CHAPTER 5

Projections

This chapter focuses on the principles connected to projections—the transformation of spherical coordinates to planar coordinates—you will need for work with GI for making maps and other purposes. Chapter 4 also presents some historical background and specific details of various projections.

Projections occupy one of the most essential roles in cartography for geography and GI. For some people, this role may arguably be perhaps the most essential, because most GI is "projected," even if the projection information never shows up on a map. This has started to change as more and more GI is collected and stored in latitude and longitude coordinates, which are not projected and commonly used in online mapping. But even if all the data you need and want is available in latitude and longitude coordinates, you will probably need to project it to make the sort of map that people are familiar with, or combine it with data collected using another projection.

Maps without Projections

Some people would claim that if a thing or event is shown on a map, it must be projected. In most cases this is true—and for good reasons. But there are exceptions. These exceptions are important enough to pay attention to. The first exception was already mentioned: locations stored in latitude and longitude coordinates are not projected—they are spherical coordinates. It's even possible to make a planar (flat) map with these coordinates in a grid, but such a map is greatly distorted and as a result can be misleading. The second exception is the maps drawn following artistic or design criteria rather than scientific concerns. Usually these maps are used for advertisements, but they can also be used to show transportation networks, to illustrate tourist destinations, and to serve other popular forms of communication. The third exception, globes, is a nonprojected way of showing things, events, and relationships without the distortion of projections. A global **tesselation** using hexagons or octahedrons to subdivide the sphere is another nonprojected way of representing a round surface.

A Brief History of Projections

The reasons for using projections go back to desires to accurately represent the spherical surface of the earth on flat maps. For GI and a map to be useful, the locations and relationships must be accurate. The uses of a nonprojected map using latitude and longitude coordinates (historically these were determined by the use of sextants, cross-staff, etc.; today they're mostly determined by GPS—see Figure 5.1), or an advertising map showing simple directions, are limited by their inaccuracy. You can use such a map with directions to the new amusement park to find your way there, even if you're not from the area, but you can't use it in most cases to navigate to the beach or swimming pool. Its limits mean you won't be able to discover and understand the relationship of the amusement park to things and events not shown on that map. Importantly, because of the curved surface of the earth, nonprojected maps of larger areas showing locations and sizes of things and events would be inaccurate. The Euclidean geometrical measure of the distances on the earth's surface, which is the most common geometry—already practiced by the ancient Egyptians—cannot take its curvature into account, but instead uses a Cartesian coordinate system and a projection that have already taken the spherical shape of the earth into account. Thinking about other possible uses, projections make it easier to compare GI and maps of the same area because they provide a framework for people and organizations to systematically locate things and events.

As you can probably already imagine, it is no surprise that the first maps were based on work by geographers who were locating things and figuring

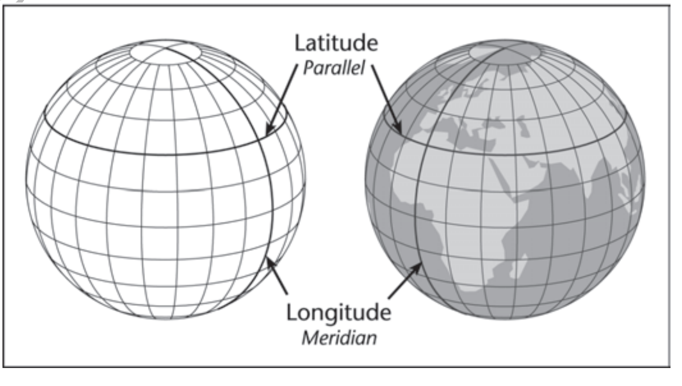

FIGURE 5.1. Latitude and longitude lines.

out relationships among events. Ptolemy (c. 100–168) wrote the book *Geography* with the location of cities, coasts, and other important places of the world known to the ancient Greeks. The Romans may have used this book for making a map that showed, in a greatly distorted manner, Europe, North Africa, the Near East, and India, even indicating China. The original and all copies of this map were lost, except for one, a re-creation done in the 15th century that is now known as the Tabula Peutingeriana.

While maps such as the Tabula Peutingeriana and many others had been used for a very long time, maps that show location accurately have only been around for some 400 years, once it was discovered how to determine longitude. Cartographers until then could only accurately determine the latitude of places. This means that while the equator could commonsensically be calculated as the halfway place between the north and south poles, the 0° starting measure for longitude was only agreed to in the late 19th century and placed in Greenwich, England. Up until then, 0° longitudes started from different locations including Paris and the Faro Islands. Knowing where the starting measure of longitude is located is crucial for accurate navigation. Before there was widespread agreement about where the starting (0°) longitude is for all people and countries, an arbitrary starting longitude was fine so long as it was used systematically.

Roles of Projections

One of the key roles of projections has been in the production of maps for navigation, naval or aeronautical in most cases, which are called **charts**. The development of accurate ways to determine location went hand-in-hand with the growth of European naval powers. However, because these are spherical coordinates, and mariners needed flat maps to take with them, projections became crucial. The Mercator projection is perhaps so commonplace because a straight line in this projection shows a constant compass bearing. You should remember that there are many other projections, but the Mercator projection possesses the quality that lines of a constant direction are straight lines.

Because of this character, the Mercator projection was very important for navigation on water by compass, but other modes of transportation can better use other projections. More recently, since airplanes began to fly regularly across and between continents, another type of projection was needed for their navigation. A line of a constant compass direction may be straight in the Mercator projection, but this line does not show the shortest distance. The shortest route for an airplane high above the earth's surface is not a straight line, but a line on a sphere, called the **great circle distance** (see Figure 5.2).

Different projections are used for maps with different roles. The size of the area to be mapped, the desired **projection properties**, and the characteristics of the GI and map are the key determinants. The size of the area

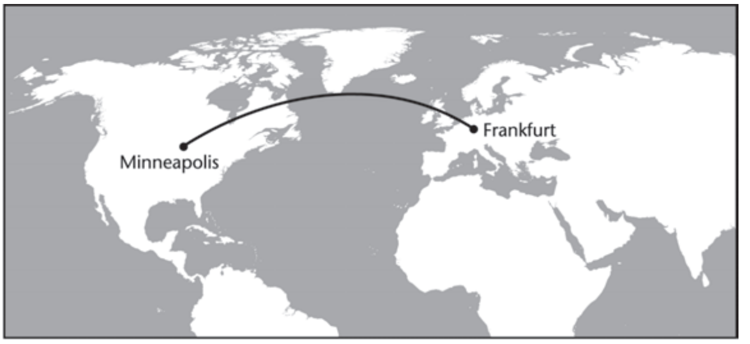

FIGURE 5.2. Great circle path between Minneapolis, USA, and Frankfurt, Germany. The great circle distance is 4,392 miles.

distinguishes basically between the whole world, a continent, a state or province, a region, a county or city, and still smaller units. Different projections fit different areas better or worse, depending on their use. The Mercator projection is quite inaccurate for comparing the sizes of areas because of distortion near the poles, but is quite useful for maps used in navigation. A projection property refers to whether the projection represents angles, areas, or distances (from one or two points) as they are found on the surface. No projection retains both angles and areas. A projection can retain one projection property—for example, the Mercator projection preserves angles. All **transverse** (turned 90° to be oriented north–south) Mercator-projected GI and maps are useful for mapping north–south oriented small areas because this projection is conformal and also preserves shapes over small areas along the line of tangency where the projection theoretically touches the earth's surface. The characteristic of the GI or map indicates how the projection should show geographic relationships and scales. Choosing a projection that preserves one projection property often leads to other **distortions**. For an individual state or province, a projection that maintains constant area to make comparisons of areas possible is beneficial, even if some shapes over a larger area may begin to look distorted. Indeed, larger areas are hard to show without distortion in any case; many projections commonly used for world maps compromise and distort both area and shapes. Why distortion is commonplace for projections, what are the projection properties and characteristics, and how to choose a projection is discussed later in this chapter.

Making Projections

Even if you are only going to use maps and will never work with GI, you need to know some important things about projections. The first is that projections make use of different models of the earth. Generally, projections for the entire earth use a simple spheroid. When dealing with maps or GI of the entire world, the loss of accuracy is slight compared to the resolution of the

Making Projections with Light

Although most projections are calculated mathematically, the underlying transformation from a three-dimensional to a two-dimensional representation of all projections can be physically constructed with the aid of a few common items: a light (flashlight or lamp), a two-liter plastic bottle, a lampshade, and a piece of wax paper or flat plastic you can draw on. You will write on all of these items, so you need to be sure they are no longer needed.

To make the construction surfaces, you will need to prepare the plastic bottle by cutting off the top and bottom carefully with a scissors or knife. The lampshade and the flat wax paper or plastic are ready to be used as they are. On each of these objects you should mark a series of horizontal and vertical lines. On the lampshade and piece of wax paper or plastic, they should radiate from the center. On the lampshade, they should, if extended, meet each other at an imaginary point above the top of the lampshade; on the wax paper or plastic, they should radiate from a circle located at the center.

The construction surfaces you made correspond to the developable surfaces used in cylindrical, conic, and planar types of projections. To show how each developable surface is used, take a flashlight or light placed at the middle of the bottle or lampshade or behind the wax paper or plastic surface and shine the light source at a nearby wall or piece of paper. (It usually helps to dim the room lights when you do this.)

What you see on the wall or paper is the projected surface that corresponds to each type of projection. Try moving the light, the paper, and the construction surface to see how the changes affect each projection. These changes correspond to parameters used in the construction of map projections discussed in this chapter.

GI or detail of the map. Projections needed for more detailed purposes or smaller areas of the earth use an *ellipsoid* (see Figure 5.3), with even more axes, that generally fits the actual shape of the earth. For very detailed purposes and the highest levels of accuracy, people use a *geoid,* often optimized for the shape of the earth in one particular and relatively small area (see Figure 5.4).

While this may seem needlessly complex, you should remember that because the earth is constantly changing shape (and not only from volcanoes

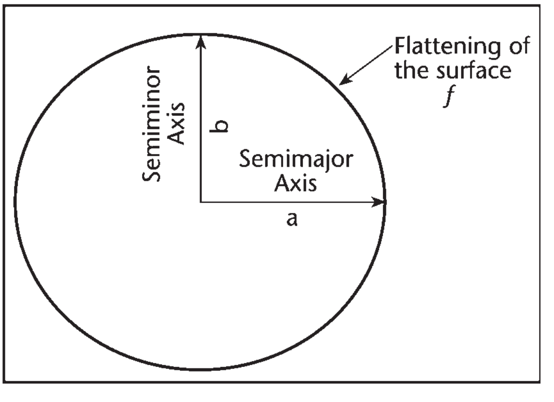

FIGURE 5.3. Reference ellipsoid showing major parameters.

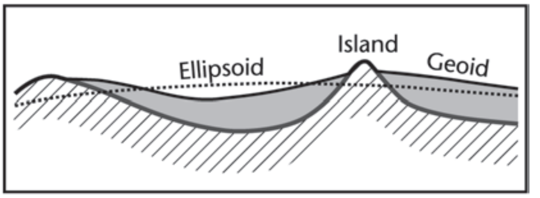

FIGURE 5.4. In this schematic drawing, an ellipsoid and a geoid represent the earth's surface. The ellipsoid is less accurate than the geoid, but both may not properly align with actual locations distant from their geospatial optimizations.

and earthquakes, but gradual movements that are imperceptible to people), different uses need different levels of positional accuracy. On one extreme, a map showing worldwide the most visited tourist sites for the last 10 years needs very little accuracy; on the other extreme, an engineer's plan of a 2-km tunnel for a new railroad needs extremely great accuracy. Most GI and mapping activities need a level of accuracy somewhere in between—often the cost of preparing the GI and the available budget determine the accuracy.

What makes all this complicated for working with GI and maps is that there is no standard earth model, or geoid model, or spheroid model, or ellipsoid used to represent locations on earth. The use of different models makes it paramount for GI users to know the model used for projecting the GI, which is often called a "datum" (see below for more information about datums). Work on specific models of the earth is done by geographers known as geodesists, and information about geoids and datums is geodetic information.

The Geoid Model

The most accurate model of the earth's surface is the geoid. The earth, because of its constantly changing shape due to tectonic movements and undulations of its gravity field, can be described in the most detailed fashion through sets of measurements that are used to produce a geoid. The expert geodesists who determine geoids and their constants put the geoid model into relationship with the planetary body or extremely detailed information about elevations in a particular area. Geodesists describe a *geoid* as the equipotential surface of the earth, which means the known earth's surface under consideration of different local strengths of gravity resulting from different masses of the earth's geological makeup, fluctuations in the earth's core, and other factors. For example, the Marianna Trench in the Pacific Ocean and the large bodies of iron ore found in parts of South Asia, Sweden, and many other places both locally affect the shape of the earth's surface because of the lessened or increased pull of gravity due to the lesser or greater mass at those locations. Basically, what geodesists consider is how differences in the earth's gravity affect the shape and size of the earth. For instance, denser material in the earth's crust, such as iron, influences gravity more than lighter sedimentary rocks do. The geoid takes these and other differences into account. These differences are measured in millionths of the earth's normal gravity, which seems small, but the effects on the shape of the earth and location measurements can be large. You can think of the geoid as a collection of many gravity vectors, individual gravity forces, each of which is perpendicular to the pull of gravity, that distort the earth's shape from the ideal, yet imaginary, sphere.

Practically, the geoid was until recently only used for specialized purposes for relatively small areas. Geoids were almost always calculated for smaller areas because of the complexity and cost of collecting the necessary data. With the advent of satellites and improved measuring devices, however, data collection has become much easier and geoids have become more common. They are the reference standard when working with global positioning systems (see Chapter 7). The geoid provides vertical location control. Geoid positions usually refer to a reference ellipsoid for horizontal location control and vertical location control. Differences between the ellipsoid positions and the geoid positions are called "geoid undulations," "geoid heights," or "geoid separations." The horizontal and vertical locations of the projection surface based on an ellipsoid can be adjusted to the irregular shape of the geoid compared to the regular mathematical surface of an ellipsoid through geodetic techniques.

The Ellipsoid Model

The ellipsoid (also called a spheroid in some cases) is the most commonly used model for projections of GI and maps. It includes the noticeable distortion between the length of the earth's north–south axis and its equator, which bulges a small amount due to the centrifugal force of the earth's rotation. In the simplest mathematical form, it consists of three parameters:

- An equatorial semimajor axis a
- A polar semiminor axis b
- The flattening f

Mapmakers and geodesists have produced many ellipsoids. John Snyder wrote that between 1799 and 1951 twenty-six ellipsoid determinations of the earth's size were made. Each of these ellipsoids has a history and sheds light onto the science, culture, politics, and personalities involved in establishing the ellipsoid through complicated and challenging field survey coupled with exhaustive calculations. Ellipsoids were developed to have a more accurate reference for mapping, to satisfy individual ambition, to serve national goals, to make more accurate measurements, and so on. The surveys conducted to create ellipsoids were often ambitious expeditions into the remote areas of the world and continue to provide the material for many stories. Multiple ellipsoids were developed and refined as measurements improved, and ellipsoids have often been specially defined for specific areas—for example, for U.S. counties. Working with data or maps from different periods often involves determining if different ellipsoids were used in collecting data; data from different coordinate systems, even if in the same area, may also have different ellipsoids (see Table 5.1).

TABLE 5.1. Selected Ellipsoids Parameters

Name	Semimajor axis	Semiminor axis	Flattening
Bessel (1841)	6,377,483.865 m	6,356,079.0 m	1/299.1528128
Clarke (1866)	6,378,206 m	6,356,584 m	1/294.98
Krassovsky (1940)	6,378,245 m	6,356,863.03 m	1/298.3
Australian (1960)	6,378,160 m	6,356,774.7 m	1/298.25
WGS (1984)	6,378,137 m	6,356,752.31425 m	1/298.257223563

The Spheroid Model

The *spheroid* is the simplest model of the earth's surface, using only a single measurement to approximate the shape of the earth's surface for GI and maps. This measurement is the distance from the hypothetical center of the earth to the surface, or, in geometrical terms, the radius. The mean earth radius *RE* is 3,959 miles (6,371.3 km). It is very inaccurate for many GI uses and you should only use the spheroid for scales smaller than 1:5,000,000,000. The inaccuracies of this model of the earth's surface are not apparent at these small scales. It is much easier to calculate the projections using a spherical model but using the spheroid for projecting GI and making maps for scales larger can lead to grave inaccuracies.

Putting the Models Together: Demythologizing the Datum

Datum is the term used to refer to the calibration of location measurements including the vertical references, horizontal references, and particular projections or versions of projections—for example, the North American Datum 1927 (NAD 27) or the North American Datum 1983 (NAD 1983). Datums constitute one of the most confounding aspects of working with projections for many GI and map users. For most intents, this term simply specifies the model of the shape of the earth at a particular point in time and often for a particular area—for example, North America, Europe, or Australia. A horizontal datum is often the basis for determining an ellipsoid used in a projection for a coordinate system (see Chapter 6). A datum can be used with different projections—for example, the NAD 1927 is used with both the Lambert and the transverse Mercator projections. For GI users, datums are references to a set of parameters needed for measuring locations and the basis for projections. Because there are many parameters and the mathematics for transforming datums is highly complex, many people have been stymied by datums. But it is really, for most general purposes, quite simple: the datum refers to a reference surface for making positional measurements. While most datums in North America are described in technical guidelines or even laws, theoretically a datum can be defined by any government agency or private group as it sees fit.

Datums distinguish between horizontal and vertical references and local and geocentric datums. A datum should (but might not) contain both horizontal and vertical references. Horizontal references are used to measure the location of positions on the earth and vertical datums are used to measure the elevation of a position. You can think of a vertical datum as the base level used in recording elevations or the mean height of tides. All elevations using the vertical datum are related to this zero elevation. Local datums, in fact, are used for areas up to the size of continents—for example, the NAD of 1927, which made a location on Meades Ranch in Kansas the starting point of the triangulation that measured the earth's undulations and put them into relationship with the Clarke 1866 ellipsoid. Geocentric datums—for example, the World Geodetic System Datum of 1984—take the entire earth into consideration and lack an origin point; they don't have a defined datum point, but are calculated from a network of geodetic observations. The difference between local datums can be several hundred meters—for instance, between NAD 1927 and NAD 1983 in some areas of the United States. Conversions of measurements between the two systems can become quite complex. Fortunately, programs are widely available to transform between popular datums—for example, between NAD 1927 and NAD 1983—for most areas. Datums are constantly being changed and updated. Currently the most up-to-date U.S. horizontal datum is the North American Datum of 1983 (NAD 1983) after the U.S. National Geodetic Survey's (NGS) National Adjustment of 2011 and the most up-to-date U.S. vertical datum is the North American Vertical Datum of 1988 (NAVD 1988). The NGS has further plans to replace these two datums with new geometric and vertical datums including changes in position with time, adding a fourth dimension to the system. A few important datums in North America and globally are listed in Table 5.2.

Types of Projection and Their Characteristics

Theoretically, the number of possible projections to transform coordinates from a spheroid, ellipsoid, or geoid to a planar coordinate system or flat map is unlimited; practically, the number is limited only by the creativity of

TABLE 5.2. Selected Datums

Horizontal datum name	Ellipsoid	Local/ geocentric	Where used
NAD 1927	Clarke 1866	Local	North America
NAD 1983	GRS 1980	Geocentric	North America
WGS 1984	GRS 1980 with additional measurements	Geocentric	World
New Zealand Geodetic Datum (NZGD) 2000	GRS 1980	Geocentric	New Zealand

mathematicians and geodesists and the needs of organizations to coordinate their creation, maintenance, and use of GI for the many public and private uses. To start understanding projections, you should familiarize yourself with the three basic developable surfaces, also called "projection families," used to create map projections. **Developable surfaces**, which are an actual or imaginary drawing of the projection, were used to help cartographers visualize the projection process (see Figure 5.5). They are no longer used to project maps, but they are helpful in understanding projections.

Developable surfaces can be drawn, but many projections are created without them. Projections created with developable surfaces can be demonstrated using a light hung in the middle of a transparent globe or by shining a flashlight through a portion of a globe onto the developable surface. For example, a two-liter plastic bottle, cut off at both ends and marked with a constant interval of vertical lines, with a light bulb hung in the middle to project the lines on a wall, will show how a cylindrical projection projects latitude and longitude on a flat surface. All projections using pseudo-developable surfaces can only be described mathematically. They cannot be created in any mechanical manner.

A key characteristic of all projections, whether developable surface or pseudo-developable, is called aspect (see Figure 5.6). The **projection aspect**

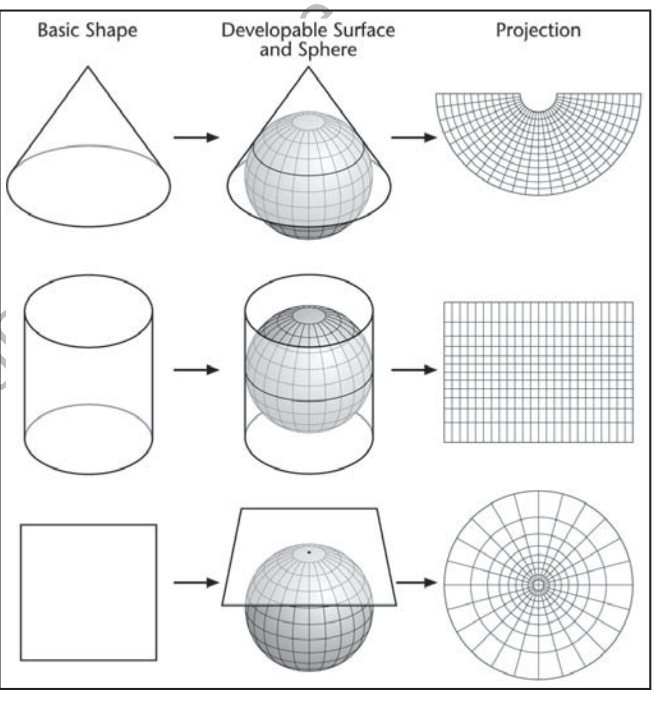

FIGURE 5.5. Basic geometric shapes (cone, cylinder, and plane) serve as developable surfaces, shown here with a reference globe. The resulting projections of latitude and longitude lines are shown in the rightmost column.

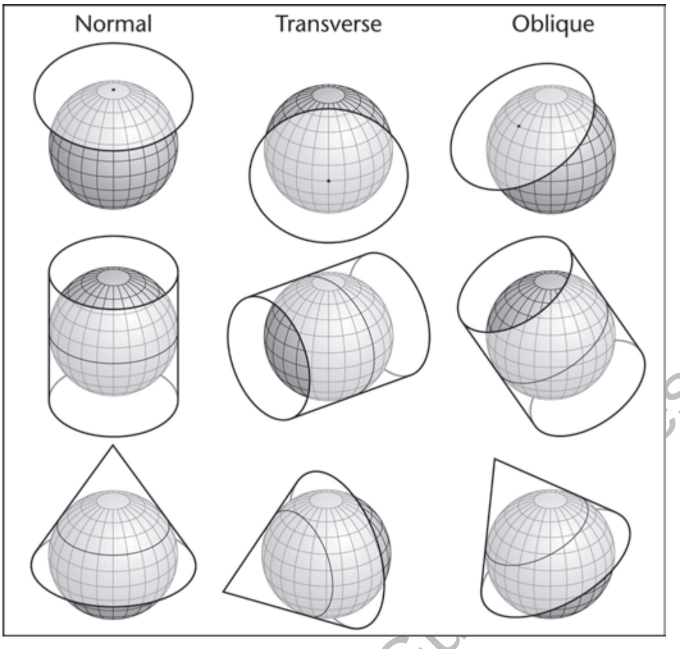

FIGURE 5.6. Some possible aspects for conical, cylindrical, and planar projections.

refers to the orientation of the developable surface to the earth. Various conventions have come and gone in cartography over time. For future users of GI, I think it is most pragmatic to distinguish among equatorial, transverse, oblique, and polar aspects. The differences refer either to the orientation of the projection to a region of the earth (equatorial or polar) or to the developable surface of that type of projection—for example, transverse Mercator projections are rotated 90° from the Mercator projection's usual equatorial orientation. The basic differences are best visualized in a figure showing the different aspect for each developable surface (see Figure 5.6). The consequences for distortion and accuracy are discussed later in this chapter.

Some possible aspects for conical, cylindrical, and planar projection include equatorial and polar. Equatorial orientation has the projection's center positioned somewhere along the equator. Polar aspect occurs only with planar projections. All three projections may have an oblique aspect (based on Jones, 1997, p. 75).

Tangent/Secant

Figure 5.7 illustrates differences in how projections "touch" the developable surface of a reference globe, another important characteristic of projections. These places of contact between the developable surface and spheroid, ellipsoid, or geoid are the most accurate for any projection and are called *standard parallels* or *standard lines*. Tangent projections "touch" the reference

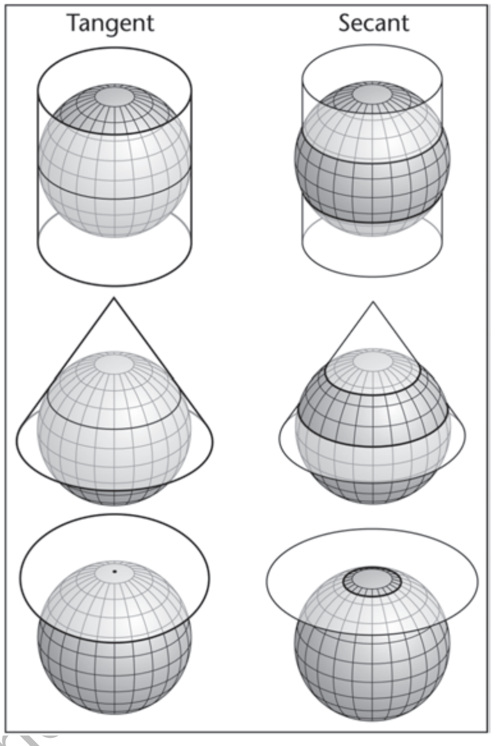

FIGURE 5.7. Examples of tangent and secant projection surfaces.

globe at one point or along one line. Secant projections "touch" the reference globe along two lines or in an area.

Projection Properties

Projections alter the four spatial relationships (angles, areas, distances, and direction) found on a three-dimensional object. As mentioned earlier in the overview of projections, most projections only maintain one of the properties in a specific manner—for example, equidistant projections preserve distance from *one* point to all other points. Many projections, especially projections used for larger areas, compromise all these properties.

The projections that preserve angular relationships from one point are called *conformal*, but you should remember that conformal refers to the preservation of angles only, never shapes. Figure 5.8 includes a Lambert conformal conic projection, which preserves angles, but not areas. If a projection preserves areas in the projection by a constant scaling factor, it is called an *equivalent* projection. Equivalent projections preserve areas, but not shapes. The shapes of continents or countries can change in an equivalent projection, but their areas correspond to the actual areas on the earth (Figure 5.8, Sinusoidal projection). Projections that preserve distances from one or two points to other points are called *equidistant* (Figure 5.8, Stereographic projection). The projections that preserve directions are called *azimuthal*, or true direction projections. Directions are only preserved from the center of the map in azimuthal projections.

Projections that are neither conformal nor equivalent are called *compromise* projections. They are usually developed to make more graphically pleasing maps and do this by finding a balance between areal and angular distortion (Figure 5.8, Robinson projection).

Some Common Projections, Characteristics, and Uses

With so many projections, it is certainly possible to find a projection for every occasion. Fortunately, for most GI uses, the projections are already determined. The choices for maps, especially maps of large areas, are much broader. The following examples highlight a few widely used projections for each of the four projection properties.

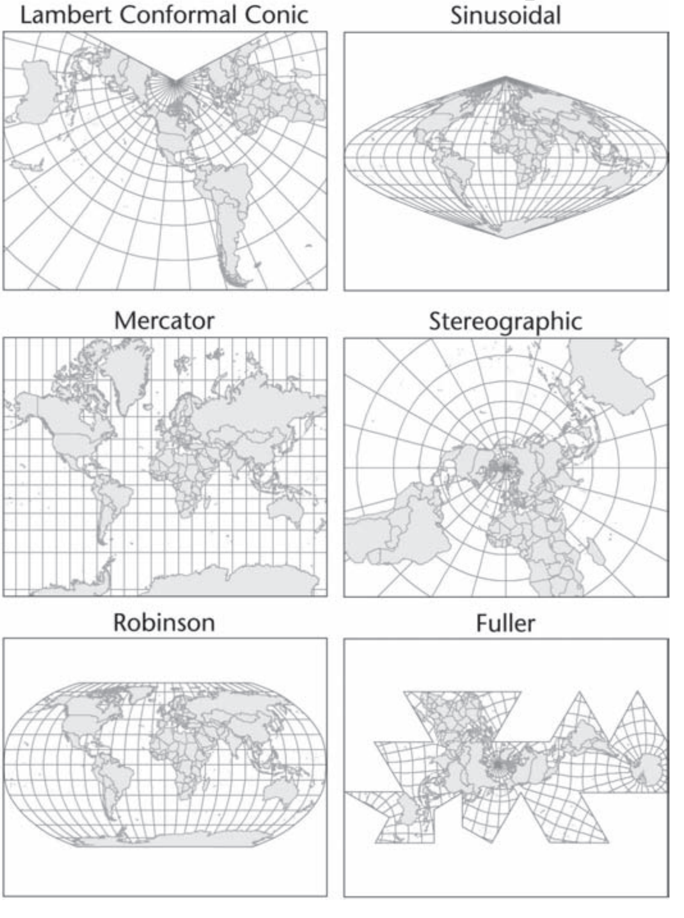

FIGURE 5.8. These six different projections show the countries of the world.

Lambert	The Lambert conformal conic projection preserves only angles. Used for mapping continents or similar areas, it is commonly used for areas with an east–west orientation—for example, the continental United States.
Sinusoidal	The sinusoidal equal area projection preserves areas, but distorts angles and shapes. It is used for maps showing distribution patterns.
Mercator	The very common Mercator projection is a conformal projection with the very unusual quality of showing lines of constant bearing (called *loxodromes* or *rhumb lines*) as straight lines. This made the Mercator projection very valuable for sailors, who could use one single compass heading to determine the direct route between two points. Transverse Mercator projections are widely used for areas with north–south primary orientations (see Figure 5.9).
Stereographic	The widely used stereographic projection is an azimuthal projection developed in the 2nd century B.C.E. that preserves directions; it is a further development of much older stereographic projections. It additionally has the particular quality of showing all great circle routes as straight lines; however, directions are true from only one point on the projection. It is used usually to show airplane navigation routes.

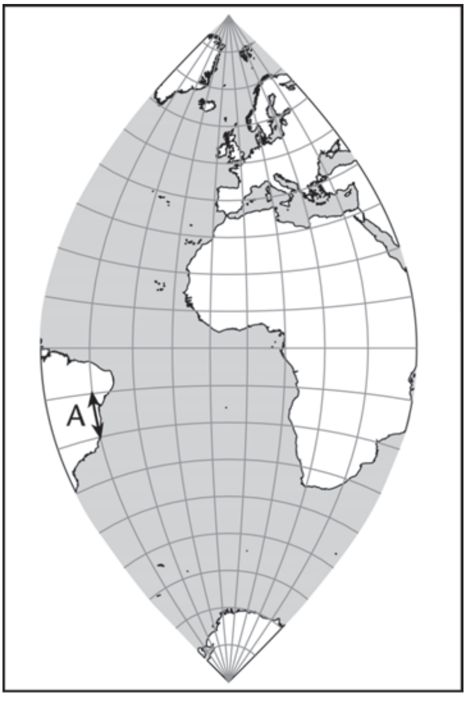

FIGURE 5.9. A transverse Mercator projection.

Robinson	The Robinson projection is a compromise projection that fails to preserve any projection properties. It is graphically attractive; it was adopted by *National Geographic* in 1988 and is widely used elsewhere.
Fuller	The Fuller projection was introduced in 1954 by Buckminster Fuller. It transforms spherical latitude and longitude coordinates to a 20-sided figure called the *icosahedron*.

Calculating Projections

Examining the mathematics of projections is helpful for grasping how a projection transforms locations measured in three dimensions to two-dimensional locations. You should always note that projections are never transformations between two two-dimensional coordinate systems, but between locations found on or near the surface of the three-dimensional planet earth to a two-dimensional coordinate system.

The three examples examined here are widely used. The sinusoidal projection is a pseudo-cylindrical projection developed in the 16th century; the Lambert conformal conic projection is widely used around the world for east-to-west-oriented areas; the Mercator projection is very common. However, the mathematics for each map projection discussed here are quite straightforward, especially since these examples are based on spheroids.

Sinusoidal Projection

The sinusoidal projection is a simple construction that shows areas correctly, but shapes are increasingly distorted away from the central meridian. Parallels of latitude are straight, and longitudinal meridians appear as sine or cosine curves.

The equations for calculating the sinusoidal projection are quite simple. You only need to remember to use radians for the angle measures of longitude and latitude and to place a negative sign in front of longitude values from the western hemisphere.

Equations for calculating a sinusoidal projection:

$$x = R\lambda(\cos \phi)$$
$$y = R\phi$$

where ϕ is the latitude, λ is the longitude, and R is the radius of the earth measured at the scale of map.

Lambert Projection

The cylindrical equal-area projection shown here is one of several projections that Lambert developed in the 18th century. It remains a widely used projection, especially in atlases showing comparisons between different countries or regions of the world. Polar areas have strongly distorted shapes, but most continents evidence only minor distortion.

Below are equations for calculating a Lambert cylindrical equal-area projection:

$$x = R\lambda$$
$$y = R \sin \phi$$

where ϕ is the latitude, λ is the longitude, and R is the radius of the earth measured at the scale of map.

IN DEPTH Calculating Projections with Radians

You may need to use radians for an exercise calculating projections or for other angular measures. The sinusoidal projection, many other projections, and other measures involving angles are often calculated with radians, which is another form of angle measures: $1^\circ = \pi/180$ radians, $360^\circ = 2\pi$ radians. Radians indicate the length of that part of the circle cut off by the angle, and make it easy to determine distances on circular edges or round surfaces.

$$\text{radians} = (\text{degrees} \cdot \pi)/180$$

The length of part of a circle (called an arc) is determined by multiplying the number of radians by the radius. For example, the length of an arc defined by an angle of 10° on a circle with a 100-m radius is 0.1745.

1. Determine radian measure of angle:

$$n \text{ radians} = (10^\circ \times \pi/180)$$
$$n \text{ radians} = 0.1745$$

2. Calculate length of the arc:

$$\text{arc length} = n \text{ radians} \times \text{radius}$$
$$\text{arc length} = 0.1745 \times 100 \text{ m}$$
$$\text{arc length} = 17.45$$

Some common angle measures in degrees and their equivalents in radians are listed here.

Degrees	**Radians**
90°	π/2
60°	π/3
45°	π/4
30°	π/6

Mercator Projection

The very common Mercator projection uses only slightly more complicated equations.

Below are equations for a Mercator projection (Snyder, 1993):

$$x = R\lambda$$
$$y = R \ln \tan(\pi/4 + \phi/2)$$

where λ is the longitude (– if in the western hemisphere) for determining values of the y axis and ϕ is the latitude (+ if north, – if south of the equator) and R is the radius of the earth measured at the scale of the map. The term *ln* refers to the natural logarithm to the base *e*. All angles again are measured in radians.

Distortions

Distortions arising through projections are unavoidable. They have significant consequences for accuracy, so it helps to know more about distortions in order to choose the best projection for different purposes and to be able to take distortions into account.

As a general place to start out, we can categorize distortions in terms we have already seen: the four projection properties of angles, areas, distances, and direction. Many projections distort one or two of these projection properties. Distortion of angles (including shapes) is sometimes easy to detect, especially for large areas when familiar shapes of states, continents, or even provinces are distorted; but projections of small areas may lack readily visible evidence of distortions and require the use of special graphics or statistical measures to determine the distortions. The same applies to areas. The distortions arising related to distance can be significant because, as you know now, no projection for large areas accurately shows distances for all points, but can only be accurate for a few points. Small areas are another matter, but you still should check to see what distortion a projection creates. Direction can likewise be distorted in a subtle fashion that is not visually noticeable, but is of significance should the map be used for navigation purposes.

One easily overlooked source of distortions is the difference between the datums, geoids, and ellipsoids used in creating different GI or maps. Even if GI or a map is made by the same agency or company using the same projection, a change in the datum, geoid, or ellipsoid can lead to distortions when compared with other GI or maps for the same area.

Describing Distortions

To describe and assess distortions, it is useful to determine the scale factor (SF) at different places on a map. By comparing scale factors with map scale

The Case of Changing Projections and Datums in Belgium

Not only famous for beer and chocolate, and as the site of much of the European Union administration, Belgium has a rich history in geography and cartography. Datums specify the earth's size and shape as well as specifying the origin and orientation of a coordinate system. This brief history of changes made to projections and datums in Belgium shows why datums change. When two geographic areas (countries) use different projections and datums, points on the border between the two areas will not match up.

After World War II Belgium adopted a new geodetic system, which was based on the Hayford ellipsoid (1924), with its initial point located at the Brussels Observatory. The Lambert conformal conic projection with two standard parallels was the projection chosen to create Belgium's cartographic grid. On this basis a cartographic series at the scale of 1:25,000 was established. By 1972, it had become necessary to create a new geodetic datum, Belgium Datum 1972, which was based on a new global compensation. This datum also used the Hayford ellipsoid, but the initial point had shifted since 1950. New project parameters were defined, constituting the Belgian Lambert 72 (BD72) system. Most current topographical maps in Belgium use the BD72 coordinates. The system can only be used in Belgium.

Spatial geodesy's advances in the second half of the 20th century have made it possible to determine and track over time the shape of the geoid and position of the earth's center of mass with great accuracy. (They are constantly changing, notably due to plate tectonics.) The center of mass is the starting point for a system with three perpendicular axes (*x, y, z*). Two are on the plane of equator and the third corresponds to the direction of the poles. A point may be localized by any triplet of Cartesian coordinates and may be below, above, or on the surface of the earth. The International Terrestrial Reference System (ITRS) and World Geodetic System (WGS) are based on this principle. A reference framework based on satellite instruments and earth observations determines the *x, y, z* coordinates of several hundred points on earth. The resulting "Frame" is referred to by the acronym (initialism) of the reference system (e.g., ITRS) followed by the year of the observations. In Belgium and other countries, geodesists globally adjust ellipsoids to the shape of the geoid and center them on one center of mass, which guides how longitude and latitude are determined. The Geodetic Reference System (GRS80) of 1980 was followed by the World Geodetic System of 1984 (WGS84); it had a slightly different flattening coefficient and definition of the geodetic datum and ellipsoid. The problem of different countries having different geodetic systems resulted in a push to standardize the use of WGS84. More recently, Belgium has begun to implement European Terrestrial Reference System 1989 (ETRS89). This is the EUREF system recommended for European cartographical and topographical activities. In Belgium a GPS collection of point locations for around 4,200 points using ETRS89 coordinates was used to create the Belgian geodetic frame, or BEREF. The GRS80 ellipsoid is associated with BEREF and was used as the basis for the new 2008 Belgian Lambert projection. Once the BEREF was completed, a new version of Belgian Lambert went into use with the coordinates of the central meridian and standard parallels (49-50 N and 51-10 N) defined on the global ellipsoid GRS80 and lined with ETRS89. To avoid confusion with the 1972 version, the coordinates were given a false origin in the 2008 version by adding approximately 500 km. Though projections and datums must change due to changes in the earth's crust, discretionary decisions can be made, to help avoid confusion among the different systems in use.

Source: National Committee of Geography of Belgium. (2012). *A concise geography of Belgium.* Ghent, Belgium: Academia Press.

at the standard point or standard lines you can assess the scale distortions using this formula:

$$\text{Scale factor} = \frac{\text{Local scale}}{\text{Principle scale}}$$

where *local scale* is the scale calculated at a particular place and *principle scale* is the scale computed at the standard point or a standard line (see Table 5.3).

For example, the scale factor of a transverse Mercator projection with a principle scale of 1:400,000,000 calculated between 20° and 30° S will indicate how much distortion the projection introduces. First, calculate the local scale by measuring along the meridian between 20° and 30° S. This gives you the map distance, which is 3.1 cm (1.2 in.) (shown in Figure 5.9 with the letter A). You compute the ground distance between the same portion of the meridian by consulting a table showing the lengths of a degree of latitude along a meridian. At 20° a degree of latitude is 110,704.278 m long. Multiplying the 10° of latitude to 30° S would measure approximately 1,107,042.78 m or 1,107.04 km. Second, by substituting the 3.1 cm and 1,107.04 km into the map scale equation, you can calculate the local scale:

Map scale = earth distance/map distance
Map scale = 1,107.04 km/3.1 cm

The units in the equation must be equal, so you first need to convert kilometers to centimeters by multiplying by 100,000.

1,107.04 km × 100,000 = 110,704,000 cm

Calculate the local map scale:

Map scale = 110,704,000 cm/3.1 cm
Map scale = 1:35,710,967.74

TABLE 5.3. Table of Meridian Distances for Various Latitudes

Latitude (°)	Miles	Kilometers
0	68.71	110.57
10	68.73	110.61
20	68.79	110.70
30	68.88	110.85
40	68.99	111.04
50	69.12	111.23
60	69.23	111.41
70	69.32	111.56
80	69.38	111.66
90	69.40	111.69

This map scale is considerably larger than the map scale along the standard line of 1:30,000,000. You can now compute the scale factor using the scale factor equation:

$$\text{Scale factor} = \frac{35{,}710{,}967.74}{30{,}000{,}000}$$

$$\text{Scale factor} = 1.19$$

This scale factor suggests that the distances in the transverse Mercator projection increase away from the central meridian. A visual check of the projected map supports this conclusion.

Tissot Indicatrix

A visual analytic to display and examine projection distortions was developed by the mathematician Nicholas Tissot in the 19th century. The concept is simply that any small circle on a spheroid or ellipsoid, when projected to the same point on the flat map, will show the distortion created by the map projection for that area through the projected shape and size of the circle. When the circles are plotted at various points on a map, they allow for a visual comparison of distortion. You should note that the changed shapes and sizes of the indicatrix refer to individual points and cannot be used in evaluating distortion of continents or water bodies.

The indicatrix has two characteristics that can be used to evaluate distortion. The first is the two radii, semimajor (a) and semiminor (b), which are perpendicular to each other (see Figure 5.10). The semimajor axis is aligned in the direction of the maximum SF and the semiminor axis is aligned in the direction of the minimum SF. The second is the angle between two lines l and m that intersect the center of the indicatrix circle, but are turned 45° in respect to the center, if there is no angular distortion. The distances of the semimajor and the semiminor axes, respectfully, indicate the scale factor distortion along each axis. The angle between two lines l and m indicates the amount of angular distortion. For example, a circle where l and m intersect

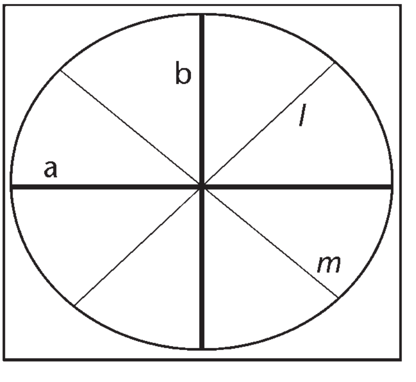

FIGURE 5.10. Tissot's indicatrix circle indicating no areal and no angular distortion.

FIGURE 5.11. Two Tissot indicatrix circles shown on a Mercator projection with the standard line of the equator.

at right angles indicates no distortion. If the shape of the circle is distorted into an ellipse, but the area is the same as the circle and the two lines l and m intersect at angles greater or less than 90°, there is no areal distortion, but there is angular distortion.

A map showing multiple Tissot indicatrix circles is a valuable aid to determine projection distortion (see Figure 5.11). The revealed patterns of distortion help in choosing the appropriate projection for a particular area.

Combining GI from Different Projections

The large number of projections available means that great care must be taken when working with GI from different sources. Projections for GIS provide a great deal of flexibility, but easily introduce problems when working with data created using different projections. You should note that projections used for GI differ from maps in an important way. When a map is made, one single projection is used with a single scale for the entire map. The same thing applies for GI with one important difference: the coordinate system of the GI usually is much larger than a piece of paper used for a map. The GI must be scaled another time when a map is made, which can introduce some distortion. Obviously, if the GI is stored in the coordinates of a piece of paper, it is much harder to use it with other data, so this makes sense.

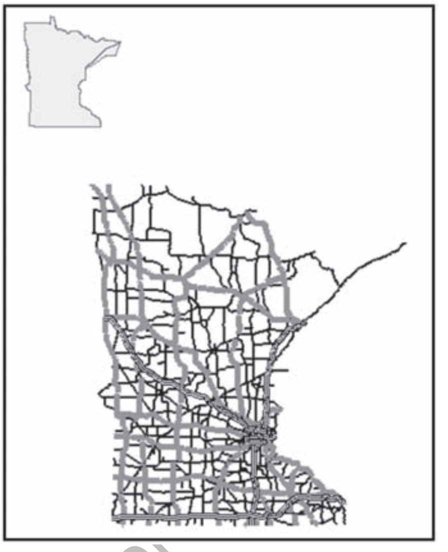

FIGURE 5.12. An example of an obvious error resulting from using data sources for the same area (Minnesota) but with different projections. Diagnosing the causes of such errors and resolving them can be very time-consuming if information about the choices made regarding the respective projections is unavailable.

Assuming that different data sets of GI for the same area use the same projection can lead to vast problems (see Figure 5.12). Usually the problems when combining GI from different projections are so obvious that they can't be missed. Sometimes the distortions are slight and may seem inexplicable: a road from one data source is 2 m away from the property that runs along it from another data source. If care is not taken, it is possible to create great errors by combining data prepared from different projections. The same concern applies to coordinate systems, the topic for the next chapter, where we will look at these issues in more detail.

Summary

Projections have been the core of cartography and the basis for representing GI. For millennia people have developed projections to find ways to represent the three-dimensional world humans live on in two dimensions—a format much better suited for recording observations and measurements. While it is possible to make maps without a projection, unprojected GI or maps are inaccurate and distorted for all larger areas and many small areas. A projection can be applied to different areas and at different scales. The smaller the area, the more accurate a projection can be. How accurate the projection is depends on how the projection is constructed and what underlying model of the earth's form it uses. Basic characteristics of a projection are its orientation, tangency, and form. Projections have several properties. The most important properties are the preservation of angles (conformality) and the preservation of areas (equivalent). Only one of these two properties can be preserved in any one projection. Some projections distort both properties and are called compromise projections. The resulting distortions can be ascertained and described using a Tissot indicatrix. Because of the number of differences, it is important to assess the characteristics and properties of projections when working with GI, especially when combining GI from different sources.

REVIEW QUESTIONS

1. Identify the type (equal angle, equal area, compromise) of the following projections:

   Mercator   Lambert   Mollweide   sinusoidal   azimuthal   Robinson

2. What is the difference between a secant and a tangent projection?
3. What is a transverse projection?
4. Why is a transverse Mercator projection better for north–south oriented areas and states (e.g., Illinois) than a Lambert conformal conic projection?
5. What are the three important characteristics of projections?
6. Why is most GI projected to a two-dimensional, Cartesian coordinate system?
7. Why should you never combine GI from different projections?
8. How can positional distortion be measured?
9. What is the difference between a geoid and a spheroid?
10. Why are Mercator and Peters projections technically satisfactory? Why do people consider the Mercator projection to be a bad projection?

ANSWERS

1. Identify the type (equal angle, equal area, compromise) of the following projections:

Mercator	Lambert	Mollweide	sinusoidal	azimuthal	Robinson
(Equal shape)	(Equal area)	(Equal area)	(Equal area)	(Equal distance)	(Compromise)

2. What is the difference between a secant and a tangent projection?

   A secant projection surface "touches" the earth's surface in two places; a tangent projection "touches" only at one.

3. What is a transverse projection?

   A transverse projection is a cylindrical projection, which is normally oriented east–west, rotated 90 degrees to a north–south orientation.

4. Why is a transverse Mercator projection better for north–south oriented areas and states (e.g., Illinois) than a Lambert conformal conic projection?

   The conic projection works best for areas with an east–west orientation; its line(s) of tangency run east–west. The transverse Mercator projection's line of tangency runs north–south, providing a more accurate positional reference than the Lambert conformal conic projection of the same area.

5. What are the three important characteristics of projections?

   Equal shape: preservation of shapes; equal area: preservation of areas; equal distance; preservation of distances

6. Why is most GI projected to a two-dimensional, Cartesian coordinate system?

   Several reasons need to be considered. Much GI comes from maps with such coordinate systems. Most GI is used to make planar maps. Most GIS are designed to store two-dimensional coordinate locations.

7. Why should you never combine GI from different projections?

   GI from different projections for the same area will be in different coordinate systems that do not align properly.

8. How can positional distortion be measured?

   For small-scale maps, Tissot's indicatrix provides a good graphical indicator. Large-scale maps, showing small areas, require the use of statistical measures.

9. What is the difference between a geoid and a spheroid?

   A geoid is a more accurate representation of the earth's surface, accounting for local variations. A spheroid is a more round form that fails to account for local variations and the oblateness of the earth resulting from its spin.

10. Why are Mercator and Peters projections technically satisfactory? Why do people consider the Mercator projection to be a bad projection?

    The Mercator projection is well suited for compass navigation at sea. The Peters projection is a compromise that offers a different way of representing the world. The overuse and ill-suited use of the Mercator projection to show regions of the world has led to the Mercator acquiring a bad reputation.

Chapter Readings

Jones, C. (1997). *Geographical information systems and computer cartography*. Upper Saddle River, NJ: Prentice Hall.

For a fascinating, if wide-reaching, biography and study of a person who was instrumental in determining the elliptical shape of the earth, see

Terrall, M. (2002). *The man who flattened the earth: Maupertuis and the sciences in the Enlightenment*. Chicago: University of Chicago Press.
For information about the basic mathematical principles of cartography, see
Cotter, C. H. (1966). *The astronomical and mathematical foundations of geography*. New York: Elsevier.
For a history of projections, see
Montgomery, S. (1996). Naming the heavens: A brief history of earthly projections. *Science as Culture, 5*(25), 546–587.
For a very thorough history of projections, see
Snyder, J. P. (1993). *Flattening the earth: Two thousand years of map projections*. Chicago: University of Chicago Press.

Web Resources

- More specific information about projection parameters and accuracy is provided by government agencies, for example, the California Department of Fish and Game: *www.dfg.ca.gov/biogeodata/gis/pdfs/DFG_Projection_and_Datum_Guidelines.pdf*.
- A good resource for fundamentals of geodesy is provided by the U.S. National Geospatial-Intelligence Agency, *Geodesy for the Layman*, available online at *www.ngs.noaa.gov/PUBS_LIB/Geodesy4Layman/toc.htm*.
- For information about homemade map projections using plastic bottles and the like, see *http://octopus.gma.org/surfing/imaging/mapproj.html*.
- This first of four articles offers a very well written and detailed discussion of the new U.S. datums: Minkel, D. H., & Dennis, M. L. (2012). Frames for the future: New datum definitions for modernization of the U.S. NSRS (Part 1 of 4). *The American Surveyor, 9*(1). Available online at *www.amerisurv.com/content/view/9609*.

EXERCISES

1. Projections for Different Needs

If you collect maps from magazines and newspapers for a few weeks, you will have a pretty sizable collection of different kinds of maps and different kinds of projections.

Come up with a list of different uses of maps and the projections used for each.

Think about how projections can preserve the shape of things on the earth, their size, or the distance from a point or along a line, or must compromise between these three projection properties.

Knowing what you do now about the different qualities of projections, what do you think about newspaper maps that do not indicate the projection? Are they common? What kind of errors do you think can arise?

If possible, you also can explore the collection of maps and atlases in a nearby library.

2. Questions for Map Projections

1. Is the map whole or broken up?
2. What shape does the projection make the map?
3. How are features (continents and islands) arranged?
4. Are gridlines curved or straight?
5. Do parallels and meridians cross at right angles?

EXTENDED EXERCISE

3. Sinusoidal Projection

Overview

In this exercise you will calculate values for a sinusoidal projection that you produce.

Concepts

The location of a point (x, y) in a sinusoidal equal area projection is calculated for this exercise in two steps. First, the longitude value is transformed to east–west values (x) by multiplying the longitude value times the radius and times the cosine of the latitude. Multiplying the longitude values by a cosine of latitude creates the gradually increasing distortion of areas farther away from the equator. The north–south values (y) of the projection are calculated through a linear relationship between the radius and the latitude. Second, you will scale the calculated x and y values to fit a map on a piece of paper by determining a scale ratio that transforms the radius of the sphere (6,371 km).

Exercise Steps and Questions

Preparation

In this exercise you will be calculating a projection of a graticule. You will have to do the calculations and show that you have done them, but you can work with other people to check your answers and determine the process. Before the calculating part of this exercise, let's look at the fundamental problems of projecting a spherical object on a plane.

Part 1. Angle Measures: Degrees and Radians

In Part 2 of this exercise, you will need to make the calculations in radians. Radians are one of three ways to measure angles. They are mainly used for engineering and science. We won't spend much time getting into the mathematics of angular measures. For this exercise, you only need to understand the relationship between degree and radian measures of angles.

If you know an angle measure in degrees, you can easily convert it to radians, another measure for angles used in engineering and scientific calculations:

$$\text{radians} = (\text{degrees} \cdot \pi)/180$$

For example, 180 degrees equals 3.14 radians; 90 degrees equals 1.57 radians; 45 degrees equals 0.785 radians. As the examples show, radians express angular measures in relation to the radius.

Part 2. Construct a Sinusoidal Projection of a Graticule

STEP 1: CALCULATE THE PROJECTION.

Use the table below for recording the results of your calculations. The rows indicating latitude are on the left and the columns indicating longitude are on the top. You will be calculating the sinusoidal projection for latitudes 0°, 30°, 60°, and 90°, and for longitudes 0°, 30°, 60°, 90°, 120°, 150°, and 180°. Your results will be in kilometers, or, for an idealized projection surface, about 10,000 km in length and height.

The equations you will use are:

$$x = \text{radius} \cdot \text{longitude} \cdot \text{cosine (latitude)}$$
$$y = \text{radius} \cdot \text{latitude}$$

where radius = 6,371 km. Remember: Convert all angle measures from degrees to radians by multiplying by pi and dividing by 180 degrees. For example, 30° corresponds to $\pi/6$ *using the conversion equation from above.*

Table of Projected Values (Step 1)

Latitude	Longitude						
	0°	30°	60°	90°	120°	150°	180°
0°	0,0						
30°							
60°							
90°							

STEP 2: SCALE THE *X, Y* VALUES AND THEN GRAPH THEM.

The *x, y* values calculated in Step 1 are in kilometers; therefore they are certainly too large to fit on a piece of paper. As with creating any other map, the values need to be converted to map units by determining a ratio that fits the *x, y* values on the sheet of paper you use (8.5 × 11 inches, or approximately 22 × 33 cm). Scale can be determined by putting the ground values and map values in the same units, here cm, and calculating the ratio between the shortest ground value distance and the longest map value distance.

Determine this value and fill it in here:

Scale factor: ____________________

With the scale factor, convert your original projected values to map units. Use the table below for those calculations.

Table of Projected Values (Step 2)

Latitude	Longitude						
	0°	30°	60°	90°	120°	150°	180°
0°	0,0						
30°							
60°							
90°							

Using a ruler, graph each coordinate pair on the *x* and *y* axis on a separate piece of paper. The graph should look like the northeastern quadrant of a sinusoidal projection. When this is completed, label the axis with tick marks that indicate the corresponding degree value from 0° to 90° latitude and 0° to 180° longitude. This is a map projected to a sinusoidal projection.

Questions

1. The sinusoidal projection is an example of an equal-area projection. What are the major differences between this type of projection and conformal projections?
2. Why do the *x* values lack two-dimensional scaling at 0° longitude in the sinusoidal projection?
3. What are the major differences between the Mercator and sinusoidal projections? How big is a pole in each projection?
4. Minneapolis/St. Paul is located at approximately –93° longitude, 45° latitude. What are the *x, y* coordinates in the sinusoidal projection?