PDHonline Course L105G (12 PDH)

GPS Surveying

Instructor: Jan Van Sickle, P.L.S.

2020

PDH Online | PDH Center

5272 Meadow Estates Drive
Fairfax, VA 22030-6658
Phone: 703-988-0088
www.PDHonline.com

An Approved Continuing Education Provider

Module 5

A Few Ideas about Geodesy for GPS

GPS satellites orbit around the center of mass of the earth. It is not possible to survey with GPS without taking the figure of the earth into account. When you do surveying with GPS, you are doing geodetic surveying.

Plane Surveying

Plane surveying has relied on an imaginary flat reference surface, or *datum*, with Cartesian axes. This rectangular system is used to describe measured positions by ordered pairs, usually expressed in northings and eastings, or *x*- and *y*- coordinates. Even though surveyors have always known that this assumption of a flat earth is fundamentally unrealistic, it provided, and continues to provide, an adequate arrangement for small areas. The attachment of elevations to such horizontal coordinates somewhat acknowledges the topographic irregularity of the earth, but the whole system is undone by its inherent inaccuracy as the area over which surveys extend grow large.

State Plane Coordinates

In the 1930s, an engineer in North Carolina's highway department, George F. Syme, appealed to the then United Coast and Geodetic Survey (C&GS, now NGS) for help. He had found that the stretching and compression inevitable in the representation of the curved earth on a plane was so severe over his long-route surveys that he could not check into the C&GS geodetic control stations across his state within reasonable limits. To alleviate the problem, Dr. O.S. Adams of the Division of Geodesy designed the first state plane coordinate system in 1933. The approach was so successful in North Carolina similar systems were devised for all the states in the Union within a year or so.

The purpose of the state plane coordinate system was to overcome some of the limitations of the horizontal plane datum while avoiding the imposition of geodetic methods and calculations on local surveyors. Using the conic and cylindrical models of the Lambert and Mercator map projections, the flat datum was curved, but only in one direction. By curving the datums and limiting the area of the zones, Dr. Adams managed to limit the distortion to a scale ratio of about 1 part in 10,000 without disturbing the traditional system of ordered pairs of Cartesian coordinates familiar to surveyors.

The state plane coordinate system was a step ahead at that time. To this day, it provides surveyors with a mechanism for coordination of surveying stations that approximates geodetic accuracy more closely than the commonly used methods of small-scale plane surveying. However, the state plane coordinate systems were organized in a time of generally lower accuracy and efficiency in surveying measurement. Its calculations were designed to avoid the lengthy and complicated mathematics of geodesy. It was an understandable compromise in an age when such computation required sharp pencils, logarithmic tables, and lots of midnight oil.

Today, GPS has thrust surveyors into thick of geodesy, which is no longer, the exclusive realm of distant experts. Thankfully, in the age of the microcomputer, the computational drudgery can be handled with software packages. Nevertheless, it is unwise to venture into GPS believing that knowledge of the basics of geodesy is, therefore, unnecessary. It is true that GPS would be impossible without computers, but blind reliance on the data they generate eventually leads to disaster.

Some Geodetic Coordinate Systems

A spatial Cartesian system with three axes lends itself to describing the terrestrial positions derived from space-based geodesy. Using three rectangular coordinates instead of two, one can unambiguously define any position on the earth, or above it for that matter. But such a system is only useful if its origin (0,0,0) and its axes (x,y,z) can be fixed to the planet with certainty, something easier said than done (Figure 5.1).

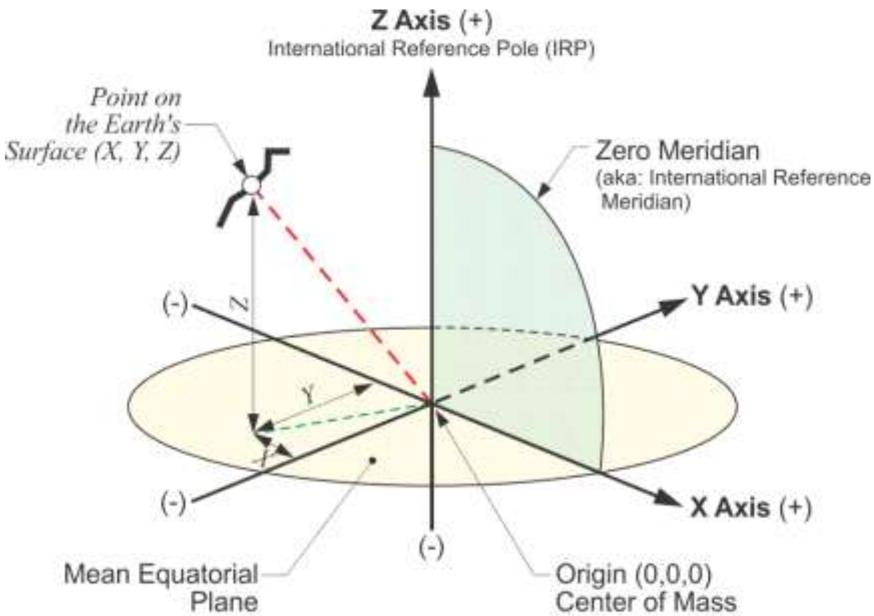


Figure 5.1

The usual arrangement is known as the *conventional terrestrial system (CTS)*. The origin is the center of mass of the earth, the *geocenter*. The *x*-axis is a line from the geocenter through the intersection of the Greenwich meridian, the zero meridian, with the equator. The *y*-axis is extended from the geocenter along a line perpendicular from the *x*-axis in the same mean equatorial plane. They both rotate with the earth as part of a *right-handed* orthogonal system.

A three-dimensional Cartesian coordinate system is right-handed if the following model can describe it: the extended forefinger of the right hand symbolizes the positive direction of the *x*-axis. The middle finger of the same hand extended at right angles to the forefinger symbolizes the positive direction of the *y*-axis. The extended thumb of the right hand, perpendicular to them both, symbolizes the positive direction of the *z*-axis. In applying this model to the earth, the *z*-axis is imagined to nearly coincide with the earth's axis of rotation, and therein lies a difficulty.

The earth's rotational axis will not hold still. It actually wanders slightly with respect to the solid earth in a very slow oscillation called *polar motion*. The largest component of the movement relative to the earth's crust has a 430-day cycle known as the *Chandler period*. The actual displacement caused by the wandering generally does not exceed 12 meters. Nevertheless, the conventional terrestrial system of coordinates would be useless if its third axis was constantly wobbling. Therefore, an average stable position was chosen for the position of the pole and the *z*-axis.

Between 1900 and 1905, the mean position of the earth's rotational pole was designated as the *Conventional International Origin (CIO)*. It has since been refined by the *International Earth Rotation Service (IERS)* using very long baseline interferometry (VLBI) and satellite laser ranging (SLR). The name of the z-axis has been changed to the *Conventional Terrestrial Pole (CTP)*. But its role has remained the same: also the CTP provides a stable and clear definition on the earth's surface for the z-axis. So, by international agreement, the z-axis of the *Conventional Terrestrial System (CTS)* is a line from the earth's center of mass through the CTP.

The three-dimensional Cartesian coordinates (x,y,z) derived from this system are sometimes known as *earth-centered-earth-fixed (ECEF)* coordinates. They are convenient for many types of calculations, such as the single-baseline or single-vector solution in GPS. In fact, most modern GPS software provides data that express vectors as the difference between the *x*, *y* and *z* coordinates at each end of the baselines. The display of these differences as *DX*, *DY* and *DZ* is a typical product of these postprocessed calculations.

Latitude and Longitude

Despite their utility, such 3-D Cartesian coordinates are not the most common method of expressing a geodetic position. Latitude and longitude have been the coordinates of choice for centuries. The application of these angular designations relies on the same two standard lines as 3-D Cartesian coordinates: the mean equator and the Greenwich meridian. Unlike the CTS, they require some clear representation of the terrestrial surface. In modern practice, latitude and longitude cannot be said to uniquely define a position without a clear definition of the earth itself.

Elements of a Geodetic Datum

How can latitude, *phi*, and longitude, *lambda*, be considered inadequate in any way for the definition of a position on the earth? The reference lines - the mean equator and the Greenwich meridian - are clearly defined. The units of degrees, minutes, seconds, and decimals of seconds, allow for the finest distinctions of measurement. Finally, the reference surface is the earth itself.

Despite the certainty of the physical surface of the earth, the *lithosphere* remains notoriously difficult to define in mathematical terms. The dilemma is illustrated by the ancient struggle to represent its curved surface on flat maps. There have been a whole variety of map projections developed over the centuries that rely on mathematical relationships between positions on the earth's surface and points on the map. Each projection serves a particular application well, but none of them can represent the earth without distortion. For example, no modern surveyor would presume to promise a client a high-precision control network with data scaled from a map.

As the technology of measurement has improved, the pressure for greater exactness in the definition of the earth's shape has increased. Even with electronic tools that widen the scope and increase the precision of the data, perfection is nowhere in sight.

The Shape of the Earth

Despite the fact that local topography is the most obvious feature of the lithosphere to an observer standing on the earth, efforts to grasp the more general nature of the planet's shape and size have been occupying scientists for at least 2,300 years. There have, of course, been long intervening periods of unmitigated nonsense on the subject. Ever since 200 B.C. when Eratosthenes almost calculated the planet's circumference correctly, geodesy has been getting ever closer to expressing the actual shape of the earth in numerical terms. A leap forward occurred with Newton's thesis that the earth was an ellipsoid rather than a sphere in the first edition of his ***Principia*** in 1687.

Newton's idea that the actual shape of the earth was slightly ellipsoidal was not entirely independent. There had already been some other suggestive observations. For example, 15 years earlier astronomer J. Richter had found that to maintain the accuracy of the one-second clock he used in his observations in Cayenne, French Guiana, he had to shorten its pendulum significantly. The clock's pendulum, regulated in Paris, tended to swing more slowly as it approached the equator. Newton reasoned that the phenomenon was attributable to a lessening of the force of gravity. Based on his own theoretical work, he explained the weaker gravity by the proposition, "the earth is higher under the equator than at the poles, and that by an excess of about 17 miles" (Philosophiae naturalis principia mathematica, Book III, Proposition XX).

Although Newton’s. model of the planet bulging along the equator and flattened at the poles was supported by some of his contemporaries, notably Huygens, the inventor of Richter's clock, it was attacked by others. The director of the Paris Observatory, Jean Dominique Cassini, for example, took exception with Newton's concept. Even though the elder Cassini had himself observed the flattening of the poles of Jupiter in 1666, neither he nor his equally learned son Jacques were prepared to accept the same idea when it came to the shape of the earth. It appeared they had some empirical evidence on their side.

For geometric verification of the earth model, scientists had employed arc measurements at various latitudes since the early 1500s. Establishing the latitude of their beginning and ending points astronomically, they measured a cardinal line to discover the length of one degree of longitude along a meridian arc. Early attempts assumed a spherical earth and the results were used to estimate its radius by simple multiplication. In fact, one of the most accurate of the measurements of this type, begun in 1669 by the French abbé J. Picard, was actually used by

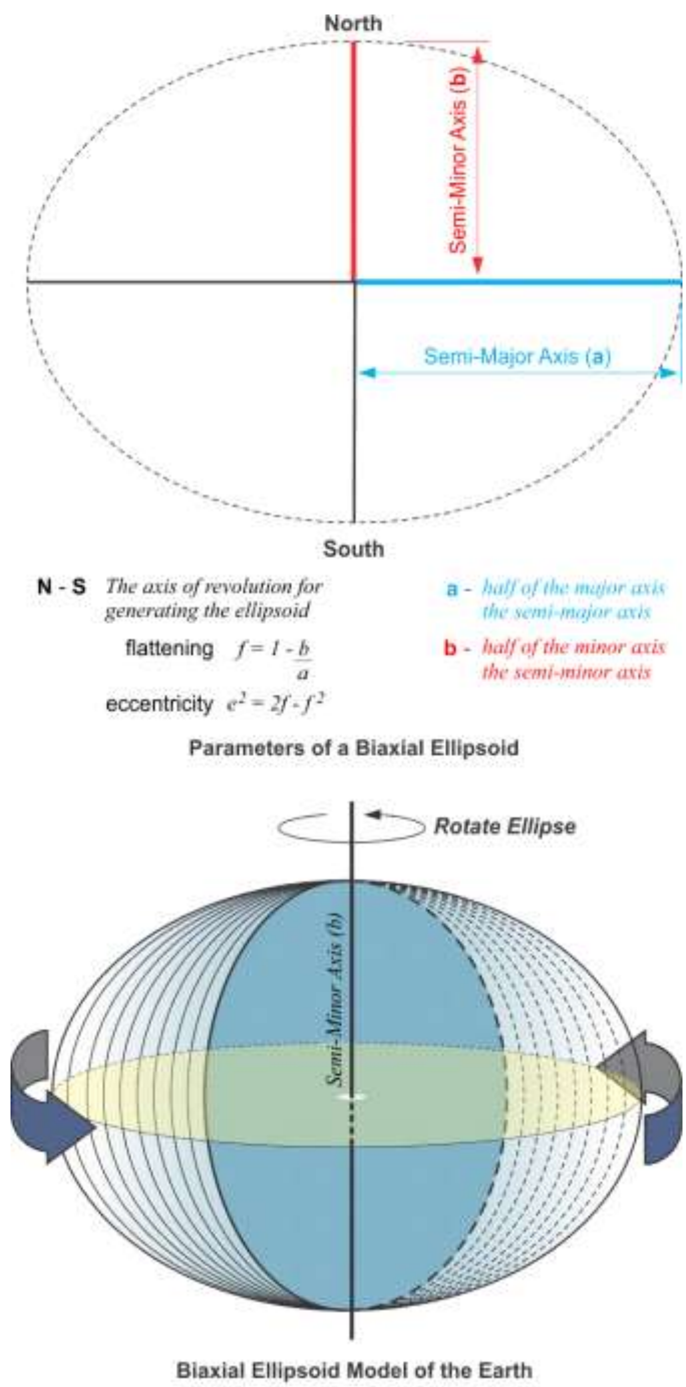


Figure 5.2

Newton in formulating his own law of gravitation. However, Cassini noted that close analysis of Picard's arc measurement, and others, seemed to show the length of one degree of longitude actually *decreased* as it proceeded northward. He concluded that the earth was not flattened as proposed by Newton, but was rather elongated at the poles.

The argument was not resolved until two expeditions begun in 1733 and completed in 1744. They were sponsored by the Paris Académie Royale des Sciences produced irrefutable proof. One group which included Clairaut and Maupertuis was sent to measure a meridian arc near the Arctic Circle, 66 degrees20 minutes N Latitude, in Lapland. Another expedition with Bouguer and Godin, to what is now Ecuador, measured an arc near the equator, 01 degrees 31 minutes S Latitude. Newton's conjecture was proved correct, and the contradictory evidence of Picard's arc was charged to errors in the latter's measurement of the astronomic latitudes.

Ellipsoids

The ellipsoidal model (Figure 5.2), bulging at the equator and flattened at the poles, has been used ever since as a representation of the general shape of the earth's surface. In fact, several reference ellipsoids have been established for various regions of the planet. They are precisely defined by their semimajor axis and flattening. The relationship between these parameters are

$$f = \frac{a - b}{a}$$

expressed in the formula:

Where f = flattening, a = semimajor axis, and b = semiminor axis.

The Role of an Ellipsoid in a Datum

The semimajor and flattening can be used to completely define an ellipsoid. However, six additional elements are required if that ellipsoid is to be used as a *geodetic datum:* three to specify its center and three more to clearly indicate its orientation around that center. The *Clarke 1866 spheroid* is one of many reference ellipsoids. Its shape is completely defined by a semimajor axis, *a*, of 6378.2064 km and a flattening, *f,* of 1/294.9786982. It is the reference ellipsoid of the datum known to surveyors as the *North American Datum of 1927 (NAD27),* but it is not the datum itself.

For the Clarke 1866 spheroid to become NAD27, it had to be attached at a point and specifically oriented to the actual surface of the earth. However, even this ellipsoid, which fits North America the best of all, could not conform to that surface perfectly. Therefore, the initial point was chosen near the center of the anticipated geodetic network to best distribute the inevitable distortion. The attachment was established at Meades Ranch, Kansas, 39 degrees 13 minutes 26.686 seconds N Latitude, 98 degrees 32 minutes 30. 506 seconds W Lonigtude and *geoidal height* zero (we will discuss geoidal height later). Those coordinates were not sufficient, however. The establishment of directions from this initial point was required to complete the orientation. The azimuth from Meades Ranch to station Waldo was fixed at 75 degrees 28 minutes 09. 64 seconds and the deflection of the vertical set at zero.

Once the initial point and directions were fixed, the whole orientation of NAD27 was established, including the center of the reference ellipsoid. Its center was imagined to reside somewhere around the center of mass of the earth. However, the two points were certainly not coincident, nor were they intended to be. In short, NAD27 does not use a geocentric ellipsoid.

Datums

In the period before space-based geodesy was tenable, such a regional datum was not unusual. The *Australian Geodetic Datum 1966*, the *Datum Eurpeén 1950*, and the *South American Datum 1969*, among others, were also designed as nongeocentric systems. Achievement of the minimum distortion over a particular region was the primary consideration in choosing their ellipsoids, not the relationship of their centers to the center of mass of the earth (Figure 5.3). For example, in the Conventional Terrestrial System (CTS) the 3-D Cartesian coordinates of the center of the Clarke 1866 spheroid as it was used for NAD27 are about X = -4 m, Y = +166 m and Z = +183 m.

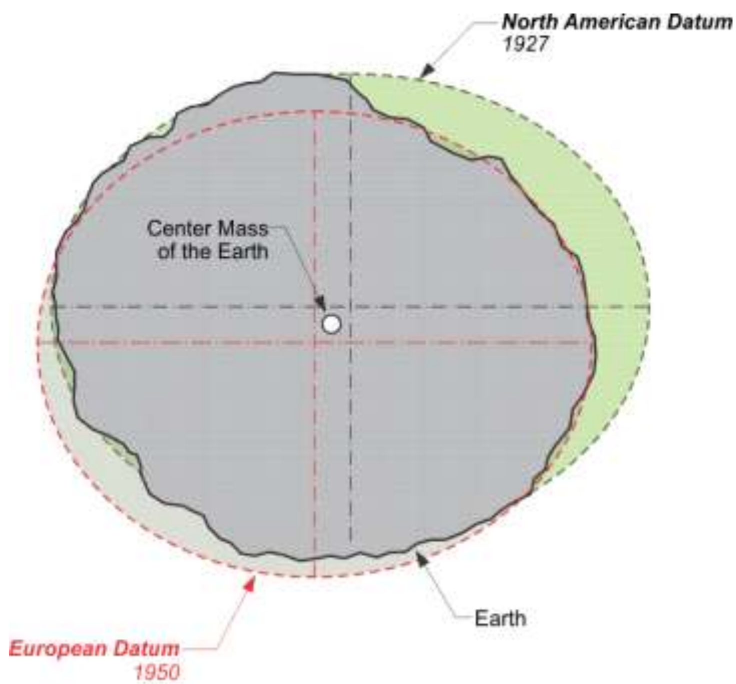


Figure 5.3

This approach to the design of datums was bolstered by the fact that the vast majority of geodetic measurements they would be expected to support were of the classical variety. That is, the work was done with theodolites, towers, and tapes. They were, in short, earth-bound. Even after the advent of electronic distance measurement, the general approach involved the determination of horizontal coordinates by measuring from point to point on the earth's surface and adding heights, otherwise known as *elevations*, through a separate leveling operation. As long as this methodological separation existed between the horizontal and vertical coordinates of a station, the difference between the ellipsoid and the true earth's surface was not an overriding concern. Such circumstances did not require a geocentric datum.

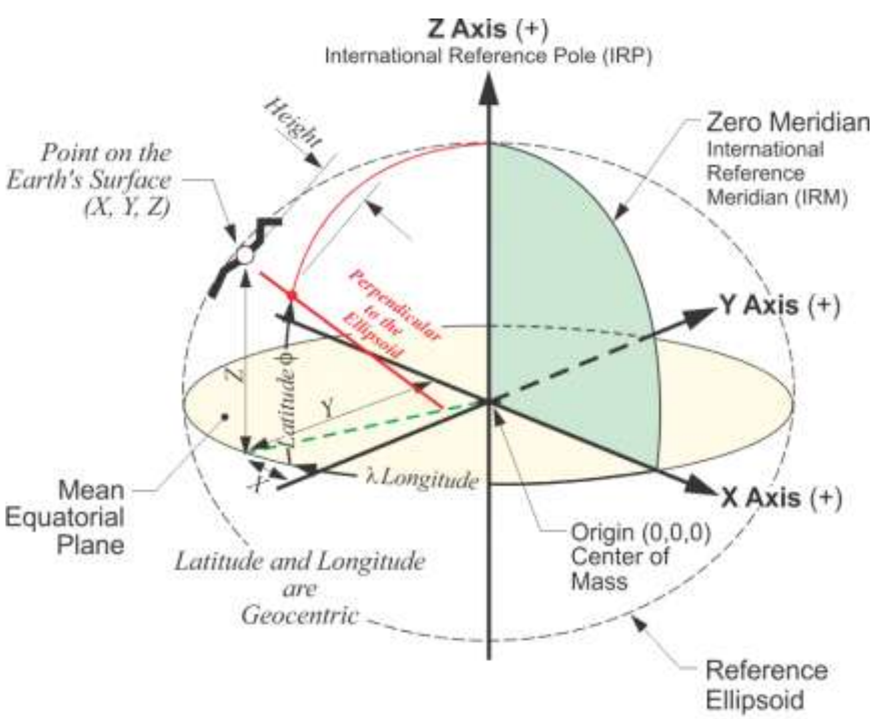


Figure 5.4

However, as the sophistication of satellite geodesy increased, the need for a truly global, geocentric datum became obvious. The horizontal and vertical information were no longer separate. Since satellites orbit around the center of mass of the earth, a position derived from space-based geodesy can be visualized as a vector originating from that point.

So, today, not only are the horizontal and vertical components of a position derived from precisely the same vector, the choice of the coordinate system used to express them is actually a matter of convenience. The position vector can be transformed into the 3D Cartesian system of CTS, the traditional latitude, longitude and height, or virtually any other well-defined coordinate system. However, since the orbital motion and the subsequent position vector derived from satellite geodesy are themselves earth-centered, it follows that the most straightforward representations of that data are earth-centered as well (Figure 5.4).

A Geocentric Model

Satellites have not only provided the impetus for a geocentric datum; they have also supplied the means to achieve it. In fact, the orbital perturbations of man-made near-earth satellites have probably brought more refinements to the understanding of the shape of the earth in a shorter span of time than was ever before possible. For example, the analysis of the precession of Sputnik 2 in the late 1950s showed researchers that the earth's semiminor axis was actually 85 meters shorter than had been previously thought. In 1958, while studying the tracking data from the orbit of Vanguard I, Ann Bailey of the Goddard Spaceflight Center discovered that the planet is shaped a bit like a pear. There is a slight protuberance at the North Pole, a little depression at the South Pole, and a small bulge just south of the equator.

These formations and others have been discovered through the observation of small distortions in satellites otherwise elliptical orbits, little bumps in their road, so to speak. The deviations are caused by the action of earth's gravity on the satellites as they travel through space. Just as Richter's clock reacted to the lessening of gravity at the equator and thereby revealed one of the largest features of the earth's shape to Newton, small perturbations in the orbits of satellites, also responding to gravity, reveal details of earth's shape to today's scientists. The common aspect of these examples is the direct relationship between direction and magnitude of gravity and the planet's form. In fact, the surface that best fits the earth's gravity field has been given a name. It is called the *geoid*.

The Geoid

An often-used description of the geoidal surface involves idealized oceans. Imagine the oceans of the world utterly still, completely free of currents, tides, friction, variations in temperature and all other physical forces, except gravity. Reacting to gravity alone, these unattainable calm waters would coincide with the figure known as the geoid. Admitted by small frictionless channels or tubes and allowed to migrate across the land, the water would then, theoretically, define the same geoidal surface across the continents, too.

Of course, the 70 percent of the earth covered by oceans are not so cooperative, nor is there any such system of channels and tubes. In addition, the physical forces eliminated from the model cannot be avoided in reality. These unavoidable forces actually cause mean sea level to deviate up to 1, even 2, meters from the geoid, a fact frequently mentioned to emphasize the inconsistency of the original definition of the geoid as it was offered by J.B. Listing in 1872. Listing thought of the geoidal surface as equivalent to mean sea level. It is important to remember that mean sea level (MSL) and the geoid are not the same. Even though his idea does not stand up to scrutiny today, it can still be instructive.

An Equipotential Surface

Gravity is not consistent across the topographic surface of the earth. At every point it has a magnitude and a direction. In other words, anywhere on the earth, a mathematical vector can describe gravity. Along the solid earth, such vectors do not have all the *same* direction or magnitude, but one can imagine a surface of constant gravity potential. Such an *equipotential* surface would be *level* in the true sense. It would coincide with the top of the hypothetical water in the previous example. Despite the fact that real mean sea level does not define such a figure, the geoidal surface is not just a product of imagination. For example, the vertical axis of any properly leveled surveying instrument and the string of any stable plumb bob are perpendicular to the geoid. Just as pendulum clocks and earth-orbiting satellites, they clearly show that the geoid is a reality.

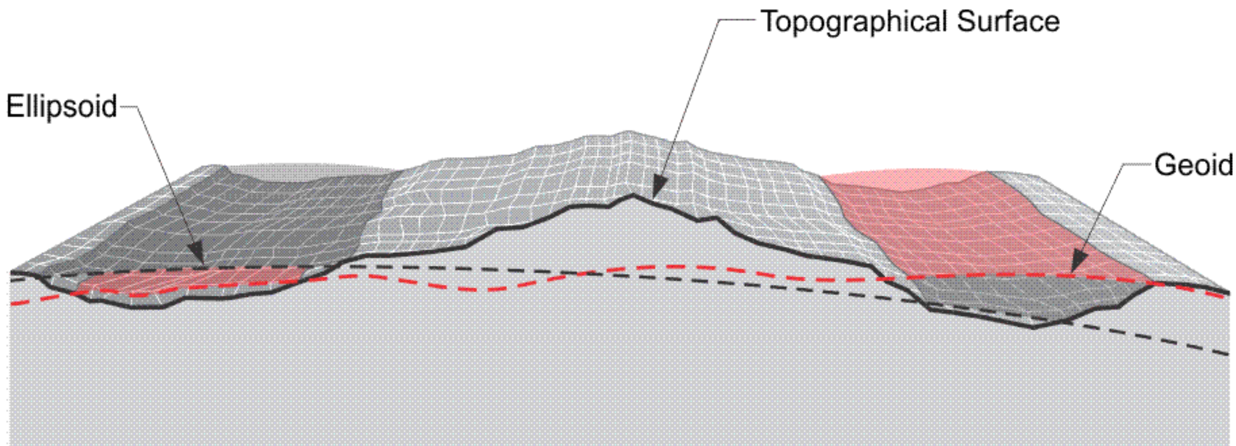


Figure 5.5

Just as the geoid does not precisely follow mean sea level, neither does it exactly correspond with the topography of the dry land. However, it is similar to the terrestrial surface. It has similar peaks and valleys. It is bumpy. Uneven distribution of the mass of the planet makes it maddeningly so. Maddening because if the solid earth had no internal anomalies of density, the geoid would be smooth and almost exactly ellipsoidal. In that case, the reference ellipsoid could fit the geoid to near perfection and the lives of geodesists would be much simpler. But like the earth itself, the geoid defies such mathematical consistency and departs from true ellipsoidal form by as much as 100 meters in places (Figure 5.5).

The Modern Geocentric Datum

Three distinct figures are involved in a geodetic datum for latitude, longitude and height: the geoid, the reference ellipsoid, and the earth itself. Due in large measure to the ascendancy of satellite geodesy, it has become highly desirable that they share a common center.

While the level surface of the geoid provides a solid foundation for the definitions of heights (more about that later) and the topographic surface of the Earth is necessarily where measurements are made, neither can serve as the reference surface for geodetic positions. From the continents to the floors of the oceans, the solid earth's actual surface is too irregular to be represented by a simple mathematical statement. The geoid, which is sometimes under, and sometimes above, the surface of the earth, has an overall shape that also defies any concise geometrical definition. But the ellipsoid not only has the same general shape as the earth, but, unlike the other two figures can be described simply and completely in mathematical terms.

Therefore, a global geocentric system has been developed based on the ellipsoid adopted by the *International Union of Geodesy and Geophysics (IUGG)* in 1979. It is called the *Geodetic Reference System 1980 (GRS80)*. Its semimajor axis, *a*, is 6378.137 km and is probably within a very few of meters of the earth's actual equatorial radius. Its flattening, *f*, is 1/298.25722 and likely deviates only slightly from the true value, a considerable improvement over Newton's calculation of a flattening ratio of 1/230. But then he did not have orbital data from near-earth satellites to check his work.

With very slight changes, GRS80, is the reference ellipsoid for the coordinate system, known as the *World Geodetic System 1984 (WGS84)*. This datum has been used by the U.S. military since January 21, 1987, as the basis for the GPS Navigation message computations. Therefore, coordinates provided directly by GPS receivers are based in WGS84. However, most available GPS software can transform those coordinates to a number of other datums as well. The one that is probably of greatest interest to surveyors in the United States today is the *North American Datum 1983 (NAD83)*. But the difference between WGS84 and NAD83 coordinates is so small, usually about the 0.1-mm level that transformation is unnecessary and they can be considered equivalent for most applications.

North American Datum 1983

NAD27

The Clarke 1866 ellipsoid was the foundation of NAD27, and the blocks that built that foundation were made by geodetic triangulation. After all, an ellipsoid, even one with a clearly stated orientation to the earth, is only an abstraction until physical, identifiable control stations are available for its practical application. During the tenure of NAD27, control positions were tied together by triangulation. Its measurements grew into chains of figures from Canada to Mexico and coast to coast, with their vertices perpetuated by bronze disks set in stone, concrete, and other permanent media.

These tri-stations, also known as brass caps and their attached coordinates have provided a framework for all types of surveying and mapping projects for many years. They have served to locate international, state, and county boundaries. They have provided geodetic control for the planning of national and local projects, development of natural resources, national defense, and land management. They have enabled surveys to be fitted together, provided checks, and assisted in the perpetuation of their marks. They have supported scientific inquiry, including crustal monitoring studies and other geophysical research. But even as application of the nationwide control network grew, the revelations of local distortions in NAD27 were reaching unacceptable levels.

Judged by the standards of newer measurement technologies, the quality of some of the observations used in the datum were too low. That, and its lack of an internationally viable geocentric ellipsoid, finally drove its positions to obsolescence. The monuments remain, but it was clear as early as 1970 that the NAD27 coordinates of the national geodetic control network were no longer adequate.

A New Datum

Work on the new datum, NAD83, did not really begin until 1975. Leading the charge was an old agency with a new name, the same one that had administered NAD27. Once the *United States Coast and Geodetic Survey (USC&GS)*, then the Coast and Geodetic Survey (CGS) and now known as the *National Geodetic Survey (NGS)* is within the *National Oceanic and Atmospheric Administration (NOAA)*. The first ancestor of today's NGS was established back in 1807 and was known as the *Survey of the Coast*. Its current authority is contained in United States Code, Title 33, USC 883a.

The NGS and the Geodetic Survey of Canada set about the task of attaching and orienting the

GRS80 ellipsoid to the actual surface of the earth, as it was defined by the best positions available at the time. It took more than 10 years to readjust and redefine the horizontal coordinate system of North America into what is now NAD83. More than 1.75 million positions derived from classical surveying techniques throughout the Western Hemisphere were involved in the least-squares adjustment. They were supplemented by approximately 30,000 EDM measured baselines, 5,000 astronomic azimuths and 800 Doppler stations positioned by the TRANSIT satellite system. Very Long Baseline Interferometry (VLBI) vectors were also included. But GPS, in its infancy, contributed only five points.

GPS was growing up in the early 1980s and some of the agencies involved in its development decided to join forces. NOAA, the *National Aeronautics and Space Administration (NASA)*, the *United States Geological Survey (USGS)*, and the Department of Defense coordinated their efforts. As a result each agency was assigned specific responsibilities. NGS was charged with the development of specifications for GPS operations, investigation of related technologies, and the use of GPS for modeling crustal motion. It was also authorized to conduct its subsequent geodetic control surveys with GPS. So, despite an initial sparseness of GPS data in the creation of NAD83, the stage was set for a systematic infusion of its positions as the datum matured.

With the surveying capability of GPS and the new NAD83 reference system in place, NGS began the long process of a nationwide upgrade of their control networks. Now known as the *National Geodetic Reference System (NGRS)*, it actually includes three networks. A horizontal network provides geodetic latitudes and longitudes in the North American Datums. A vertical network furnishes heights, also known as elevations, in the *National Geodetic Vertical Datums (NGVD)*. A gravity network supplies gravity values in the U.S. absolute gravity reference system. Any particular station may have its position defined in one, two, or all three networks.

NGS is computing and publishing NAD83 values for monumented stations, old and new, throughout the United States. Gradually, the new information will provide the common- coordinate basis that is so important to all surveying and mapping activities. But the pace of such a major overhaul must be deliberate, and a significant number of stations will still have only NAD27 positions for some time to come. This unevenness in the upgrade from NAD27 to NAD83 causes a recurrent problem to GPS surveyors across the country.

Since geodetic accuracy with GPS depends on relative positioning, surveyors continue to rely on NGS stations to control their work just as they have for generations. Today, it is not unusual for surveyors to find that some NGS stations have published coordinates in NAD83 and others, perhaps needed to control the same project, only have positions in NAD27. In such a situation, it is often desirable to transform the NAD27 positions into coordinates of the newer datum. But, unfortunately, there is no single-step mathematical approach that can do it accurately.

The distortions between the original NAD27 positions are part of the difficulty. The older coordinates were sometimes in error as much as 1 part in 15 000. Problems stemming from the deflection of the vertical, lack of correction for geoidal undulations, low-quality measurements, and other sources contributed to inaccuracies in some NAD27 coordinates that cannot be corrected by simply transforming them into another datum.

Transformations from NAD27 to NAD83

Nevertheless, various approximate methods are used to transform NAD27 coordinates into supposed NAD83 values. For example, the computation of a constant local translation is sometimes attempted using stations with coordinates in both systems as a guide. Another technique is the calculation of two translations, one rotation and one scale parameter, for particular locations based on the latitudes and longitudes of three or more common stations. Perhaps the best results derive from polynomial expressions developed for coordinate differences, expressed in Cartesian ($\Delta x, \Delta y, \Delta z$) or ellipsoidal coordinates ($\Delta\varphi, \Delta\lambda, \Delta h$), using a 3-D Helmert transformation. However, besides requiring seven parameters (three shift, one scale and three rotation components) this approach is at its best when ellipsoidal heights are available for all the points involved. Where adequate information is available, software packages such as the NGS programs LEFTI or NADCON can provide geodetic quality coordinates.

Even if a local transformation is modeled with these techniques, the resulting NAD27 positions might still be plagued with relatively low accuracy. The NAD83 adjustment of the national network is based on nearly 10 times the number of observations that supported the NAD27 system. This larger quantity of data, combined with the generally higher quality of the measurements at the foundation of NAD83, can have some rather unexpected results. For example, when NAD27 coordinates are transformed into the new system, the shift of individual stations may be quite different from what the regional trend indicates. In short, when using control from both NAD83 and NAD27 simultaneously on the same project, surveyors have come to expect difficulty.

In fact, the only truly reliable method of transformation is not to rely on coordinates at all, but to return to the original observations themselves. It is important to remember, for example, that geodetic latitude and longitude, as other coordinates, are specifically referenced to a given datum and are not derived from some sort of absolute framework. But the original measurements, incorporated into a properly designed least-squares adjustment, can provide most satisfactory results.

Densification and Improvement of NAD83

The inadequacies of NAD27 and even NAD83 positions in some regions, are growing pains of a fundamentally changed relationship. In the past, relatively few engineers and surveyors were employed in geodetic work. Perhaps the greatest importance of the data from the various geodetic surveys was that they furnished precise points of reference to which the multitude of surveys of lower precision could then be tied. This arrangement was clearly illustrated by the design of state plane coordinates systems, devised to make the national control network accessible to surveyors without geodetic capability.

However, the situation has changed. The gulf between the precision of local surveys and national geodetic work is virtually closed by GPS, and that has changed the relationship between local surveyors in private practice and geodesists. For example, the significance of state plane coordinates as a bridge between the two groups has been drastically reduced. Today's surveyor has relatively easy and direct access to the geodetic coordinate systems themselves through GPS. In fact, the 1- to 2-ppm probable error in networks of relative GPS-derived positions frequently exceed the accuracy of the first-order NAD83 positions intended to control them.

Fortunately, GPS surveyors have a chance to contribute to the solution of these difficulties. NGS will accept GPS survey data submitted in the correct format with proper supporting documentation. The process, known as *blue-booking*, requires strict adherence to NGS specifications. GPS measurements that can meet the criteria are processed, classified, and incorporated into the NGRS for the benefit all GPS surveyors.

Other significant work along this line is underway in the state-by-state supernet programs. The creations of High Accuracy Reference Networks (HARN) are cooperative ventures between NGS and the states, and often include other organizations as well. With heavy reliance on GPS observations, these networks are intended to provide extremely accurate, vehicle-accessible, regularly spaced control points with good overhead visibility. To ensure coherence, when the GPS measurements are complete, they are submitted to NGS for inclusion in a statewide readjustment of the existing NGRS covered by the state. Coordinate shifts of 0.3 to 1.0 m from NAD83 values have been typical in these readjustments.

High Accuracy Reference Networks

Other significant work along this line was accomplished in the state-by-state super-net programs. The creation of High Accuracy Reference Networks (HARN) were cooperative ventures between NGS and the states, and often include other organizations as well. The campaign was originally known as High Precision Geodetic Networks (HPGN).

A station spacing of not more than about 62 miles and not less than about 16 miles was the objective in these statewide networks. The accuracy was intended to be 1 part-per-million, or better between stations. In other words, with heavy reliance on GPS observations, these networks were intended to provide extremely accurate, vehicle-accessible, regularly spaced control point monuments with good overhead visibility. These stations were intended to provide control superior to the vectors derived from the day-to-day GPS observations that are tied to them. In that way the HARN points provide the user with a means to avoid any need to warp vectors to fit inferior control. That used to sometime happen in the early days of GPS. To further ensure such coherence in the HARN, when the GPS measurements were complete, they were submitted to NGS for inclusion in a statewide readjustment of the existing NGRS covered by the state. Coordinate shifts of 0.3 to 1.0 m from NAD83 values were typical in these readjustments which were concluded in 1998.

The most important aspect of HARN positions was the accuracy of their final positions. Entirely new orders of accuracy were developed for GPS relative positioning techniques by the Federal Geodetic Control Committee (FGCC). Its 1989 provisional standards and specifications for GPS work include Orders AA, A, and B, which are now defined as having minimum geometric accuracies of 3mm ± 0.01 ppm, 5mm ± 0.1 ppm and 8mm ± 1 ppm, respectively, at the 95%, or 2σ, confidence level. The publication of up-to-date geodetic data, always one of the most important functions of NGS, is even more crucial today. Today the Federal Geodetic Control Subcommittee is within the Federal Geographic Data Committee and has published accuracy standards for geodetic networks in part 2 of the Geospatial Positioning Standards (FGDC-007-1998)

The original NAD83 adjustment is indicated with a suffix including the year 1986 in parentheses, that is, NAD83 (1986). However, when a newer realization is available the year in the parentheses will be the year of the adjustment. The most recent realization is NAD83 (2011).

Continuously Operating Reference Stations, CORS

From 1998 to 2004 NGS introduced another series of observations in each state designed to tie the network to the Continuously Operating Reference Stations, CORS. This work resulted in the Federal Base Network (FBN), which is a nationwide network of monumented stations. These spatial reference positions are among the most precise available and are particularly dense in crustal motion areas. In general these points are spaced at approximately 100 kilometers apart. The accuracies intended are: 1 cm-latitudes and longitudes, 2 cm-ellipsoidal heights, and 3 cm-orthometric heights. These stations are few compared to the much more numerous Cooperative Base Network (CBN). This is a high-accuracy network of monumented control stations spaced at 25 to 50 km apart throughout the United States and its territories. The CBN was created and is maintained by state and private organizations with the help of NGS.

In about 1992 the NGS began establishing a network of Continuous Operating Reference Stations (CORS) throughout the country. The original idea was to provide positioning for navigational and marine needs. There were about 50 CORS in 1996. Their positional accuracies are 3 cm horizontal and 5 cm vertical. They also must meet NOAA geodetic standards for installation, operation, and data distribution. Today there are nearly 2000 Continuously Operating Reference Stations (CORS) online.

The Continuously Operating Reference Stations in the NGS network are mostly to provide support for carrier phase observations. Information is available for postprocessing on the Internet.

NAD83 Positions and Plane Coordinates

The newly published data also include state plane coordinates in the appropriate zone. As before, the easting and northing are accompanied by the mapping angle and grid azimuths, but a scale factor is also included for easy conversions. *Universal Transverse Mercator (UTM)* coordinates are among the new elements offered by NGS in the published information for NAD83 stations.

These plane coordinates, both state plane and UTM, are far from an anachronism. The UTM projection has been adopted by the IUGG, the same organization that reached the international agreement to use GRS80 as the reference ellipsoid for the modern geocentric datum. NATO and other military and civilian organizations worldwide also use UTM coordinates for various mapping needs. UTM coordinates are often useful to those planning work that embraces large areas. In the United States, state plane systems based on the transverse Mercator projection, an oblique Mercator projection, and the Lambert conic map projection, grid every state, Puerto Rico, and the U.S. Virgin Islands into their own plane rectangular coordinate system. And GPS surveys performed for local projects and mapping are frequently reported in the plane coordinates of one of these systems.

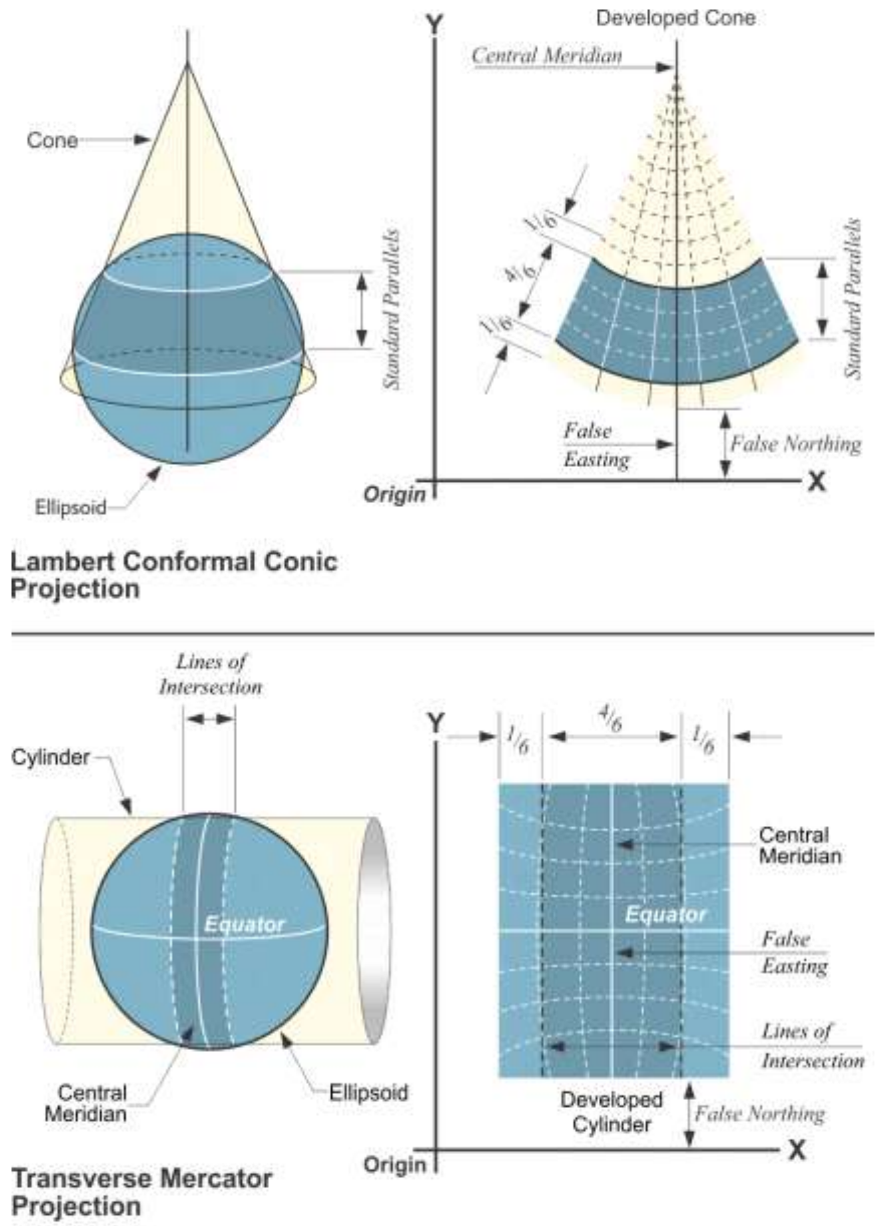


Figure 5.6

For states with large east-west extent, the Lambert conic projection is used, this system uses a projection cone that is imagined to intersect the ellipsoid at standard parallels. When the cone is *developed*, that is, opened to make a plane, the ellipsoidal meridians become straight lines that converge at the cone’s apex. The apex is also the center of the circular lines that represent the projections of the parallels of latitude.

Some states use both the Lambert conic and the transverse Mercator projections for individual zones within the state system (Figures 5.6 and 5.7). Some rely on the transverse Mercator projection alone. The transverse Mercator projection uses a projection cylinder whose axis is imagined to be parallel to the earth's equator and perpendicular to its axis of rotation. It intersects the ellipsoid along standard lines parallel to a central meridian. However, after the cylinder is developed all the projected meridians and parallels become curved lines.

Coordinates from these developed projections are given in reference to a Cartesian grid with two axes. Eastings are reckoned from an axis placed far west of the coordinate zone, adding a large constant value so all remain positive. Northings are reckoned from a line far to the south for the same reason.

The *x*-coordinate, the easting, and the *y*-coordinate, the northing, are expressed in either survey feet or international feet, depending on the state. NAD83 has required a redefinition of the state plane coordinate systems for the updated latitudes and longitudes. The constants now published by NGS are given in meters.

Both of these projections may be said to be *conformal*. Conformality means that an angle on the ellipsoid is preserved after mapping it onto the plane. This feature allows the shapes of small geographical features to look the same on the map as they do on the earth.

The UTM projection divides the world into 60 zones that begin at λ 180E, each with a width of 6 E of longitude, extending from 84E Nφ and 80E Sφ. Its coverage is completed by the addition of two polar zones. The coterminus United States are within UTM zones 10 to 20.

The UTM grid is defined in meters. Each zone is projected onto a cylinder that is oriented in the same way as that used in the transverse Mercator state plane coordinates described above. The radius of the cylinder is chosen to keep the scale errors within acceptable limits. Coordinates of points from the reference ellipsoid within a particular zone are projected onto the UTM grid.

The intersection of each zone's central meridian with the equator defines its origin of coordinates. In the Southern Hemisphere, each origin is given the coordinates: easting, X_0 = 500 000 meters, and northing, Y_0 = 10 000 000 meters, to ensure that all points have positive coordinates. In the Northern Hemisphere, the values are: easting, X_0 = 500 000 meters, and northing, Y_0 = 0 meters, at the origin.

The scale factor grows from 0.9996 along the central meridian of a UTM zone to 1.00000 at 180 000 meters to the east and west. The state plane coordinate zones in the United States are limited to about 158 miles and so embrace a smaller range of scale factors than do the UTM zones. In state plane coordinates, the variance in scale is usually no more than 1 part in 10,000. In UTM coordinates the variance can be as large as 1 part in 2,500.

The distortion of positions attributable to the transformation of NAD83 geodetic coordinates into the plane grid coordinates of any one of these projections is generally less than a centimeter. Most GPS and land surveying software packages provide routines for automatic transformation of latitude and longitude to and from these mapping projections. Similar programs can also be purchased from the NGS. Therefore, for most applications of GPS, there ought to be no technical compunction about expressing the results in grid coordinates. However, given the long traditions of plane surveying, it can be easy for some to lose sight of the geodetic context of the entire process that produced the final product of a GPS survey is presented in plane coordinates

The Deflection of the Vertical

Other new elements in the information published by NGS for NAD83 positions include deflection of the vertical. The deflection of the vertical can be defined as the angle made by a

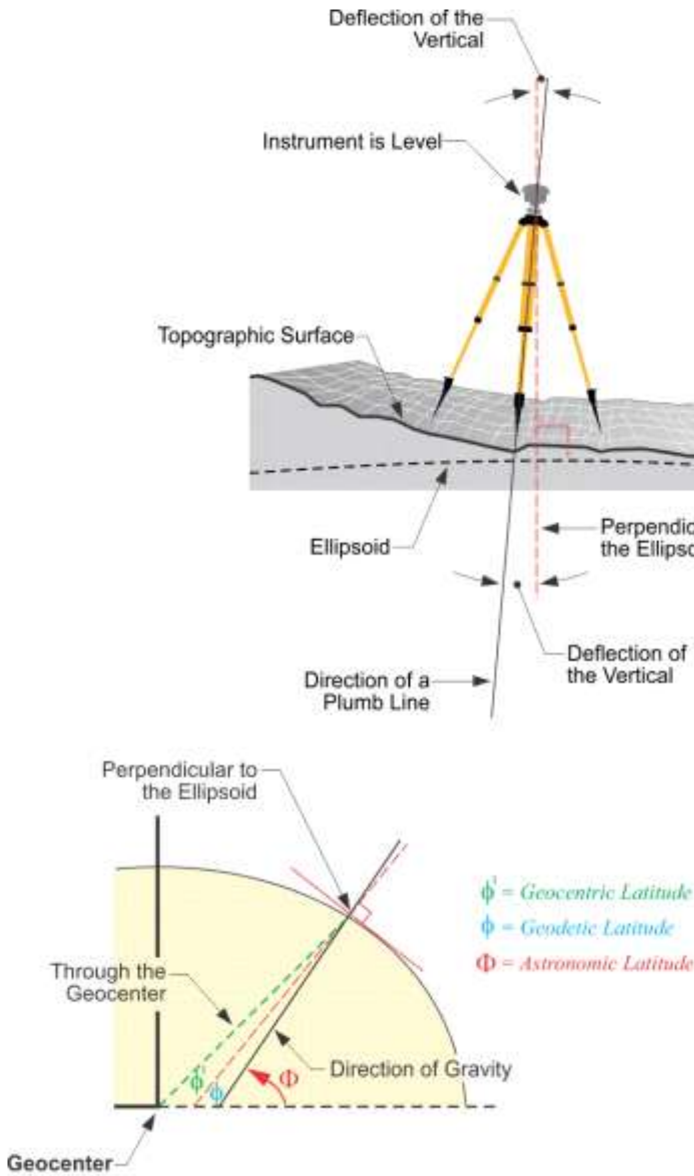


Figure 5.8

line perpendicular to the geoid that passes through a point on the earth's surface with a line that passes through the same point but is perpendicular to the reference ellipsoid (Figure 5.8).

Described another way, the deflection of the vertical is the angle between the direction of a plumb line with the ellipsoidal normal through the same point. The deflection of the vertical is usually broken down into two components, one in the plane of the meridian through the point and the other perpendicular to it. The first element is illustrated above.

When surveyors relied on astronomical observations for the determination of latitude, longitude, and azimuth, calculation of the deflection of the vertical was critical to deriving the corresponding ellipsoidal coordinates from the work. Today, if the orientation of a GPS vector is checked with an astronomic azimuth, some discrepancy should be expected. The deflection of the vertical is a result of the irregularity of the geoid and is mathematically related to separation between the geoid and the reference ellipsoid.

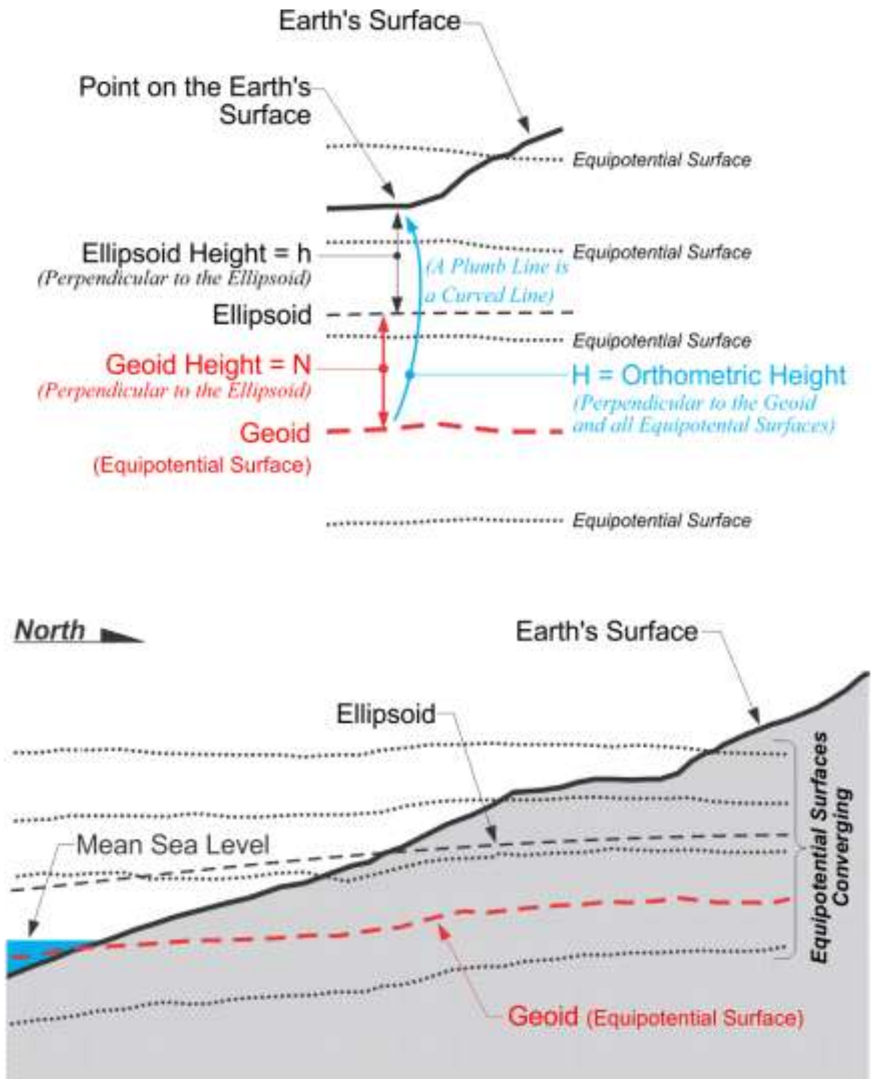
Heights

A point on the earth's surface is not completely defined by its latitude and longitude. In such a context there is, of course, a third element, that of height. Surveyors have traditionally referred to this component of a position as its elevation. One classical method of determining elevations is spirit leveling. As stated earlier, a level, correctly oriented at a point on the surface of the earth, defines a line parallel to the geoid at that point. Therefore, the elevations determined by level circuits are *orthometric;* that is, their vertical distance above the geoid defines them, as it would be measured along a plumb line.

However, orthometric elevations are not directly available from the geocentric position vectors derived from GPS measurements. The vectors are not difficult to reduce to ellipsoidal latitude, longitude, and height because the reference ellipsoid is mathematically defined and clearly oriented to the earth. But the geoid defies such certain definition. As stated earlier, the geoid undulates with the uneven distribution of the mass of the earth and has all the irregularity that implies. In fact, the separation between the bumpy surface of the geoid and the smooth GRS80 ellipsoid varies from 0 up to +/–100 meters. Therefore, the only way a surveyor can convert an ellipsoidal height from a GPS observation on a particular station into a useable orthometric elevation is to know the extent of geoid-ellipsoid separation, also called the *geoid height*, at that point.

Toward that end, major improvements have been made over the past quarter century or so in the mapping the geoid on both national and global scales. This work has gone a long way toward the accurate determination of the geoid-ellipsoid separation or geoid height, known as *N*. The formula for transforming ellipsoidal heights, *h*, into orthometric elevations, *H*, is (Figure 5.9):

$$H = h - N$$



Earth's Surface
Point on the Earth's Surface
Equipotential Surface
Equipotential Surface
Ellipsoid Height = h
(Perpendicular to the Ellipsoid)
(A Plumb Line is a Curved Line)
Ellipsoid
Equipotential Surface
Geoid Height = N
(Perpendicular to the Ellipsoid)
H = Orthometric Height
(Perpendicular to the Geoid and all Equipotential Surfaces)
Geoid
(Equipotential Surface)
Equipotential Surface
North
Earth's Surface
Ellipsoid
Mean Sea Level
Equipotential Surfaces Converging
Geoid (Equipotential Surface)

Figure 5.9