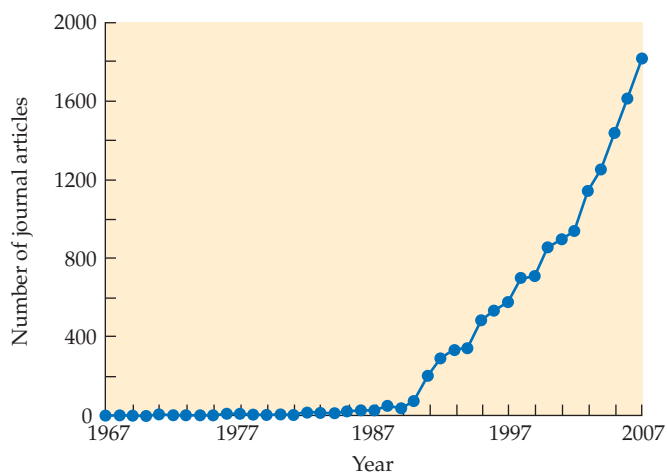


FIGURE 3.23 Since the first general accessibility to geographic information systems (GIS) in the 1980s, their use by scientists has continued to accelerate. Shown here is the number of journal articles with either geographic information system or GIS as keywords. (After Sastre et al. 2009.)

or arcs. Because each system has its advantages (e.g., raster-based GIS's ease of computation and representation of continuous features, and vector-based GIS's relationship to traditional cartography, precision, and ability to represent discontinuous features), modern GIS software programs are sometimes hybrids of these complementary platforms.

This is a highly dynamic and innovative area of technological development, such that any more detailed description—even if it were within the scope of this very general introduction—would likely be outdated before the ink on this page dried. We strongly urge those genuinely serious about studying the geographic variation in nature to add GIS to their tool kits through courses and self-directed studies of more comprehensive accounts of the strategies for utilizing GIS in biogeography (e.g., Burrough 2001; Bernhardsen 2002; Millington et al. 2002; Vogiatzakis 2003; McMaster and Usery 2005; Anselin et al. 2006; Stigall and Lieberman 2006; Harvey 2008; Kidd and Liu 2008).

Cartograms and strategic “distortions”

While it may seem counter-intuitive, one of the most informative means of visualizing geographic variation is through purposeful distortion (e.g., vertical exaggeration of elevations in three-dimensional maps are relatively simple cases of purposeful distortion that have been used in cartography for centuries). This again emphasizes our earlier assertion that maps are far more than just dutiful works of accuracy and precision; they are potentially powerful tools for conceptualizing the interplay among a variety of spatial patterns and their underlying, causal forces. **Cartograms** are examples of such strategic distortion where mapping units (particular grid cells or polygons such as those representing countries or biogeographic regions) are scaled (and distorted), not according to their actual surface area, but in proportion to another theme, such as population density or species diversity.

One implicit, but important, consideration is that the cartogram be paired up with a reference map with geographic units drawn to more standard projections (i.e., equal area projections), or in proportion to another theme to be compared with the first.

The utility and versatility in potential applications of cartograms are best appreciated by example. In an illustrative case study, Wake and Vredenburg (2008) use cartograms to demonstrate the highly uneven diversity and endangerment of amphibians across the globe, along with the geographic variation in rates of discovering new species (**Figure 3.24**). In these examples, the highly uneven nature of these variables (i.e., highly disproportionate to the areas of the countries) is also emphasized by monochrome shading and use of alternative colors. The result is a series of visualizations that, although at first somewhat otherworldly in appearance, illustrate geographic variation in biodiversity more clearly than could be done using conventional maps.

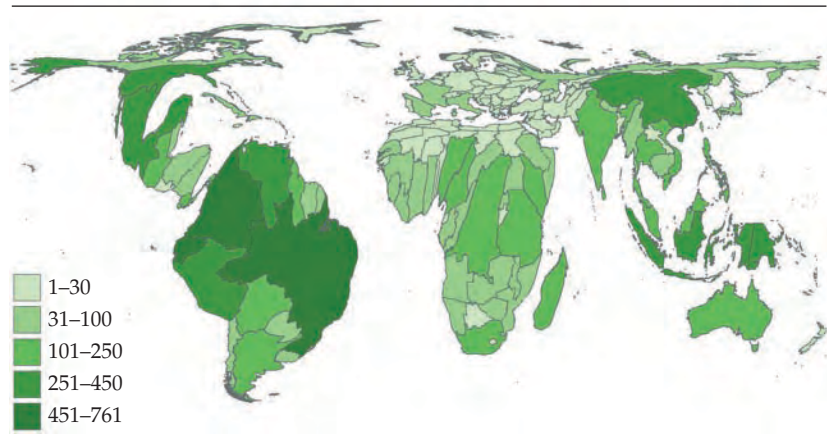
Obtaining Geo-Referenced Data

Humboldt's legacy: A global system of observatories

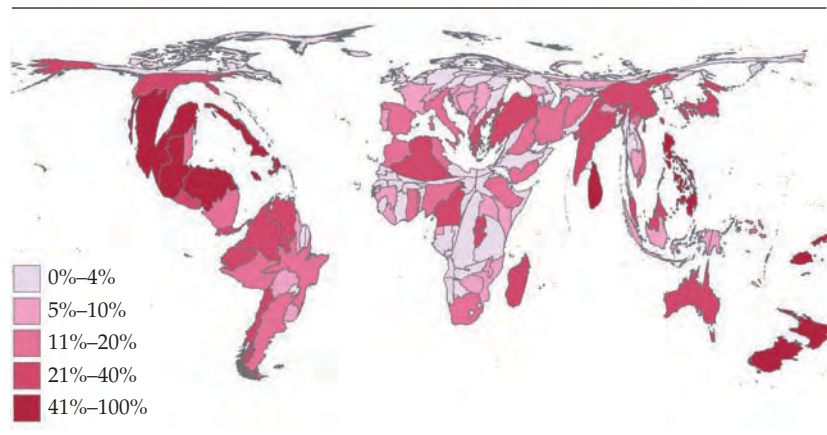
We have one remaining tribute to Alexander von Humboldt (see Chapter 2), for his remarkable legacy also includes a prescient call in 1836 for the establishment of a global system of climatological and geomagnetic observations. This would not only provide the data requisite to generations of scientists who pursued Humboldt's vision of the unity of nature, in its broadest sense, but ultimately would provide invaluable information on the temporal dynam-

FIGURE 3.24 Cartograms utilize special projections to purposefully distort geographic units to emphasize geographic variation in characteristics (themes) of those units. Here, Wake and Vredenburg (2008) utilize cartograms to visualize the geographic variation in (A) diversity (number of species), (B) endangerment (percentage of fauna in the top three categories of threat—critically endangered, endangered, and threatened), and (C) discovery rates (number of species discovered and named during the period of 2004–2007) of amphibians (per area of each country) across the globe. (For a description of the actual projection [density-equalization] methods used to develop these cartograms, see Gastner and Newman 2004.)

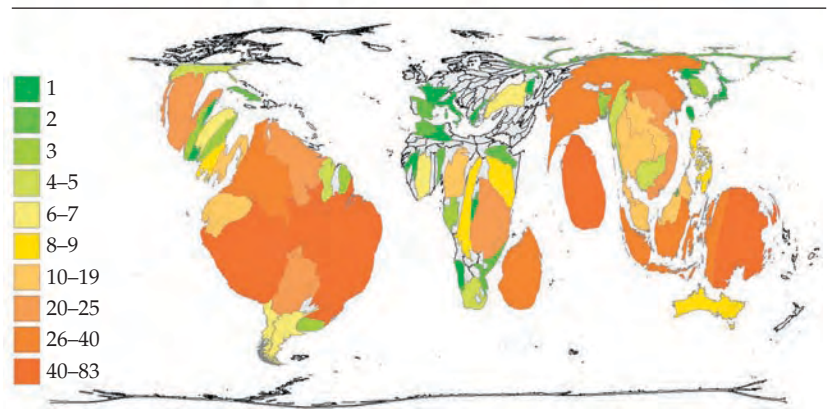
(A) Diversity



(B) Endangerment



(C) Discovery rates



ics of a diversity of natural and anthropogenic phenomena—not the least of which being ongoing global changes in climates, landscapes, and seascapes.

The current wealth of on-site, globally integrated observatories—for example, Smithsonian Institution Global Earth Observatories (SIGEO; www.sigeo.si.edu/), Ocean Observatories Initiative (OOI; www.oceanleadership.org/programs-and-partnerships/ocean-observing/ooi), and Terrestrial En-

vironmental Observatory (TERENO; www.tereno.net/)—provide critical information to biogeographers using GIS and spatial analyses to investigate the geographic dynamics of biotas and their underlying, environmental drivers. Of course, it is essential that these in situ measurements follow standardized protocols and that resultant databases include **metadata**—detailed descriptions of the methods used to record and store data and the relevant characteristics of those data. Contemporary analyses in biogeography, however, increasingly utilize other systems of complementary databases generated by off-site technology—in particular, via continually advancing innovations in remote sensing.

Remote sensing and satellite imagery

Remote sensing refers to any means of collecting data where the recorder or sensing device is not in direct contact with the area or objects of interest. In biogeographic applications, this typically includes the use of technologically sophisticated sensing devices operated from distant platforms, including aircraft, ships, and satellites. As a result, these systems can generate massive data sets covering areas that are much more extensive and often more continuous in coverage than even the most extensive in situ global observatories, which are by definition point-based and discontinuous.

Modern remote sensing systems are capable of recording an incredible variety of characteristics about focal biota, landscapes, and seascapes, depending on the particular types of electromagnetic radiation (typically ranging through radio, microwave, infrared, visible, and ultraviolet bands) being emitted or reflected from those targets (see overview by Gillespie et al. 2008). Once detected, the sensory information must first be processed, which typically requires using complex systems of mathematical models to interpret the spectral (wavelength) information in terms of relevant features of the biota and regions under study (e.g., the locations, temperatures, or activity states of organisms; the nature of relatively deep geological formations and the bathymetric features of the oceans; the depth and volume of ice sheets; or the rates of productivity of entire ecosystems; see Figure 5.30A).

Interpolation over space and time

As we indicated above, although in situ observations and, for that matter, many forms of remote sensing yield point-based (discontinuous) data, biogeographers often require information that covers a broader span of sites and areas, including those between locations of existing observations. Similarly, information from these sources, especially very remote sensing programs such as the Landsat and Envisat satellite programs, generate images and other information that are limited in temporal extent—often recorded during a time period far removed from that of other data to be used or the period most relevant to intended applications of that imagery.

Fortunately, we can solve this problem of observational discordance by creating more continuous coverages and estimates of values at unmeasured points in space and time using the method of **interpolation**. Spatial and temporal interpolations are equivalent procedures that provide estimates of the expected value of a variable at an unmeasured point in space (or time), based on statistical models that take into account the values of recorded variables at actual observation sites, and their distances to the site whose value is to be interpolated. These estimates can then be weighted such that, rather than being the simple average of all measured values in the vicinity, the estimated value is more strongly influenced by values of sites closer to the site to be

interpolated. The result is tremendous variety of spatially and temporally concordant data (measured or derived from interpolations) that can then be subjected to a powerful battery of analytical tools now available to contemporary biogeographers.

Analysis of Biogeographic Patterns

Comprehensive coverage, even at the level of an introductory primer of all of the modern approaches for analyzing geographic patterns, would require many volumes on a variety of topics in geo-statistics (i.e., spatial statistics; see Webster and Oliver 2001; Fortin and Dale 2005; Kent et al. 2006; Rangel et al. 2006; Webster and Oliver 2007; Bivand et al. 2008). It is important, however, that we introduce some of the fundamental considerations and concepts that are especially relevant to, and sometimes confound, analyses of spatial patterns.

As with any statistical analyses, biogeographers must carefully consider the special nature of their data prior to evaluating the significance of any empirical patterns. Because of the extensive spatial and temporal scales inherent in most biogeographic studies, their research approach is often comparative and opportunistic. As a result, requisite data are typically collected by different scientists, often utilizing a variety of sampling protocols across different locations and time periods. As we described above, the latter two challenges can be solved by interpolation of values over space and time, respectively. Interpolations, however, are not observations but derived estimates—thus introducing another level of error to spatial analyses. Of more fundamental concern to nearly all forms of spatial analyses is that geographically explicit data, whether interpolated or genuine observational data, often violate a fundamental assumption of almost all statistical analyses—*independence of observations*. Ironically, this tendency for points closer in space or time to be more similar (**spatial** and **temporal autocorrelation**) can be viewed as an asset or an actual pattern of interest, as well as a potentially confounding problem (Fortin et al. 1989; Legendre 1993). That is, autocorrelation of adjacent observations is why interpolations work, and indeed the primary objective of many scientific investigations (e.g., those in landscape ecology) is determining the spatial extent of this interdependence (or the distance beyond which observations and relevant phenomena become independent). On the other hand, multiple data points taken within the critical windows of space and time where observations are not independent may repeatedly sample the same observations or phenomena—a problem termed **pseudoreplication** (inflation of the functional sample size). Fortunately, geo-statistical tool kits include methods both for estimating the extents of empirical autocorrelation in space and time and for adjusting statistical analyses to control for these potential problems (e.g., by adjusting inferred probabilities of statistical tests to those based on a smaller number [N] of independent observations, or by adjusting the values of data to be analyzed based on empirical trends in autocorrelation over space and time).

These and other recent advances in spatial analyses and remote sensing, combined with what are sometimes centuries' worth of empirical information on species distributions and their ecological associations, have enabled biogeographers to rigorously explore a wonderful diversity of patterns (see Turner et al. 2003; Loarie et al. 2007; Kark et al. 2008). One of the most fundamental yet challenging of these applications is estimation of what can be viewed as the fundamental unit of biogeography—the geographic range (see Chapter 4). Regardless of how ranges are represented (i.e., as outline,

dot, or contour maps; see Figures 4.5–4.8), it is important to realize that all distributional maps are *estimates* of the actual distributions. Fortunately, although biogeographers will never reach perfect consensus on how best to estimate species distributions, alternative approaches share some key features in common: They utilize both empirical data (e.g., on recorded occurrences of the focal species) and remotely acquired and processed information (on soils, climates, or cover of habitats); they employ sophisticated GIS and statistical software to estimate environmental associations of the focal species; they use this information to develop models predicting distributions over a particular area and time period; they test the efficacy of those predictive models using empirical data not utilized in previous steps; and they modify the predictive models until some target efficacy is achieved (see pages XX–XX in Chapter 4).

Finally, it would be misleading for us to imply that the solutions to the challenges of rigorously exploring and assessing patterns in the geography of nature—in features of individuals, ranges of species, or diversity of entire biotas—are to be solved simply by advances in technology. Granted, these advances have provided and will continue to provide us with abilities to visualize and investigate biogeographic phenomena in ways unimagined, even by the most visionary founders of the field, including Humboldt, Darwin, and Wallace. Yet we continue to wrestle with one of the most fundamental questions in biogeography: What actually limits the range of a species? This question is the focal topic of the next chapter, and one relevant to nearly all patterns in the geography of nature.

