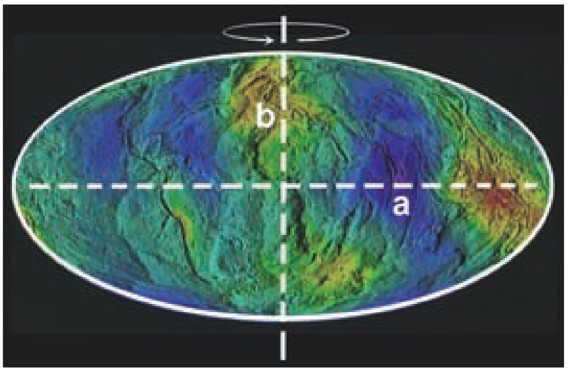
Part 2: The Ellipsoid
Practical Geodesy

In the previous article it was stated that the earth could be approximated by a sphere or ellipsoid. The sphere is often used for simple navigation calculations, but for more accurate positioning the ellipsoidal shape of the earth has to be taken into account.

By Huibert-Jan Lekkerkerk



Ellipsoid parameters, a: semi-major axis; b: semi-minor axis (source: denali.gsfc.nasa.gov - adapted).

The ellipsoidal shape of the earth is caused by the rotation of the earth around an imaginary axis running through both poles. Physics tells us that an object in a circular orbit experiences so-called centrifugal force. The equator, which follows a larger orbit than the poles, will therefore experience a force that is greater than that at the poles. As a result the earth is 'flattened' at the poles.

The Ellipsoid

The true shape of the earth is, of course, the geoid (see previous article), but since it is hard to calculate on this surface it is approximated by the ellipsoid for day-to-day geodetic calculations. The actual flattening of the ellipsoid when compared to a perfect sphere is not very large. On average the flattening is said to be around 1/300 on a radius of approximately 6370 kilometers. This results in a difference of 'only' 21 kilometers between both axes. Only two parameters are needed to describe the shape of the ellipsoid: the semi-major axis and the flattening.

Which Ellipsoid?

The ancient Greeks did some measurements and calculations on the true shape of the earth, but only in the 19th century were the techniques advanced enough to attain sufficient precision. Moreover, it became increasingly important for countries to determine their exact borders. As a result, various countries staged expeditions to determine the shape of the ellipsoid. Some geodesists took only a few countries into account in their calculations, while others made observations over vast areas. Consequently, quite a number of ellipsoids are used, with every ellipsoid matched as closely as possible with a certain piece of the earth. Since the ellipsoids were usually directly employed in land survey work, each became anchored to a certain country, making it nearly impossible to change it at a later date. Worldwide, some dozens of ellipsoids are in use today. Internationally the GRS80 ellipsoid, the calculations for which were finished in 1980, is currently the most accurate.

Geodetic Datum

It is not enough to determine the size of the ellipsoid that best matches our piece of the earth. In the example, the semi-major axis was approximately in the direction of the equator, but the true orientation of the ellipsoid can vary. In geodesy the ellipsoid is therefore never used by itself, but always as part of the so-called horizontal or geodetic datum. The geodetic datum is therefore a combination of an ellipsoid and the coupling of that ellipsoid to the surface of the earth. The ellipsoid is coupled at the datum point, of which both the coordinates and the orientation relative to the local gravity are determined. The North American Datum (NAD23), for example, is coupled to the earth at Meade's Ranch (Kansas).

Which Geodetic Datum?

We have already seen that there are a great many ellipsoids in use. There are, however, even more geodetic datums, since almost every country has its own datum point. Furthermore, geodetic datums were developed covering areas larger than a single country. Directly after the Second World War, for example, the measurements for a European Datum, which was published in 1950 (ED50), were started. Shortly thereafter oil and gas fields were found in the North Sea. The determination of concessions and the positioning of the various platforms were (and still are) done on this newly-developed ED50. With the advent of spacecraft, the need for a worldwide geodetic datum became apparent. However, for a worldwide system it is no longer possible to couple the ellipsoid to a terrestrial reference point since such a point will slowly move due to the movement of the continents. A local datum does not have this problem since the point moves with the continent. For this reason the ITRS (International Terrestrial Reference System) was developed, taking into account the drifting of the continents. The ITRS is determined every year and as such is not a practical reference frame.

One Datum for Everyone?

Since GPS uses satellites, a geodetic datum, the World Geodetic System 1984 (WGS84), was needed to cover the entire earth. In the early years the drifting of the continents was not taken into account, resulting in gradually increasing differences between the local datum for a certain country and WGS84.

In 1996 it was decided to update WGS84 to the ITRS on a yearly basis. Since then there has been only a slight difference at any one time between WGS84 and the ITRS. Since both the positions of the satellites and the position of the receiver are determined on the same datum, there are no practical problems within the system. For geodetic work, however, the precision is not

The datum point of the North American Datum (NAD27) is located near Meade's Ranch (source: www.photolib.noaa.gov & Google Earth - adapted)

The continental plates are slowly drifting apart (source: denali.gsfc.nasa.gov).

good enough. In order to obtain a more practical solution, Europe chose to 'fix' the 1989 ITRS to the Eurasian continental plate. This resulted in ETRS1989 (European Terrestrial Reference System 1989), which is now the official geodetic datum for use within the European Union.

ETRS89 versus WGS84

Since most data acquisition nowadays takes place using GPS systems, it is important to understand how the various geodetic datums are related to one another. Usually the local geodetic datum such as NAD23 is related to another datum like WGS84. In Europe, however, the local geodetic datums are related to ETRS89 and not to WGS84. GPS measurements are, however, always related to WGS84 in the same way that Galileo measurements will be related to ETRS89. With the continuous drift of the continents, the relation between ETRS89 and WGS84 is constantly changing. The differences are small enough to be of no concern when using stand-alone or code phase dGPS. When using more accurate carrier phase GPS systems such as RTK dGPS, though, problems can occur. However, since all carrier-phase systems employ relative positioning techniques, the errors are usually relatively small, to within a few millimeters. In practice the coordinates of the base station can be entered in ETRS89, resulting in rover coordinates referenced to the same system, even with its geodetic datum set to WGS84. The resulting WGS84 coordinates can then be transformed to a local geodetic datum using, for example, the NADCON conversion as stated for converting WGS84 coordinates to NAD23.

Coordinates

A position determined relative to a geodetic datum is always expressed as a longitude and latitude (on that datum). When needed, it can be augmented with the so-called ellipsoid height. The position of a point (or GPS receiver) is thus always related to a certain horizontal datum. When multiple datums are used on a certain project, deviations may occur between the 'true' position and the calculated position.
If, for example, a GPS receiver giving coordinates referenced to WGS84 is used to determine the ED50 coordinates of an offshore platform in the North Sea, an error of approximately 180 meters may be found.

Finally

The ellipsoid is the basis for all geodetic calculations. When using satellite positioning, or when using charts of larger areas, it is relevant to know the geodetic datum on which the positions are computed.

Huibert-Jan Lekkerkerk (hlekkerkerk@geoinformatics.com) is a freelance writer and trainer in the fields of positioning and hydrography.