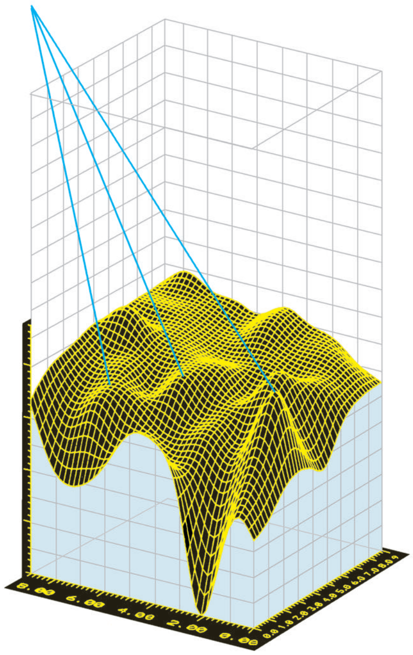
The UTM Grid System

Richard B. Langley

University of New Brunswick



All GPS receivers can provide position information in terms of latitude, longitude, and height, and usually in a variety of selectable geodetic datums. For many purposes, position information in this format is more than adequate. However, when plotting position information on maps or carrying out supplemental calculations using the position coordinates, it can be advantageous to work instead with the corresponding grid coordinates on a particular map projection.

One of the most widely used map projection and grid systems is the Universal Transverse Mercator (UTM) system. Many GPS receivers can directly output position information in UTM coordinates. In this month's column, we'll take a look at the UTM system, see how UTM grid coordinates are related to geodetic coordinates, and indicate the corrections to be applied to grid distances and bearings to get the actual true quantities on the earth's surface.

"Innovation" is a regular column featuring discussions about recent advances in GPS technology and its applications as well as the fundamentals of GPS positioning. The column is coordinated by Richard B. Langley of the Department of Geodesy and Geomatics Engineering at the University of New Brunswick, who appreciates receiving your comments as well as topic suggestions for future columns. To contact him, see the "Columnists" section on page 4 of this issue.

As we discussed in a recent "Innovation" column, one can employ a variety of equivalent ways to mathematically describe the position of a point on or near the earth's surface. For example, we can use the point's *x, y,* and *z* Cartesian coordinates in a system that is fixed to the earth in some manner. Such a system typically has its center, or *origin,* at the earth's center of mass, with the coordinate axes aligned so that the z-axis coincides with the earth's rotation axis as it was situated early in this century and the x-axis lies in the plane of the Greenwich meridian as defined by the International Earth Rotation Service or its predecessor, the Bureau International de l'Heure. The y-axis completes the right-handed coordinate system.

This definition accounts for the fact that the earth's rotation pole is not fixed with respect to the earth's material body. Instead, it moves in a roughly circular fashion with respect to an axis fixed in the earth, with dominant periodicities of about 12 months and 14 months. The latter period is called the Chandler period after Seth Carlo Chandler, the American astronomer who first identified it late in the last century using astronomical observations. In addition, the rotation pole exhibits a secular drift toward Canada. The total motion of the pole is referred to as *polar motion* or *wobble.*

The geodetic community refers to a coordinate system defined in this way as a conventional terrestrial reference system (CTRS). The navigation community typically calls it an earth-centered, earth-fixed (ECEF) system. A CTRS is realized, or made accessible, through an adopted set of reference station coordinates. These coordinates constitute a conventional terrestrial reference *frame.* Such frames include the International Terrestrial Reference Frame (ITRF) and the reference frame of the World Geodetic System 1984 (WGS 84).

COORDINATES AND PROJECTIONS

More familiar to us than Cartesian *x, y, z* values are the geodetic coordinates (sometimes referred to by the more inclusive term *geographic coordinates*) of latitude, longitude, and height (ϕ, λ, h). Such coordinates actually refer to a particular biaxial reference ellipsoid — a precisely defined mathematical figure used to represent the roughly ellipsoidally shaped earth — and so are sometimes called ellipsoidal coordinates.

Although in the past it was common to position the reference ellipsoid so that it best approximated the geoid (the equipotential surface of the earth's gravity field that is closest to mean sea level) in a specific region of interest — for example, Europe — it is now common practice to center the ellipsoid at the earth's center of mass and select the ellipsoid's dimensions so that it best approximates the geoid over the whole earth. The WGS 84 and the North American Datum 1983 (NAD 83), for example, both use geocentric ellipsoids. In fact, the two ellipsoids are essentially identical in shape and orientation.

A position's quoted latitude and longitude, therefore, actually refer to a point on the surface of the given system's reference ellipsoid. A point's geodetic height — which one measures along a line perpendicular to the ellipsoid surface — takes us from the ellipsoid to the actual position. This height, though, is not the one we are most familiar with. We more frequently talk about heights above mean sea level, or more technically correct, heights above the geoid. We can determine such *orthometric* heights from geodetic ones if we know the geoidal undulation — the separation of the ellipsoid and the geoid.

Down-to-Earth Coordinates. Navigators, at least those confined to land or the ocean surface, are usually not as much interested in their three-dimensional position as in their two-dimensional position — they know their vertical coordinate quite well. To relate such surface or horizontal positions to each other and to geographic features, we use a map — a flat representation of all or part of the earth's curved surface. One obtains a map (or *chart,* the navigator's term for a map) by projecting the earth's surface onto a plane — the mapping plane. This cannot be done with absolute fidelity, though, and some distortion will occur. The shapes and areas of large land masses or ocean boundaries may be distorted, directions between two points may be incorrect, and the map's scale may not be constant.

In attempts to overcome such problems, cartographers and others have devised a variety of map projections. Some are more appropriate for specific uses than others. One of the most useful for navigation and general positioning purposes is the Mercator projection.

MERCATOR'S WORLD

The Mercator projection was popularized by Gerardus Mercator with an atlas he published in 1569. It projects the earth onto a concentric cylinder tangent to it along the equator. The meridians, or lines of equal longitude, on a Mercator map are equally spaced and straight. The parallels, or lines of equal latitude, are also straight but unequally spaced, closest together at the equator and cutting the meridians at right angles. A Mercator projection's poles are at infinity, which greatly distorts area in the polar regions — hence, the reason Greenland appears larger than South America, when in reality it has only about one-eighth the area. Nevertheless, the shapes of small features are preserved, because the Mercator is in a class of projections known as *conformal*. Such projections are also known as *orthomorphic* — at any given point, their scale is the same in all directions. However, a Mercator projection's scale is exact or true — that is, equal to a projection's specified scale that relates map distances to distance on the ground — only along the equator or along two specific parallels equidistant from the equator.

A significant feature of a map using this projection is that any straight line on the map is a line of constant direction; that is, as one travels along this line, the oblique angle at which the line crosses meridians (or the azimuth) is constant. This line, called a *rhumb line* or *loxodrome*, spirals toward the pole.

An important variation of the standard or regular Mercator projection is obtained by changing the orientation of the cylinder upon which the earth is projected. Rotating the cylinder 90 degrees so that its axis is perpendicular to the earth's poles results in a constant scale along any chosen central meridian (see Figure 1). Also conformal, this projection is frequently used to map regions with large north–south extent. One can also orient the cylinder at any arbitrary angle to map an area with greater extent in a particular oblique direction. In fact, this represents the projection's most general form, with the regular and transverse projections being special cases. We shall return to the transverse Mercator projection shortly, but first a word about the earth's assumed shape.

For many general-purpose map projections, the earth is treated as if it were a sphere. However, as we noted earlier, the earth is more nearly an oblate ellipsoid and, although the flattening totals only about 1 part in 300, it's large enough to be significant when plotting accurate maps at a scale of 1:100,000 or larger.

Adopting the Ellipsoid. The Alsatian mathematician Johann Heinrich Lambert invented and subsequently published, in 1772, the transverse Mercator projection using the earth's spherical approximation. Carl Friedrich Gauss developed the ellipsoidal version in 1822, which Leonhard Krüger further studied in the early 1920s. Consequently, one sometimes hears of the Gauss conformal or Gauss-Krüger projection, particularly in Europe.

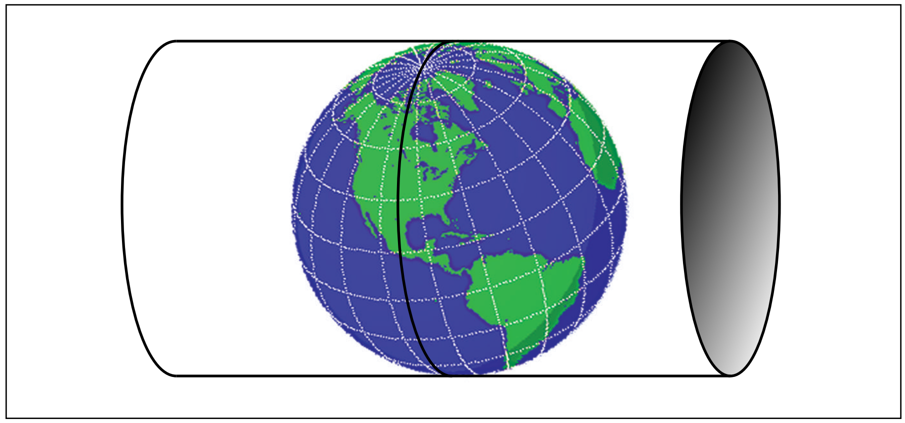


Figure 1. The transverse Mercator projection is developed on a cylinder tangent to the ellipsoid along a central meridian. The cylinder is then opened and flattened.

Whereas all meridians and parallels are straight lines on a map constructed using the regular Mercator projection, on a transverse Mercator map only the central meridian, meridians 90 degrees from the central meridian, and the equator are straight lines. The other meridians and parallels form complex curves. Without modification, the scale on a standard transverse Mercator projection is true only along the central meridian. If the scale along the central meridian is deliberately reduced slightly such that the mean scale over the map's area is more nearly correct, then scale will be true along two given lines on either side of the central meridian and equidistant from it.

A UNIVERSAL PROJECTION

In 1947, the United States Army adopted the ellipsoidal transverse Mercator projection and an associated grid system for designating rectangular coordinates on large-scale military maps covering almost the whole world. This Universal Transverse Mercator (UTM) system has since become very widely used not only for military purposes but for general surveying and mapping and for navigation both on land and at sea.

In the UTM system, the earth, between 80 degrees south and 84 degrees north, is divided into 60 zones, each generally 6 degrees wide in longitude. With minor exceptions, the zones are numbered from 1 to 60 starting at 180 degrees longitude and proceeding east. The zones are narrow enough to keep the distortion and scale variation within the zone to an acceptable level. Between 80 degrees south and 72 degrees north, the zones are divided into quadrangles, or *grid* zones, that are 8 degrees in latitudinal extent. The grid zones from 72 to 84 degrees north span 12 degrees of latitude. The latitude bands are lettered from C (between 80 and 72 degrees south) through X (between 72 and 84 degrees north) with I and O omitted to avoid possible confusion with numerals 1 and 0.

Each UTM zone's central meridian lies midway between the two bounding meridians, and its scale is reduced to 0.9996 of true scale. This value keeps the scale variation within a zone to less than about 1 part in 1,000. The lines of true scale are approximately parallel to the central meridian and about 180 kilometers to either side of it (see Figure 2).

The Grid. Although very easy to use in calculations, three-dimensional ECEF Cartesian coordinates are not immediately interpretable in terms of where one is on or near the planet's surface. Calculations involving geodetic coordinates, on the other hand, can be quite complex. For these reasons, rectangular grid systems have been developed that permit both ready interpretation and ease of computation.

Using such a system, any point may be described by its distance from two perpendicular axes, X and Y. The Y-axis normally coincides with the central meridian, with y-values increasing toward the north. The X-axis crosses the Y-axis at a latitude of origin on the central meridian, with x-values increasing toward the east. The x- and y-coordinates are frequently referred to as *eastings* and *northings*, respectively. To avoid negative values, the coordinates may be offset by adding false eastings and northings.

Most countries have implemented grid systems for their national mapping, each with a particular reference ellipsoid, map projection, central meridians, scale factors, and so forth. For example, the Ordnance Survey of Great Britain uses the British National Grid, a system based on the transverse Mercator projection, Airy ellipsoid, a central meridian at 2 degrees west with a grid origin on the central meridian at 49 degrees north, and a central meridian scale factor of 0.99960 12717.

For easy referencing, grids are usually divided into a sequence of nested square areas. For example, the British National Grid is divided into 500-kilometer squares and each of those into 100-kilometer squares. Each large and small square is designated by a letter. Within each 100-kilometer square, positions are given as x- and y-coordinates with respect to the southwest corner of the square. A grid reference is then formed by combining the square letters with the coordinates, omitting decimal points and listing *x* before *y*. For example, the grid reference for the Tower of London is TQ 336805 or 33.6 kilometers east and 80.5 kilometers north of the southwest corner of grid square TQ. One can easily convert such grid references into standard eastings and northings and obtain higher precisions by adding more digits.

The origin of x- and y-coordinates in the UTM grid system is the intersection of the equator with the central meridian. But the x-coordinate at the origin is assigned a false easting of 500,000 meters. This eliminates negative eastings within any zone. For the northern hemisphere, the origin has a y-coordinate of 0 but for the southern hemisphere, the origin is assigned a false northing of 10,000,000 meters. Note that in the UTM grid system, eastings and northings always appear in meters. When listed, normally the easting precedes the northing.

Military Grid Reference. The U.S. Department of Defense established the Military Grid Reference System (MGRS) for worldwide use with the UTM grid (and also the Universal Polar Stereographic grid that is used for the polar regions). Each UTM grid zone is divided into 100-kilometer squares designated by a pair of letters. Within a 100-kilometer square, a sequence of digits denotes a point's numerical grid coordinates to the desired level of precision. The first half of the total number of digits (always an even number) represents the easting and the second half the northing. For example, the UTM grid coordinates for a point on the roof of Gillin Hall at the University of New Brunswick in Fredericton (to a precision of 1 meter and based on the WGS 84 datum) are 682,725 meters E, 5,091,225 meters N in UTM zone 19. The corresponding MGRS grid reference is 19TFL8272591225.

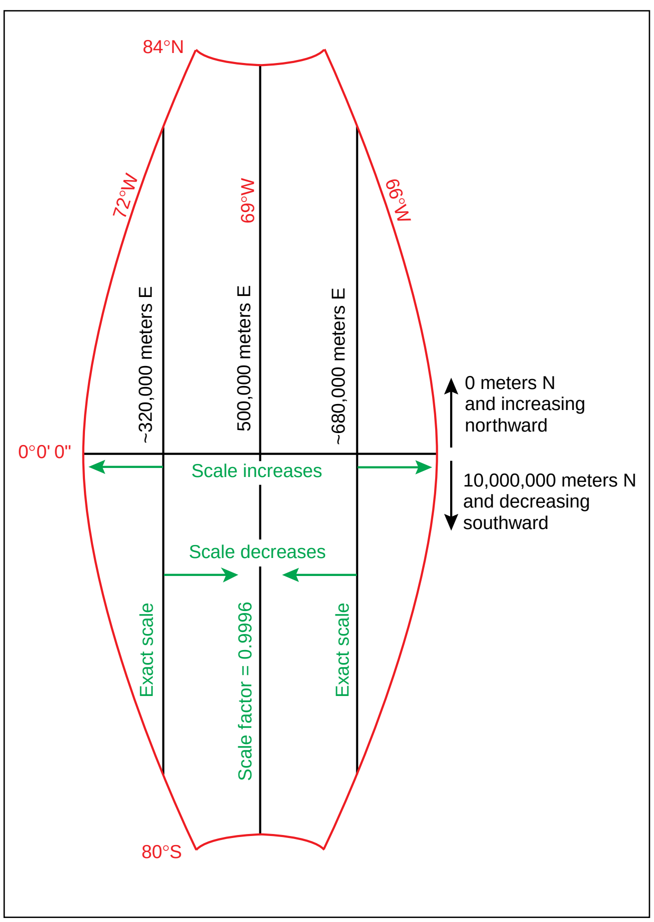


Figure 2. UTM zone 19 extends from 66 degrees to 72 degrees west longitude. As with all UTM zones, the scale factor is 0.9996 on the central meridian and true (unity) on two slightly curved lines approximately 180 kilometers to either side. The shape of the zone has been exaggerated for clarity.

Geodetic coordinates can be readily converted to UTM eastings and northings or vice versa with a series expansion (involving, in addition to the latitude and longitude, the reference ellipsoid's dimensions and the central meridian's longitude) that, depending on truncation, can give an accuracy exceeding 1 centimeter. The accuracy of grid coordinates, therefore, should not be considered as necessarily inferior to that of geodetic coordinates.

Although the UTM projection and grid requires an ellipsoidal earth, the particular ellipsoid used can vary from one region or political administration to another. Whenever UTM grid coordinates are given, the chosen geodetic datum (including ellipsoid) should be stated. An incorrectly specified ellipsoid can result in large errors when converting between geodetic and UTM grid coordinates.

The advantage of grid coordinates is that they permit easy calculations using plane trigonometry. For example, the bearing, β, of a point P_2 with grid coordinates (N_2, E_2) from a point P_1 with grid coordinates (N_1, E_1) is given by

$$\beta = \arctan\left[\frac{E_2 - E_1}{N_2 - N_1}\right]$$

and the grid length or distance, *d*, between P_1 and P_2 is given by

$$d = \sqrt{\left[\left(E_2 - E_1\right)^2 + \left(N_2 - N_1\right)^2\right]}.$$

Simple, right? But there are some caveats.

First of all, the grid distance as calculated using the aforementioned equation is not equal to the true shortest distance over the ground between the points — for three reasons. First, we must consider the projection's variable scale within the UTM zone. It is equal to 0.9996 on the central meridian and increases to more than 1 at the zone boundaries. However, over relatively short distances (say, less than 10 kilometers), the scale is more or less constant and readily calculated so that the grid distances can be corrected for scale error. Note that the correctly scaled grid distance is a chord distance, which is not the path providing the shortest distance between two points on an ellipsoid — the geodesic. However, within a UTM zone, the difference between a geodesic length and the corresponding scaled grid distance is minute.

Second, the grid coordinates refer to positions on the reference ellipsoid's surface. In general, the two points will have arbitrary heights above or below the ellipsoid, which requires that an elevation factor be applied to a grid distance (corrected for scale) to obtain

the true distance between the points. For example, if two points were, say, in central Canada at 300 meters above the ellipsoid, the factor 1/0.99995 would have to be applied to the grid distance to determine the true ground distance.

Third, the grid distance doesn't take into account the undulating terrain that would have to be traversed in traveling between the points. In mountainous areas, this could add 40 percent to a grid distance.

Let's now turn to the grid bearing. On the UTM mapping plane, parallels of latitude are arcs that cross the central meridian at right angles but tend toward north on either side of this meridian. The central meridian is a straight line, but the meridians on either side of it are curved lines that converge to the geographical poles (they would meet the central meridian at the north and south poles). So, there is a distinction between grid north (a constant direction for any particular grid) and true north (the tangent to a meridian and consequently in a different direction from meridian to meridian — and which, of course, is also different from magnetic north). In general, a correction must be applied to a grid bearing to get a true bearing or azimuth because of the meridian convergence. At the central meridian, it is zero and increases with the difference in longitude of the point and the central meridian.

A small correction must also be made to account for the fact that the tangent to the geodesic at P_1, which defines the geodetic azimuth, is not the same as the angle that the straight line between the two points on the mapping plane makes with the meridian through P_1. This small correction — known as the arc-to-chord angle, or t–T correction — is related to the inverse of the geodesic's radius of curvature on the mapping plane and the geodesic's length.

AN EXAMPLE

Let's illustrate these important differences between grid and geodetic values using an example. Consider two points exactly on the western border between Canada and the United States — the 49th parallel — one at a (west) longitude of 123 degrees, 0 minutes, 0 seconds and one about 6.1 kilometers to the west at a (west) longitude of 123 degrees, 5 minutes, 0 seconds. Both of these points are in UTM grid zone 10-U, with the first point lying directly on the central meridian of the zone.

We'll assume that these coordinates are given in the old North American Datum of 1927 (NAD 27) and both of these points are on the nongeocentric NAD 27 ellipsoid (an ellipsoid whose dimensions were determined by the British geodesist Alexander Ross Clarke in 1866). The length of the geodesic between the two points is 6,097.845 meters, and the geodetic azimuth of Point 2 measured from Point 1 is 270 degrees, 1 minute, 53.2 seconds. Notice that the shortest route between the points on the ellipsoid follows a path that arcs slightly to the north of the 49th parallel!

Now let's look at the UTM coordinates. Converting the geodetic coordinates provides the following values.

- Point 1: 5,427,236.772 meters N; 500,000.000 meters W
- Point 2: 5,427,240.117 meters N; 493,904.594 meters W

We can use plane trigonometry on the coordinates to determine the distance between the points and the grid azimuth or bearing of Point 2 from Point 1. We get 6,095.407 meters and 270 degrees, 1 minute, 53.2 arc-seconds. The bearing is exactly the same as the geodetic azimuth, but the distance is shorter by more than 2 meters. Why? As we've already noted, the grid distance is rarely equal to the length of the geodesic because of the distortion in trying to represent the ellipsoid's surface on a plane. A scale correction must be applied to grid distances to get accurate distances on the ellipsoid.

We know that on the central meridian of a UTM zone, the scale factor is 0.9996 and increases to either side. Because Point 1 is on such a meridian and Point 2 lies only about 6.1 kilometers away, we can use this factor to scale the grid distance. When we do this, we get a distance on the ellipsoid of 6,097.846 meters — equal, within the round-off error, to the distance we computed directly from the geodetic coordinates! In addition, because Point 1 is on the UTM zone's central meridian, the convergence correction is zero, and for this short line, the t–T correction is insignificant. So the bearing is equal to the geodetic azimuth.

CONCLUSION

In this brief article, we have introduced only the basic concepts involved with the UTM grid system. For further details about the system, including algorithms for converting between geodetic and UTM grid coordinates, interested readers should consult the references listed in the "Further Reading" sidebar.

ACKNOWLEDGMENT

Thanks to Andrew Harris, Vancouver, British Columbia, for suggesting the UTM coordinate conversion example. ■

Further Reading

For an introduction to geodetic datums and map projections, see

- "Basic Geodesy for GPS," by R.B. Langley, *GPS World*, Vol. 3, No. 2, February 1992, pp. 44–49.
- "Coordinates and Datums and Maps! Oh My!" by W. Featherstone and R.B. Langley, *GPS World*, Vol. 8, No. 1, January 1997, pp. 34–41.

For a thorough discussion of the mathematics of map projections, see

- *Map Projections — A Working Manual,* by J.P. Snyder, U.S. Geological Survey Professional Paper 1395, United States Government Printing Office, Washington, D.C., 1987.

For an overview of map grid systems including UTM and MGRS, see

- *Datums, Ellipsoids, Grids and Grid Reference Systems*, by J.W. Hager, L.L. Fry, S.S. Jacks, and D.R. Hill, Defense Mapping Agency Technical Memorandum TM 8358.1, 1983. Available from the National Technical Information Service, Springfield, Virginia, and in a Web version at <http://164.214.2.59/geospatial/products/GandG/tm83581/toc.htm>.

For an introduction to the use of UTM grid coordinates and MGRS with GPS for land navigation, see

- *A Comprehensive Guide to Land Navigation with GPS*, by N.J. Hotchkiss, 2nd edition, published by Alexis Publishing, Herndon, Virginia, 1996.

For details about the conversion between geodetic and UTM grid coordinates, see, for example,

- *Global Positioning System: Theory and Practice*, by B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, 4th revised edition, published by Springer-Verlag, Vienna and New York, 1997.

For other Web-based resources about geodetic datums and grid systems, see

- <http://www.netonecom.net/~rburtch/geodesy/datums.html>.

For an opinion piece calling for the United States to adopt the UTM grid system as the standard for large-scale maps, see

- "One Nation, One Map Grid," by N.G. Terry, Jr., *GPS World*, Vol. 7, No. 4, April 1996, pp. 32–37.

Subsequent *GPS World* "Letters to the Editor" respond to that article and address various aspects of this issue: see July and September 1996 and January 1997.

For articles about maps and map projections, old and new, see *Mercator's World* magazine at your local newsstand.