Geoviz
Where in the World?
Maps, Ellipsoids, Datums, and Projections

Connor

Introduction

Geologists and geophysicists make and use maps. These maps must have a common coordinate system. Otherwise, it becomes very difficult to compare information and/or collate observations. In the past, a common coordinate system for most maps simply was not easy to achieve. During the 1990's, the advent of widely available global positioning system (GPS) receivers made it possible to achieve commonality and cross-referencing among coordinate systems. The downside, if there is one, is that we have to be much more careful about the way we gather and use data that is destined to be displayed on maps.

Ellipsoids, Datums, and Geoids

Earth is a geometrically complex object. The easiest way to approximate the shape of Earth is with a sphere. A sphere has constant radius and therefore departs from the actual shape of Earth fairly dramatically. The equatorial radius of Earth is approximately 6378.139 km, the polar radius of Earth is approximately 6356.75 km, a difference of 21.38 km.

Consequently, the figure of the Earth is more precisely approximated by an oblate ellipsoid of revolution. The planet is flattened along its axis of rotation, its equatorial bulge a result of rotation. The amount of flattening of the ellipsoid can be expressed in terms of equatorial and polar radii:

$$f = \frac{R_{eq} - R_{po}}{R_{eq}} \tag{1}$$

where R_{eq} is the radius of Earth at the equator and R_{po} is the radius at the pole. In many map projection formulas, the eccentricity of the ellipsoid is used, rather than flattening.

$$e^2 = 2f - f^2 \tag{2}$$

where e is the eccentricity. If Earth were a body of homogeneous density rotating in space, the ellipsoid of revolution should describe its shape perfectly. This is not the case. There are substantial regional departures from the ideal ellipsoidal shape. A best-fit ellipsoid can be used to approximate the shape of the Earth. Such an ellipsoid is called a reference ellipsoid. As discussed in the following, different reference ellipsoids are in common use.

Departures of the figure of the Earth from the reference ellipsoid occur primarily because of large scale density differences within Earth. These density differences cause undulations in the shape of Earth. So a further surface can be defined that accounts for these regional undulations. This surface is called the geoid. The two main characteristics of the geoid are:

- Earth's gravity field is perpendicular to the geoid everywhere - making the geoid an equipotential surface.
- The geoid coincides with the theoretical position of the surface of Earth's oceans at rest.

The geoid could be any equipotential surface. On Earth, it is simply convenient to refer the geoid to mean sea-level. The elevation of the geoid commonly deviates from the reference ellipsoid by up to 40 m. In some locations, such as off the coast of India, the geoid differs from the ellipsoid by more than 100 m. This means that if you sail on a ship from Madagascar to Bombay, you travel through a trough 100 m deep with respect to the reference ellipsoid. Actual topography varies with respect to both the ellipsoid and the geoid. Usually, elevation is referenced to the geoid, as in *meters above sea-level*, but it is possible to reference topography to the reference ellipsoid.

Any geographic location on the surface of Earth can be expressed as a latitude and longitude. While elevation is usually referenced to the geoid, latitude and longitude are referenced to the ellipsoid. Two things are needed to express a position on the surface of Earth in terms of latitude and longitude:

- An origin for the coordinate system
- An equation for the reference ellipsoid

The origin is agreed upon by convention. Zero latitude corresponds to the equator (the plane orthogonal to the axis of Earth's rotation). Values of latitude vary from 0° N at the equator to 90° N at the north pole and 90° S at the south pole. Often, especially in many computer programs, latitude is represented as positive in the northern hemisphere and negative in the southern hemisphere. That is latitude varies from $+90°$ at the north pole to $-90°$ at the south pole. Longitude is more arbitrary. Zero longitude is designated as the meridian of the astronomical observatory in Greenwich, England. Longitude increases to the east (just like on a familiar *xy* -plot with positive values increasing to the right). The eastern hemisphere extends to a longitude of 180° E. Three different ways are used to represent longitude in the western hemisphere. The simplest approach, and most rarely used, is to simply let longitude values increase eastward until reaching 360° as the meridian at Greenwich, England, is approached from the west. In this system, longitude is uniquely expressed as any number, $0° - 360°$. Alternatively, longitude in the western hemisphere can be assigned a negative number, in which case longitude varies from $-180°$ to $+180°$. The most common means of expressing longitude, and the least useful for computations and in computer programs, is to express longitude using E and W to designate eastern and western hemisphere, respectively. In this system, longitude varies from 180° W to 180° E. Many computational errors have been made by forgetting which system is being used to designate longitude.

Various equations for the reference ellipsoid are in use. These equations differ in the average radius of the Earth chosen and the eccentricity (or flattening) of the ellipse. In the context of map coordinates, these different reference ellipsoids are referred to as different map datums. Three common map datums are described in Table 1.

Table 1: Common map datums and their parameters

Ellipsoid (Datum)	R_{eq} (m)	$1/f$
Clarke 1866 (NAD27)	6378206.4	294.9786982
WGS84	6378137.0	298.257223563
GRS 1980 (NAD83)	6378137.0	298.257222101

NAD27 refers to the North American Datum of 1927; NAD83 refers to the North American Datum of 1983; WGS refers to the World Geodetic System; GRS refers to Geodetic Reference Systems. These are the most common ellipsoids (sometimes called spheroids) and map datums that you are likely to encounter, but there are very many more. Prior to the mid-1980s, map datums and reference ellipsoids were derived from regional surveys. These ellipsoids were optimized, naturally enough to fit the figure of Earth in the regional area of interest. So, for example, the Clarke 1866 ellipsoid was derived for North America and its datum referenced to a location in central Kansas (meaning that the fit is optimal at this location). Departures from the geoid using Clarke 1866 NAD27 are minimized for North America, but are quite large for other parts of the globe. In the mid-1980s a global best-fit ellipsoid could be calculated using satellite data. The datum for this ellipsoid is Earth-centered, in other words, the center of Earth is essentially the datum's *origin*, and departure of the ellipsoid from the geoid is minimized globally.

The WGS84 datum is now commonly used, worldwide. Nevertheless, a large number of US maps in print were constructed using the Clarke 1866 ellipsoid and the NAD27 datum. In other parts of the world, other local ellipsoids and datums were used. When different datums are in use, it is crucial to understand that *a given location at the surface of the Earth will have a different latitude and longitude depending on the datum used*. The difference between WGS84 and NAD27 is large in a state like Florida, that is far from the NAD27 origin. In Florida, the difference between NAD27 and WGS84 is about one second of latitude.

All this has two practical results for people who gather and use map data. First, the ellipsoid and datum used must be reported when map coordinates are given. Otherwise the data cannot be associated to a single specific location on the surface of Earth. Second, GPS receivers are very often used to determine position. Most GPS receivers will report position (*e.g.,* latitude and longitude) using one of any number of datums. If these data are to be plotted on a map, the datum used by the GPS receiver must match the datum used to create the map, otherwise the location of the data will be incorrectly plotted on the map. It is a very simple procedure to note the map datum - surprisingly this information often is not noted, resulting in errors that can be very difficult to identify and fix later.

Grids

Latitude and longitude coordinates provide a reasonable way to report position on the reference ellipsoid. But this coordinate system has serious drawbacks. Just how far is it between 56°N, 47°W and 56.1°N, 46.9°W? In what direction would you walk to go from 56°N, 47°W toward 56.1°N, 46.9°W? Such distances and directions are difficult to calculate on a sphere, much less on an oblate ellipsoid.

Rectangular grid systems have been developed to simplify distance calculations on maps. These are Cartesian grids. The Y-axis is oriented N–S and Y-coordinates are usually referred to as the Northing, or sometimes as distance north. The X-axis is oriented E–W and X coordinates are referred to as the Easting or distance east. Usually the units of these grids are meters, or occasionally (unfortunately) feet.

The most important grid in use today is the Universal Transverse Mercator (UTM) grid - not to be confused with the Universal Transverse Mercator map projection. The Northing for the UTM grid in the northern hemisphere is given in meters north of the equator. Negative numbers are avoided at all costs in the UTM grid. So, in the southern hemisphere, the Northing is given in meters south of the equator (a negative number) $+10\,000\,000$ m. For example, the Northing of a point located 2000 km north of the equator is 2 000 000 N. The Northing of a point located 2000 km south of the equator is 8 000 000 N. So in both the northern and southern hemispheres, UTM Northing is a positive number that increase as one walks north and decrease as one walks south.

The X-axis is perpendicular to the Y-axis at the origin (central meridian) of the coordinate system. Far from the central meridian, this results in considerable distortion of the grid. Consequently, the UTM grid breaks up Earth's surface into 60 zones, each with its own central meridian. Again, negative numbers are avoided in the UTM grid. Consequently the value of the Easting along the central meridian (the origin) of any zone is 500 000 E, rather than 0 E. West of the central meridian the Easting is less than 500 000 E; east of the central the Easting is more than 500 000 E. There are enough zones so that, using this scheme, no negative Easting coordinates occur on Earth. Note that there are exactly 60 places on the surface of Earth that have the UTM grid coordinate 2010222 N, 342343 E. When UTM coordinates are used, the zone must be reported. These zones are uniquely numbered from 1–60 (see the following URL for some useful images of UTM zone maps `http://tmackinnon.com/utm-rows-and-zones-map.php`). As with latitude and longitude coordinates, the ellipsoid and datum must also be known. Like latitude and longitude, *the UTM coordinate of a given point on the surface of Earth is different for different ellipsoids and datums*.

The UTM grid is ideally suited for maps of relatively small areas (large scale maps). Large regions are more likely to cross UTM zones and this leads to problems. GPS receivers are normally capable of reporting position using the UTM grid or a variety of alternative grids. Different grids are sometimes reported on maps - particularly old maps. These alternative grids include the British Grid, and the State Plane Coordinate system. These are all constructed with the same basic idea, but use different origins. For science, use a UTM grid.

Converting from Latitude/Longitude to UTM coordinates and the like

Given the differences in map datums and in coordinate systems, it is essential for geologists and geophysicists to easily convert coordinates from one datum to another, or from one grid to another. Fortunately there are tools available for doing this conversion. One of the most reliable tools was developed by staff at the US Geological Survey. This conversion tool is the **Proj.4 Cartographic Projections Library**. It is freely available; check out the Proj.4 website: `http://proj4.org/` for current versions and downloads. Once installed, Proj.4 is useful for solving coordinate system issues.

For example, suppose a colleague gives you the latitude and longitude of 82° W, 27° N (NAD27 datum) for a work site near Tampa, Florida. The *cs2cs* program can be run to convert this coordinate to the NAD83 datum:
type:

```
cs2cs +proj=latlong +datum=NAD27 +to +proj=latlong +datum=NAD83 -f "%0.6f"
```

then type:

```
-82 27
```

Output of the *cs2cs* code using these parameters will be:

```
-81.999809      27.000336
```

That is, 82° W, 27° N (NAD27 datum) converts to 81.999809° W, 27.000336° N (NAD83 datum). Proj.4 can also convert from latitude/longitude to the UTM grid system using the program, *proj*:
type:

```
proj +proj=utm +datum=WGS84 +zone=12 -f "%.0f"
```

then type:

```
111d17'55"W 38d34'N
```

Note that the datum (in this case WGS84) and UTM zone (in this case zone 12) must be specified. Output of *proj* is:

```
473985     4268733
```

corresponding to 473985 E, and 4268733 N in the UTM grid for zone 12 (WGS84 datum). Similarly, to convert UTM coordinates to Latitude and Longitude, one can use:
type:

```
proj -I +proj=utm +ellips=WGS84 +zone=12  -f "%.6f"
```

then type:

```
473985 4268733
```

this outputs:

```
-111.298620    38.566665
```

which corresponds to 111.298620° W, 38.566665° N (WGS84 datum). The **Proj.4 Cartographic Projections Library** contains vast resources for making such conversions, and can easily be applied to whole data files. Note, the *cs2cs* and *proj* programs can understand numbers in decimal degree format and in degrees/minutes/seconds format.

Map Projections

A map projection transfers the coordinates of points located on the reference ellipsoid to a flat surface - a map. A huge number of map projections are in frequent use and essentially an infinite number can be derived. Snyder (1982) discusses the mathematical basis of various map projections in detail. He broadly states that a map projection can be chosen on the bases of:

- Area: many types of map projection are equal area. A given area on one part of the map corresponds to an actual area on the surface of the Earth. This ratio is constant across the entire map. This is done at some expense - scale and angles are distorted on equal area map projections.
- Shape: Many map projections preserve shape and angle. These map projections are often called conformal
- Scale: No map projection preserves scale. That is, scale is not the same from one part of the map to another.

For large scale maps showing small areas, the differences between projections are fairly trivial. For large areas - like continents or ocean basin, shown on small scale maps, the differences are obvious and important to take into account. Snyder (1982) suggests appropriate map projections for specific uses. Here is a brief synopsis:

- To show the whole Earth in conformal projection - Mercator
- To show the whole Earth in equal area projection - Hammer or Eckert IV or VI
- To show a continent or ocean scale near the equator with predominant E-W extent in conformal projection - Mercator
- To show a continent or ocean scale along the equator with predominant E-W extent in equal-area projection - Cylindrical equal area
- To show a continent or ocean scale along the equator with predominant N-S extent in conformal projection - Transverse Mercator
- To show a continent or ocean scale along the equator with predominant N-S extent in equal area projection - Transverse Cylindrical Equal Area
- To show a continent or ocean scale away from the equator with predominant E-W extent in conformal projection - Lambert Conformal Conic
- To show a continent or ocean scale away from the equator with predominant E-W extent in equal-area projection - Albers Equal-Area Conic

...and there are many more!

Reference

Snyder, John Parr, Map Projections Used by the U. S. Geological Survey, 2nd edition, Geol. Survey Bulletin 1532, 313 p., U. S. Government Printing Office, Washington, D. C., 1982.