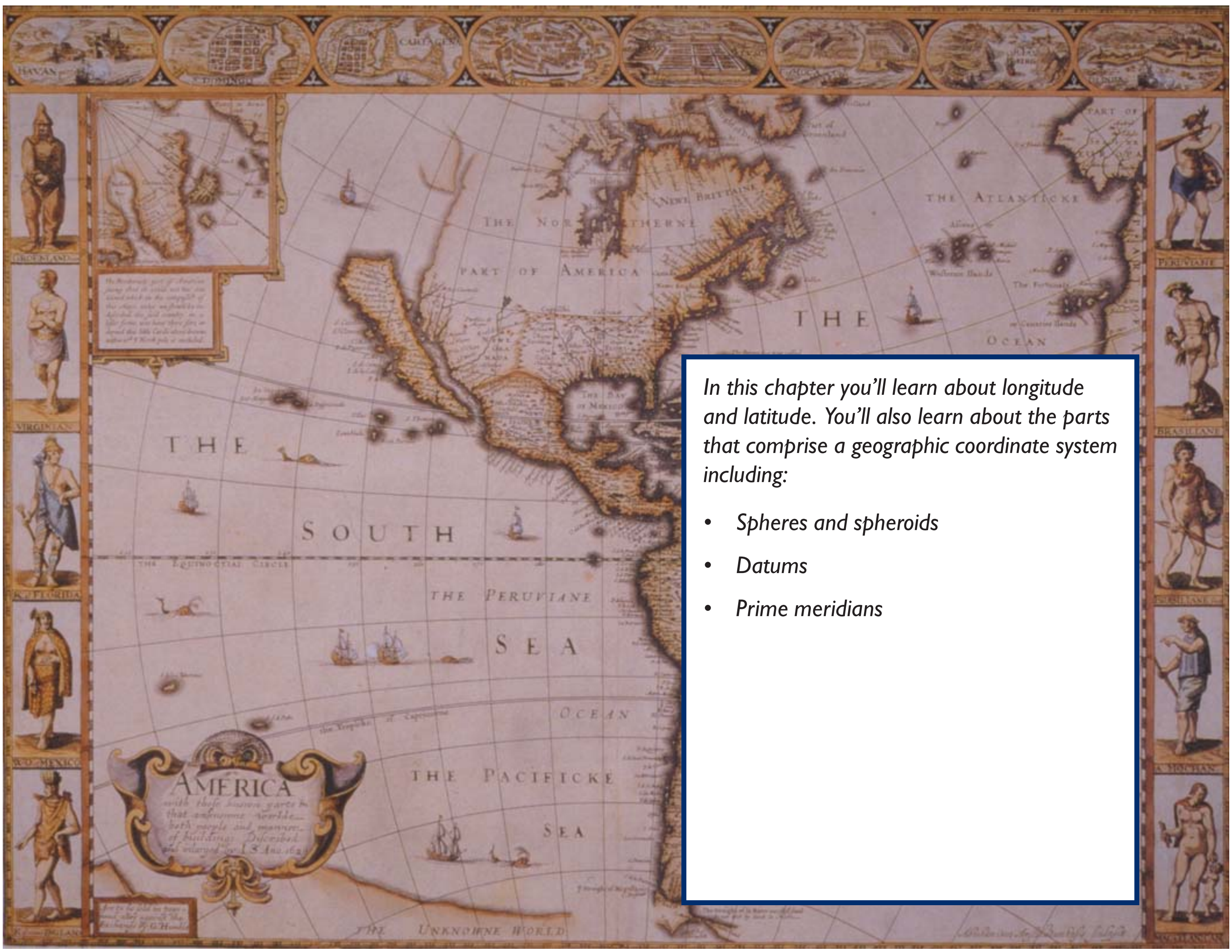
Understanding Map Projections

GIS by ESRI®

Contents

1 Geographic coordinate systems

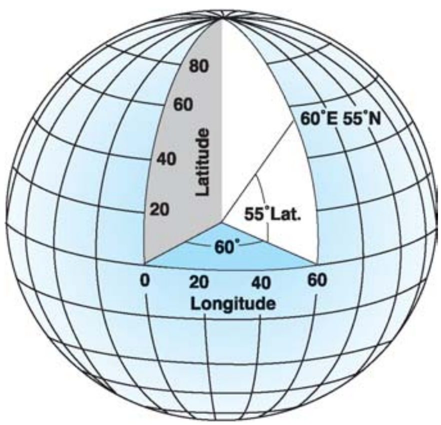


In this chapter you'll learn about longitude and latitude. You'll also learn about the parts that comprise a geographic coordinate system including:

- *Spheres and spheroids*
- *Datums*
- *Prime meridians*

A *geographic coordinate system* (GCS) uses a three-dimensional spherical surface to define locations on the earth. A GCS is often incorrectly called a datum, but a datum is only one part of a GCS. A GCS includes an angular unit of measure, a prime meridian, and a datum (based on a spheroid).

A point is referenced by its *longitude* and *latitude* values. Longitude and latitude are angles measured from the earth's center to a point on the earth's surface. The angles often are measured in degrees (or in grads).



The world as a globe showing the longitude and latitude values.

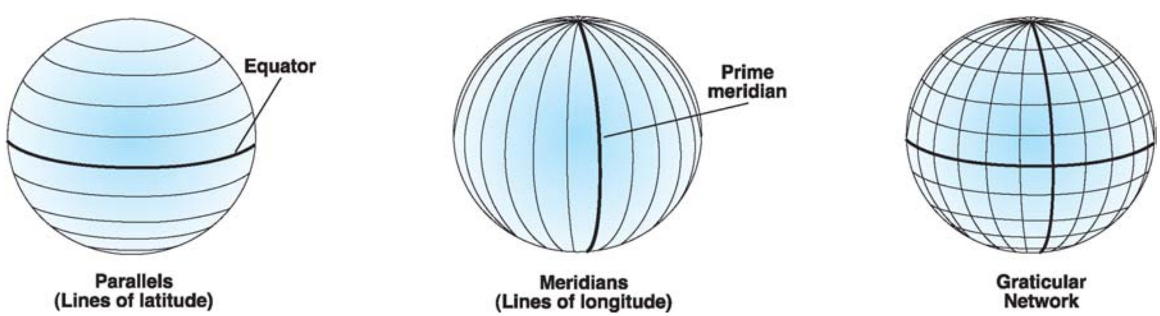
In the spherical system, horizontal lines, or east–west lines, are lines of equal latitude, or *parallels*. Vertical lines, or north–south lines, are lines of equal longitude, or *meridians*. These lines encompass the globe and form a gridded network called a *graticule*.

The line of latitude midway between the poles is called the equator. It defines the line of zero latitude. The line of zero longitude is called the prime meridian. For most geographic coordinate systems, the prime meridian is the longitude that passes through Greenwich, England. Other countries use longitude lines that pass through Bern, Bogota, and Paris as prime meridians.

The origin of the graticule (0,0) is defined by where the equator and prime meridian intersect. The globe is then divided into four geographical quadrants that are based on compass bearings from the origin. North and south are above and below the equator, and west and east are to the left and right of the prime meridian.

Latitude and longitude values are traditionally measured either in decimal degrees or in degrees, minutes, and seconds (DMS). Latitude values are measured relative to the equator and range from -90° at the South Pole to +90° at the North Pole. Longitude values are measured relative to the prime meridian. They range from -180° when traveling west to 180° when traveling east. If the prime meridian is at Greenwich, then Australia, which is south of the equator and east of Greenwich, has positive longitude values and negative latitude values.

Although longitude and latitude can locate exact positions on the surface of the globe, they are not uniform units of measure. Only along the equator does the distance represented by one degree of longitude

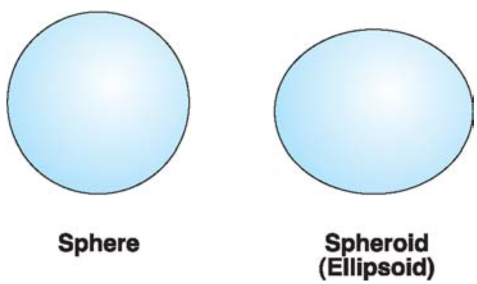


The parallels and meridians that form a graticule.

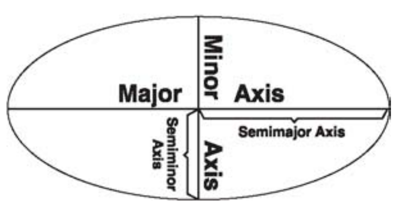
approximate the distance represented by one degree of latitude. This is because the equator is the only parallel as large as a meridian. (Circles with the same radius as the spherical earth are called *great circles*. The equator and all meridians are great circles.)

Above and below the equator, the circles defining the parallels of latitude get gradually smaller until they become a single point at the North and South Poles where the meridians converge. As the meridians converge toward the poles, the distance represented by one degree of longitude decreases to zero. On the Clarke 1866 spheroid, one degree of longitude at the equator equals 111.321 km, while at 60° latitude it is only 55.802 km. Since degrees of latitude and longitude don't have a standard length, you can't measure distances or areas accurately or display the data easily on a flat map or computer screen.

The shape and size of a geographic coordinate system's surface is defined by a sphere or spheroid. Although the earth is best represented by a spheroid, the earth is sometimes treated as a sphere to make mathematical calculations easier. The assumption that the earth is a sphere is possible for small-scale maps (smaller than 1:5,000,000). At this scale, the difference between a sphere and a spheroid is not detectable on a map. However, to maintain accuracy for larger-scale maps (scales of 1:1,000,000 or larger), a spheroid is necessary to represent the shape of the earth. Between those scales, choosing to use a sphere or spheroid will depend on the map's purpose and the accuracy of the data.

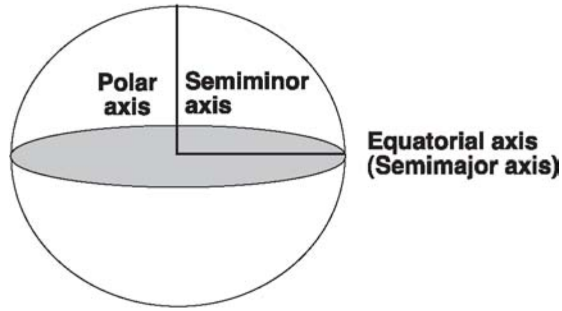


A sphere is based on a circle, while a spheroid (or ellipsoid) is based on an ellipse. The shape of an ellipse is defined by two radii. The longer radius is called the semimajor axis, and the shorter radius is called the semiminor axis.



The major and minor axes of an ellipse.

Rotating the ellipse around the semiminor axis creates a spheroid. A spheroid is also known as an oblate ellipsoid of revolution.



The semimajor axis and semiminor axis of a spheroid.

A spheroid is defined by either the semimajor axis, *a,* and the semiminor axis, *b,* or by *a* and the *flattening*. The flattening is the difference in length between the two axes expressed as a fraction or a decimal. The flattening, *f,* is:

```
f = (a - b) / a
```

The flattening is a small value, so usually the quantity *1/f* is used instead. The spheroid parameters for the World Geodetic System of 1984 (WGS 1984 or WGS84) are:

```
a  =  6378137.0  meters
1/f  =  298.257223563
```

The flattening ranges from zero to one. A flattening value of zero means the two axes are equal, resulting in a sphere. The flattening of the earth is approximately 0.003353.

Another quantity, that, like the flattening, describes the shape of a spheroid, is the square of the *eccentricity,* e^2. It is represented by:

$$e^2 = \frac{a^2 - b^2}{a^2}$$

DEFINING DIFFERENT SPHEROIDS FOR ACCURATE MAPPING

The earth has been surveyed many times to help us better understand its surface features and their peculiar irregularities. The surveys have resulted in many spheroids that represent the earth. Generally, a spheroid is chosen to fit one country or a particular area. A spheroid that best fits one region is not necessarily the same one that fits another region. Until recently, North American data used a spheroid determined by Clarke in 1866. The semimajor axis of the Clarke 1866 spheroid is 6,378,206.4 meters, and the semiminor axis is 6,356,583.8 meters.

Because of gravitational and surface feature variations, the earth is neither a perfect sphere nor a perfect spheroid. Satellite technology has revealed several elliptical deviations; for example, the South Pole is closer to the equator than the North Pole. Satellite-determined spheroids are replacing the older ground-measured spheroids. For example, the new standard spheroid for North America is the Geodetic Reference System of 1980 (GRS 1980), whose radii are 6,378,137.0 and 6,356,752.31414 meters.

Because changing a coordinate system's spheroid changes all previously measured values, many organizations haven't switched to newer (and more accurate) spheroids.

While a spheroid approximates the shape of the earth, a datum defines the position of the spheroid relative to the center of the earth. A datum provides a frame of reference for measuring locations on the surface of the earth. It defines the origin and orientation of latitude and longitude lines.

Whenever you change the datum, or more correctly, the geographic coordinate system, the coordinate values of your data will change. Here's the coordinates in DMS of a control point in Redlands, California, on the North American Datum of 1983 (NAD 1983 or NAD83).

```
-117 12 57.75961   34 01 43.77884
```

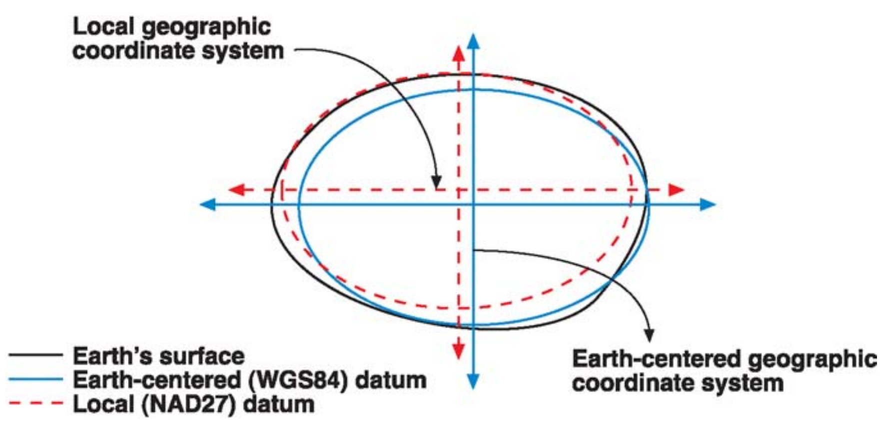
Here's the same point on the North American Datum of 1927 (NAD 1927 or NAD27).

```
-117 12 54.61539   34 01 43.72995
```

The longitude value differs by about three seconds, while the latitude value differs by about 0.05 seconds.

In the last 15 years, satellite data has provided geodesists with new measurements to define the best earth-fitting spheroid, which relates coordinates to the earth's center of mass. An earth-centered, or geocentric, datum uses the earth's center of mass as the origin. The most recently developed and widely used datum is WGS 1984. It serves as the framework for locational measurement worldwide.

A local datum aligns its spheroid to closely fit the earth's surface in a particular area. A point on the surface of the spheroid is matched to a particular position on the surface of the earth. This point is known as the origin point of the datum. The coordinates of the origin point are fixed, and all other points are calculated from it. The coordinate system origin of a local datum is not at the center of the earth. The center of the spheroid of a local datum is offset from the earth's center. NAD 1927 and the European Datum of 1950 (ED 1950) are local datums. NAD 1927 is designed to fit North America reasonably well, while ED 1950 was created for use in Europe. Because a local datum aligns its spheroid so closely to a particular area on the earth's surface, it's not suitable for use outside the area for which it was designed.



The two horizontal datums used almost exclusively in North America are NAD 1927 and NAD 1983.

NAD 1927

NAD 1927 uses the Clarke 1866 spheroid to represent the shape of the earth. The origin of this datum is a point on the earth referred to as Meades Ranch in Kansas. Many NAD 1927 control points were calculated from observations taken in the 1800s. These calculations were done manually and in sections over many years. Therefore, errors varied from station to station.

NAD 1983

Many technological advances in surveying and geodesy—electronic theodolites, Global Positioning System (GPS) satellites, Very Long Baseline Interferometry, and Doppler systems—revealed weaknesses in the existing network of control points. Differences became particularly noticeable when linking existing control with newly established surveys. The establishment of a new datum allowed a single datum to cover consistently North America and surrounding areas.

The North American Datum of 1983 is based on both earth and satellite observations, using the GRS 1980 spheroid. The origin for this datum is the earth's center of mass. This affects the surface location of all longitude–latitude values enough to cause locations of previous control points in North America to shift, sometimes as much as 500 feet. A 10-year multinational effort tied together a network of control points for the United States, Canada, Mexico, Greenland, Central America, and the Caribbean.

The GRS 1980 spheroid is almost identical to the WGS 1984 spheroid. The WGS 1984 and NAD 1983 coordinate systems are both earth-centered. Because both are so close, NAD 1983 is compatible with GPS data. The raw GPS data is actually reported in the WGS 1984 coordinate system.

HARN OR HPGN

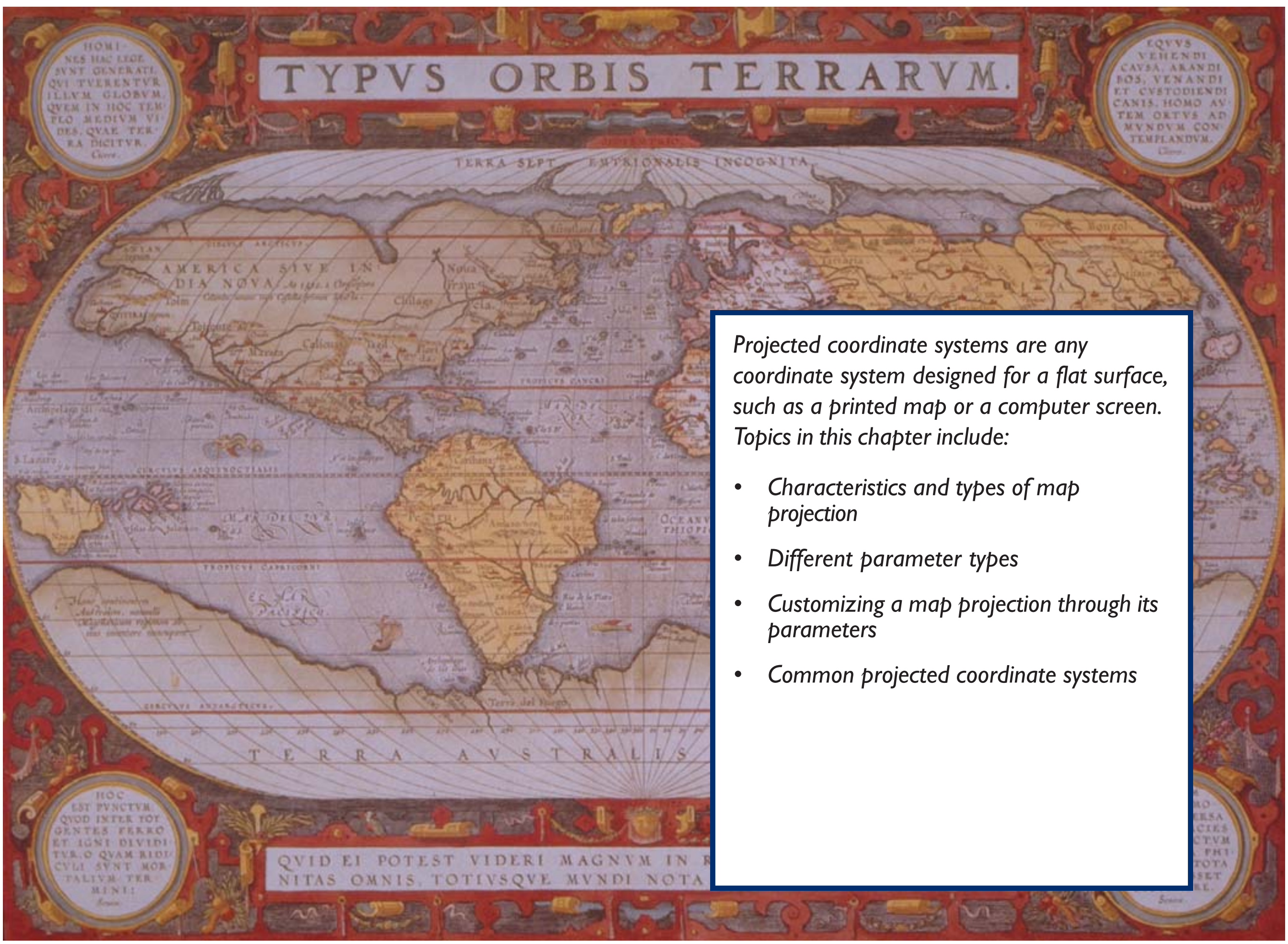
There is an ongoing effort at the state level to readjust the NAD 1983 datum to a higher level of accuracy using state-of-the-art surveying techniques that were not widely available when the NAD 1983 datum was being developed. This effort, known as the High Accuracy Reference Network (HARN), or High Precision Geodetic Network (HPGN), is a cooperative project between the National Geodetic Survey and the individual states.

Currently, all states have been resurveyed, but not all of the data has been released to the public. As of September 2000, the grids for 44 states and two territories have been published.

OTHER UNITED STATES DATUMS

Alaska, Hawaii, Puerto Rico and the Virgin Islands, and some Alaskan islands have used other datums besides NAD 1927. See Chapter 3, 'Geographic transformations', for more information. New data is referenced to NAD 1983.

2 Projected coordinate systems



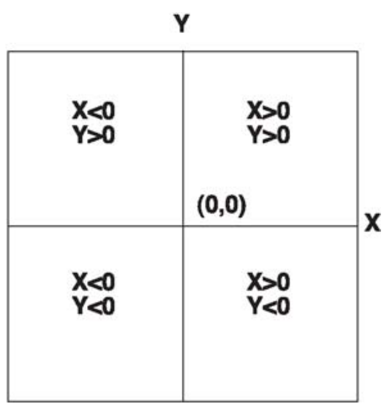
Projected coordinate systems are any coordinate system designed for a flat surface, such as a printed map or a computer screen. Topics in this chapter include:

- *Characteristics and types of map projection*
- *Different parameter types*
- *Customizing a map projection through its parameters*
- *Common projected coordinate systems*

A projected coordinate system is defined on a flat, two-dimensional surface. Unlike a geographic coordinate system, a projected coordinate system has constant lengths, angles, and areas across the two dimensions. A projected coordinate system is always based on a geographic coordinate system that is based on a sphere or spheroid.

In a projected coordinate system, locations are identified by x,y coordinates on a grid, with the origin at the center of the grid. Each position has two values that reference it to that central location. One specifies its horizontal position and the other its vertical position. The two values are called the x-coordinate and y-coordinate. Using this notation, the coordinates at the origin are x = 0 and y = 0.

On a gridded network of equally spaced horizontal and vertical lines, the horizontal line in the center is called the x-axis and the central vertical line is called the y-axis. Units are consistent and equally spaced across the full range of x and y. Horizontal lines above the origin and vertical lines to the right of the origin have positive values; those below or to the left have negative values. The four quadrants represent the four possible combinations of positive and negative x- and y-coordinates.



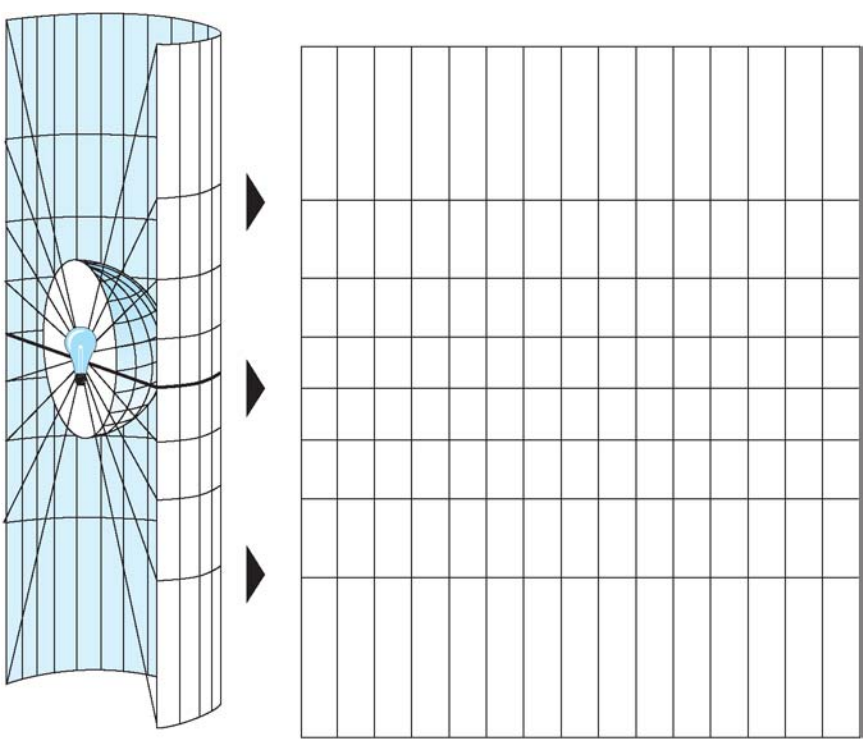
The signs of x,y coordinates in a projected coordinate system.

Whether you treat the earth as a sphere or a spheroid, you must transform its three-dimensional surface to create a flat map sheet. This mathematical transformation is commonly referred to as a *map projection*. One easy way to understand how map projections alter spatial properties is to visualize shining a light through the earth onto a surface, called the projection surface. Imagine the earth's surface is clear with the graticule drawn on it. Wrap a piece of paper around the earth. A light at the center of the earth will cast the shadows of the graticule onto the piece of paper. You can now unwrap the paper and lay it flat. The shape of the graticule on the flat paper is very different than on the earth. The map projection has distorted the graticule.

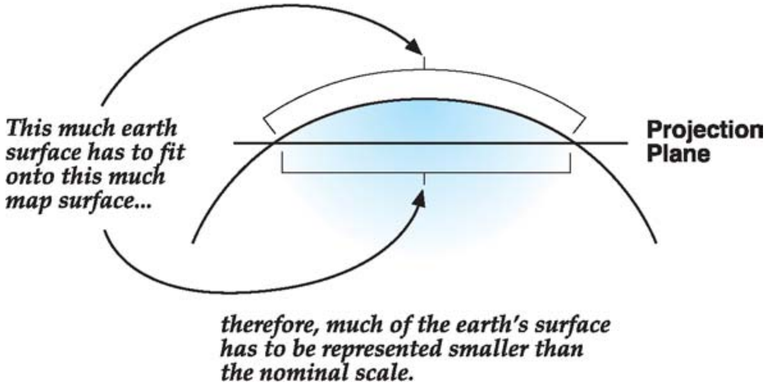
A spheroid can't be flattened to a plane any easier than a piece of orange peel can be flattened—it will rip. Representing the earth's surface in two dimensions causes distortion in the shape, area, distance, or direction of the data.

A map projection uses mathematical formulas to relate spherical coordinates on the globe to flat, planar coordinates.

Different projections cause different types of distortions. Some projections are designed to minimize the distortion of one or two of the data's characteristics. A projection could maintain the area of a feature but alter its shape. In the graphic below, data near the poles is stretched. The diagram on the next page shows how three-dimensional features are compressed to fit onto a flat surface.



The graticule of a geographic coordinate system is projected onto a cylindrical projection surface.



Map projections are designed for specific purposes. One map projection might be used for large-scale data in a limited area, while another is used for a small-scale map of the world. Map projections designed for small-scale data are usually based on spherical rather than spheroidal geographic coordinate systems.

Conformal projections

Conformal projections preserve local shape. To preserve individual angles describing the spatial relationships, a conformal projection must show the perpendicular graticule lines intersecting at 90-degree angles on the map. A map projection accomplishes this by maintaining all angles. The drawback is that the area enclosed by a series of arcs may be greatly distorted in the process. No map projection can preserve shapes of larger regions.

Equal area projections

Equal area projections preserve the area of displayed features. To do this, the other properties—shape, angle, and scale—are distorted. In equal area projections, the meridians and parallels may not intersect at right angles. In some instances, especially maps of smaller regions, shapes are not obviously distorted, and distinguishing an equal area projection from a conformal projection is difficult unless documented or measured.

Equidistant projections

Equidistant maps preserve the distances between certain points. Scale is not maintained correctly by any projection throughout an entire map; however, there are, in most cases, one or more lines on a map along which scale is maintained correctly. Most equidistant projections have one or more lines for which the length of the line on a map is the same length (at map scale) as the same line on the globe, regardless of whether it is a great or small circle or straight or curved. Such distances are said to be *true*. For example, in the Sinusoidal projection, the equator and all parallels are their true lengths. In other equidistant projections, the equator and all meridians are true. Still others (e.g., Two-Point Equidistant) show true scale between one or two points and every other point on the map. Keep in mind that no projection is equidistant to and from all points on a map.

True-direction projections

The shortest route between two points on a curved surface such as the earth is along the spherical equivalent of a straight line on a flat surface. That is the great circle on which the two points lie. True-direction, or *azimuthal*, projections maintain some of the great circle arcs, giving the directions or azimuths of all points on the map correctly with respect to the center. Some true-direction projections are also conformal, equal area, or equidistant.

Projection types

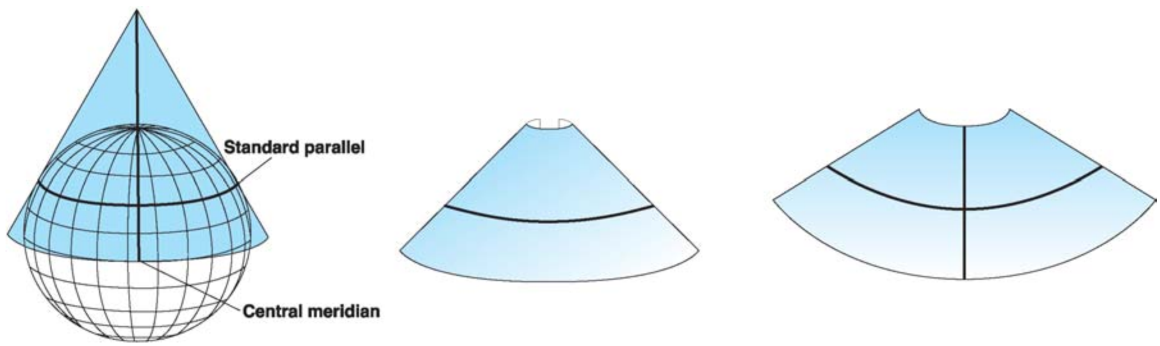
Because maps are flat, some of the simplest projections are made onto geometric shapes that can be flattened without stretching their surfaces. These are called developable surfaces. Some common examples are cones, cylinders, and planes. A map projection systematically projects locations from the surface of a spheroid to representative positions on a flat surface using mathematical algorithms.

The first step in projecting from one surface to another is creating one or more points of contact. Each contact is called a point (or line) of tangency. As illustrated in the section about 'Planar projections' below, a planar projection is tangential to the globe at one point. Tangential cones and cylinders touch the globe along a line. If the projection surface intersects the globe instead of merely touching its surface, the resulting projection is a secant rather than a tangent case. Whether the contact is tangent or secant, the contact points or lines are significant because they define locations of zero distortion. Lines of true scale are often referred to as *standard lines*. In general, distortion increases with the distance from the point of contact.

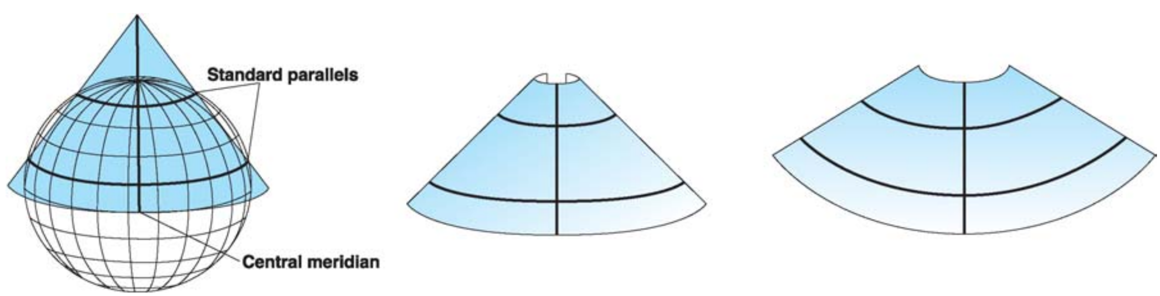
Many common map projections are classified according to the projection surface used: conic, cylindrical, or planar.

Conic projections

The most simple conic projection is tangent to the globe along a line of latitude. This line is called the *standard parallel*. The meridians are projected onto the conical surface, meeting at the apex, or point, of the cone. Parallel lines of latitude are projected onto the cone as rings. The cone is then cut along any meridian to produce the final conic projection, which has straight converging lines for meridians and concentric circular arcs for parallels. The meridian opposite the cut line becomes the *central meridian*.



In general, the further you get from the standard parallel, the more distortion increases. Thus, cutting off the top of the cone produces a more accurate projection. You can accomplish this by not using the polar region of the projected data. Conic projections are used for midlatitude zones that have an east–west orientation.



Somewhat more complex conic projections contact the global surface at two locations. These projections are called *secant projections* and are defined by two standard parallels. It is also possible to define a secant projection by one standard parallel and a scale factor. The distortion pattern for secant projections is different between the standard parallels than beyond them. Generally, a secant projection has less overall distortion than a tangent projection. On still more complex conic projections, the axis of the cone does not line up with the polar axis of the globe. These types of projections are called *oblique*.

The representation of geographic features depends on the spacing of the parallels. When equally spaced, the projection is equidistant north–south but neither conformal nor equal area. An example of this type of projection is the Equidistant Conic projection. For small areas, the overall distortion is minimal. On the Lambert Conic Conformal projection, the central parallels are spaced more closely than the parallels near the border, and small geographic shapes are maintained for both small-scale and large-scale maps. On the Albers Equal Area

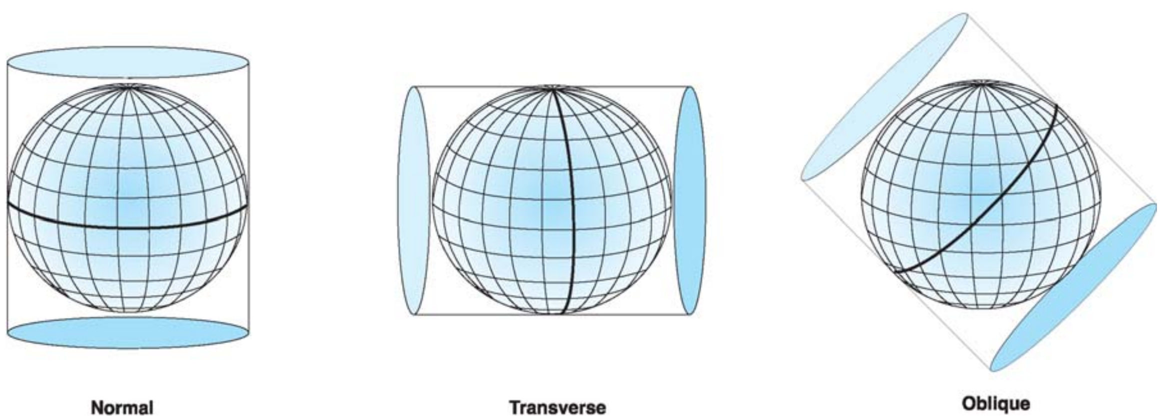
Conic projection, the parallels near the northern and southern edges are closer together than the central parallels, and the projection displays equivalent areas.

Cylindrical projections

Like conic projections, cylindrical projections can also have tangent or secant cases. The Mercator projection is one of the most common cylindrical projections, and the equator is usually its line of tangency. Meridians are geometrically projected onto the cylindrical surface, and parallels are mathematically projected. This produces graticular angles of 90 degrees. The cylinder is cut along any meridian to produce the final cylindrical projection. The meridians are equally spaced, while the spacing between parallel lines of latitude increases toward the poles. This projection is conformal and displays true direction along straight lines. On a Mercator projection, *rhumb lines*, lines of constant bearing, are straight lines, but most great circles are not.

For more complex cylindrical projections the cylinder is rotated, thus changing the tangent or secant lines. Transverse cylindrical projections such as the Transverse Mercator use a meridian as the tangential contact or lines parallel to meridians as lines of secancy. The standard lines then run north–south, along which the scale is true. Oblique cylinders are rotated around a great circle line located anywhere between the equator and the meridians. In these more complex projections, most meridians and lines of latitude are no longer straight.

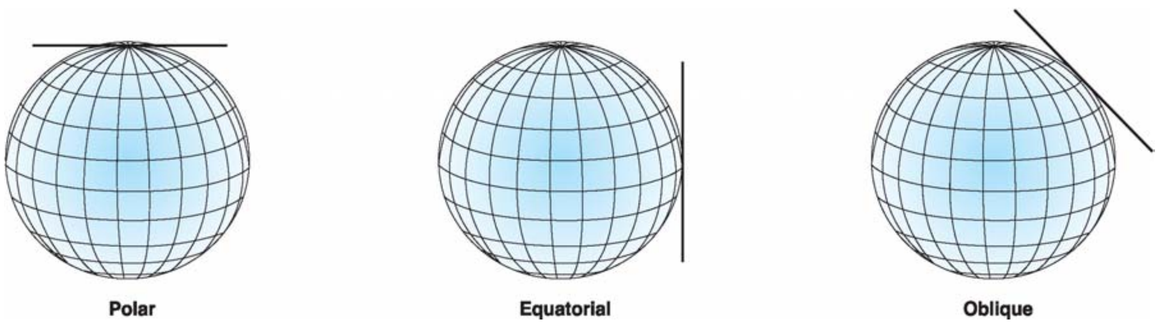
In all cylindrical projections, the line of tangency or lines of secancy have no distortion and thus are lines of equidistance. Other geographical properties vary according to the specific projection.



Planar projections

Planar projections project map data onto a flat surface touching the globe. A planar projection is also known as an azimuthal projection or a zenithal projection. This type of projection is usually tangent to the globe at one point but may be secant, also. The point of contact may be the North Pole, the South Pole, a point on the equator, or any point in between. This point specifies the aspect and is the focus of the projection. The focus is identified by a central longitude and a central latitude. Possible aspects are *polar*, *equatorial*, and *oblique*.

Azimuthal projections are classified in part by the focus and, if applicable, by the perspective point. The graphic below compares three planar projections with polar aspects but different perspectives. The Gnomonic projection views the surface data from the center of the

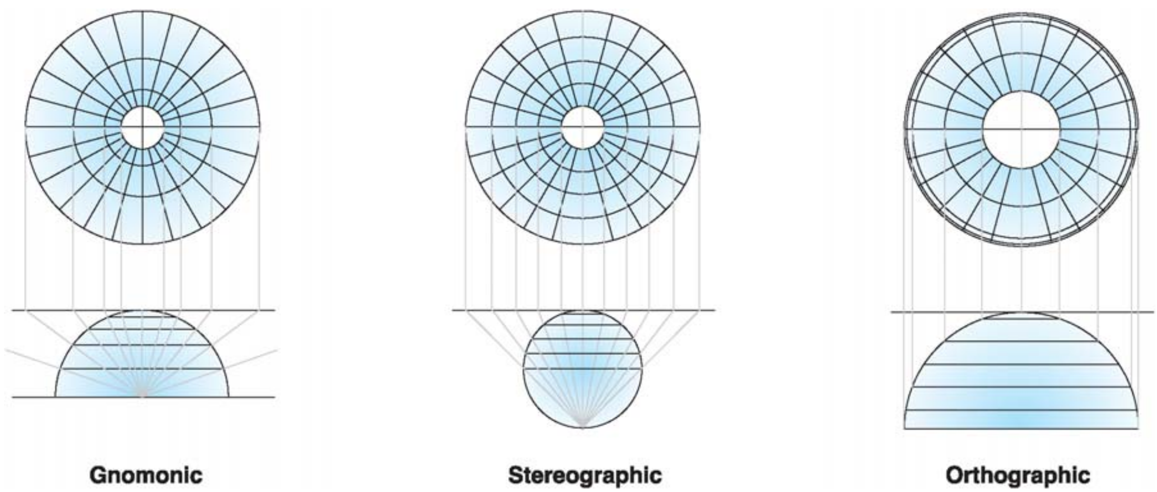


Polar aspects are the simplest form. Parallels of latitude are concentric circles centered on the pole, and meridians are straight lines that intersect with their true angles of orientation at the pole. In other aspects, planar projections will have graticular angles of 90 degrees at the focus. Directions from the focus are accurate.

Great circles passing through the focus are represented by straight lines; thus the shortest distance from the center to any other point on the map is a straight line. Patterns of area and shape distortion are circular about the focus. For this reason, azimuthal projections accommodate circular regions better than rectangular regions. Planar projections are used most often to map polar regions.

Some planar projections view surface data from a specific point in space. The point of view determines how the spherical data is projected onto the flat surface. The perspective from which all locations are viewed varies between the different azimuthal projections. The perspective point may be the center of the earth, a surface point directly opposite from the focus, or a point external to the globe, as if seen from a satellite or another planet.

earth, whereas the Stereographic projection views it from pole to pole. The Orthographic projection views the earth from an infinite point, as if from deep space. Note how the differences in perspective determine the amount of distortion toward the equator.



OTHER PROJECTIONS

The projections discussed previously are conceptually created by projecting from one geometric shape (a sphere) onto another (a cone, cylinder, or plane). Many projections are not related as easily to a cone, cylinder, or plane.

Modified projections are altered versions of other projections (e.g., the Space Oblique Mercator is a modification of the Mercator projection). These modifications are made to reduce distortion, often by including additional standard lines or changing the distortion pattern.

Pseudo projections have some of the characteristics of another class of projection. For example, the Sinusoidal is called a pseudocylindrical projection because all lines of latitude are straight and parallel and all meridians are equally spaced. However, it is not truly a cylindrical projection because all meridians except the central meridian are curved. This results in a map of the earth having an oval shape instead of a rectangular shape.

Other projections are assigned to special groups, such as circular or star.

A map projection by itself isn't enough to define a projected coordinate system. You can state that a dataset is in Transverse Mercator, but that's not enough information. Where is the center of the projection? Was a scale factor used? Without knowing the exact values for the projection parameters, the dataset can't be reprojected.

You can also get some idea of the amount of distortion the projection has added to the data. If you're interested in Australia but you know that a dataset's projection is centered at 0,0, the intersection of the equator and the Greenwich prime meridian, you might want to think about changing the center of the projection.

Each map projection has a set of parameters that you must define. The parameters specify the origin and customize a projection for your area of interest. Angular parameters use the geographic coordinate system units, while linear parameters use the projected coordinate system units.

Linear parameters

False easting—A linear value applied to the origin of the x-coordinates.

False northing—A linear value applied to the origin of the y-coordinates.

False easting and northing values are usually applied to ensure that all x or y values are positive. You can also use the false easting and northing parameters to reduce the range of the x- or y-coordinate values. For example, if you know all y values are greater than five million meters, you could apply a false northing of -5,000,000.

Height—Defines the point of perspective above the surface of the sphere or spheroid for the Vertical Near-side Perspective projection.

Angular parameters

Azimuth—Defines the center line of a projection. The rotation angle measures east from north. Used with the Azimuth cases of the Hotine Oblique Mercator projection.

Central meridian—Defines the origin of the x-coordinates.

Longitude of origin—Defines the origin of the x-coordinates. The central meridian and longitude of origin parameters are synonymous.

Central parallel—Defines the origin of the y-coordinates.

Latitude of origin—Defines the origin of the y-coordinates. This parameter may not be located at the center of the projection. In particular, conic projections use this parameter to set the origin of the y-coordinates below the area of the interest. In that instance, you don't need to set a false northing parameter to ensure that all y-coordinates are positive.

Longitude of center—Used with the Hotine Oblique Mercator Center (both Two-Point and Azimuth) cases to define the origin of the x-coordinates. Usually synonymous with the longitude of origin and central meridian parameters.

Latitude of center—Used with the Hotine Oblique Mercator Center (both Two-Point and Azimuth) cases to define the origin of the y-coordinates. It is almost always the center of the projection.

Standard parallel 1 and standard parallel 2—Used with conic projections to define the latitude lines where the scale is 1.0. When defining a Lambert Conformal Conic projection with one standard parallel, the first standard parallel defines the origin of the y-coordinates.

For other conic cases, the y-coordinate origin is defined by the latitude of origin parameter.

Longitude of first point
Latitude of first point
Longitude of second point
Latitude of second point

The four parameters above are used with the Two-Point Equidistant and Hotine Oblique Mercator projections. They specify two geographic points that define the center axis of a projection.

Pseudo standard parallel 1—Used in the Krovak projection to define the oblique cone's standard parallel.

XY plane rotation—Along with the X scale and Y scale parameters, defines the orientation of the Krovak projection.

Unitless parameters

Scale factor—A unitless value applied to the center point or line of a map projection.

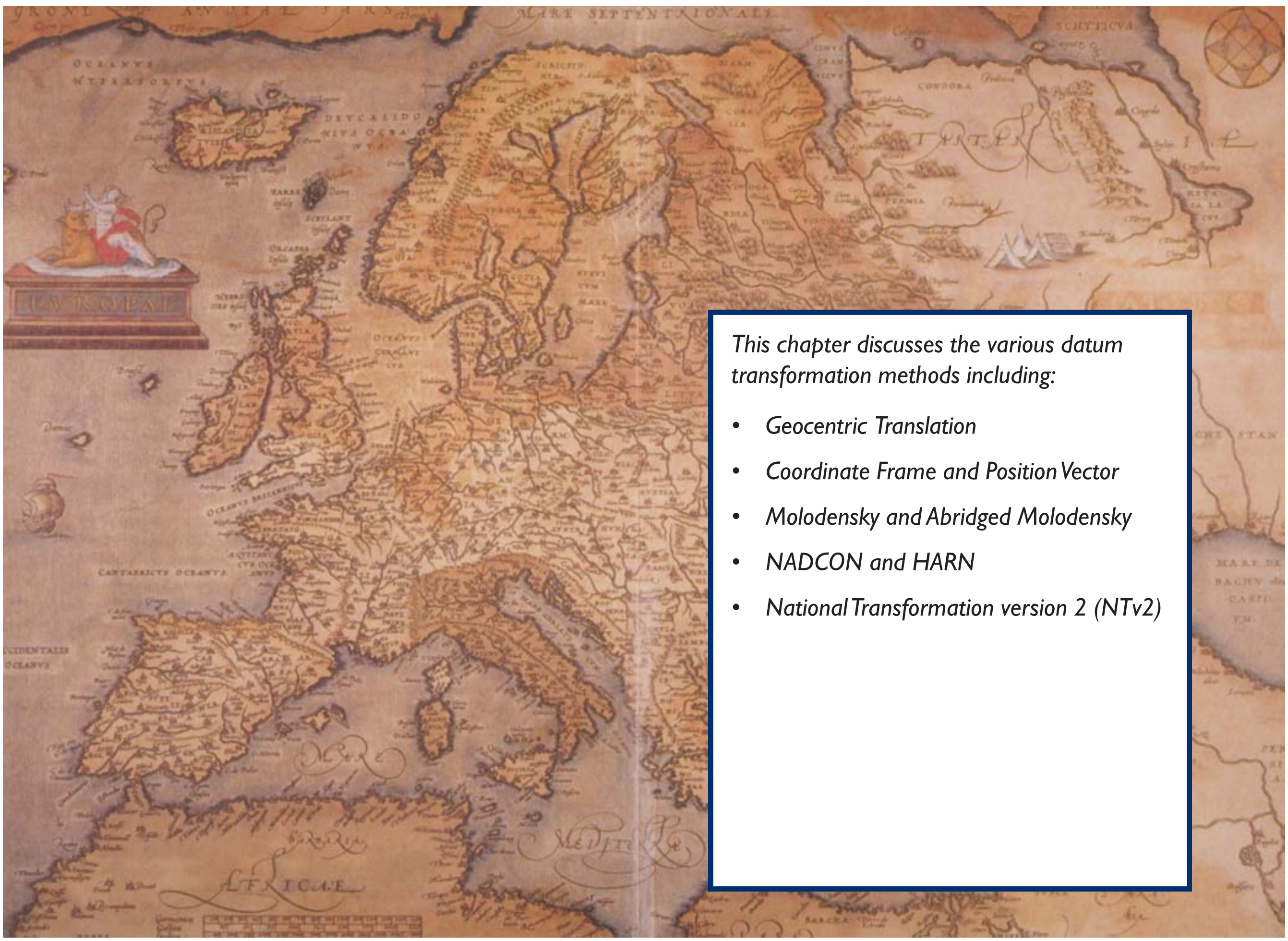
The scale factor is usually slightly less than one. The UTM coordinate system, which uses the Transverse Mercator projection, has a scale factor of 0.9996. Rather than 1.0, the scale along the central meridian of the projection is 0.9996. This creates two almost parallel lines approximately 180 kilometers away, where the scale is 1.0. The scale factor reduces the overall distortion of the projection in the area of interest.

X scale—Used in the Krovak projection to orient the axes.

Y scale—Used in the Krovak projection to orient the axes.

Option—Used in the Cube and Fuller projections. In the Cube projection, option defines the location of the polar facets. An option of 0 in the Fuller projection will display all 20 facets. Specifying an option value between 1–20 will display a single facet.

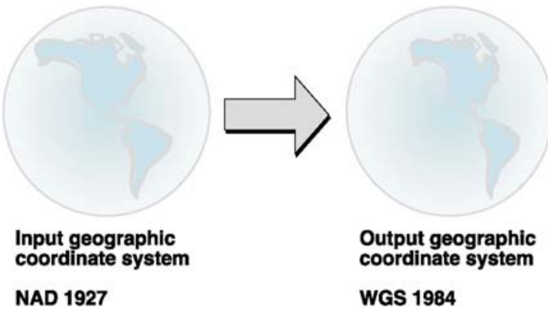
3 Geographic transformations



This chapter discusses the various datum transformation methods including:

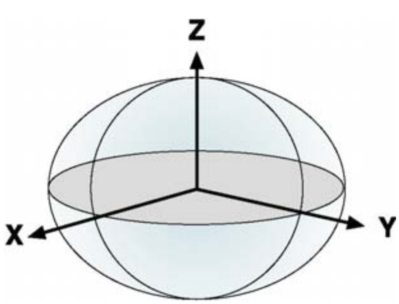
- *Geocentric Translation*
- *Coordinate Frame and Position Vector*
- *Molodensky and Abridged Molodensky*
- *NADCON and HARN*
- *National Transformation version 2 (NTv2)*

Moving your data between coordinate systems sometimes includes transforming between the geographic coordinate systems.



Because the geographic coordinate systems contain datums that are based on spheroids, a geographic transformation also changes the underlying spheroid. There are several methods, which have different levels of accuracy and ranges, for transforming between datums. The accuracy of a particular transformation can range from centimeters to meters depending on the method and the quality and number of control points available to define the transformation parameters.

A geographic transformation always converts geographic (longitude–latitude) coordinates. Some methods convert the geographic coordinates to geocentric (X,Y,Z) coordinates, transform the X,Y,Z coordinates, and convert the new values back to geographic coordinates.



The X,Y,Z coordinate system.

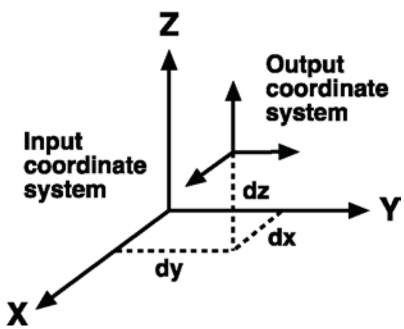
These include the Geocentric Translation, Molodensky, and Coordinate Frame methods.

Other methods such as NADCON and NTv2 use a grid of differences and convert the longitude–latitude values directly.

A geographic transformation is always defined with a direction. The transformation parameters describe how to convert from the input geographic coordinate to the output geographic coordinate system. All supported methods are invertible. Given a geographic transformation, you can apply it in the opposite direction. Generally, applications will automatically apply the transformation in the appropriate direction. As an example, if you wish to convert data from WGS 1984 to Adindan but a list of available geographic transformations shows Adindan_To_WGS_1984, you can choose this transformation and the application will apply it properly.

Three-parameter methods

The simplest datum transformation method is a geocentric, or three-parameter, transformation. The geocentric transformation models the differences between two datums in the X,Y,Z coordinate system. One datum is defined with its center at 0,0,0. The center of the other datum is defined at some distance (ΔX,ΔY,ΔZ) in meters away.



Usually the transformation parameters are defined as going from a local datum to WGS 1984 or another geocentric datum.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{new} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} + \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{original}$$

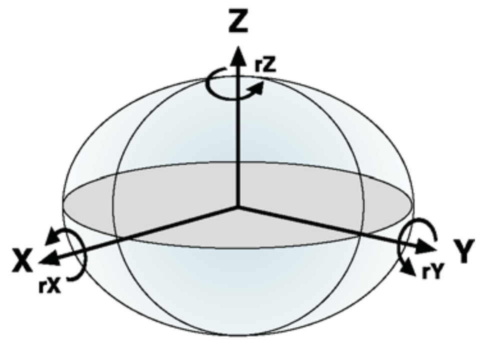
The three parameters are linear shifts and are always in meters.

Seven-parameter methods

A more complex and accurate datum transformation is possible by adding four more parameters to a geocentric transformation. The seven parameters are three linear shifts (ΔX,ΔY,ΔZ), three angular rotations around each axis (r_x,r_y,r_z), and scale factor(s).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{new} = \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} + (1+s) \cdot \begin{bmatrix} 1 & r_z & -r_y \\ -r_z & 1 & r_x \\ r_y & -r_x & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{original}$$

The rotation values are given in decimal seconds, while the scale factor is in parts per million (ppm). The rotation values are defined in two different ways. It's possible to define the rotation angles as positive either clockwise or counterclockwise as you look toward the origin of the X,Y,Z systems.



The Coordinate Frame (or Bursa–Wolf) definition of the rotation values.

The equation in the previous column is how the United States and Australia define the equations and is called the Coordinate Frame Rotation transformation. The rotations are positive counterclockwise. Europe uses a different convention called the Position Vector transformation. Both methods are sometimes referred to as the Bursa–Wolf method. In the Projection Engine, the Coordinate Frame and Bursa–Wolf methods are the same. Both Coordinate Frame and Position Vector methods are supported, and it is easy to convert transformation values from one method to the other simply by changing the signs of the three rotation values. For example, the parameters to convert from the WGS 1972 datum to the WGS 1984 datum with the Coordinate Frame method are (in the order, ΔX, ΔY,ΔZ,r_x,r_y,r_z,s):

```
(0.0, 0.0, 4.5, 0.0, 0.0, -0.554,
0.227)
```

To use the same parameters with the Position Vector method, change the sign of the rotation so the new parameters are:

```
(0.0, 0.0, 4.5, 0.0, 0.0, +0.554, 0.227)
```

Unless explicitly stated, it's impossible to tell from the parameters alone which convention is being used. If you use the wrong method, your results can return inaccurate coordinates. The only way to determine how the parameters are defined is by checking a control point whose coordinates are known in the two systems.

Molodensky method

The Molodensky method converts directly between two geographic coordinate systems without actually converting to an X,Y,Z system. The Molodensky method requires three shifts ($\Delta X, \Delta Y, \Delta Z$) and the differences between the semimajor axes (Δa) and the flattenings (Δf) of the two spheroids. The Projection Engine automatically calculates the spheroid differences according to the datums involved.

$$
\begin{aligned}
(M + h)\Delta\varphi = & -\sin\varphi\cos\lambda\Delta X - \sin\varphi\sin\lambda\Delta Y \\
& + \cos\varphi\Delta Z + \frac{e^2\sin\varphi\cos\varphi}{(1-e^2\sin^2\varphi)^{1/2}}\Delta a \\
& + \sin\varphi\cos\varphi(M\frac{a}{b} + N\frac{b}{a})\Delta f
\end{aligned}
$$

$$(N + h)\cos\varphi\Delta\lambda = -\sin\lambda\Delta X + \cos\lambda\Delta Y$$

$$
\begin{aligned}
\Delta h = & \cos\varphi\cos\lambda\,\Delta X + \cos\varphi\sin\lambda\,\Delta Y \\
& + \sin\varphi\,\Delta Z - (1-e^2\sin^2\varphi)^{1/2}\Delta a \\
& + \frac{a(1-f)}{(1-e^2\sin^2\varphi)^{1/2}}\sin^2\varphi\,\Delta f
\end{aligned}
$$

h	ellipsoid height (meters)
φ	latitude
λ	longitude
a	semimajor axis of the spheroid (meters)
b	semiminor axis of the spheroid (meters)
f	flattening of the spheroid
e	eccentricity of the spheroid

M and N are the meridional and prime vertical radii of curvature, respectively, at a given latitude. The equations for M and N are:

$$M = \frac{a(1-e^2)}{(1-e^2\sin^2\varphi)^{3/2}}$$

$$N = \frac{a}{(1-e^2\sin^2\varphi)^{1/2}}$$

You solve for $\Delta\lambda$ and $\Delta\varphi$. The amounts are added automatically by the Projection Engine.

Abridged Molodensky method

The Abridged Molodensky method is a simplified version of the Molodensky method. The equations are:

$$
\begin{aligned}
M\Delta\varphi = & -\sin\varphi\cos\lambda\Delta X - \sin\varphi\sin\lambda\Delta Y \\
& + \cos\varphi\Delta Z + (a\Delta f + f\Delta a)\cdot 2\sin\varphi\cos\varphi
\end{aligned}
$$

$$N\cos\varphi\Delta\lambda = -\sin\lambda\Delta X + \cos\lambda\Delta Y$$

$$
\begin{aligned}
\Delta h = & \cos\varphi\cos\lambda\Delta X + \cos\varphi\sin\lambda\Delta Y \\
& + \sin\varphi\Delta Z + (a\Delta f + f\Delta a)\sin^2\varphi - \Delta a
\end{aligned}
$$

NADCON and HARN methods

The United States uses a grid-based method to convert between geographic coordinate systems. Grid-based methods allow you to model the differences between the systems and are potentially the most accurate method. The area of interest is divided into cells. The National Geodetic Survey (NGS) publishes grids to convert between NAD 1927 and other older geographic coordinate systems and NAD 1983. We group these transformations into the NADCON method. The main NADCON grid, CONUS, converts the contiguous 48 states. The other NADCON grids convert older geographic coordinate systems to NAD 1983 for

- Alaska
- Hawaiian islands
- Puerto Rico and Virgin Islands
- St. George, St. Lawrence, and St. Paul Islands in Alaska

The accuracy is around 0.15 meters for the contiguous states, 0.50 for Alaska and its islands, 0.20 for Hawaii, and 0.05 for Puerto Rico and the Virgin Islands. Accuracies can vary depending on how good the geodetic data in the area was when the grids were computed (NADCON, 1999).

The Hawaiian islands were never on NAD 1927. They were mapped using several datums that are collectively known as the Old Hawaiian datums.

New surveying and satellite measuring techniques have allowed NGS and the states to update the geodetic control point networks. As each state is finished, the NGS publishes a grid that converts between NAD 1983 and the more accurate control point coordinates. Originally, this effort was called the High Precision Geodetic Network (HPGN). It is now called the High Accuracy Reference Network (HARN). More than 40 states have published HARN grids as of September 2000. HARN transformations have an accuracy around 0.05 meters (NADCON, 2000).

The difference values in decimal seconds are stored in two files: one for longitude and the other for latitude. A bilinear interpolation is used to calculate the exact difference between the two geographic coordinate systems at a point. The grids are binary files, but a program, NADGRD, from the NGS allows you to convert the grids to an American Standard Code for Information Interchange (ASCII) format. Shown at the bottom of the page is the header and first row of the CSHPGN.LOA file. This is the longitude grid for Southern California. The format of the first row of numbers is, in order, the number of columns, number of rows, number of z values (always one), minimum longitude, cell size, minimum latitude, cell size, and not used.

The next 37 values (in this case) are the longitude shifts from -122° to -113° at 32° N in 0.25° intervals in longitude.

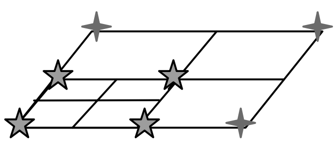
```
NADCON EXTRACTED REGION                                    NADGRD
  37   21    1   -122.00000        .25000      32.00000        .25000        .00000
      .007383       .004806       .002222      -.000347      -.002868      -.005296
     -.007570      -.009609      -.011305      -.012517      -.013093      -.012901
     -.011867      -.009986      -.007359      -.004301      -.001389       .001164
      .003282       .004814       .005503       .005361       .004420       .002580
      .000053      -.002869      -.006091      -.009842      -.014240      -.019217
     -.025104      -.035027      -.050254      -.072636      -.087238      -.099279
     -.110968
```

A portion of a HARN grid file.

National Transformation version 2

Like the United States, Canada uses a grid-based method to convert between NAD 1927 and NAD 1983. The National Transformation version 2 (NTv2) method is quite similar to NADCON. A set of binary files contains the differences between the two geographic coordinate systems. A bilinear interpolation is used to calculate the exact values for a point.

Unlike NADCON, which can only use one grid at a time, NTv2 is designed to check multiple grids for the most accurate shift information. A set of low-density base grids exists for Canada. Certain areas such as cities have high-density local subgrids that overlay portions of the base, or parent, grids. If a point is within one of the high-density grids, NTv2 will use the high-density grid; otherwise, the point 'falls through' to the low-density grid.



A high-density subgrid with four cells overlaying a low-density base grid, also with four cells.

If a point falls in the lower-left part of the above picture between the stars, the shifts are calculated with the high-density subgrid. A point whose coordinates are anywhere else will have its shifts calculated with the low-density base grid. The software automatically calculates which base or subgrid to use.

The parent grids for Canada have spacings ranging from five to 20 minutes. The high-density grids usually have cell sizes of 30 seconds.

Unlike NADCON grids, NTv2 grids list the accuracy of each point. Accuracy values can range from a few centimeters to around a meter. The high-density grids usually have subcentimeter accuracy.

Australia and New Zealand adopted the NTv2 format to convert between datums as well. Australia has released grids that convert between either Australian Geodetic Datum of 1966 (AGD 1966) or AGD 1984 and Geocentric Datum of Australia of 1994 (GDA 1994). New Zealand has released a countrywide grid to convert between New Zealand Geodetic Datum of 1949 (NZGD 1949) and NZGD 2000.

National Transformation version 1

Like NADCON, the National Transformation version 1 (NTv1) uses a single grid to model the differences between NAD 1927 and NAD 1983 in Canada. This version is also known as CNT in ArcInfo™ Workstation. The accuracy is within 0.01 m of the actual difference for 74 percent of the points and within 0.5 m for 93 percent of the cases.