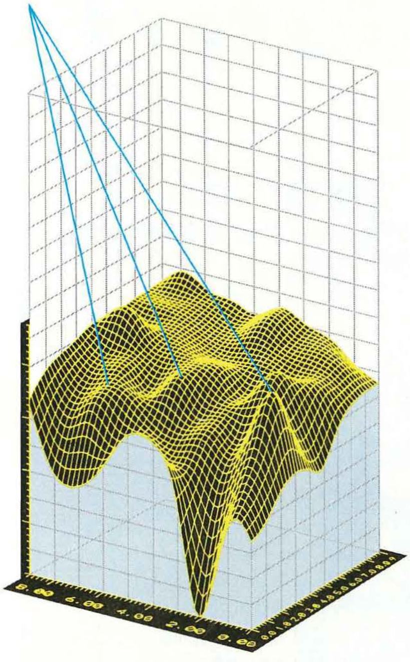


Coordinates and Datums and Maps! Oh My!

Will Featherstone
Curtin University of Technology

Richard B. Langley
University of New Brunswick

The walk through the enchanted forest of Oz, with its lions and tigers and bears, was a pretty scary proposition for Dorothy Gale and her friends. Some GPS users find themselves in a similar predicament when they try to understand the enchanted forest of geodesy and the relationship among coordinates, datums, and maps.

In this month's column, Dr. Will Featherstone, senior lecturer in geodesy at the School of Surveying and Land Information, Curtin University of Technology, Perth, Western Australia, and Professor Richard Langley of the Department of Geodesy and Geomatics Engineering, University of New Brunswick in Fredericton, Canada, sketch the relationships among the coordinate systems used worldwide for GPS and the coordinate systems and map projections used in various countries. They also discuss how these differences can affect the GPS user when employed incorrectly.

"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 Langley, 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 GPS users, we are, of course, all aware that GPS coordinates are based on the World Geodetic System 1984 (WGS84), which was developed by the former U.S. Defense Mapping Agency (DMA). But, what if we want to use a GPS receiver for positioning or navigation and relate the coordinates it provides to features on a map of, for example, Perth, Australia?

This may present some difficulties because geodesy developed well before the advent of satellite positioning systems and each country, group of countries, or continent had already defined its own coordinate system for surveying and mapping. In geodesy and related sciences, these systems are referred to as *geodetic datums*. Associated with each datum is a reference ellipsoid, the surface of which approximates the earth's shape (as represented by the geoid — the equipotential surface of the earth's gravity field that best matches, in a least-squares sense, mean sea level) over the region covered by the datum. Because most of these *local* datums were developed using exclusively terrestrial techniques, they are typically quite different from each other and from the WGS84 datum used by GPS. Such differences did not pose any significant problems in the past because each region was using its own coordinate system. Rarely was there any need to know the relationships among the geodetic datums in each region of the world. Then satellite positioning arrived.

PUTTING GPS ON THE MAP

GPS allows us to determine our position anywhere on or near the earth's surface in a single global coordinate system — that of WGS84. This system does not always coincide with the local coordinate datums that have been adequately serving different countries or continents for many years. Obviously, this creates the potential for a great deal of confusion. So, we must ask ourselves: What local geodetic datum do we need and how do we acquire it?

This question applies not only to GPS-provided latitudes and longitudes, but also to the coordinates found on local maps. A map is an attempt to represent the curved surface of the earth on a flat piece of paper or on a computer screen. As anyone who has ever cut an orange in half, removed the fruit sections, and then tried to flatten the orange peel onto a level surface knows, accurately completing such a task presents quite a challenge. The orange skin tears and does not completely cover the flat surface. In order for it to do so, we must stretch or compress the orange skin, which distorts its original shape.

The same phenomenon occurs when we try to represent or *project* the earth's surface on paper or a computer screen. This distortion dilemma has puzzled cartographers and geodesists — including the likes of Leonardo da Vinci, Gerardus Mercator, and Johann Heinrich Lambert — for nearly two millennia. Their goal has been to produce a map that accurately represents the three-dimensional relationships among features on the earth's surface when viewed on a flat map. To reduce these distortions as much as possible, more than 250 different *map projections* have been developed and proposed through the years.

Because an exact solution to this difficult problem has not yet been found, we continue to use a variety of map projections, some of which are better suited to certain parts of the world than others. If we look at a world map in an atlas, for example, Antarctica appears to be a huge continent, almost as large as North America. When we view Antarctica on the more representative globe, however, we see that it is really much smaller than North America. The appearance in the atlas is, therefore, purely a result of the distortion problem.

The rectangular or map coordinates of a point will vary depending on which map projection is selected. As a result, the same map coordinates from two different map projections do not necessarily coincide with the same point on the earth's surface. When expressed in terms of meters, map coordinates are frequently referred to as grid coordinates (typically called "eastings" and "northings"), and either grid lines or graticules of meridians and parallels or both may appear on a map.

So, before we can use GPS for positioning and navigating in parts of the world that do not use WGS84 or a compatible system, such as North American Datum 1983, we must make our coordinates compatible with the local coordinate system using processes called *transformation* and *projection*. But first, we must have basic information about the geodetic datum, reference ellipsoid, and map projection used in that region.

CHOOSE WISELY

The processes of transformation and projection are extremely important because the difference between WGS84 coordinates and local map coordinates and the corresponding coregistration errors can exceed 1 kilometer in some parts of the world. For example, the difference between the eastings and northings of points referenced to the geodetic datum used in Japan and WGS84 is almost 1,500 meters (see Figure 1). Perhaps in the future, when addresses are given purely in latitude and longitude or grid coordinate form, you will have to be careful to specify the datum and map projection when you phone for delivery of a pizza (or sushi, perhaps). Otherwise, your order could end up in the wrong neighborhood altogether.

The story of a mariner who was navigating offshore Perth, near an island resort called Rottnest, provides a present-day, and less light-hearted, example of the difference between geodetic datums. The chart (which by definition is just a map designed for navigation) he was using to avoid the shallow reefs near the island was based on the Australian Geodetic Datum. His GPS receiver, however, was delivering WGS84 coordinates. The difference between the geographic coordinates of these two systems is approximately 200 meters, which was enough to cause him to run aground on a reef.

Such a situation could occur when navigating an airplane if the plane were to take off from an airport in one country and land in another country that uses a different geodetic datum. If the pilot were using incorrect datums to navigate, the airplane could miss the runway by hundreds of meters. Don't be alarmed; to our knowledge, this situation has never occurred for two reasons. First, the cockpit crew are aware of the differences in geodetic datums and second, the airplane relies on approach navigation aids based at the airport where they will land. With the move toward worldwide use of GPS for air navigation, however, these types of datum difficulties could conceivably occur.

Equally important to the geodetic datum is the map projection used and its associated parameters. As described earlier, a specific map projection is used for a state or province, country, or continent to minimize feature distortions on the map. Most map projections used today project geographic coordinates from the ellipsoid associated with the local geodetic datum. Therefore, in addition to knowing what projection is being used, we must also know the parameters of the associated ellipsoid.

TRANSFORMING COORDINATES

Thankfully, the process of converting between the different datums and map projections used throughout the world is relatively straightforward. The former DMA, which was recently absorbed into the new National Imagery and Mapping Agency, published a technical report entitled *Department of Defense World Geodetic System 1984 — Its Definition and Relationships with Local Geodetic Systems*. Every person who works with

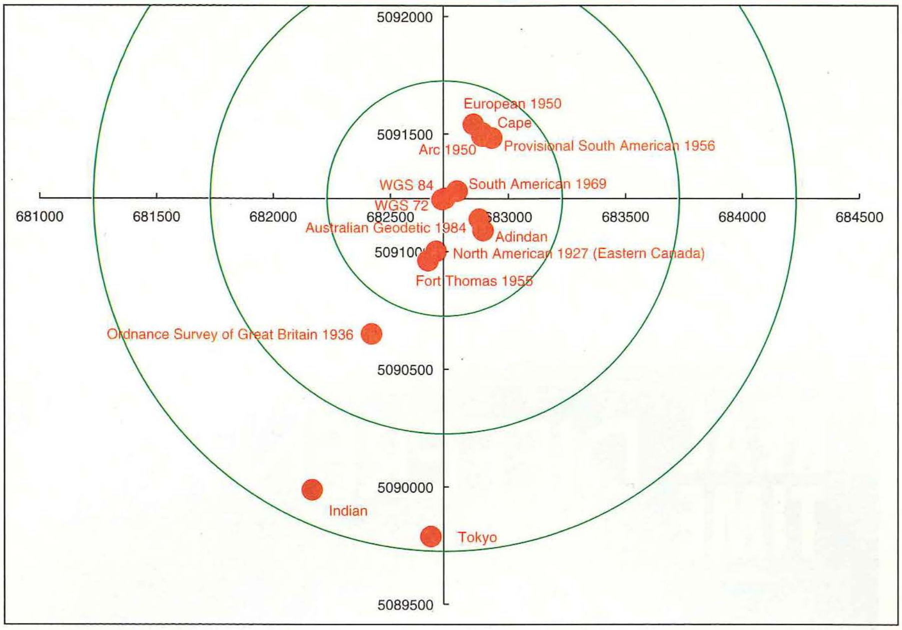


Figure 1. Using the wrong datum and associated ellipsoid when transforming and projecting latitudes and longitudes to map grid coordinates can result in large errors. This plot shows the corresponding UTM eastings and northings (in meters) of a point with WGS84 geographical coordinates of 45° 57' 0.96" N, 66° 38' 32.22" W, for a variety of geodetic datums. The circles have radii of 500, 1,000, and 1,500 meters.

GPS in more than one country should have this report on his or her bookshelf. The report's second edition, published in 1991, provides the numerical relationships between more than 100 local datums and WGS84.

Several procedures are available to help the GPS user convert latitude and longitude from WGS84 to a local datum and vice versa. These procedures can be divided into three categories: simple block shifts, similarity transformations, and projective transformations.

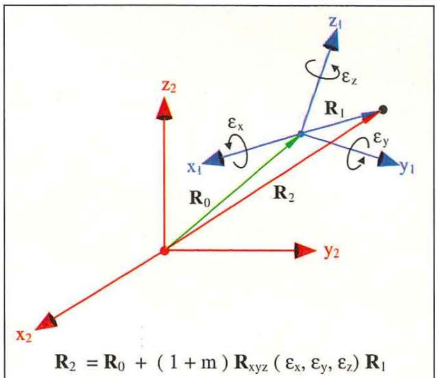
Block Shift. The block shift is conceptually the simplest transformation approach and is also the coarsest. A simple bias in latitude and longitude is applied to the WGS84 coordinates to coregister them with the local datum. Because of its low accuracy, however, this approach is suited only to small areas.

Similarity Transformations. Similarity transformations can be applied by one of two main approaches. The first, known as the Burša-Wolf transformation, is suitable for conducting transformations between satellite-derived datums. The second similarity transformation is the Molodenskiy-Badekas transformation, generally called Molodenskiy, and is more appropriate for completing transformations between satellite and terrestrial datums. The only computational and conceptual differences between these methods is which origin point is used for the transformation.

Seven parameters are normally required to carry out similarity transformations using either of the two approaches. These comprise three origin shifts between the geocenter (the earth's center of mass, which is the origin of WGS84) and the center of the local system; three rotations about the axes that pass through the poles, and the intersections of the Greenwich and 90 degrees east meridians with the equator; and a scale change between the two systems. In addition, the semimajor axis lengths and flattenings of the two datum's reference ellipsoids are required (see Figure 2). The exact implementation of these similarity methods usually varies from country to country, and the GPS user should rely on the parameters and methods adopted in the country of interest.

DMA produced a set of standard Molodenskiy transformation parameters, which can be used in those countries where specific procedures are unavailable or have yet to be developed. DMA's standard Molodenskiy formulas do not, however, use all seven parameters. Instead, a three-parameter approach is used in which the axial rotations and the scale changes are neglected. The accuracies of these DMA-derived transformation parameters vary from datum to datum but are typically in the order of 5–10 meters and are thus acceptable for most code-based GPS positioning and navigation.

To achieve the higher accuracy usually required for surveying and geodesy applications, GPS users can compute their own transformations for each particular survey area. To do this, GPS measurements are made at local geodetic control stations to pro-



$$\mathbf{R}_2 = \mathbf{R}_0 + (1 + m)\,\mathbf{R}_{xyz}(\varepsilon_x, \varepsilon_y, \varepsilon_z)\,\mathbf{R}_1$$

$$\text{where } \mathbf{R}_{xyz} = \begin{bmatrix} 1 & \varepsilon_z & -\varepsilon_y \\ -\varepsilon_z & 1 & \varepsilon_x \\ \varepsilon_y & -\varepsilon_x & 1 \end{bmatrix}$$

Figure 2. A seven-parameter similarity transformation is used to transform the components of vector R_1 — which describes the position of a point in one coordinate system — to R_2, describing the position in a second system. This transformation consists of a three-axis translation of the system's origins given by the vector R_0 and involves three small rotations about the coordinate axes (ε_x, ε_y, and ε_z) and a scale difference (m). The sign convention for the rotations used here considers the rotation to be positive if it is counterclockwise when viewed from the positive end of the axis (the right-hand rule).

vide common points in both the local and WGS84 geodetic datums. With a sufficient number of common points, users can derive their own transformation parameters to attain GPS coordinates that are fully compatible with the local geodetic control.

Projective Transformations. An alternative approach is to use a projective transformation, which is often necessary for larger areas because of errors that occurred during the original surveys to establish the local geodetic datum. These errors can now be detected using the more precise GPS surveys. In this case, transformation size varies depending on the location, where part of the transformation accounts for the change in datum and part accounts for the errors in the realization of the local datum through published control station coordinates. Again, DMA produced such projective transformations, called multiple-regression equations, for the datums used in the contiguous land areas of seven continental-sized regions (Australia, Brazil, Argentina, western Europe, Canada, the United States, and the rest of South America). These multiple-regression equations can provide better accuracy than the similarity transformations because of their ability to rectify datum errors, and because they are typically accurate to a few meters for these continental areas. These transformations can also be improved for local areas using the methods described for similarity transformations.

After applying one of these transformation approaches, the GPS-derived geographic coordinates are compatible with the local geodetic datum and can then be integrated with existing coordinates. In many cases, however, these existing coordinates are expressed as map grid coordinates — eastings and northings — on a map projection for that particular part of the world. As a result, one further conversion must be applied — namely, the map projection. Most of the algorithms required to transform geographical latitude and longitude to easting and northing can be found in a monograph called *Map Projections — A Working Manual*, by John P. Snyder of the U.S. Geological Survey.

MAP PROJECTIONS

Basically, a map projection can be likened to a movie or slide projection. A beam of light shines at the center of the earth, projecting the features of the earth's surface (or actually, the representative ellipsoid) onto a screen.

In mapping, the shape of that screen partly determines the type of map projection: cylindrical, conical, or azimuthal. In a cylindrical projection, the screen is a cylinder straddling the earth; a conical projection corresponds to a cone-shaped screen placed over the earth; and an azimuthal projection is associated with a flat screen.

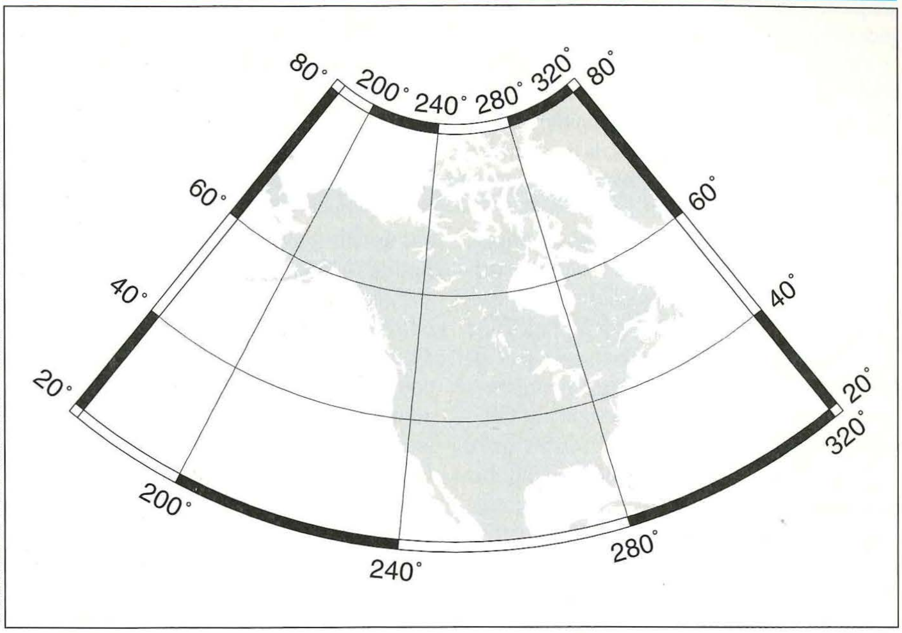
In each of these types, the amount of distortion increases as the screen moves away from the earth's projected surface. As described earlier, a different class of projection may be chosen to best represent the earth's surface over a particular region.

In addition, the map projection properties can be chosen so that the map is one of equal area (preserving a constant ratio between the area of a region on the ellipsoid and the area of the corresponding region on the map). Or, it can be equidistant (preserving an exact scale along a set of lines, perpendicular to the line along which distortion is zero, for example). It can also be equiangular (preserving the angle between any two lines on the ellipsoid). An equal-angle projection is also known as conformal. These properties determine whether a map is more suitable for showing, for example, demographics (equal area) or for navigating (equal angle).

The third distinguishing feature of a map projection is its aspect. This characteristic is analogous to the position of a movie viewer with respect to the screen. Choices include normal, oblique, and transverse. A normal aspect corresponds to a viewer situated directly in front of the screen; an oblique aspect corresponds to a viewer to one side of the screen's center; and a transverse aspect corresponds to a viewer sitting sideways.

Some of the more common map projections are Lambert Conical (see Figure 3), Mercator, and Azimuthal.

These rather basic descriptions of projections only scratch the surface of map projection theory. Again, the interested reader is referred to Snyder's manual. The principal concern for GPS users is that GPS-derived



GENERATED BY WILL FEATHERSTONE USING GENERIC MAPPING TOOLS SOFTWARE

Figure 3. The Lambert Conformal Conical map projection is one of many projections devised in the eighteenth century by J.H. Lambert. All meridians are represented by equally spaced straight lines, and the parallels of latitude are represented by a series of unequally spaced arcs of concentric circles. This projection has been used extensively for mapping regions with a predominantly east–west expanse.

map coordinates are consistent with the projection used on their maps.

THE LINKS

One common misconception, most probably delivered by school textbooks, is that latitude and longitude are unique. But, according to the differences between datums and their associated ellipsoids used around the world, this is not exactly true. Before satellite positioning, when surveyors, geodesists, and cartographers were using only the latitudes and longitudes of positions in their own part of the world, this school-book assumption was taken as valid. With the proliferation of GPS, however, the coordinate differences have become more readily apparent, proving this assumption to be wrong. Any single point on or near the earth's surface can have numerous coordinate values based on the datum and associated ellipsoid chosen.

This argument can be extended when using different map projections. The map grid coordinates depend on the map projection to make them unique. Therefore, those who produce coordinates and maps must now provide information about the datum, ellipsoid, and map projection used. Only when this vital information is supplied do the latitudes and longitudes become unique and unambiguous.

One caveat applies when transforming and projecting WGS84 latitudes and longitudes to eastings and northings on a local map. The local datum, ellipsoid, and map projection are usually linked to one another by definition. A geodetic datum, in addition to the size, shape, and orientation of the reference ellipsoid with respect to the earth, comprises a set of coordinates in a particular region or country. These result from a least-squares adjustment of geodetic observations. The latitudes and longitudes are referred to the surface of the selected reference ellipsoid, which has been oriented for that part of the world. Therefore, the local ellipsoid's defining parameters must be used during the coordinate transformations and projections.

The Australian Geodetic Datum, for example, is based on the Australian National Spheroid (some countries use the word *spheroid* as a synonym for *ellipsoid*), whose semimajor-axis length is 6 378 160 meters and flattening is 1/298.25, compared with an axis length of 6 378 137 meters and flattening of 1/298.257 223 563 for WGS84. The local ellipsoid's defining parameters must also be used during the map projection. Again using the Australian example, the Australian Map Grid is based on a Universal Transverse Mercator (UTM) projection but uses the geometrical parameters associated with the Australian National Spheroid. Using the incorrect ellipsoid parameters can cause errors in the range of hundreds of meters in the projected coordinates in some parts of the world.

GPS RECEIVER FEATURES

Today, a plethora of GPS receivers are available on the market. The *GPS World* receiver survey in this issue of the magazine lists more than 375 receivers from more than 60 manufacturers. In addition to all the bells and whistles, GPS users — especially those residing or working outside North America — should look for GPS receivers and postprocessing software that can provide datum transformations and map projections. This would provide users with a truly "field-to-finish" system that could be used worldwide.

Such a feature is probably most appropriate for handheld and vehicle-mounted GPS receivers that are used for general navigation. A receiver display of WGS84 latitude and longitude might not be compatible with coordinates from a local map, making it of little practical use. Moreover, it would require the user to perform the relevant transformations and projections in the field. Such a time-consuming manual procedure would probably be prone to computational blunders.

In addition, GPS receivers that offer a transformation and projection option must be properly configured to transform to the right datum and to project to the correct map grid using the correct parameter values.

AND FINALLY . . .

Before GPS users operate a receiver outside North America or other areas that use WGS84-compatible datums, they should determine which geodetic datum, ellipsoid, and map projection they will use. Because different datums, ellipsoids, and map projections are used around the world, the WGS84 coordinates provided by the GPS receiver must be made compatible with local map coordinates. As stated earlier, omitting or incorrectly completing this procedure can result in errors sometimes exceeding 1 kilometer.

This discussion has only scratched the surface of the sciences of geodesy and cartography. Those interested in venturing further into the enchanted forest of coordinates, datums, and maps should consult the references listed in the "Further Reading" sidebar. ■

Further Reading

For an introduction to geodesy, see

■ "Basic Geodesy for GPS," by R.B. Langley, in *GPS World*, February 1992, pp. 44–49.

■ *Basic Geodesy: An Introduction to the History and Concepts of Modern Geodesy without Mathematics,* by J.R. Smith, published by Landmark Enterprises, Rancho Cordova, California, 1988.

For an in-depth discussion of geodesy, see the "bible" of geodesy

■ *Geodesy: The Concepts,* 2d edition, by P. Vaníček and E.J. Krakiwsky, published by Elsevier Science Publishers, Amsterdam, the Netherlands, 1986.

For a discussion about coordinate systems, see

■ "Coordinate Systems: How to Get Your Position Very Precise, But Completely Wrong," by V. Ashkenazi, in *The Journal of Navigation,* Vol. 39, No. 2, pp. 269–278.

■ "Coordinate Systems Used in Geodesy: Basic Definitions and Concepts," by T. Soler and L. Hothem, in *Journal of Surveying Engineering,* Vol. 114, pp. 84–97.

A full description of WGS84 can be found in

■ *DoD World Geodetic System 1984 — Its Definition and Relationships with Local Geodetic Systems,* 2d edition, NIMA Technical Report 8350.2, National Imagery and Mapping Agency, Washington, D.C., 1991.

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, U.S. Government Printing Office, Washington, D.C., 1987.

For a Web-based discussion of geodetic datums, see

■ "Geodetic Datum Overview" one of Peter Dana's contributions to the University of Texas at Austin's "The Geographer's Craft": <http://www.utexas.edu/depts/grg/gcraft/notes/datum/datum.html>.

Good ol' algebra

I was working with a U.S. Geological Survey (USGS) 7.5-minute topographic (topo) quad map just as the January 1997 *GPS World* arrived at my office. In it was a letter from Captain Bill Brogdon about plane coordinate grids. Captain Brogdon complained that most GPS users are not surveyors, don't know any trigonometry, and have never heard of plane coordinate grids.

Plane coordinates should not be that big a mystery. They are nothing but good old *x* and *y*, the abscissa and the ordinate, which made us adolescents grind our teeth in basic algebra class when I was in high school in the late 1940s. When I was a boy, as we old codgers are fond of saying, we had to take algebra, plane and solid geometry, and plane and spherical trigonometry. Perhaps the educational system has changed radically since then, and requiring such rigors of today's students would be considered damaging to their delicate psyches.

Which reminds me of the old joke:

Q: Tell me, Mr. Principal, how many students do you have in your high school?

A: Oh, about one out of forty.

Looking at my topo map, I see that the generous USGS has provided me with four options in grid systems. The spherical grid is geographic coordinates (latitude and longitude). The plane grids are (1) state plane coordinate system; (2) Universal Transverse Mercator; and (3) the Bureau of Land Management land net of townships and ranges with their sections (this can be considered a grid system, too).

Captain Brogdon's suggestion of a bar scale for geographic coordinates is appealing, but it is impractical from a map-publishing standpoint — the problem is convergence with a spherical system. However, drafting and engineering supply houses do sell clear, plastic templates that can handle the convergence and are good for any USGS 1:24,000 map, north or south.

I do not know the basis of Captain Brogdon's contention that *most* GPS users are not surveyors, but I do know that a very large number of them are surveyors, civil engineers, designers, architects, and urban planners, who all use both plane and spherical grid systems every day. It is not that difficult.

Richard Nagle
Chief, Cadastral Section
Real Estate Division
U.S. Army Engineer
District of Los Angeles, California

Coordinates, datums, indeed!

The January 1997 "Innovation" column entitled "Coordinates and Datums and Maps! Oh My!" about datum pitfalls said a lot that needed saying. But you don't have to go to Australia (authors Will Featherstone and Richard Langley's example in the first paragraph) or anywhere overseas to run into serious datum and coordinate problems using GPS and maps. You can come up with awful results right here in the United States, especially when you put GPS and geographic information system (GIS) data together, as I know from first-hand experience.

I teach GIS in the University of Kentucky's Geography Department. Several years ago, I borrowed a GPS receiver from Trimble Navigation to use in a class. It had occurred to me that our departmental faculty directory listed names, office and home addresses, phone numbers, and so on, but not geographic coordinates. So I sent students to park in my colleagues' driveways and collect GPS data, which we converted to an ARC/INFO coverage based on World Geodetic System of 1984 (WGS84) Universal Transverse Mercator (UTM) coordinates.

We then digitized the major roads from Lexington-area U.S. Geological Survey (USGS) 7.5-minute topographic quad maps based on the North American Datum of 1927 (NAD27). Two of the four maps indicated that to convert NAD27 to NAD83 (that is, WGS84), one should move the grid lines 4 meters south and 6 meters west — hardly an issue for us, because our accuracy was limited by selective availability to perhaps 50 meters when calculated over a number of measurements. It was, therefore, distressing when a graphic overlay of these two coverages put the current chair's house smack in the median of Lexington's limited-access beltway. (He, in fact, lives a couple of football fields north of there.)

What I learned, the embarrassing way, was that while the area's latitude and longitude coordinates were adjusted by a few meters in the 1927-to-1983 datum change, the UTM coordinates were adjusted more than 200 meters in the north–south direction.

The moral of this story — which partly prompted my writing a textbook relating GPS and GIS — is that datum and coordinate issues can be tricky anywhere, and one must be especially careful when combining data from different sources.

Michael Kennedy
Associate Professor
Geography Department
University of Kentucky, Lexington