Datums, Coordinate Systems and GPS

By Ahmed El-Rabbany[1] and Dan Lymer[2]

[1] Dept. of Civil Engineering (Geomatics Option), Ryerson University

[2] City of Toronto

The ability of GPS to determine the precise coordinates of a user anywhere and under any weather condition, attracted millions of users worldwide from various fields and backgrounds. With the advances in the GPS and the computer technologies, GPS manufacturers were able to come up with very user-friendly systems. One common problem, however, that many GPS users face is the issue of datums and coordinate systems. This article tackles this problem and attempts to provide possible solutions to overcome it.

What is a Datum?

The fact that the topographic surface of the earth is highly irregular makes it difficult for the geodetic calculations, for example the determination of the user's location, to be performed. To overcome this problem, geodesists adopted a smooth *mathematical* surface, called the reference surface, to approximate the irregular shape of the earth (more precisely to approximate the *global mean sea level, the geoid*). For high accuracy positioning such as GPS positioning, the best mathematical surface to approximate the earth and at the same time keep the calculations as simple as possible, was found to be the biaxial ellipsoid (see Figure 1). The biaxial reference ellipsoid, or simply the reference ellipsoid, is obtained by rotating an ellipse around its minor axis, *b* [5].

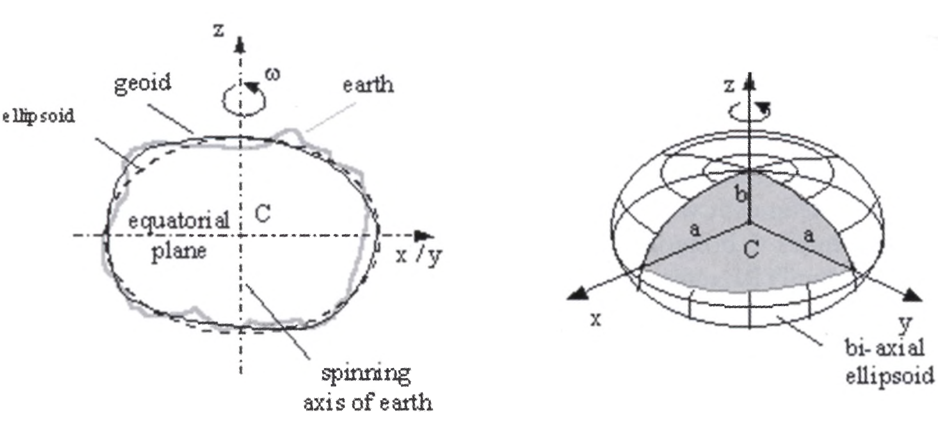
An appropriately positioned reference ellipsoid is known as the *geodetic datum* [5]. In other words, a geodetic datum is a mathematical surface, or a reference ellipsoid, with a well-defined origin (center) and orientation. For example, a geocentric geodetic datum is a geodetic datum with its origin coinciding with the center of the earth. It is clear that there are an infinite number of geocentric geodetic datums with different orientations. A geodetic datum is, therefore, uniquely determined by specifying 8 parameters: two parameters to define the dimension of the reference ellipsoid; three parameters to define the position of the origin; and three parameters to define the orientation of the three axes with respect to the earth. Table 1 shows some examples of three common reference systems and their associated ellipsoids [3].

Table 1.
Examples of Reference Systems and Associated Ellipsoids

Ref. Sys.	Ellipsoid	a (m)	$1/f$
WGS 84	WGS 84	6378137.0	298.257223563
NAD 83	GRS 80	6378137.0	298.257222101
NAD 27	Clarke 1866	6378206.4	294.9786982

In addition to the geodetic datum, the so-called *vertical datum* is also used in practice as a reference surface to which the heights (elevations) of points are referred [5]. Because the height of a point directly located on the vertical datum is zero, such a vertical reference surface is commonly known as the surface of zero height. The vertical datum is often selected to be the *geoid*; the surface that best approximates the mean sea level on a global basis (see Figure 1a).

In the past, positions with respect to horizontal and vertical datums have been determined independent of each other [5]. However, with the advent of space geodetic positioning systems like GPS, it is possible to determine the three-dimensional positions with respect to a three-dimensional reference system.



(a) (b)

Figure 1. (a) Relationship Between the Physical Surface of the Earth, the Geoid and the Ellipsoid. (b) Ellipsoidal Parameters.

Geodetic Coordinate System

A coordinate system is defined as a set of rules for specifying the coordinates of points [3]. This usually involves specifying an origin of the coordinates as well as a set of reference lines (axes) with known orientation. Coordinate systems may be classified as one-dimensional (1-D), two-dimensional (2-D) or three-dimensional (3-D) coordinate systems, according to the number of coordinates required to identify the location of a point. For example, a 1-D coordinate system is needed to identify the height of a point above the sea surface.

Coordinate systems may also be classified according to the reference surface, the orientation of the axes and the origin. In the case of a 3-D geodetic coordinate system, the reference surface is selected to be the ellipsoid. The orientation of the axes and the origin are specified by two planes: the meridian plane through the polar or z-axis (a meridian is a plane that passes through the north and south poles) and the equatorial plane of the ellipsoid.

Of particular importance to the GPS users is the 3-D geodetic coordinate system. In this system, the coordinates of a point are identified by the geodetic latitude (ϕ), the geodetic longitude (λ) and the height above the reference surface (h). Geodetic coordinates (ϕ, λ, and h) can be easily transformed to Cartesian coordinates (x, y and z). It is also possible to transform the geodetic coordinates (ϕ and λ) into a rectangular grid coordinate (e.g. Northing and Easting) for mapping purposes.

Conventional Terrestrial Reference Systems

The Conventional Terrestrial Reference System (CTRS) is a three-dimensional geocentric coordinate system; that is its origin coincides with the center of the earth. The CTRS system is rigidly tied to the earth; i.e. it rotates with the earth [5]. It is therefore known also as Earth-Centered, Earth-Fixed (ECEF) coordinate system.

The orientation of the axes of the CTRS is defined as follows: The Z-axis points towards the Conventional Terrestrial Pole (CTP), which is defined as the average location of the pole during the period 1900-1905 [5]. The X-axis is defined by the intersection of the *terrestrial* equatorial plane and the meridianal plane that contains the mean location of Greenwich observatory (known as the mean Greenwich meridian). It is clear from the definition of the X and Z axes that the XZ-plane contains the mean Greenwich meridian. The Y-axis is selected to make the coordinate system right-handed (i.e. 90° east of the X-axis, measured in the equatorial plane). The three axes intersect at the center of the earth.

The CTRS must be *positioned with respect to the earth* (known as *realization*) to be of practical use in positioning. This is done by assigning coordinate values to a selected number of well-distributed reference stations. One of the most important CTRS is the International Terrestrial Reference System (ITRS), which is realized as the International Terrestrial Reference Frame (ITRF). The ITRF solution is based on the measurements from globally distributed reference stations using GPS and other space geodetic systems. It is therefore considered to be the most accurate coordinate system [1]. The ITRF is updated every one to three years to achieve the highest possible accuracy. The most recent version at the time of this writing is the ITRF2000.

The WGS 84 and NAD 83 Systems

The World Geodetic System of 1984 (WGS 84) is a three-dimensional, earth-centered reference system developed by the former US Defense Mapping Agency (now incorporated into a new agency, NIMA). It is the official GPS reference system. In other words, a GPS user who employs the broadcast ephemeris in the solution process will get his/her coordinates in the WGS 84 system. The WGS 84 utilizes the CTRS combined with a reference ellipsoid that is identical, up to a very slight difference in flattening, with the ellipsoid of the Geodetic Reference System of 1980 (GRS 80); see Table 1. WGS 84 was originally established (realized) using a number of Doppler stations. It was then updated several times to bring it as close as possible to the ITRF reference system. With the most recent update, WGS 84 is coincident with the ITRF at the sub-decimeter accuracy level [3].

In North America, another *nominally* geocentric datum, the North American Datum of 1983 (NAD 83), is used as the legal datum for spatial positioning. NAD 83 utilizes the ellipsoid of the GSR 80, which means that the size and shape of both WGS 84 and NAD 83 are almost identical. Original realization of NAD 83 was done in 1986, by adjusting: "primarily" classical geodetic observations that connected a network of *horizontal* control stations spanning North America; and several hundred observed Doppler positions. Initially, NAD 83 was designed as an earth-centered reference system [2]. However, with the development of more accurate techniques, it was found that the origin of NAD 83 is shifted by about 2 m from the true earth's center. In addition, access to NAD 83 was provided mainly through a horizontal control network, which has a limited accuracy due to the accumulation of errors. To overcome these limitations, NAD 83 was tied to ITRF using 12 common VLBI stations located in both Canada and USA (VLBI is a highly accurate, yet complex, space positioning system). This resulted in an improved realization of the NAD 83, which is referred to as NAD 83 (CSRS) and NAD 83 (NSRS) in both Canada and the USA, respectively [2]. The acronyms CSRS and NSRS refer to the Canadian Spatial Reference System and National Spatial Reference System, respectively. It should be pointed out that, due to the different versions of the ITRF, it is important to define the epoch to which the ITRF coordinates refer.

Datum Transformations

As stated earlier, in the past positions with respect to horizontal and vertical datums have been determined independent of each other. In addition, horizontal datums were non-geocentric and were selected to best fit certain regions of the world. As such those datums were commonly called *local datums*. Over 150 local datums are used by different countries of the world. An example of the local datums is the North American datum of 1927 (NAD 27). With the advent of space geodetic positioning systems like GPS, it is now possible to determine global three dimension geocentric datums.

Old maps were produced with the local datums while the new maps are mostly produced with the geocentric datums. Therefore, to ensure consistency, it is necessary to establish the relationships between the local datums and the geocentric datums, such as WGS 84. Such a relationship is known as the *datum transformation*. The former US Defense Mapping Agency (NIMA) has published the transformation parameters between WGS 84 and the various local datums used in many countries. Many GPS manufacturers currently use these parameters within their processing software packages. It should be clear, however, that these transformation parameters are only approximate and should not be used for precise GPS applications. In Toronto, for example, a difference as large as several meters in the horizontal coordinates is obtained when applying the NIMA's parameters (WGS 84 to NAD 27) as compared to the more precise National Transformation software (NTv.2) produced by Geomatics Canada. Such a difference could be even larger in other regions. The best way to obtain the transformation parameters is by comparing the coordinates of well-distributed common points in both datums.

Moving Towards NAD 83 - The City of Toronto's Initiative

Metropolitan Toronto and the Province of Ontario began establishing geodetic control frameworks to support the growth seen during the 1970s. Although built on the Federal First Order Geodetic Network, the two networks grew independently of each other. Integration between the networks usually took place along the edges of the City. In addition to the networks being constructed independently, the adjustments relating the NAD 27 began to differ. In 1974 and 1976 the Province readjusted the network by holding the perimeter points around Toronto fixed leaving the City on the 1968 Adjustment of NAD 27 with coordinates expressed in the 3-degree Modified Transverse Mercator projection. The Province adopted the NAD 27, 1974 and 1976 adjustments with a 6-degree Universal Transverse Mercator projection.

The NAD 27 system was found to be distorted due to a number of factors, including the geometrical weakness in the network, the unavailability of an accurate geoid, and non-rigorous estimation methods [4]. The ever-growing need for spatial data referred to a continuous and consistent reference system has created the problems associated with transforming data between systems to satisfy the needs of the Geomatics community and government agencies. To overcome these problems, the City, in partnership with Provincial Geo-referencing of the Ministry of Natural Resources (MNR), is taking steps to harmonize the networks in NAD 83 (CSRS) and migrate the City's control records to the COSINE database. The conversion will create grid shift parameter files, enabling reliable conversions between the existing NAD 27 coordinates and the corresponding NAD 83 (CSRS) values. This will provide a conversion that will precisely reflect the relationship between the two reference surfaces. The grid shift files will be produced in a format consistent with the grid shift files used in the current datum transformation application NTv.2. Currently, NTv.2 transforms coordinates between NAD 83 adopted and NAD 27 based on a nationwide grid. The grid shift files generated from the City's project will cover the Toronto area and provide a more detailed model for the distortions in NAD 27.

References:

[1] Boucher, C. and Z. Altamimi (1996). "International Terrestrial Reference Frame." *GPS World*, September, Vol. 7, No. 9, pp.71-74.

[2] Craymer, M., R. Ferland, R. Snay (1998). "Realization and Unification of NAD83 in Canada and the U.S. via the ITRF." *Proceedings of the International Symposium of the International Association of Geodesy*, Section II, "Towards an Integrated Geodetic Observing System (IGGOS)", Munich, October 5-9.

[3] El-Rabbany, A. (2001). "*An Introduction to GPS: The Global Positioning System*". Artech House Publishing.

[4] Pinch, M.C. (1990). "Differences Between NAD 27 and NAD 83." *NAD 83 Implementation Seminar*, The Canadian Institute of Geomatics.

[5] Vanicek, P. and E.J. Krakiwsky (1986). *Geodesy: The Concepts*. 2nd Edition, North Holland, Amsterdam.

Calendar of Events

* * * * *

February 20th to 22nd, 2002

AOLS' Annual Meeting
Toronto, Ontario
www.aols.org

April 8th to 11th, 2002

GEOTech Event
The Global Conference and Exhibition on Geospatial Technology, Tools and Solutions
Toronto, Ontario
www.GEOTechEvent.com

April 19th to 26th, 2002

The International Federation of Surveyors (FIG) Congress 2002
Washington, D.C.
www.fig2002.com

June 10th to 12th, 2002

Geospatial World 2002
Atlanta, Georgia
www.intergraph.com/geospatialworld