Global Positioning System: The Mathematics of GPS Receivers

RICHARD B. THOMPSON
University of Arizona
Tucson, AZ 85721

Introduction

GPS satellite navigation, with small hand-held receivers, is widely used by military units, surveyors, sailors, utility companies, hikers, and pilots. Such units are even available in many rental cars. We will consider the mathematical aspects of three questions concerning satellite navigation.

How does a GPS receiver use satellite information to determine our position?

Why does the determined position change with each new computation, even though we are not moving?

What is done to improve the accuracy of these varying positions?

We will see that receivers use very simple mathematics, but that they use it in highly ingenious ways.

Being able to locate our position on the surface of the earth has always been important for commercial, scientific, and military reasons. The development of navigational methods has provided many mathematical challenges, which have been met and overcome by some of the best mathematicians of all time.

Navigation by means of celestial observation, spherical trigonometry, and hand computation had almost reached its present form by the time of Captain James Cook's 1779 voyage to the Hawaiian Islands. For the next 150 years these methods were used to determine our location on land or sea. In the 1940s electronic navigation began with the use of fixed, land-based, radio transmitters. The present-day *LOng RAnge Navigation* (LORAN-C) system uses sequenced chains of such transmitters.

The use of satellites in navigation became common in the 1970s, with the introduction of the *Navy Navigation Satellite System* (NAVSAT or TRANSIT). This system uses the Doppler shift in radio frequencies to determine lines of position and locations.

The Satellites

Almost all satellite navigation now uses the *Global Positioning System* (GPS). This system, operated by the United States Department of Defense, was developed in the 1980s and became fully operational in 1995. The system uses a constellation of satellites transmitting on radio frequencies, 1227.60 mHz and 1575.42 mHz.

The original design of the system provided for eighteen satellites, with three satellites in each of six orbits. Currently, there are four satellites in each orbit. In the basic plan, the six orbits are evenly spaced every 60° around the Earth, in planes that

FIGURE 1
The System of Satellites.

are inclined at 55° from the Equator. Orbits are circular, at a rather high altitude of 20,200 kilometers above the surface of the Earth, with periods of twelve hours. FIGURE 1 displays one configuration of the basic eighteen satellites. Although not drawn to scale, it gives the correct feeling that we are living inside a cage of orbiting satellites, several of which are "visible" from any point on the surface of the Earth at any given time.

Receivers

Current GPS receivers are electronic marvels. They are hand-held, run on small batteries, weigh as little as nine ounces, and can cost under $150. We can turn on a receiver at any point on or above the surface of the Earth and, within a few minutes, see a display showing our latitude, longitude, and altitude. The indicated surface position is usually accurate to within 100 meters, and the altitude is usually in error by no more than 160 meters.

How does a small radio receiver listen to a group of satellites, and then compute our position, with great accuracy? We start by noting exactly what sort of information is received from the satellites. Each satellite sends signals, on both of its frequencies, giving (i) its position and (ii) the exact times at which the signals were transmitted.

The receiver also picks up time signals from the satellites, and uses them to maintain its own clock. When a signal comes in from a satellite, the receiver records the difference, Δt, in the time at which the signal was transmitted and the time at which it was received. If the Earth had no atmosphere, the receiver could use the speed, c, of radio waves in a vacuum to compute our distance $d = c \cdot \Delta t$ from the known position of the satellite. This information would suffice to show that we are located at some point on a huge sphere of radius d, centered at the point from which the satellite transmitted. However, the layer of gasses surrounding the Earth slows down radio waves and, therefore, distorts the measurement of distance. Receivers can partially correct for this by allowing for the effect of mean atmospheric density and thickness. Information from several satellites is combined to give the coordinates—latitude, longitude, and altitude—of our position in any selected reference system.

Several factors restrict the accuracy of this process, including: (i) errors in the determined positions of the satellites; (ii) poor satellite positioning; (iii) limitations on the precision with which times and distances can be measured; and (iv) the varying density of Earth's atmosphere and the angles at which the radio signals pass through the atmosphere. Some of these difficulties are overcome by the use of an ingenious plan that provides the key to GPS technology. It is rather complicated to explore this method in the actual setting of positioning on the Earth: The distances are large, the time differences are small, and the geometry is all in three dimensions. Fortunately, we can capture most of the salient features of GPS receiver operation in a simple two-dimensional model.

A Simple Model

Suppose that you are standing somewhere in a circular lot, with a radius of 100 ft. The lot is paved, except for an irregularly-shaped gravel plot that surrounds you. The mean distance from your position to the edge of the gravel is 20 ft. Cars circle the lot on a road. To determine your position, messengers leave from cars on the road and walk straight toward you. When such a messenger arrives, he tells you where and at what time he left the road. You have a watch and know that all messengers walk at a rate of 5 ft/sec on pavement but slow down to 4 ft/sec on gravel. Our model is shown in FIGURE 2.

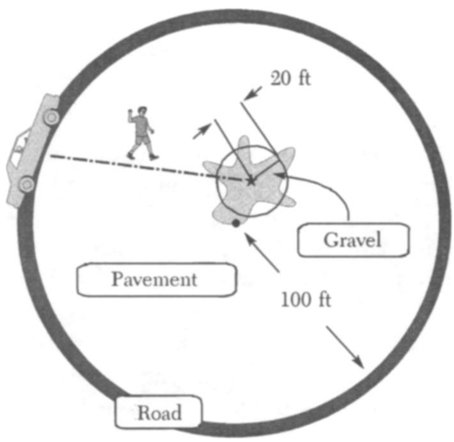

FIGURE 2
The Model.

Consider a rectangular coordinate system with its origin at the center of the lot. Distances will be measured to tenths of a foot, and time will be measured to tenths of a second. The location of a point on the road will be described by its angular distance from due north, measured in a clockwise direction.

At noon a messenger leaves a position 45° from north. When he arrives, your watch shows that it is 20.2 seconds after noon. Since you have no way to know the exact distance that he walked on the gravel, you assume that he covered the mean distance of 20 ft. At 4 ft/sec, this took him 5 sec. For the remaining 15.2 sec he walked on pavement, covering $5\frac{\text{ft}}{\text{sec}} \times 15.2 \text{ sec} = 76.0$ ft. Allowing for the assumed distance of 20 ft on the gravel, you know that you are located at some point on a circle of radius 96.0 ft, centered at the starting location of the messenger.

A second messenger leaves the road at a point 135° from north at 12:01 pm and walks to your position. On his arrival, your watch shows that it is 29.5 sec after he started. Assuming that he took 5 sec to walk 20 ft on the gravel, he walked 5 ft/sec 24.5 sec = 122.5 ft on the pavement. Hence, you are on a circle of radius 142.5 ft, centered at this messenger's point of departure.

The coordinates of the departure points for the two messengers are $P_1 = (100 \cdot \sin 45°, 100 \cdot \cos 45°)$ and $P_2 = (100 \cdot \sin 135°, 100 \cdot \cos 135°)$, respectively. Using our precision of one tenth of a foot, these are rounded to (70.7 70.7) and (70.7, − 70.7). Thus, the coordinates (x_0, y_0), of your position satisfy

$$\left\{ \begin{array}{l} (x_0 - 70.7)^2 + (y_0 - 70.7)^2 = 96.0^2 \\ (x_0 - 70.7)^2 + (y_0 + 70.7)^2 = 142.5^2 \end{array} \right\}.$$

The system has two solutions, $(-20.0, 39.2)$ and $(161.4, 39.2)$, rounded to tenths. Since the latter point is outside of the lot, you can conclude that you are located 20.0 ft west and 39.2 ft north of the center of the lot. The situation is shown in FIGURE 3.

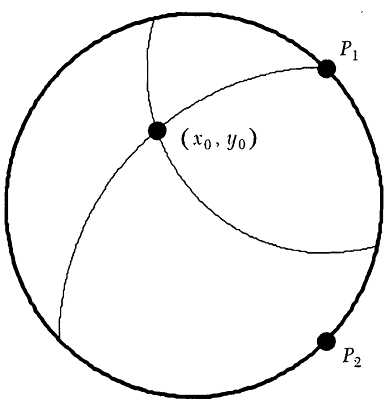

FIGURE 3
Two Messengers.

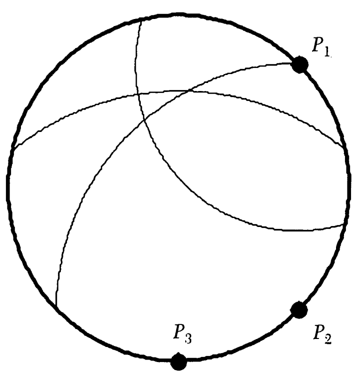

FIGURE 4
Three Messengers.

So far so good. Suppose that, just to be careful, you decide to check your position by having a third messenger leave the road at a point 180° from north and walk to your location. He leaves at 12:02 pm and, according to your watch, arrives 32.2 sec later. As before, you compute your distance from this departure point P_3. FIGURE 4 shows the result of adding information from the third messenger to your picture.

What has happened? *The most likely problem is that your watch does not agree with the times used at the departure points on the road.* Suppose that your watch runs steadily but has a fixed error of ε seconds, where a positive ε means that your watch is ahead of the road times and a negative ε means that your watch is behind the road times. If we let Δt be the time difference between departure and arrival, as shown on your watch, then the estimate for the distance traveled is

$$d(\Delta t, \varepsilon) = 20 \text{ ft} + (\Delta t \text{ sec} - \varepsilon \text{ sec} - 5 \text{ sec}) 5 \frac{\text{ft}}{\text{sec}}.$$

Thus, *the radius of each circle is in error by the same amount,* $-5\ \varepsilon$ *ft, and there must be a value of* ε *for which the three circles have a common point.* FIGURE 5 shows the effect of various watch errors.

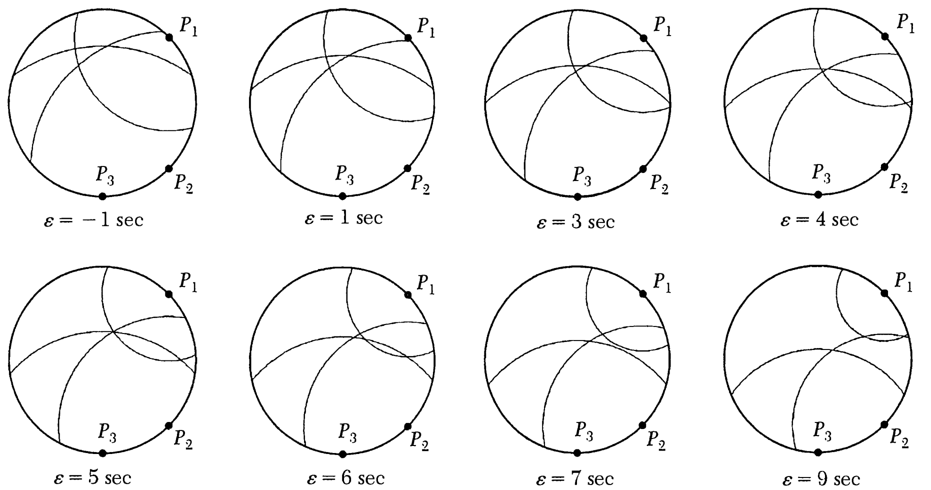

FIGURE 5
Effect of Watch Error.

It appears that your watch has an error of approximately 5 sec. The error and the coordinates of your position are a solution for the following system of equations:

$$\begin{Bmatrix} (x_0 - 70.7)^2 + (y_0 - 70.7)^2 = d(20.2, \varepsilon)^2 \\ (x_0 - 70.7)^2 + (y_0 + 70.7)^2 = d(29.5, \varepsilon)^2 \\ (x_0 - 0.0)^2 + (y_0 + 100.0)^2 = d(32.2, \varepsilon)^2 \end{Bmatrix}.$$

The system can be solved numerically, starting with seed values of 0 for ε and estimated coordinates of your position for x_0 and y_0. There is only one solution giving a location inside of our lot. Rounding this to our level of precision yields $(x_0, y_0, \varepsilon) = (10.9, 31.2, 4.9)$. You conclude that you are 10.9 ft east and 31.2 ft north of the center of the lot, and that your watch is 4.9 sec fast. You note the coordinates of your position, and discard the watch error, which is of no further interest to you.

As this example of our GPS model shows, you can use time difference information from three messengers to determine your position, relative to a coordinate system in the lot. *The only tools needed for this effort are a steady, but not necessarily accurate, watch and the ability to approximate the solution of a system of three equations in three unknowns.*

Back to the Satellites

Our "lot" is now the region inside of the satellite orbits (including the Earth), "cars on the road" are satellites, "messengers" are radio waves, and "gravel" is the Earth's atmosphere. We take the center of the Earth as the origin in our coordinate system. Working in three dimensions, we need information from four satellites. Call these S_1, S_2, S_3, and S_4; and suppose that S_i is located at (X_i, Y_i, Z_i) when it transmits a signal at time T_i. If the signals are received at times T_i', according to the clock in our receiver, we let $\Delta t_i = T_i' - T_i$, and let ε represent any error in our clock's time. The receiver allows for the mean effects of passage through the Earth's atmosphere and computes distances $d(\Delta t_i, \varepsilon)$ that indicate how far we are from each of the satellites. Our position (x_0, y_0, z_0) is located on each of four huge spheres. In most situations, there will be only one sensible value of ε that allows the spheres to have a point in common. Our location is determined by solving a system equations.

$$\begin{Bmatrix} (x_0 - X_1)^2 + (y_0 - Y_1)^2 + (z_0 - Z_1)^2 = d(\Delta t_1, \varepsilon)^2 \\ (x_0 - X_2)^2 + (y_0 - Y_2)^2 + (z_0 - Z_2)^2 = d(\Delta t_2, \varepsilon)^2 \\ (x_0 - X_3)^2 + (y_0 - Y_3)^2 + (z_0 - Z_3)^2 = d(\Delta t_3, \varepsilon)^2 \\ (x_0 - X_4)^2 + (y_0 - Y_4)^2 + (z_0 - Z_4)^2 = d(\Delta t_4, \varepsilon)^2 \end{Bmatrix}$$

When a numerical solution is found, the rectangular coordinates (x_0, y_0, z_0) are converted into the essentially spherical coordinates of latitude, longitude, and altitude above sea level.

As a practical matter, there are times and locations when a GPS receiver can receive usable data from only three satellites. In such cases, a position at sea level can still be found. The receiver simply substitutes the surface of the Earth for the missing fourth sphere.

To summarize our results so far, the receiver is expected to (i) receive time and position information from the satellites, (ii) maintain a steady (but not necessarily accurate clock), (iii) select four satellites with a good range of positions, (iv) find an approximate numerical solution for a system of four equations, and (v) make a transformation of coordinates. Given the current state of electronics, these are easy tasks for a small hand-held instrument.

Variability of Positions

Our second question about GPS positioning causes a great deal of discussion and confusion among those who use the system. *If a person stands in one fixed location and determines repeated positions with a receiver, the coordinates of these positions will vary over time.* Since the observer's location has not changed, the changing positions are often attributed to alteration of the satellite signals by the Department of Defense. The Department does, at times, degrade the satellite data and cause a loss of GPS accuracy. This *Selective Availability* (SA) will be phased out within the next few years. (It is stated that SA is used for reasons of national security.) However, manipulation of the signals explains very little of the variation in positioning. The variation is primarily caused by random errors in measurement, the selection of different satellites, and by the effects of the atmosphere. We will illustrate these problems by returning to our simple 2-dimensional model.

In our example, you determined that the coordinates of your position were (10.9, 31.2) and that your watch was 4.9 seconds fast. *Suppose that these values are exactly correct.* After a couple of minutes you again use three messengers to determine your location. This time the first messenger leaves from the road at a point that is 47.2° from north. Rounding to your level of precision, you record the departure point as $P_1 = (73.4, 67.9)$. In this case we will assume that your information on the location of the departure point is not quite correct, and that the messenger actually left from $Q_1 = (74.1340, 67.3568)$. This is only a 1.0% error in the first coordinate and a 0.8% error in the second coordinate. Suppose also that the messenger actually encountered 25.9 ft of gravel on his way to your position.

Keeping track of 6 places, the distance from Q_1 to your location is 72.841286 ft. Covering the 25.9 ft of gravel at 4 ft/sec took the messenger 6.475000 sec, and covering the 46.941286 ft of pavement at 5 ft/sec took 9.388257 sec. The actual walking time was 15.863287 sec, which with your watch error of 4.9 sec, is 20.763257 sec. Using the allowed one place of precision, you would note $\Delta t = 20.8$ sec. Recall that, as you stand in the lot, you have no way of knowing the amount of gravel over which a messenger has walked. Hence, you always assume the mean distance of 20 ft. Under this assumption, the messenger would take 5 sec to cover the gravel, leaving 15.8 sec to walk on the pavement. At 5 ft/sec he would cover 79.0 ft. You conclude that the messenger has traveled 99.0 ft, and that you are at that distance from P_1.

To find your position, messengers leave from the road at points 138.5° and 8.1° from north. You record these departure points as $P_2 = (66.3, -74.9)$ and $P_3 = (14.1, 99.0)$. Now suppose that your information is slightly incorrect, and that the departure points are actually $Q_2 = (66.8404, -75.6490)$ and $Q_3 = (13,9731, 98.0100)$. In addition, assume that the second messenger walked over 22.1 ft of gravel and that the third messenger walked over 12.0 ft of gravel. Working in the same way as you did for the first messenger, you record time differences of 30.1 sec and 18.9 sec for the second and third messengers, and solve the following system of equations.

$$\begin{Bmatrix} (x_0 - 73.4)^2 + (y_0 - 67.9)^2 = d(20.8, \varepsilon)^2 \\ (x_0 - 66.3)^2 + (y_0 + 74.9)^2 = d(30.1, \varepsilon)^2 \\ (x_0 - 14.1)^2 + (y_0 - 99.0)^2 = d(18.9, \varepsilon)^2 \end{Bmatrix}$$

The solution, when rounded, gives your location as $(x_0, y_0) = (5.4, 32.3)$ and your watch error as 4.4 sec. Small errors in the location of the departure points, variation in the amount of gravel covered, and the rounding of numbers to one-place have produced a "position" that is 5.61 ft from your actual location of (10.9, 31.2).

We can let a computer simulate what happens if you stay in your fixed location and make repeated computations of your position. Each determination of a position is made with the following assumptions. (i) Three points of departure for messengers are picked at random, assuming that the angle between any two points of departure is at least 30°, but not more than 150°. (ii) The distance over which a messenger must walk on gravel is a normal random variable with a mean of 20 ft and a standard deviation of 5 ft. (iii) The relative error in each coordinate of the point of departure is a normal random variable, with a mean of 0 and a standard deviation of 0.3%.

It is common to discuss accuracy of positioning in terms of *circular errors of probability* (c.e.p.). The $n\%$ c.e.p. is the distance, d_n, such that the probability of an error that is less than or equal to d_n is $n\%$. A set of 1,000 simulated positions allowed us to estimate c.e.p.'s for our model. We found $d_{50} = 2.11$ ft and $d_{95} = 7.69$ ft. (It is interesting to note that one simulated position was 16.9 ft from the correct location.) The positions computed in a run of 50 simulations are plotted on the left side of FIGURE 6, along with circles of radii d_{50} and d_{95}. *Our probabilistic model yields results that agree quite well with plots of successive positions found with an actual GPS receiver from a fixed location.*

Commercially available GPS units operate under what is called the *Standard Positioning Service* (SPS), measuring distances using satellites' 1575.42 mHz frequency. Under the best circumstances, the 50% c.e.p. for the SPS is 40 meters. As we mentioned, selective availability adds a small amount of random error into the SPS. At almost all times the 50% c.e.p. is no more than 100 meters, with a common value being around 50 meters. Under these conditions, the 95% c.e.p. for SPS is

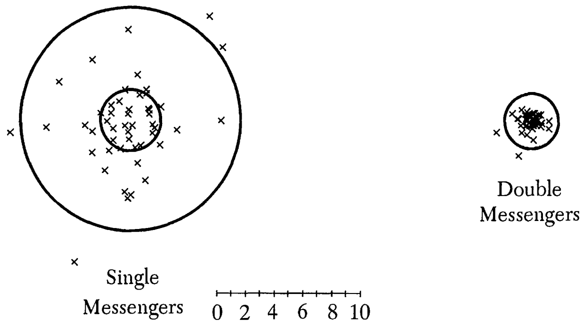

FIGURE 6
Simulations.

approximately 100 meters. As in our model, almost all of the variability in SPS positions comes from random errors that are inherent in the components of the system.

PPS and Differential GPS

What is done to remove some of the random errors from GPS positions? At the present time, there are two common methods of improving GPS accuracy. One of these is the *Precise Positioning Service* (PPS), which is available for governmental use only. This uses signals transmitted on both of the GPS frequencies to eliminate much of the variability caused by the Earth's atmosphere. Just as with various colors of light in the visible spectrum, the reduction in the speed of a radio wave as it passes through the atmosphere depends upon its frequency. Hence, measurements of the arrival times of two signals of different frequencies can be used to greatly improve the accuracy of our distance estimates.

As before, the situation is most easily understood in terms of our simple two-dimensional model. To model the PPS we will suppose that each messenger is accompanied by an assistant, who also walks at 5 ft/sec over pavement. However, while the messenger walks at 4 ft/sec over gravel, the assistant is slowed to 3 ft/sec when walking on gravel. We will return to our first example of the variability of positions and see what improvement in accuracy results from knowledge gained with assistant messengers.

Recall that your location in the lot has coordinates (10.9, 31.2), and that your watch error is 4.9 sec. The first messenger departed from $Q_1 = (74.1340, 67.3568)$ and walked over 25.9 ft of gravel while covering the 72.841286 ft to your position. With your watch error, you recorded a time difference of $\Delta t = 20.8$ sec.

The assistant messenger will require 8.633333 sec to cover the 25.9 ft of gravel at 3 ft/sec and 9.388257 sec to cover the paved part of the route at 5 ft/sec. Hence, his total walking time will be 18.021591 sec. Due to the error in your watch and the allowed level of precision, you record an elapsed time of $\Delta s = 22.9$ sec for the assistant messenger. The time differences for the messenger and the assistant messenger give you enough information to estimate the amount of gravel that lies between you and the point of departure on the road, and to estimate the total distance from your location to the point of departure. If we let G be the number of feet of gravel and let D be the total distance, in feet, then we have the following system of linear equations:

$$\left\{\begin{aligned} 20.8 &= \frac{D-G}{5} + \frac{G}{4} \\ 22.9 &= \frac{D-G}{5} + \frac{G}{3} \end{aligned}\right\}.$$

Solving the system at your level of precision, you conclude that the messenger and his assistant crossed 25.2 ft of gravel, and came 97.7 ft from their point of departure. Thus, as best you can tell, you are at some point on a circle of radius 97.7 ft, centered at the nominal point of departure $P_1 = (73.4, 67.9)$.

Similarly, suppose the second and third messengers also have assistants. Computations similar to those above show that the second messenger traveled 144.8 ft, including 22.8 ft over gravel, and that the third messenger traveled 91.5 ft, including 12.0 ft over gravel. (These values include any possible watch error.) Your estimate for the actual distance that a messenger has traveled is now a function of the distance, G, of gravel covered; the time difference, Δt; and your watch error, ε.

$$d(G, \Delta t, \varepsilon) = G \text{ ft} + \left(\Delta t \text{ sec} - \varepsilon \text{ sec} - \frac{G \text{ ft}}{4\frac{\text{ft}}{\text{sec}}}\right) 5\frac{\text{ft}}{\text{sec}}.$$

This distance formula and the three points of departure lead to a system of equations whose solution (x_0, y_0, ε) gives an estimate of the coordinates for your position and for the error of your watch.

$$\left\{\begin{aligned} (x_0 - 73.4)^2 + (y_0 - 67.9)^2 &= d(25.2, 20.8, \varepsilon)^2 \\ (x_0 - 66.3)^2 + (y_0 + 74.9)^2 &= d(22.8, 30.1, \varepsilon)^2 \\ (x_0 - 14.1)^2 + (y_0 - 99.0)^2 &= d(12.0, 18.9, \varepsilon)^2 \end{aligned}\right\}.$$

Solving this system, at your level of precision, yields a position of $(x_0, y_0) = (9.2, 31.6)$ and a watch error of 4.8 sec. Your current estimate is only 1.75 ft from your correct location of (10.9, 31.2). This compares with an error of 5.61 ft found by using single messengers.

A computer-generated set of 1,000 simulations for positions computed with messengers and assistants gave estimates of 0.50 ft and 1.90 ft for the c.e.p.'s d_{50} and d_{95}, respectively. The maximum distance of a computed position from the actual location was 3.41 ft. These simulations were based upon the same conditions that we used for our model of the SPS. The positions computed in a run of 50 simulations for our model of PPS are plotted on the right side of FIGURE 6, along with circles of radii d_{50} and d_{95}. Comparison of the two sides of FIGURE 6 shows that there is a considerable gain in accuracy when most of the variation due to distance walked over gravel is eliminated.

In the real world of satellites and positions on the Earth, the use of two radio frequencies in the PPS produces considerably more accuracy than can be obtained with the single-frequency SPS. It is believed that the PPS has a 50% c.e.p. of approximately 16 meters.

A second method for improving the accuracy of the usual SPS locations is coming into use at airports and major harbors. This is called the *Differential Global Positioning System* (DGPS). Most of the error in a GPS position is due to random variables in the atmosphere and the satellite system. Hence, within a small geographical area, the error at any instant tends to be independent of the exact location of the receiver. DGPS exploits this situation by establishing a fixed base station, whose exact location is already known. Equipment at the base station computes its current "GPS position," compares this with its known location, and continuously broadcasts a correction term. A DGPS receiver in the area receives its own satellite information and computes its position. Simultaneously it receives the current correction from its base station, and applies this to its computed position. The result is a very accurate determination of the receiver's position; 50% c.e.p.'s for GDPS run close to 9 meters.

Conclusions

The very ingenious idea of leaving clock error as a variable allows a GPS receiver to display our position on the Earth at any location and at any time, using nothing more than simple algebra. The variations in computed positions are almost entirely due to inherent limitations on precision within the system. A second clever plan allows the use of two radio frequencies to eliminate much of the variability caused by the passage of signals through the Earth's atmosphere.

REFERENCES

1. Nathaniel Bowditch, American Practical Navigator, Defense Mapping Agency Hydrographic/Topographic Center, 1984.
2. GPS Global Positioning System, a supplemental program produced by the United States Power Squadrons, 1995.
3. Product information, West Marine, Inc., 1997.
4. United States Coast Guard navigation web site, `http://www.navcen.uscg.mil/`