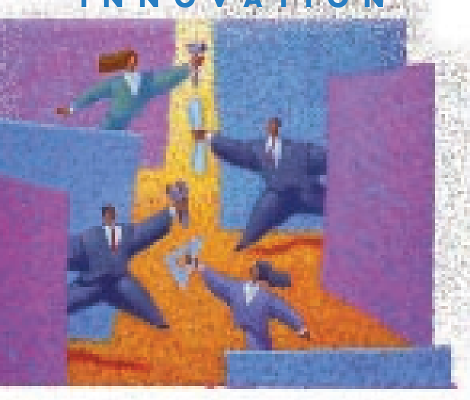


Navigation 101: Basic Navigation with a GPS Receiver

Richard B. Langley University of New Brunswick

The uses of GPS are virtually limitless, from monitoring the bulges of volcanoes to synchronizing communications over cellular-telephone networks. With GPS applications becoming more and more specialized, some users may have lost sight of the fact that, first and foremost, GPS is a navigation system — a system that anyone can use any time, and almost anywhere. In this month's column, we present a primer on this most basic use of GPS.

Navigation is the art and science of directing the movements of a person or craft of some kind — such as an automobile, vessel, or aircraft — expeditiously and safely from one point to another. This ancient endeavor goes back thousands of years. Phoenician and Greek traders plying the Mediterranean Sea and beyond used basic navigational techniques to travel safely from port to port. Although they typically hugged the coast, undoubtedly some were more adventurous and sailed out of sight of land, navigating by the Sun, the stars, and the winds. They may have had crude maps and sailing directions, but these early navigators relied mostly on their wits to get them safely to their destination and home again.

Essentially, a navigator needs to know how to steer the craft or, in other words, how to follow a route from origin to destination. To do this, he or she must know the craft's position and velocity (speed and direction of travel) at all times. Over the years, various tools have been developed to provide this information. For example, by the middle of the eighteenth century, astronomical observations with a sextant and chronometer could establish a ship's position to within 10 kilometers or so. A compass provided direction information and a log and line (a float on the end of a line knotted at precise intervals and tossed overboard) gave the ship's speed.

In the twentieth century, many developments have improved the accuracy of navigation. In addition to more refined maps and charts, various electronic navigational devices and systems have been developed, including the gyrocompass, radio-positioning systems such as Loran-C, and inertial navigation systems. These tools have allowed the navigator to answer questions such as "Where am I?" "How fast am I going?" "In what direction am I traveling?" and "How long before I reach my destination?" with greater and greater accuracy. But perhaps the most significant development of the twentieth century for navigators was the development of the Global Positioning System. GPS has revolutionized navigation on land, at sea, in the air, and in space. The system with its global coverage is available 24 hours a day every day, providing the navigator with a high-accuracy tool which operates in all weather conditions. The receivers are compact and relatively inexpensive, allowing their use by anyone from a backpacker to an airline pilot.

In this month's column, we'll examine the basics of GPS navigation. Although some aspects will deal specifically with marine navigation, we'll try to keep the discussion general enough to cover personal, vehicle, and air navigation as well.

WHERE AM I?

In answering this question, we could identify our location relative to some nearby, visible landmarks. But such a procedure is quite inaccurate and, in fact, there may be no visible landmarks available as is the case on the open sea. Since the third century B.C., navigators have expressed positions on the surface of the Earth using a system of angular coordinates — latitude and longitude. Up until the time of Newton, and even later for certain purposes, these coordinates were established by assuming the Earth to be a sphere. We require such a mathematical surface, approximating the actual shape of the Earth, to perform calculations involving positions, distances, and directions. What Newton demonstrated theoretically (and was later verified by precise surveys and astronomical observations) was that the Earth is better represented by an oblate biaxial ellipsoid than by a sphere. Both the sphere and the ellipsoid are approximations not to the physical surface of the Earth but to the *geoid*. The geoid is the equipotential surface of the Earth's gravity field that best approximates mean sea level over the whole globe. The geoid is much smoother than the actual topography but still somewhat irregular because it is influenced by variations in the Earth's mass density from place to place. However, the geoid's departure from a best-fitting biaxial ellipsoid is no more than about 100 meters. The best-fitting sphere, on the other hand, departs from the geoid by more than 10 kilometers! Accordingly, we use the ellipsoid for almost all of our positioning needs, and spherical coordinates have given way to ellipsoidal coordinates.

The geodetic latitude, ϕ, and the geodetic longitude, λ (sometimes loosely referred to as geographic coordinates) define a position on the surface of the ellipsoid. To position a point on the physical surface of the Earth, a third coordinate — the height above the ellipsoid — is required. This height, the geodetic height, h, is measured along the normal (or perpendicular line) between the ellipsoid and the point. The position of a point in geodetic coordinates is then fully defined by the triplet ϕ, λ, h. It is a straightforward procedure to convert geodetic coordinates to Cartesian coordinates (x, y, z) or vice versa.

In general, the geodetic height of a point is not the same as the orthometric height found on a topographic map. Orthometric heights are measured from the geoid or, loosely speaking, mean sea level, and differ from the geodetic height because the ellipsoid and the geoid don't coincide. However, the orthometric height can be computed if we know the geoidal undulation — the separation of the ellipsoid and the geoid.

Navigators, at least those confined to the Earth's surface (dry or wet), are usually not as interested in their three-dimensional position as in their two-dimensional one — they know their vertical coordinate quite well. So in the remainder of this article, we'll mostly talk about horizontal coordinates.

GETTING FROM A TO B

When navigating from origin to destination, we usually follow a planned route. In many cases, we would prefer to follow the shortest route between the two locations. The shortest path between two points on a mathematically defined surface such as an ellipsoid or sphere is a curve called a *geodesic*. On a sphere, the geodesic is a segment of a great circle which is formed by the intersection of the surface of the sphere and a plane through its center. It is the largest circle that can be drawn on the sphere's surface. If we plot a great-circle path between two sites on an ellipsoid, it can depart from the geodesic by many kilometers.

To illustrate the difference in the lengths of geodesics on the ellipsoid and the sphere, consider the path between Ronald Reagan Washington National Airport and Los Angeles International Airport. If we compute the distance assuming the Earth to be a sphere with the same surface area as the best-fitting ellipsoid, it works out to 3,711 kilometers or 2,004 nautical miles. Computing the length of the geodesic on the ellipsoid, we get a more exact measurement of 3,719 kilometers or 2,008 nautical miles. By the way, the *nautical mile* is a common and convenient unit in navigation. Originally, it was defined as the length of one minute of a great circle on the sphere approximating the Earth's shape. Since the dimensions of the best-approximating sphere have changed over the years, a variety of nautical miles has existed. In 1929, the International Hydrographic Bureau proposed a standard length of 1,852 meters, which is known as the International Nautical Mile. This unit has become the de facto standard.

A route which follows a single geodesic from origin to destination, although the shortest, may not be the most practical one. When navigating on land, for example, the Earth's topography or road network may prevent us from following the shortest-distance route. Even if we could follow it, it may not be necessarily the fastest route. Even on the open ocean, with no topographic barriers, it might be impractical to follow a geodesic as the ship's course would have to be adjusted continuously.

So typically a route must be broken down into a series of segments or legs — each segment a geodesic. The points at the beginning and end of each segment are called *waypoints*. A waypoint might be associated with a physical feature such as a road intersection, trail turn, navigational aid or a mooring buoy, or simply a point identified by coordinates such as a favorite fishing hole or a convenient point for changing course. Aircraft frequently follow routes from one navigational beacon to another, and the locations of these beacons serve as waypoints. Of course one of the key advantages of GPS for air navigation is that direct airport-to-airport routing is possible, obviating the need for doglegged routes following a string of beacons or other terrestrial navigational aids. Furthermore, GPS nonprecision approaches to airports use a series of waypoints which the aircraft either flies by or over as it proceeds to the runway. A special type of waypoint is a proximity waypoint that indicates the position of a hazard to navigation such as a shoal or shipwreck. A GPS receiver can provide warnings to users when they are within a certain distance of such a waypoint.

When navigating a specified route, a GPS receiver can provide the values of a number of navigational parameters including current position, bearing and distance to the next waypoint, speed, desired track, course made good, cross-track error, estimated time en route, and estimated time of arrival.

Position. The GPS receiver determines its three-dimensional position from four or more *pseudoranges*, or biased distances, between its antenna and the transmitting antennas of the GPS satellites. At least four pseudoranges

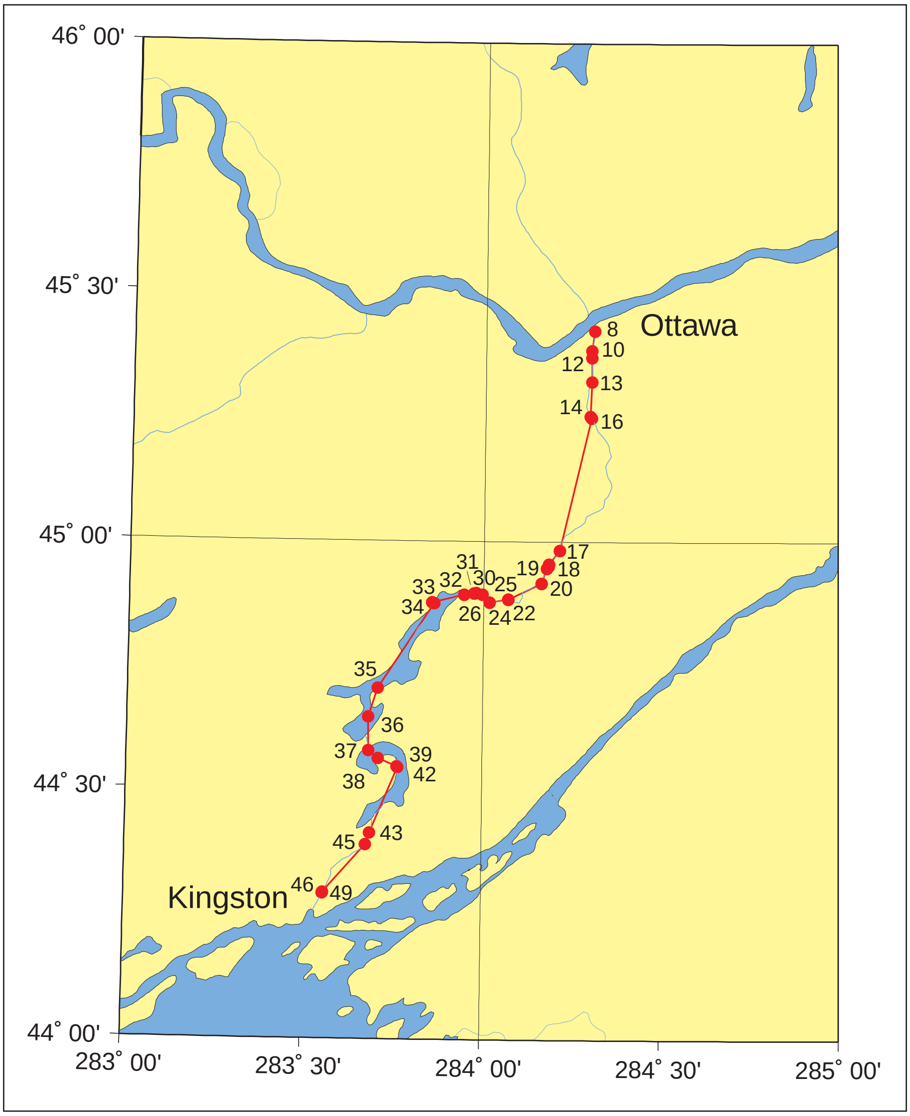


Figure 1 Navigation waypoints identify specific locations on or near a route such as the canal locks shown on this map of the Rideau Canal Waterway linking Kingston and Ottawa, Ontario. The lines connecting the waypoints are geodesics — lines with the shortest distances between the points. Of course, the actual route traveled by boats following the waterway is longer. Aircraft and vessels navigating in open seas can follow geodesic routes, although this requires frequent course changes.

are needed so that the receiver can also compute the unknown offset of its clock from GPS (System) Time. Sometimes this procedure is referred to as triangulation, but since no angles are measured, this term is technically incorrect. The correct term is *trilateration.*

If one of the receiver coordinates (say the height) or the clock offset is known, then the three remaining unknowns can be determined with only three pseudoranges. Most modern GPS receivers can track eight or more satellites simultaneously and determine their positions using all of the available observations with a least-squared-error algorithm or a Kalman filter. Depending on the size of various perturbing errors such as uncompensated atmospheric propagation delay or signal multipath and satellite-receiver geometry, horizontal position accuracy (now that selective availability has been turned off) can be better than 10 meters at the 95 percent probability level.

A receiver can usually display coordinates in a variety of selectable formats (degrees, minutes, and seconds; degrees and minutes; just degrees; or in grid coordinates — northings and eastings). The coordinates initially computed by the receiver refer to the World Geodetic System 1984 (WGS 84). However, the receiver can transform the coordinates to some other datum as selected by the user. Some receivers offer more than 100 different datums to choose from. It is also usually possible for a user to define a datum by specifying its parameters with respect to WGS 84. It is crucial that the user select the correct datum — the same datum indicated on the map or chart being used. And if using grid coordinates, the same projection must be used. Selecting the wrong datum and projection could result in displayed positions being hundreds of meters different from the position on the chart. Perhaps the most common map projection and grid system used for land navigation is the Universal Transverse Mercator (UTM) system. A key advantage of the UTM projection is its preservation of the shape of small areas on a map and its grid coordinates permit easy calculations using plane trigonometry.

Bearing. *Bearing* is the horizontal angle to a line from one point to another measured from a reference direction, usually north, clockwise from 0 through 360 degrees. The reference direction may be true, compass, magnetic (compass reading corrected for deviation), or grid north, or an arbitrary specified direction. Accordingly, bearings are designated as true, compass, magnetic, grid, or relative. Sometimes bearings are given in a range of 0 to 90 or 0 to 180 degrees, in which case the appropriate quadrant or semi-circle must be indicated. For example, the bearing N 40° W is 40 degrees west of north or equivalently 320 degrees. Occasionally the term "bearing angle" is used to designate such quadrant bearings. Frequently the term *azimuth* is used synonymously with bearing. Some navigation purists prefer to reserve the term azimuth to describe the position of an astronomical object on the celestial sphere and use the word bearing to refer to terrestrial objects.

Distance. The GPS receiver computes the geodesic distance between its current position and the next waypoint or between any two waypoints to the nearest 10 or 100 meters. Most receivers allow the user to select the distance units as statute miles, nautical miles, or kilometers. Note that the geodesic distance is not the same as the rhumb-line distance. A *rhumb line* makes the same oblique angle (has the same bearing) with all meridians. Marine navigators sometimes follow a rhumb line as the vessel does not need to change its true course while following such a route. A rhumb line appears as a straight line on a regular Mercator chart.

Course and Track. The horizontal direction in which a craft is actually moving (or its intended direction) expressed as an angular distance measured clockwise from north is called the course. In other words, it is the bearing or azimuth of a line along which a craft travels. In marine use, the term strictly refers to the direction through the water, ignoring the effects of currents and other sea motion. The direction relative to the ground is called the *track.* However, the terms course and track are often used interchangeably, particularly in land and air navigation. To remove the ambiguity, the term *course over ground* is also sometimes used to refer to the current direction of the craft. The accuracy with which a GPS receiver can compute this direction depends on its speed but is usually better than one degree for speeds greater than about 10 kilometers per hour. A course may be designated as true, compass, magnetic, or grid according to the reference direction being true, compass, magnetic, or grid north.

Note that the course is not necessarily the same as the *heading* or direction in which the vehicle or craft is pointing. Due to winds or currents, a craft may have to be pointed at an angle with respect to the desired track. A conventional GPS receiver with a single omnidirectional antenna cannot provide heading information; a multiple-antenna system is required to determine a craft's orientation using GPS signals.

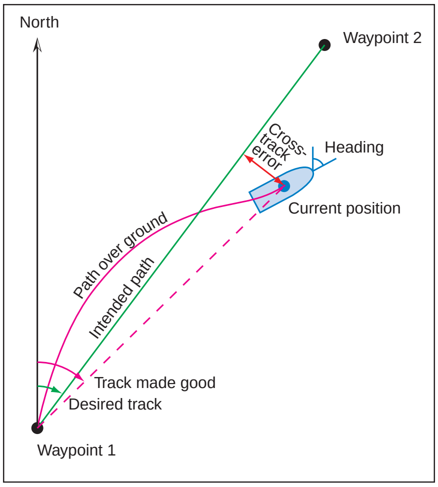


Figure 2 Track made good is the bearing of a craft's current position computed from the previous waypoint. The departure of the craft's actual path from the intended one is measured by the cross-track error.

Desired Track. This is the bearing between a route segment's starting and ending waypoints.

Course Made Good. The terms *course made good* and *track made good* are used to describe the direction from the previous waypoint to the actual current position. Once again, the terms course made good and track made good are often used interchangeably, ignoring the motion of the air or water through which the craft may be traveling.

Speed. In addition to computing its current position, the receiver computes its current speed. While in principle the receiver could determine its speed by differencing two closely spaced position fixes, a "back of the envelope calculation" readily shows that the result could be in error by tens of kilometers per hour. A receiver actually determines its speed using the measured Doppler shifts of the received signal carrier phases. In fact, the receiver determines its velocity vector, the magnitude of which is the speed and the direction of which gives the receiver's course. Both the speed and the course as directly computed by the receiver refer to the ground and not to the air or the water through which a craft may be moving. To be more precise, the terms *speed over ground* or *groundspeed* are sometimes used. As with position, speed can be expressed in different units including miles per hour, kilometers per hour, or knots (nautical miles per hour). Because of the action of the filter used to process the observations, the speed actually displayed by the receiver may be averaged over the past few seconds of travel. Speed accuracies depend on satellite-receiver geometry and other factors, but root-mean-square accuracies can be as good as 0.2 kilometers per hour or better under steady-state conditions.

Speed Made Good. This is the projection of the receiver's velocity vector along the desired track. It's the speed with which the destination waypoint is being approached. It is also called *velocity made good* with the understanding that the direction of the velocity vector is along the desired track. If the desired track is actually being followed, then the speed made good is the same as the speed over ground.

Cross-Track Error. The degree to which a craft is off course is termed the *cross-track error*. It is the distance measured perpendicularly from the desired track to the current position. A GPS receiver continuously computes the cross-track error and displays it on a course-deviation indicator (CDI). The CDI has a user-selectable linear scale and typically includes a pointer or arrow which points to the next waypoint. If the craft is off the desired track, steering in the direction of the arrow will get it back on course.

Estimated Time en Route. Based on the current speed made good, the receiver estimates how long it will take to reach the next waypoint. Using this same speed it may also calculate the estimated time to travel between each of the remaining waypoints. The sum of these times gives the time to reach the final destination.

Estimated Time of Arrival. By adding the estimated time en route to the current time as kept by the GPS receiver, the time of arrival at the next or any other waypoint (including the final destination) can be computed.

In addition to these primary navigational parameters, a GPS receiver may also provide the values of other parameters such as total distance traveled, elapsed time since start of trip or some other epoch, and average speed during the trip.

Map Displays. Even the simplest handheld receivers offer some kind of moving-map display which can portray the receiver's current position, nearby waypoints, planned route, and actual path history. It may also be able to determine the distance and bearing from the receiver's current position to any point on the map. More sophisticated receivers may have stored maps of road networks or marine charts, or be capable of displaying topographic or other features.

AUGMENTED NAVIGATION

Although the unaided accuracy of standalone GPS receivers is sufficient for many navigation uses, especially now that selective availability has been turned off, even higher accuracies are achievable through the use of differential corrections provided by an augmentation service such as the Coast Guard's Differential GPS Service or the Federal Aviation Administration's Wide Area Augmentation System. A GPS receiver might also be supplemented by a non-GPS system to provide continual navigation when GPS signals are temporarily blocked, such as when navigating a vehicle in urban canyons. Such a system might be as simple as one using a vehicle's odometer and speedometer readings and input from a digital magnetic compass, or as sophisticated as an inertial navigation system. Map matching might also be used, in which a GPS-derived position is snapped to a street or highway to improve positional accuracy and availability.

No navigation system is completely foolproof or without potential anomalies — not even GPS. A GPS receiver could become nonfunctional if its batteries are allowed to run down and there is no alternative power supply, or a handheld receiver could fall overboard when being used for boating. The GPS receiver's antenna could, unknown to the user, become disconnected, incorrect waypoint or map information could be entered into the receiver, or the wrong map datum could be used — all resulting in misleading information being presented to the navigator. These situations have actually occurred. For example, vessels have run aground on sandbars as a result of using positional information from a malfunctioning receiver. The prudent navigator should always perform checks on the data presented by the GPS receiver and use additional safeguards such as plotting positions on a paper chart. It is also a good idea to have a backup system available for use should the GPS receiver no longer function properly. For example, a marine navigator should still keep a sextant in good working order and know how to use it. A backpacker should still know how to navigate with map and compass using dead reckoning.

CONCLUSION

There have been many advances in GPS technology over the past 20 years or so. With GPS modernization already under way, we will see many more changes in the years to come. Receivers will no doubt become even smaller yet more sophisticated. They will be integrated with large, detailed map databases and high-resolution display systems with updates delivered over a real-time communications network. Receivers will be interfaced to vehicle, vessel, or aircraft guidance systems to provide automatic control. Yet the principles of navigation with a GPS receiver are unlikely to change.

In this brief article, we have overviewed the principles of GPS navigation, concentrating primarily on the basic concepts of the technology. But navigation is an art as well as a science. Like any art, it is one that can be mastered only by practice. ■

FURTHER READING

For one of the best general references on navigation, see

- *American Practical Navigator — An Epitome of Navigation,* originally by Nathaniel Bowditch and maintained continuously since it was first published in 1802. Published since 1868 by the U.S. government, the book is still commonly known as "Bowditch." The current edition was issued in 1995 by the National Imagery and Mapping Agency's (NIMA's) Marine Navigation Department. A PDF online version will soon be available from NIMA's website at <http://pollux.nss.nima.mil/pubs/>.

For guides to GPS navigation on land, at sea, and in the air, see

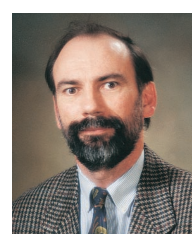
- *A Comprehensive Guide to Land Navigation with GPS,* 3rd Edition, by N.J. Hotchkiss. Published by Alexis USA, Herndon, Virginia, 1999.
- *GPS Instant Navigation,* 2nd Edition, by K. Monahan and D. Douglass. Published by Fine Edge Productions, Anacortes, Washington, 2000.
- *Aviator's Guide to GPS,* 3rd Edition, by C.W. Clarke. Published by McGraw-Hill, New York, 1998.

For a discussion of both the pitfalls and benefits of marine GPS navigation, see

- "Making Sense of GPS for Marine Navigation Training," by S.G. Shaw, in *GPS World,* Vol. 4, No. 6, June 1993, pp. 40–45.

For a review of the Universal Transverse Mercator projection and grid system, see

- "The UTM Grid System," by R.B. Langley, in *GPS World,* Vol. 9, No. 2, February 1998, pp. 46–50.



"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 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.