

Art, said Picasso, is a lie that makes us realize the truth. So is a map. We don't usually associate the precise craft of the mapmaker with the fanciful realm of art. Yet a map has many of the ingredients of a painting or a poem. It is truth compressed in a symbolic way, holding meanings that it doesn't express on the surface. And, like any work of art, it requires imaginative reading.

Map reading involves determining what the mapmaker has depicted, how he or she has gone about it, and what artifacts of the cartographic method deserve special attention. To read a map, you translate its features into a mental image of the environment. The first step is to identify map symbols. The process is usually intuitive, especially if the symbols are self-evident and the map is well designed. As obvious as this step might seem, however, you should look first at the map and its marginalia (legend, text boxes, and so on), both to confirm the meaning of familiar symbols and to make sure that you understand the logic that underlies unfamiliar or poorly designed ones. This explanation is usually found in the legend, but sometimes explanatory labels are found on the map itself or in text boxes that provide the missing information. Too many people look for the symbol explanation only after becoming confused. Such a habit is not only inefficient but potentially dangerous. A better approach is to first check the marginalia, and then attempt to interpret the map.

In addition to clarifying symbols, the map marginalia contains other information, such as scale, orientation, and data sources that are important to reading the map; and the marginalia sometimes includes unexpectedly revealing facts. But the marginalia is still only a starting point. The map reader must make a creative effort to translate the world that is represented on the map into

Part I

Map reading

an image of the real world, for there is often a large gap between the two. Much of what exists in the environment is left off the map, whereas many things on the map do not occur in reality but are instead interpretations of characteristics of the environment, such as population density or average streamflow.

Thus, map and reality are not—and cannot be—identical. No aspect of map use is so obvious and yet so often overlooked. Most map reading mistakes occur because the user forgets this vital fact and expects a one-to-one correspondence between the map and reality.

Because the exact duplication of a geographic area is impossible, a map is actually a metaphor. The mapmaker asks the map reader to believe that an arrangement of points, lines, and areas on a flat sheet of paper or a computer screen is equivalent to some facet of the real world in space and time. To gain a fuller understanding, the map reader must go beyond the graphic representation and study carefully what the symbols refer to in the real world.

A map, like a painting, is just one interpreter's version of reality. To understand a painting, you must have some idea of the techniques used by the interpreter—that is, the artist. You don't expect a watercolor to look anything like an acrylic painting or a charcoal drawing, even if the subject matter of all three is identical, because the three media are different. Therefore, the artistic techniques are varied. In the same way, the techniques used to create maps greatly influence the final portrayal. As a map reader, you must always be aware of the mapmaker's invisible hand. Never use a map without asking yourself how it is biased by the methods that were used to make it.

If the mapping process operates at its full potential, communication of environmental information takes place

between the cartographer and the map reader. The mapmaker translates reality into the clearest possible picture that the map can give, and the map reader converts this picture back into a useful mental image of the environment. For such translation to occur, the map reader must know something about how maps are created.

The complexities of mapping are easier to study if we break them up into simpler parts. Thus, we have divided part 1, "Map Reading," into 11 chapters, each dealing with a different aspect of mapping. Chapter 1 examines geographic coordinate systems for the earth as a sphere, an oblate ellipsoid, and a geoid. Chapter 2 looks at ways of expressing and determining map scale. Chapter 3 introduces different map projections and the types of geometric distortions that occur with each projection. Chapter 4 focuses on different grid coordinate systems used in maps. Chapter 5 looks at land partitioning systems and how they are mapped. Chapter 6 is an introduction to how maps are designed. Chapter 7 examines various methods for mapping qualitative information, and chapter 8 does the same for quantitative information. Chapter 9 is devoted to the different methods of relief portrayal found on maps. Chapter 10 is an overview of image maps. Finally, for part 1, chapter 11 explores various aspects of map accuracy.

These 11 chapters should give you an appreciation of all that goes into mapping and the ways that different aspects of the environment are shown in maps. As a result, you'll better understand the large and varied amount of geographic information that you can glean from a map. In addition, once you realize how intricate the mapping process is, you won't be able to help but view even the crudest map with more respect, and your map reading skill will naturally grow.

chapter
one

THE EARTH AND EARTH COORDINATES

THE EARTH AS A SPHERE

THE GRATICULE

- Parallels and meridians
- Latitude and longitude
- Prime meridians

THE EARTH AS AN OBLATE ELLIPSOID

- Different ellipsoids
- Geodetic latitude
- Geodetic longitude

DETERMINING GEODETIC LATITUDE AND LONGITUDE

PROPERTIES OF THE GRATICULE

- Circumference of the authalic sphere
- Spacing of parallels
- Converging meridians
- Quadrilaterals
- Great and small circles

GRATICULE APPEARANCE ON MAPS

- Small-scale maps
- Large-scale maps

GEODETIC LATITUDE AND LONGITUDE ON LARGE-SCALE MAPS

- Horizontal reference datums

THE EARTH AS A GEOID

- Vertical reference datums

SELECTED READINGS

1

The earth and earth coordinates

Of all the jobs that maps do for you, one stands out—they tell you where things are and allow you to communicate this information efficiently to others. This, more than any other factor, accounts for the widespread use of maps. Maps give you a superb **positional reference system**—a way to pinpoint the locations of things in space.

There are many ways to determine the position of a feature shown on a map. All begin with defining a geometric figure that approximates the true shape and size of the earth. This figure is either a **sphere** (a three-dimensional solid in which all points on the surface are the same distance from the center) or an **oblate ellipsoid** (a slightly flattened sphere), or an **ellipsoid** for short, of precisely known dimensions. Once the dimensions of the sphere or ellipsoid are defined, a **graticule** of east–west lines called **parallels** and north–south lines called **meridians** is draped over the sphere or ellipsoid. The angular distance of a parallel from the equator and of a meridian from what we call the **prime meridian** (the zero meridian used as the reference line from which east and west are measured) gives us the latitude and longitude coordinates of a feature. The locations of elevations measured relative to an average gravity or sea level surface called the **geoid** can then be defined by three-dimensional (latitude, longitude, elevation) coordinates. Next, we look at these concepts in more detail.

THE EARTH AS A SPHERE

We have known for over 2,000 years that the earth is spherical in shape. We owe this knowledge to several ancient Greek philosophers, particularly **Aristotle** (fourth century BC), who believed that the earth's sphericity could be proven by careful visual observation. Aristotle noticed that as he moved north or south, the stars were not stationary—new stars appeared in the northern horizon while familiar stars disappeared to the south. He reasoned that this could occur only if the earth was curved north to south. He also observed that departing sailing ships, regardless of their direction of travel, always disappeared from view by their hull (bottom and sides) first. If the earth was flat, the ships would simply get smaller as they sailed away. Only on a sphere do hulls always disappear first. His third observation was that a circular shadow is always cast by the earth on the moon during a lunar eclipse, something that occurs only if the earth is spherical. These arguments entered the Greek literature and persuaded scholars over the succeeding centuries that the earth must be spherical in shape.

Determining the size of our spherical earth was a daunting task for our ancestors. The Greek scholar **Eratosthenes**, head of the then-famous library and museum in Alexandria, Egypt, around 250 BC, made the first scientifically based

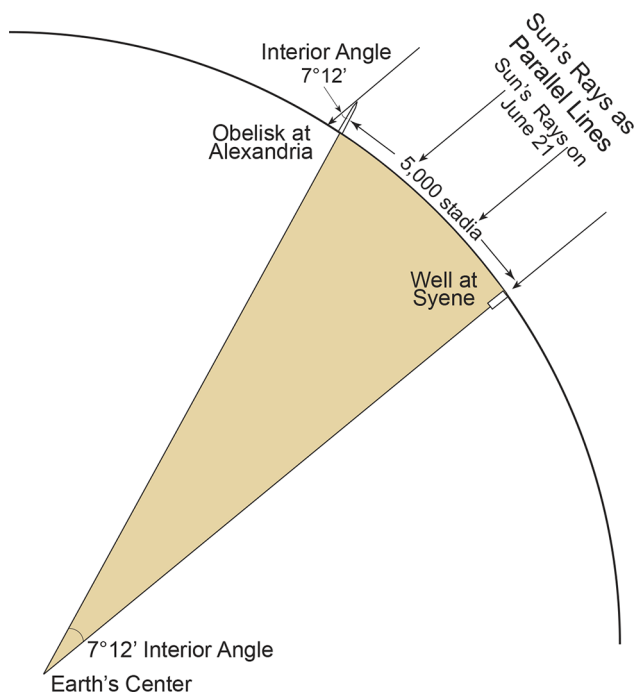


Figure 1.1. Eratosthenes' method for measuring the earth's circumference.

estimate of the earth's circumference. The story that has come down to us is of Eratosthenes reading an account of a deep well at Syene near modern Aswan, about 500 miles (800 kilometers) south of Alexandria. The well's bottom was illuminated by the sun only on June 21, the day of the summer solstice. He concluded that the sun must be directly overhead on this day, with rays perpendicular to the level ground (figure 1.1). Then he reasoned brilliantly that if the sun's rays are parallel and the earth is spherical, a vertical column such as an obelisk should cast a shadow in Alexandria on the same day. Knowing the angle of the shadow would allow the earth's circumference to be measured if the north–south distance from Alexandria to Syene could be determined. The simple geometry involved here is that if two parallel lines are intersected by a third line, the alternate interior angles are equal. From this supposition, he reasoned that the shadow angle at Alexandria equals the angular difference ($7^{\circ}12'$) at the earth's center between the two places.

The story continues that on the next summer solstice, Eratosthenes measured the shadow angle from an obelisk in Alexandria, finding it to be $7^{\circ}12'$, or 1/50th of a circle. Hence, the distance between Alexandria and Syene is 1/50th of the earth's circumference. He was told that Syene must be about 5,000 stadia south of Alexandria because camel caravans traveling at 100 stadia per day took 50 days to make the trip between the two cities. From this distance estimate, he computed the earth's circumference as $50 \times 5,000$ stadia, or 250,000 stadia. A **stadium** is an ancient Greek unit of measurement based on the length of a sports stadium at the time. A stadium varied from 200 to 210 modern yards (182 to 192 meters), so his computed circumference was somewhere between 28,400 and 29,800 modern statute miles (45,700 and 47,960 kilometers), 14 percent to 19 percent greater than the currently accepted circumference distance of 24,874 statute miles (40,030 kilometers).

We now know that Eratosthenes' error was because of an underestimate of the distance between Alexandria and Syene, and because the two cities are not exactly north–south of each other. However, his method is sound mathematically and was the best circumference measurement until the 1600s. Equally important, Eratosthenes had the idea that careful observations of the sun allowed him to determine angular differences between places on earth, an idea that you shall see was expanded to other stars and recently to the **Global Navigation Satellite System (GNSS)**, commonly known as the **Global Positioning System (GPS)**. GPS is a “constellation” of earth-orbiting satellites that make it

possible for people to pinpoint geographic location and elevation with a high degree of accuracy. GPS uses ground receivers that digitally process electronic signals from the satellite constellation to provide the position and exact time at a location (see chapter 14 for more on GPS).

THE GRATICULE

Once the shape and size of the earth were known, map-makers required a system for defining locations on the surface. We are again indebted to ancient Greek scholars for devising a system for placing reference lines on the spherical earth.

Parallels and meridians

Astronomers before Eratosthenes placed horizontal lines on maps to mark the **equator** (forming the circle around the earth that is equidistant from the North and South Poles) and the **Tropics of Cancer** and **Capricorn** (lines that mark the northernmost and southernmost positions in which the sun is directly overhead on the summer and winter solstices, respectively. (As we mentioned earlier, Syene is located almost on the Tropic of Cancer.) Later, the astronomer and mathematician **Hipparchus** (190–125 BC) proposed that a set of equally spaced east–west lines called *parallels* be drawn on maps (figure 1.2). To these

lines, he added a set of north–south lines called *meridians* that are equally spaced at the equator and converge at the North and South Poles. We now call this arrangement of parallels and meridians the *graticule*. Hipparchus’s numbering system for parallels and meridians was, and still is, called *latitude* and *longitude*. We now look at latitude and longitude on a sphere, sometimes called *geocentric latitude* and *longitude*. Later in this chapter, we explain how geocentric latitude and longitude differ from *geodetic latitude* and *longitude* on an oblate ellipsoid.

Latitude and longitude

Latitude on the spherical earth is the north–south angular distance from the equator to the place of interest (figure 1.3). The numerical range of latitude is from 0° at the equator to 90° at the poles. The letters *N* and *S*, such as 45° N for Fossil, Oregon, are used to indicate north and south latitudes, respectively. Instead of the letter *S*, you may see a minus sign (–) for south latitudes; however, a plus sign (+) is not used for north latitudes.

Longitude is the angle, measured along the equator, between the intersection of the reference meridian, called the *prime meridian*, and the point at which the meridian for the feature of interest intersects the equator. The numeric range of longitude is from 0° to 180° east and west of the prime meridian, twice as long as parallels. East and west longitudes are labeled *E* and *W*, so that Fossil, Oregon, has a longitude of 120° W. As with south latitude, west

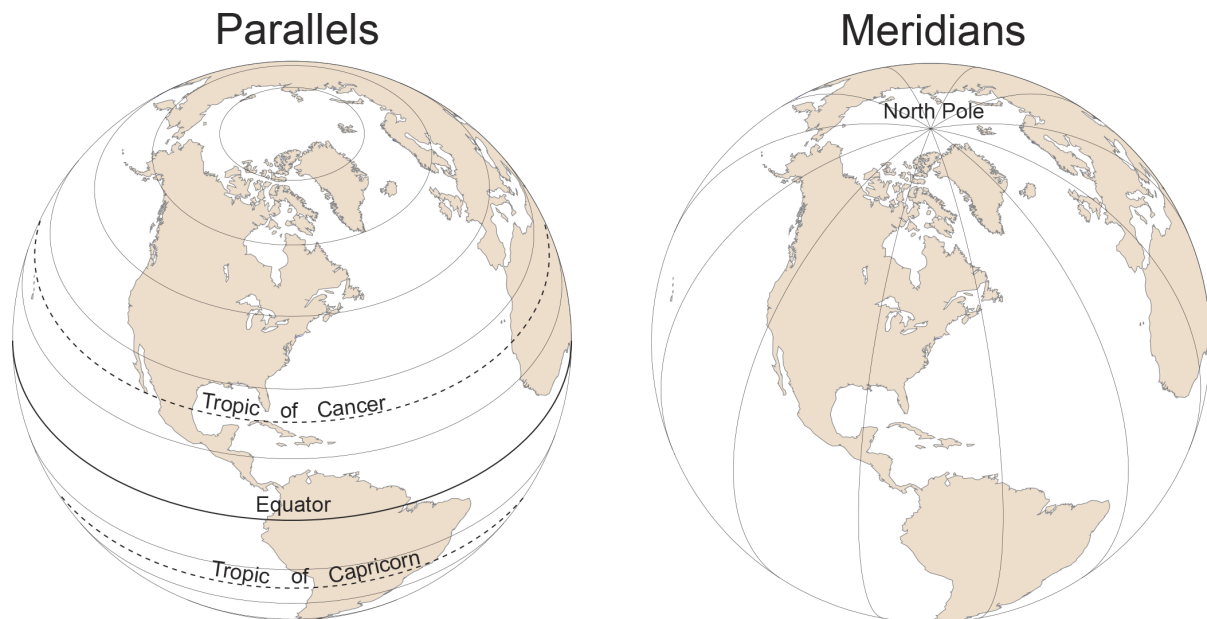


Figure 1.2. Parallels and meridians.

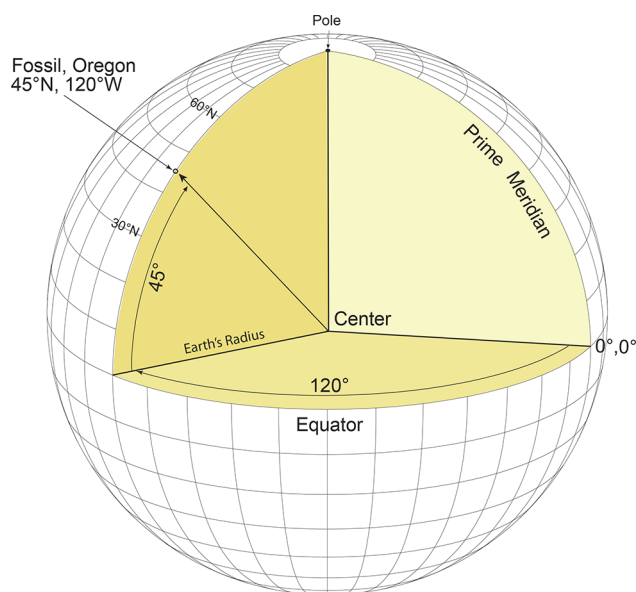


Figure 1.3. Latitude and longitude on the sphere allow explicit identification of the positions of features on the earth.

longitude may also be indicated by a minus sign, but a plus sign is not used for east longitude.

Putting latitude and longitude together into what is called a **geographic coordinate** (such as 45° N, 120° W or 45°, -120°) pinpoints a place on the earth's surface. There are several ways to write latitude and longitude values. The oldest is the **Babylonian sexagesimal system** of degrees (°), minutes (′), and seconds (″), where there are 60 minutes in a degree and 60 seconds in a minute. For example, the latitude and longitude of the capitol dome in Salem, Oregon, is 44°56′18″ N, 123°01′47″ W (or 44°56′18″, -123°01′47″).

Latitude and longitude can also be expressed in **decimal degrees** (DD) through equation (1.1):

$$DD = dd + mm / 60 + ss / 3600, \quad (1.1)$$

where dd is the number of whole degrees, mm is the number of minutes, and ss is the number of seconds. For example, in equation (1.2):

$$44^{\circ}56'18'' = 44 + 56 / 60 + 18 / 3600 = 44.9381^{\circ}. \quad (1.2)$$

Decimal degrees are often rounded to two decimal places, so the location of the Oregon state capitol dome is written in decimal degrees as 44.94, -123.03. If you can accurately define a location to the nearest one second of latitude and longitude, you can specify its location to within 100 feet (30 meters) of its true location on the earth.

Prime meridians

The choice of prime meridian (the 0° meridian used as the reference from which longitude east and west are measured) is entirely arbitrary, because there is no physically definable starting point like the equator. In the fourth century BC, Eratosthenes selected Alexandria, Egypt, as the starting meridian for longitude; and in medieval times, the Canary Islands off the west coast of Africa were used because they were then the westernmost outpost of western civilization. In the 18th and 19th centuries, many countries used their capital city as the prime meridian for their national maps (see appendix B, table B.6 for a listing of historic prime meridians). You can imagine the confusion that must have existed when trying to locate places on maps from other countries. The problem was eliminated in 1884 when the **International Meridian Conference** selected as the international standard the British prime meridian, defined by the north-south optical axis of a telescope at the Royal Observatory in Greenwich, a suburb of London. This standard is called the **Greenwich meridian**.

You may occasionally come across a historical map that uses one of the prime meridians listed in appendix B, at which time knowing the angular difference between the prime meridian used on the map and the Greenwich meridian becomes useful information. As an example, you might see in an old Turkish atlas that the longitude of Seattle, Washington, is 151°16′ W (based on the Istanbul meridian), and you know that the Greenwich longitude of Seattle is 122°17′ W. You can determine the Greenwich longitude of Istanbul through subtraction, in equation (1.3):

$$151^{\circ}16' \text{ W} - 122^{\circ}17' \text{ W} = 28^{\circ}59' \text{ E}. \quad (1.3)$$

The computation is done more easily in decimal degrees, as described earlier, in equation (1.4):

$$\begin{aligned} 122.17^{\circ} \text{ W} - 151.16^{\circ} \text{ W} &= -122.17^{\circ} - -151.16^{\circ} \\ &= 28.99^{\circ}, \text{ or } 28.99^{\circ} \text{ E}. \end{aligned} \quad (1.4)$$

THE EARTH AS AN OBLATE ELLIPSOID

Scholars assumed that the earth was a perfect sphere until the 1660s, when **Sir Isaac Newton** developed the theory of gravity. Newton thought that gravity should produce a perfectly spherical earth if it was not rotating about its **polar axis**. The earth's 24-hour rotation, however, introduces outward **centrifugal forces** that are perpendicular to the axis of rotation, or the **equatorial axis**. The amount of force varies from zero at each pole to a maximum at the equator.

Newton noted that these outward centrifugal forces counteract the inward pull of gravity, so the net inward force decreases progressively from the pole to the equator causing the earth to be slightly flattened. Slicing the earth in half from pole to pole would then reveal an ellipse with a slightly longer equatorial radius and slightly shorter polar radius; we call these radii the **semimajor** and **semiminor axes**, respectively (figure 1.4). If we rotate this ellipse 180° about its polar axis, we obtain a three-dimensional solid that we call an **oblate ellipsoid**.

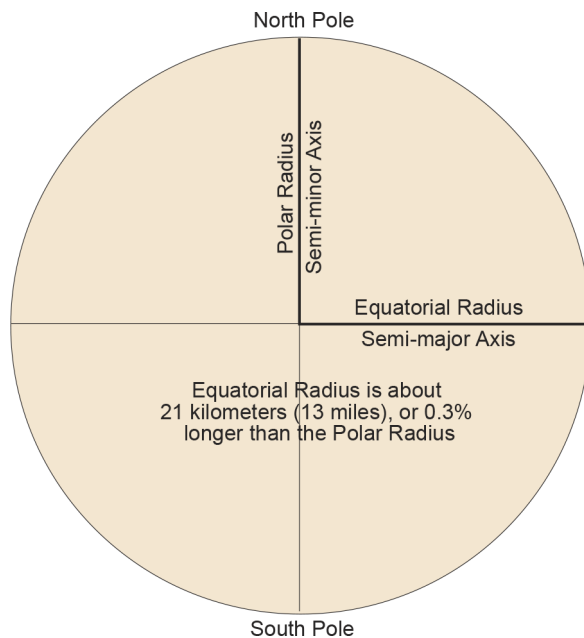


Figure 1.4. The form of the oblate ellipsoid was determined by measurements of degrees at different latitudes, beginning in the 1730s. Its equatorial radius was about 13 miles (21 kilometers) longer than its polar radius. This figure is true to scale, but our eye cannot see the flattening because the difference between the north–south and east–west axis is so small.

The oblate ellipsoid is important because parallels are not spaced equally as they are on a sphere but decrease slightly in spacing from the pole to the equator. This variation in the spacing of parallels is shown in figure 1.5, a cross section of a greatly flattened oblate ellipsoid. We say that on an oblate ellipsoid the **radius of curvature** (the measure of how curved the surface is) is largest at the pole and smallest at the equator. Near the pole, the ellipse is flatter than near the equator. You can see that there is less curvature on the ellipse with a larger radius of curvature and more curvature with a smaller radius of curvature.

The north–south distance between two points on the ellipsoidal surface equals the radius of curvature times the angular difference between the points. Because the radius of curvature on an oblate ellipsoid is largest at the pole and smallest at the equator, the north–south distance between points that are a degree apart in latitude should be greater near the pole than at the equator. In the 1730s, scientific expeditions to Ecuador and Finland measured the length of a degree of latitude at the equator and near the Arctic Circle, proving Newton to be correct. These and additional measurements in the following decades for other parts of the world allowed the semimajor and semiminor axes of the oblate ellipsoid to be computed by the early 1800s, giving about a 13-mile (21-kilometer) difference between the two, only one-third of 1 percent.

Different ellipsoids

During the 19th century, better surveying equipment was used to measure the length of a degree of latitude on different continents. From these measurements, slightly

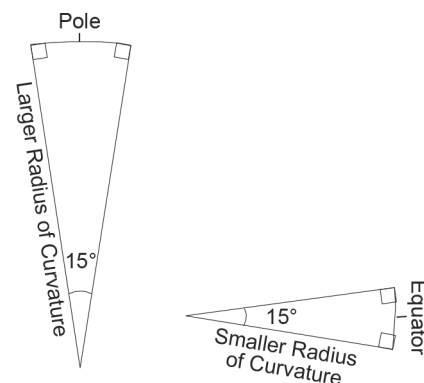


Figure 1.5. This north–south cross section through the center of a greatly flattened oblate ellipsoid shows that a larger radius of curvature at the pole results in a larger ground distance per degree of latitude relative to the equator.

Table 1.1 Historical and current oblate ellipsoids

Name	Date	Equatorial radius (km)	Polar radius (km)	Areas of use
WGS84	1984	6,378.137	6,356.75231	Worldwide
GRS80	1980	6,378.137	6,356.7523	Worldwide (NAD83)
Australian	1965	6,378.160	6,356.7747	Australia
Krasovsky	1940	6,378.245	6,356.863	Soviet Union
International	1924	6,378.388	6,356.9119	Remainder of the world not covered by older ellipsoids (European Datum 1950)
Clarke	1880	6,378.2491	6,356.5149	France; most of Africa
Clarke	1866	6,378.2064	6,356.5838	North America (NAD27)
Bessel	1841	6,377.3972	6,356.079	Central Europe, Sweden, Chile, Switzerland, Indonesia
Airy	1830	6,377.5634	6,356.2569	Great Britain, Ireland
Everest	1830	6,377.2763	6,356.0754	India and the rest of South Asia

different oblate ellipsoids, varying by only a few hundred meters in axis length, best fit the measurements. Table 1.1 is a list of these ellipsoids, along with their areas of usage. Note the changes in ellipsoid use over time. For example, the **Clarke 1866 ellipsoid** was the best fit for North America in the 19th century and was used as the basis for latitude and longitude on topographic and other maps produced in Canada, Mexico, and the United States from the late 1800s to about the late 1970s. By the 1980s, vastly superior surveying equipment, coupled with millions of observations of satellite orbits, allowed us to determine oblate ellipsoids that are excellent average fits for the entire earth. Satellite data is important because the elliptical shape of each orbit monitored at ground receiving stations mirrors the earth's shape. The most recent of these ellipsoids, called the **World Geodetic System of 1984 (WGS84)**, replaced the Clarke 1866 ellipsoid in North America and is used as the basis for latitude and longitude on maps throughout the world. In table 1.1, the WGS84 ellipsoid has an equatorial radius of 6,378.137 kilometers (3,963.191 miles) and a polar radius of 6,356.752 kilometers (3,949.903 miles). On this ellipsoid, the distance between two points that are one degree apart in latitude between 0° and 1° at the equator is 110.567 kilometers (68.703 miles), shorter than the 111.699-kilometer (69.407-mile) distance between two points at 89° and 90° north latitude. Hence, on equatorial- and polar-area maps of the same scale, the spacing between parallels is not the same, but the difference is small.

Geodetic latitude

Geodetic latitude is defined as the angle made by the horizontal equator line and a line perpendicular to the ellipsoidal surface at the parallel of interest (figure 1.6). Geodetic latitude differs from latitude on a sphere because of the unequal spacing of parallels on the ellipsoid. Lines perpendicular to the ellipsoidal surface pass through the center of the earth only at the poles and the equator, but all lines perpendicular to the surface of a sphere pass through its center. This centricity is why the latitude defined by these lines on a sphere is called **geocentric latitude**.

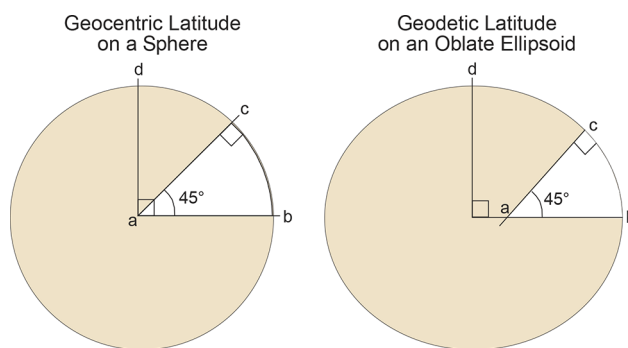


Figure 1.6. Geocentric and geodetic latitudes of 45°. On a sphere, circular arc distance b–c is the same as circular arc distance c–d. On the greatly flattened oblate ellipsoid, elliptical arc distance b–c is less than elliptical arc distance c–d. On the WGS84 oblate ellipsoid, arc distance b–c is 4,984.94 kilometers (3,097.50 miles) and arc distance c–d is 5,017.02 kilometers (3,117.43 miles), a difference of about 32 kilometers (20 miles).

Geodetic longitude

There is no need to make a distinction between **geocentric longitude** on the sphere and **geodetic longitude** on the ellipsoid. Both types of longitude are defined as the angular distance between the prime meridian and another meridian passing through a point on the earth's surface, with the center of the earth as the vertex of the angle. Because of the geometric nature of the ellipsoid, the angular distance turns out to be the same as for geocentric longitude on the sphere.

DETERMINING GEODETIC LATITUDE AND LONGITUDE

The oldest way to determine geodetic latitude is with instruments for observing the positions of celestial bodies. The essence of the technique is to establish celestial lines of position (east–west and north–south) by comparing the predicted positions of celestial bodies with their observed positions. A handheld instrument, called a **sextant**, was the tool historically used to measure the angle (or altitude) of a celestial body above the earth's horizon (figure 1.7). Before GPS, it was the tool that nautical navigators used to find their way using the moon, planets, and stars, including the sun.

Astronomers study and tabulate information on the actual motion of the celestial bodies that help pinpoint latitude. Because the earth rotates on an axis defined by the North and South Poles, stars in the Northern Hemisphere's night sky appear to move slowly in a circle centered on

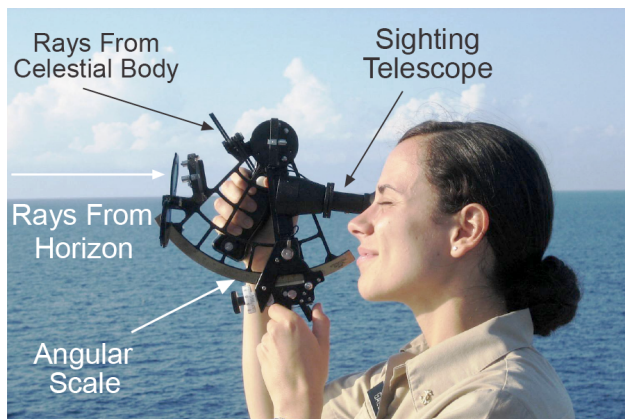


Figure 1.7. A sextant is used at sea to find latitude from the vertical angle between the horizon and a celestial body such as the sun, planet, or star. Courtesy of Dr. Bernie Bernard.

Polaris (the **North Star**), which lies almost directly above the North Pole. A navigator needs only to locate Polaris to find the approximate North Polar axis. In addition, because the star is so far away from the earth, the angle from the horizon to Polaris is approximately the same as the latitude (figure 1.8).

However, Polaris is actually 0.75 degrees from directly above the North Pole, circling the pole in a small circle that is 1.5 degrees in diameter. Precisely determining the spot near Polaris that is directly above the North Pole requires accurately knowing the time and using Polaris position data from the **Astronomical Almanac**, published jointly by the US Naval Observatory and Her Majesty's Nautical Almanac Office in the United Kingdom. Errors in determining latitude can otherwise approach three-quarters of a degree, which translates into a ground position error of around 50 miles (80 kilometers).

In the Southern Hemisphere, latitude is harder to determine by celestial measurement because there is no equivalent to Polaris directly above the South Pole. Navigators instead use a small constellation called **Crux Australis** (the **Southern Cross**) to serve the same function (figure 1.9). Finding south is more complicated because the Southern Cross is a collection of five stars that are part of the

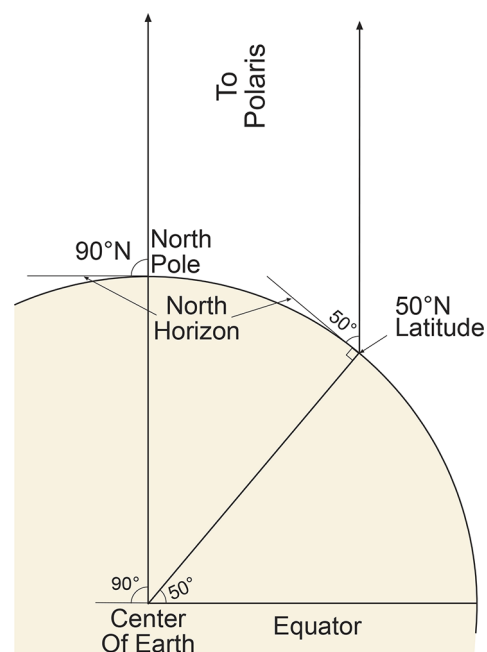


Figure 1.8. In the Northern Hemisphere, it is easy to determine your approximate latitude by observing the height of Polaris above your northern horizon. For example, at 50° N latitude, you will see Polaris at 50 degrees above the north horizon.

constellation **Centaurus**. The four outer stars form a cross, while the fifth, much dimmer star (**Epsilon**) is offset about 30 degrees below the center of the cross.

Longitude can be determined on land or sea by careful observation of time. In previous centuries, accurately determining longitude was a major problem in both sea navigation and mapmaking. It was not until 1762 that a clock that was accurate enough for longitude finding was invented by Englishman **John Harrison**. This clock, called a **chronometer**, was set to the time at Greenwich, England, now called **Greenwich Mean Time (GMT)**, before departing on a long voyage. As you saw earlier, the prime meridian at 0° longitude passes through Greenwich, England. Therefore, each hour difference between your time and that of Greenwich is equivalent to 15 degrees of longitude from Greenwich. The longitude of a distant locale was found by noting the GMT at local noon (the highest point of the sun in the sky, found with a sextant). The time difference was simply multiplied by 15 to find the longitude.

Today, you can find your longitude in the field simply by looking at the value computed by your GPS receiver or GPS unit in your smartphone. The details of how GPS receivers find your latitude and longitude are found in chapter 14, but you can rest assured that the GPS coordinates are more accurate than the most carefully measured time differences made in the past using Harrison and similar chronometers.

PROPERTIES OF THE GRATICULE

Circumference of the authalic sphere

When determining latitude and longitude, we sometimes use the spherical approximation to the earth's shape rather than the oblate ellipsoid. Using a sphere leads to simpler calculations, especially when working with small-scale maps of countries, continents, or the entire earth (see chapter 2 for more on small-scale maps). On these maps, differences between locations on the sphere and the ellipsoid are negligible. Cartographers use a value of the earth's spherical circumference called an **authalic sphere**. The **authalic** (meaning “area preserving”) sphere is a sphere that has the same surface area as the oblate ellipsoid being used. The equatorial and polar radii of the WGS84 ellipsoid are what we use to calculate the radius and circumference of the authalic sphere that is equal to the surface area of the WGS84 ellipsoid. The computations involved are moderately complex and

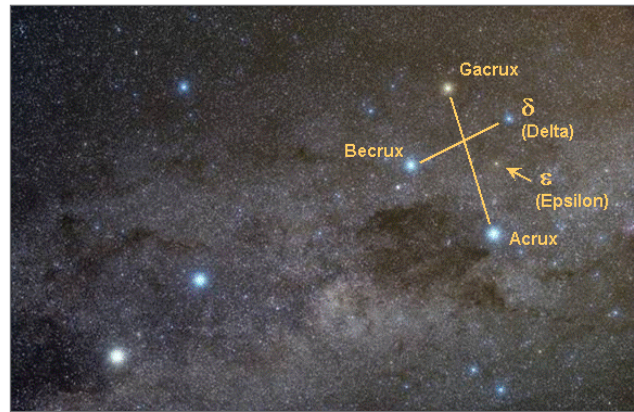


Figure 1.9. The Southern Cross is used for navigation in the Southern Hemisphere. To approximately locate the point in the sky directly above the South Pole, extend the long (Gacrux to Acrux) axis of the Southern Cross 4.5 times the length of the axis, and then draw in the sky a line perpendicular from the end of this line to the left that is the length of the short (Becrux to Delta) axis. Courtesy of Dr. Yuri Beletsky.

best left to a short computer program, but the result is a sphere of radius of 3,958.76 miles (6,371.017 kilometers) and circumference of 24,873.62 miles (40,030.22 kilometers).

Spacing of parallels

As you saw earlier, on a spherical earth the north–south ground distance between equal increments of latitude does not vary. However, it is important to know how you want to define that distance. As you will see next, there are different definitions for terms that you may take for granted, such as a “mile.”

Using the authalic sphere circumference based on the WGS84 ellipsoid, latitude spacing is always 24,873.62 miles ÷ 360°, or 69.09 statute miles per degree. **Statute miles** of 5,280 feet in length are used for land distances in the United States, whereas **nautical miles** are used around the world for maritime and aviation purposes. The original nautical mile was defined as one minute of latitude measured north–south along a meridian. The current international standard for a nautical mile is 6,076.1 feet (about 1.15 statute miles) so that there are 60.04 nautical miles per degree of latitude on the authalic sphere.

The **metric system** is used to express distances in countries other than the United States. The kilometer (1,000 meters) is, like the nautical mile, closely tied to distances along meridians, because the **meter** was initially defined

as one 10-millionth of the distance along a meridian from the equator to the North or South Pole. There are exactly 1,609.344 meters in a statute mile and 1,852 meters in a nautical mile (see table B.1 in appendix B for these and additional metric and **English unit** distance equivalents). Expressed in metric units, there are 111.20 kilometers per degree of latitude on the authalic sphere.

You have seen that parallels on an oblate ellipsoid are not spaced equally as they are on the authalic sphere, but instead decrease slightly from the pole to the equator. You can see in table B.2 in appendix B the variation in the length of a degree of latitude for the WGS84 ellipsoid, measured along a meridian at one-degree increments from the equator to the pole. The distance per degree of latitude ranges from 69.407 statute miles (111.699 kilometers) at the pole to 68.703 statute miles (110.567 kilometers) at the equator. The graph in figure 1.10 shows how these distances differ from the constant value of 69.09 statute miles (111.20 kilometers) per degree for the authalic sphere. The WGS84 ellipsoid distances per degree are about 0.3 miles (0.48 kilometers) greater than the sphere at the pole and 0.4 miles (0.64 kilometers) less at the equator. The ellipsoidal and spherical distances are almost the same in the mid-latitudes, somewhere between 45 and 50 degrees.

Converging meridians

A quick glance at any world globe (or figure 1.2) shows **converging meridians**—that is, the meridians progressively converge from the equator to a point at the North and South Poles. You will also see that the length of a degree of longitude, measured east–west along parallels, decreases from the equator to the pole. The spacing

of meridians on the authalic sphere at a given latitude is found by using equation (1.5):

$$69.09 \text{ miles (or 111.20 kilometers) / degrees} \times \cos(\text{latitude}). \tag{1.5}$$

For example, at 45 degrees north or south of the equator, $\cos(45^\circ) = 0.7071$. Therefore, the length of a degree of longitude is found in equation (1.6):

$$69.09 \times 0.7071, \text{ or } 48.85 \text{ statute miles} \\ (111.20 \times 0.7071, \text{ or } 78.63 \text{ kilometers}). \tag{1.6}$$

This spacing of meridians is roughly 20 miles (32 kilometers) shorter than the 69.09-mile (111.20-kilometer) spacing at the equator.

Quadrilaterals

Many navigational maps cover **quadrilaterals**, which are bounded by equal increments of latitude and longitude. Because meridians converge toward the poles, the shapes of the quadrilaterals vary from a square on the sphere at the equator to a narrow spherical triangle at the pole, such as the 15° × 15° quadrilaterals from equator to pole in figure 1.11. The equation cosine of center latitude, or $\cos(\text{center latitude})$, gives the **aspect ratio** (width to

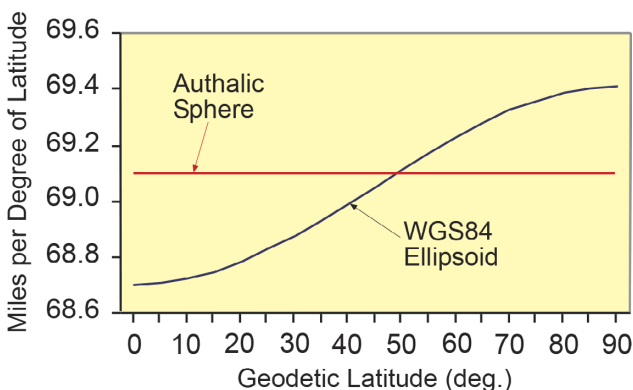


Figure 1.10. Distances along the meridian for one-degree increments of latitude from the equator to the pole on the WGS84 ellipsoid and authalic sphere.

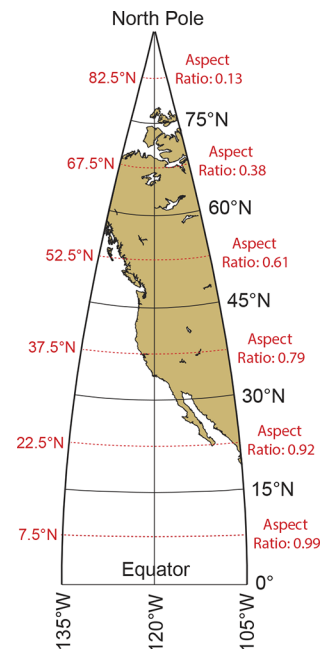


Figure 1.11. Aspect ratios for 15° × 15° quadrilaterals on the earth from the equator to the North Pole. The cosine of the mid-latitude for each quadrilateral gives its aspect ratio.

height) of a quadrilateral. A quadrilateral centered at the equator has an aspect ratio of 1.0, whereas a quadrilateral at 60° N has an aspect ratio of 0.5. A map that covers 1° × 1° or any other quadrilateral extent looks long and narrow at this high latitude, whereas a quadrilateral map at the equator is essentially square in shape.

Great and small circles

A **great circle** is the largest possible circle that can be drawn on the surface of the spherical earth and also the shortest distance between two points on the earth's surface. Its circumference is that of the sphere and its center is the center of the earth, so that all great circles divide the earth into halves. Notice in figure 1.3 that the equator is a great circle that divides the earth into the Northern and Southern Hemispheres. Similarly, the prime meridian and its **antipodal meridian**, situated at the opposite side of the earth at 180°, form a great circle that divides the earth into the Eastern and Western Hemispheres. All other pairs of meridians and their antipodal meridians are also great circles. Because a great circle is the shortest route between any two points on the earth, great circle routes are fundamental to long-distance navigation, as you will see in chapter 14.

Any circle on the earth's surface that intersects the interior of the sphere at any location other than the center is called a **small circle**, and its circumference is smaller than a great circle. You can see in figure 1.3 that all parallels other than the equator are small circles. The circumference of a particular parallel is given by equation (1.7):

$$24,874 \text{ miles (or } 40,030 \text{ kilometers)} \times \cos(\text{latitude}). \quad (1.7)$$

For example, the circumference of the 45th parallel is

$$24,874 \times 0.7071, \text{ or } 17,588 \text{ statute miles} \\ (40,030 \times 0.7071, \text{ or } 28,305 \text{ kilometers}).$$

GRATICULE APPEARANCE ON MAPS

Small-scale maps

Small-scale world or continental maps such as globes and world atlas sheets normally use geocentric latitude-longitude coordinates based on an authalic sphere. There

are several reasons for using the authalic sphere. Prior to using digital computers to make these types of maps numerically, it was much easier to construct them from spherical coordinates. Equally important, the differences in the plotted positions of geocentric and corresponding geodetic parallels become negligible on maps that cover so much area on a small page.

For example, mathematicians have computed the maximum difference between geocentric and geodetic latitude as 0.128 degrees at the 45th parallel. If you draw parallels at 45° and 45.128° on a map scaled at one inch per degree of latitude, the two parallels are drawn a very noticeable 0.128 inches (0.325 centimeters) apart. Now imagine drawing the parallels on a map scaled at one inch per 10 degrees of latitude, a scale that corresponds to a world wall map approximately 18 inches (46 centimeters) high and 36 inches (92 centimeters) wide. The two parallels are now drawn 0.013 inches (0.033 centimeters) apart, a difference that is not even noticeable considering the width of a line on a piece of paper.

Large-scale maps

Parallels and meridians are shown in different ways on large-scale maps that show small areas in great detail (see chapter 2 for more on large-scale maps). Large-scale topographic maps, which show elevations and landforms as well as other ground features, use tick marks to show

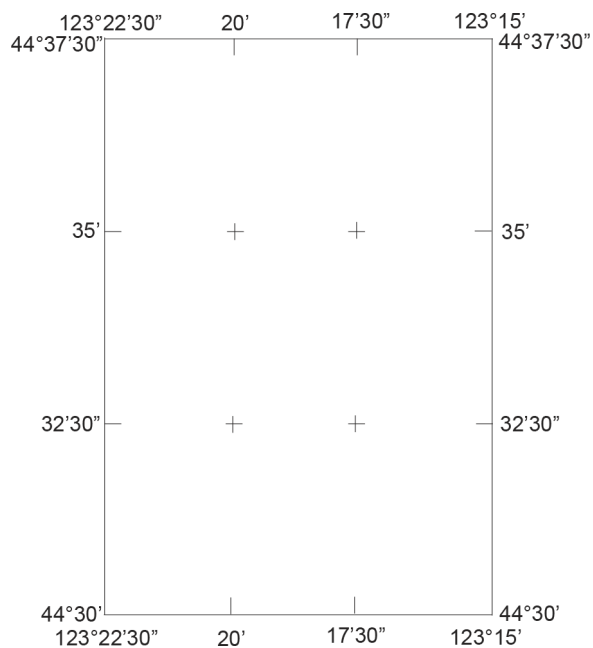


Figure 1.12 Graticule ticks on the Corvallis, Oregon, 7.5-minute topographic map.

the location of the graticule (see chapters 2, 9, and 19 for more on topographic maps). For example, all US Geological Survey 7.5-minute topographic maps have graticule ticks at 2.5-minute intervals of latitude and longitude (figure 1.12). The latitude and longitude coordinates are printed in each corner, but only the minutes and seconds of the intermediate edge ticks are shown. Note the four plus sign symbols used for the interior 2.5-minute graticule ticks in the figure.

The graticule is shown in a different way on nautical charts for marine navigation. The chart segment in figure 1.13 shows that alternating white and dark bars spaced at the same increment of latitude and longitude line the edge of the chart, along with ticks at one-minute increments. These ticks are used to find the latitude and longitude of mapped features to within a fraction of a minute. Because of the convergence of meridians, the spacing between ticks, which shows equal increments of latitude, is longer within the vertical bars on the left and right edges of the chart than the spacing between ticks within the horizontal bars at the top and bottom. The one exception

is charts of areas along the equator, in which the spacing between ticks is the same on all edges. The more closely spaced ticks beside each horizontal bar are placed every 10th of a minute.

Aeronautical charts display the graticule in yet another way. The chart segment for a portion of the Aleutian Islands in Alaska (figure 1.14) shows that parallels and meridians are drawn at 30-minute latitude and longitude intervals. Ticks are placed at one-minute increments along each graticule line, allowing features to be located easily to within a fraction of a minute.

GEODETIC LATITUDE AND LONGITUDE ON LARGE-SCALE MAPS

You will often find parallels and meridians of geodetic latitude and longitude on detailed maps of small areas.

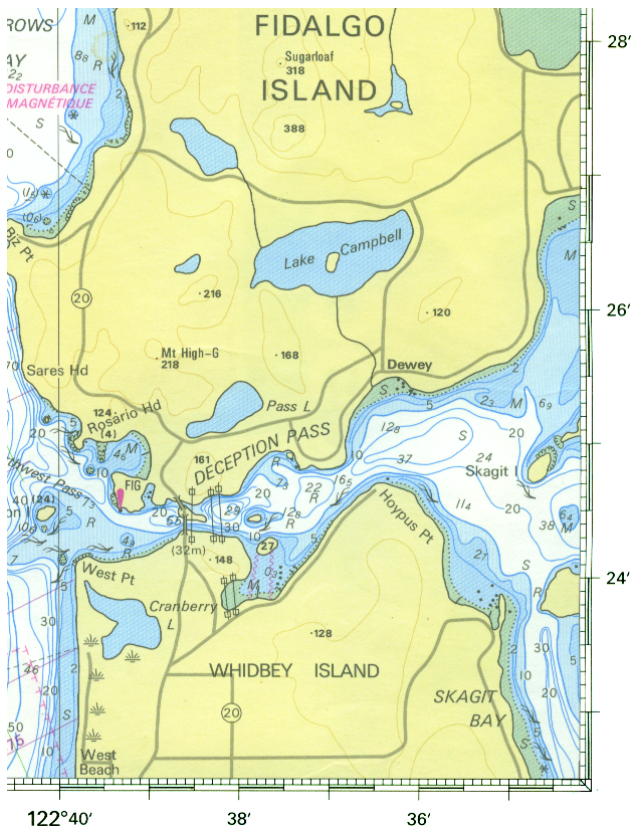


Figure 1.13. Graticule bars with ticks on the edges of a nautical chart segment. Reproduced with permission of the Canadian Hydrographic Service.

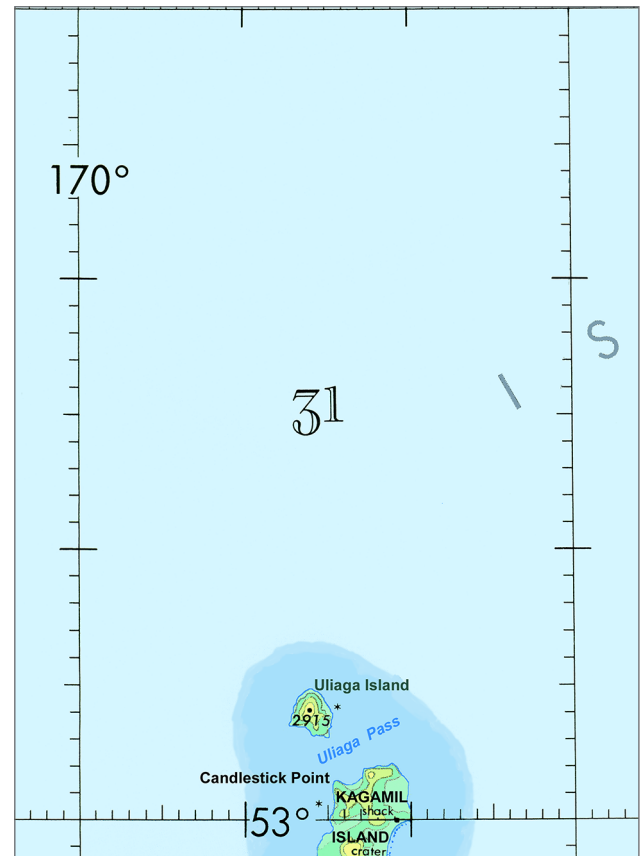


Figure 1.14. Graticule ticks on a small segment of an aeronautical chart. Courtesy of the National Aeronautical Charting Office.

Geodetic coordinates are used to make the map a close approximation to the size and shape of the part of the ellipsoidal earth that it represents. To see the perils of not doing this step, you need only examine one-degree quadrilaterals at the equator (from 0° to 1° in latitude) and at the North or South Poles (from 89° to 90° in latitude).

You can see in table B.2 in appendix B that the ground distance between these pairs of parallels on the WGS84 ellipsoid is 68.703 and 69.407 miles (110.567 and 111.699 kilometers), respectively. If the equatorial quadrilateral is mapped at a scale so that it is 100 inches (254 centimeters) high, the polar quadrilateral mapped at the same scale is 101 inches (256.5 centimeters) high. If both quadrilaterals are mapped using the authalic sphere with 69.09 miles (111.20 kilometers) per degree, both quadrilaterals are 100.6 inches (255.5 centimeters) high. Having both maps several 10ths of an inch (or around a centimeter) longer or shorter than they should be seems a small error, but it can be an unacceptably large error for maps that are used to make accurate measurements of distance, direction, or area. Yet you can see in figure 1.10 that the height differences at the equator and the poles are the extremes and that there is little difference at mid-latitudes.

Horizontal reference datums

To further understand the use of different types of coordinates on detailed maps of smaller extents, you must first look at **datums**—the collection of accurate **control points** (points with known positional accuracy) that surveyors and others use to georeference map data so that it aligns with the geodetic latitude and longitude coordinate system (see chapter 5 for more on control points and georeferencing). Surveyors determine the precise geodetic latitude and longitude of **horizontal control points** spread across the landscape. You may have seen a horizontal control point **monument** (also called a **survey marker**, *survey mark*, or sometimes a *geodetic mark*), a fixed object established by surveyors when they determine the exact position of a point. Monuments, such as the one in figure 1.15, may be placed in the ground on top of a hill or other prominent feature, but in less prominent locations such as a road or sidewalk, they may be buried so that they will not be hit by a car or bicycle.

From the 1920s to the early 1980s, these control points were surveyed relative to the Clarke 1866 ellipsoid, together forming what was called the **North American Datum of 1927 (NAD27)**. Topographic maps, nautical and aeronautical charts, and many other large-scale



Figure 1.15 Horizontal control point marker cemented in the ground. Courtesy of the National Oceanic and Atmospheric Administration.

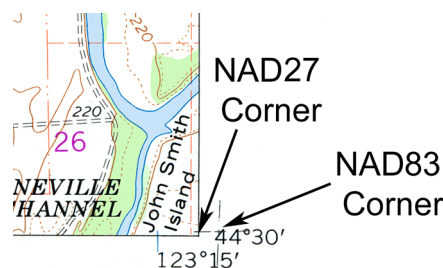


Figure 1.16 Southeast corner of the Corvallis, Oregon, topographic map showing the difference between its NAD27 and NAD83 positions. Courtesy of the US Geological Survey.

maps of this period have graticule lines or ticks based on this datum. For example, the southeast corner of the Corvallis, Oregon, topographic map first published in 1969 (figure 1.16) has a NAD27 latitude and longitude of 44°30'N, 123°15' W.

By the early 1980s, better knowledge of the earth's shape and size and far better surveying methods led to the creation of a new horizontal reference datum, the **North American Datum of 1983 (NAD83)**. The NAD27 control points were corrected for surveying errors where apparent, and then were added to thousands of more recently acquired points. The geodetic latitudes and longitudes of all these points were determined relative to the **Geodetic Reference System of 1980 (GRS80)** ellipsoid, which was essentially identical to the WGS84 ellipsoid when GRS80 was devised.

The change of horizontal reference datum from NAD27 to NAD83 meant that the geodetic coordinates for control points across the continent changed slightly in 1983, and this change had to be shown on large-scale maps published earlier but still in use. On topographic maps, the NAD83 position of the map corner is shown by a dashed plus sign, as in figure 1.16. Often, the shift is in the 100-meter range and must be taken into account when plotting on older maps the geodetic latitudes and longitudes obtained from GPS receivers and other modern position-finding devices.

Europe in the early 1900s faced another problem—separate datums for different countries that did not mesh into a single system for the continent. Military map users in World War II found different latitudes and longitudes for the same ground locations on topographic maps along the borders of France, Belgium, the Netherlands, Spain, and other countries in which major battles were fought. The **European Datum of 1950 (ED50)** was created after World War II as a consistent reference datum for most of western Europe; although Belgium, France, Great Britain, Ireland, Sweden, Switzerland, and the Netherlands continue to retain and use their own national datums. Latitudes and longitudes for ED50 were based on the **International Ellipsoid of 1924**. Users of GPS receivers will find that, moving westward through Europe from northwestern Russia, the newer **European Terrestrial Reference System 1989 (ETRS89)** longitude coordinates based on the WGS84 ellipsoid gradually shift to the west of those based on the 1924 International Ellipsoid. In Portugal and western Spain, the WGS84 longitudes are approximately 100 meters to the west of those found on topographic maps based on ED50. Moving southward, WGS84 latitudes gradually shift northward from those based on ED50, reaching a maximum difference of around 100 meters in the Mediterranean Sea.

As noted, Great Britain and Ireland are examples of countries that continue to use datums based on ellipsoids defined in the 19th century to best fit that region. Topographic maps in both nations use the **Airy 1830 ellipsoid** as the basis for the **Ordnance Survey Great Britain 1936 (OSGB36)** datum for geodetic latitude and longitude coordinates. Along the south coast of England, WGS84 latitudes are about 70 meters to the south of those based on OSGB36. This southward shift gradually diminishes to zero near the Scottish border, and then becomes a northerly difference that reaches a maximum value of around 50 meters at the northern extremes of Scotland. In Ireland, WGS84 longitudes are around 50 meters to the east of their OSGB36 equivalents and gradually increase to a maximum

difference of around 120 meters along the southeast coast of England (go to http://www.colorado.edu/geography/gcraft/notes/datum/datum_f.html for a detailed list of horizontal datums and reference ellipsoids used in foreign countries throughout the world).

THE EARTH AS A GEOID

When the earth is treated as an athermal sphere or oblate ellipsoid, mountain ranges, ocean trenches, and other surface features that have vertical relief are neglected. There is justification for this treatment, as the earth's surface is truly smooth when you compare the surface undulations to the 7,918-mile (12,742-kilometer) diameter of the earth based on the athermal sphere. The greatest relief variation is the approximately 12.3-mile (19.8-kilometer) difference between the summit of Mount Everest (29,035 feet or 8,852 meters) and the deepest point in the Mariana Trench (36,192 feet or 11,034 meters). This vertical difference is immense on the human scale, but it is only 1/640th of the earth's diameter. If you look at the difference between the earth's average land height (2,755 feet or 840 meters) and ocean depth (12,450 feet or 3,795 meters), the average roughness is only 1/2,750th of the diameter. In fact, if the earth was reduced to the diameter of a bowling ball, it would be smoother than the bowling ball.

The earth's global-scale smoothness aside, knowing the elevations and depths of features is important to us. Defining locations by their geodetic latitude, longitude, and elevation gives you a simple way to collect elevation data and display this information on maps. The top of Mount Everest, for example, is located at 27°59' N, 86°56' E, 29,035 feet (8,852 meters), but what is this elevation relative to? This question leads us to another approximation of the earth called the *geoid*, which is a surface of equal gravity used as the reference datum for elevations.

Vertical reference datums

Elevations and depths are measured relative to what is called a **vertical reference datum**, an arbitrary surface with an elevation of zero. The traditional datum used for land elevations is **mean sea level (MSL)** (see chapter 9 for more on mean sea level). Surveyors define MSL as the average of all low and high tides at a particular starting location over a **Metonic cycle** (a period of approximately 19 years or 235 lunar months, at the end of which the phases of the moon begin to occur in the same order and

on the same days as in the previous cycle). Early surveyors chose this datum because of the measurement technology of the day. Surveyors first used the method of **leveling**, in which elevations are determined relative to the point at which mean sea level is defined, using horizontally aligned telescopes and vertically aligned leveling rods. A small circular monument was placed in the ground at each surveyed benchmark elevation point. A **benchmark** is a permanent monument that establishes the exact elevation of a place.

Later, surveyors could determine elevation by making gravity measurements at different locations on the ground and relating them to the strength of gravity at the point used to define MSL. Gravity differences translate into elevation differences because the strength of gravity changes with elevation.

MSL is easy to determine along coastlines, but what about inland locations? It requires extending mean sea level across the land. Imagine that mean sea level is extended across the continental land masses on the ellipsoidal surface, which is the same thing as extending a surface that has the same strength of gravity as mean sea level (figure 1.17). However, this imaginary equal-gravity surface doesn't form a perfect ellipsoid, because differences in topography and earth density affect gravity's pull at different locations, and thus the shape of the surface.

The slightly undulating, nearly ellipsoidal surface that best fits mean sea level for all the earth's oceans is called a **global geoid**. The global geoid varies approximately 100 meters above and below the oblate ellipsoid surface in an irregular fashion. World maps that show land **topography**

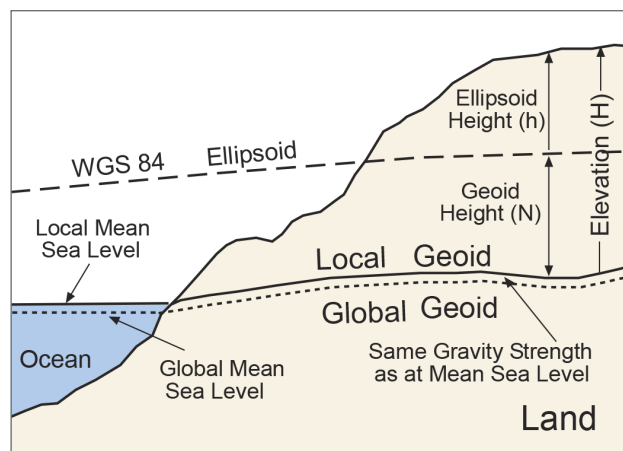


Figure 1.17 The geoid is the surface at which gravity is the same as at mean sea level. Elevations on maps are measured relative to the global geoid, but modern GPS-determined heights are relative to the WGS84 ellipsoid.

and ocean **bathymetry** use land heights and water depths relative to the global geoid surface. In the conterminous United States, geoid heights range from a low of -51.6 meters in the Atlantic Ocean to a high of -7.2 meters in the Rocky Mountains (figure 1.18). Worldwide, geoid heights vary from -105 meters just south of Sri Lanka to 85 meters in Indonesia.

The mean sea level datum based on the geoid is so convenient that it is used to determine elevations around the world and is the basis for the elevation data found on nearly all topographic maps and nautical charts. But be aware that

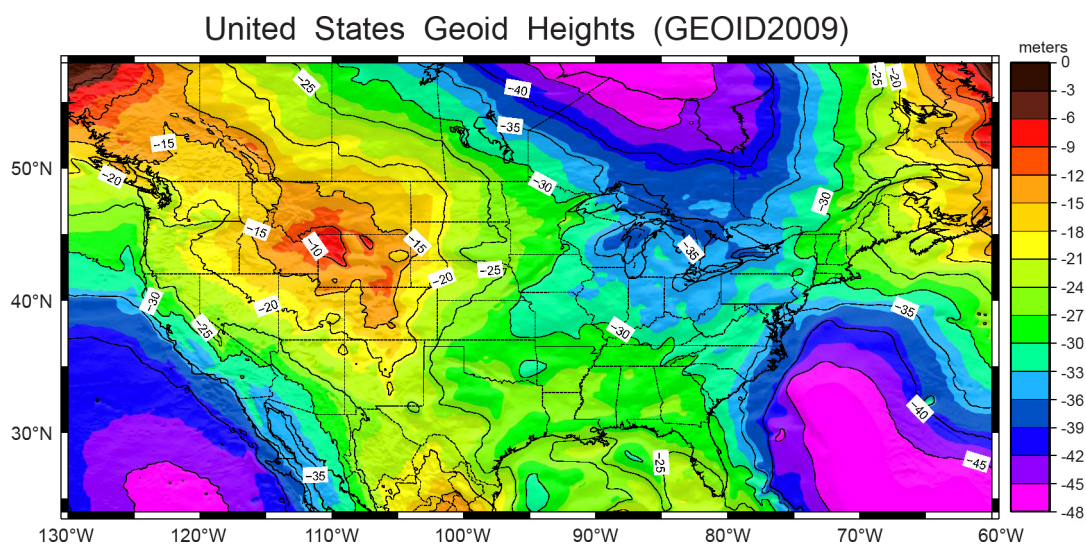


Figure 1.18 Geoid heights in the United States and vicinity. Courtesy of the National Geodetic Survey GEOID2009 model.

the **local geoid** used in your area is probably slightly above or below (usually within two meters of) the global geoid used for world maps. This difference is because the mean sea level at one or more nearby locations is being used as the vertical reference datum for your nation or continent, rather than the average sea level for all the oceans.

In the United States, for example, you may see elevations relative to the **National Geodetic Vertical Datum of 1929 (NGVD29)** on older topographic maps. This datum was defined by the observed heights of mean sea level at 26 tide gauges, 21 in the United States and 5 in Canada. It also was defined by the set of elevations of all benchmarks resulting from over 60,000 miles (96,560 kilometers) of leveling across the continent, totaling over 500,000 vertical control points. In the late 1980s, surveyors adjusted the 1929 datum with new data to create the **North American Vertical Datum of 1988 (NAVD88)**. Topographic maps, nautical charts, and other cartographic products for the United States, Canada, Mexico, and Central America made from this time forward use elevations based on NAVD88. Mean sea level for the continent was defined at one tidal station on the Saint Lawrence River at Rimouski, Quebec, Canada. NAVD88 was a necessary update of the 1929 vertical datum because about 400,000 miles (650,000 kilometers) of leveling was added to the NGVD since 1929. Additionally, numerous benchmarks were lost over the decades, and the elevations at others were affected, by vertical changes because of rising land elevations since the retreat of glaciers at the end of the last ice age or subsidence from the extraction of natural resources such as oil and water.

GPS has created a second option for measuring elevation rather than basing elevation on a vertical datum (see chapter 14 for more on GPS). GPS receivers calculate what is called the **ellipsoidal height h** , the distance above or below the surface of the WGS84 ellipsoid along a line from the surface to the center of the earth (see figure 1.17). An ellipsoidal height is not an elevation, because it is not measured relative to the mean sea level datum for your local geoid. Therefore, you must convert GPS ellipsoidal height values to mean sea level datum elevations H before you can use them with existing maps. You do this by subtracting the geoid height N at each point from the ellipsoid height h measured by the GPS receiver using equation (1.8):

$$H = h - N. \quad (1.8)$$

Ellipsoidal height is the same thing as **height above the ellipsoid (HAE)**, and this is the elevation most often found by GPS receivers. The lookup table to make this conversion is usually stored in your GPS receiver's computer.

SELECTED READINGS

- Greenhood, D. 1964. *Mapping*. Chicago: University of Chicago Press.
- Illife, J. C. 2000. *Datums and Map Projections for Remote Sensing, GIS and Surveying*. Caithness, Scotland: Whittles.
- La Condamine, C.M. de. 1747. *A Succinct abridgement of a Voyage Made with the Inland Parts of South America, from the Coasts of the South-Sea, to the Coasts of Brazil and Guiana, down the River of Amazons: As It Was Read in the Public Assembly of the Academy of Sciences at Paris*. London: E. Withers, at the Seven Stars.
- Maling, D. H. 1992. *Coordinate Systems and Map Projections*, 2nd ed. New York: Pergamon.
- Maupertius, P. L. M. de. 1738. *The Figure of the Earth Determined from Observations Made by Order of the French King, at the Polar Circle* (translation from French). London.
- Meade, B. K. 1983. "Latitude, Longitude, and Ellipsoidal Height Changes NAD-27 to Predicted NAD-83." *Surveying and Mapping* 43:65–71.
- Robinson, A. H., J. L. Morrison, P. C. Muehrcke, A. J. Kimerling, and S. C. Guptill. 1995. "Basic Geodesy." Chap. 4 in *Elements of Cartography*, 6th ed. New York: John Wiley & Sons.
- Smith, J. R. 1988. *Basic Geodesy: An Introduction to the History and Concepts of Modern Geodesy without Mathematics*. Rancho Cordova, CA: Landmark Enterprises.
- Snyder, J. P. 1987. "Map Projections: A Working Manual." US Geological Survey Professional Paper 1395. Washington, DC: US Government Printing Office.
- Sobel, D. 1995. *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. New York: Walker.
- Wallis, H. M., and A. H. Robinson, eds. 1987. *Cartographical Innovations*. London: Map Collector.
- Wilford, J. N. 1981. *The Mapmakers*. New York: Alfred A. Knopf.