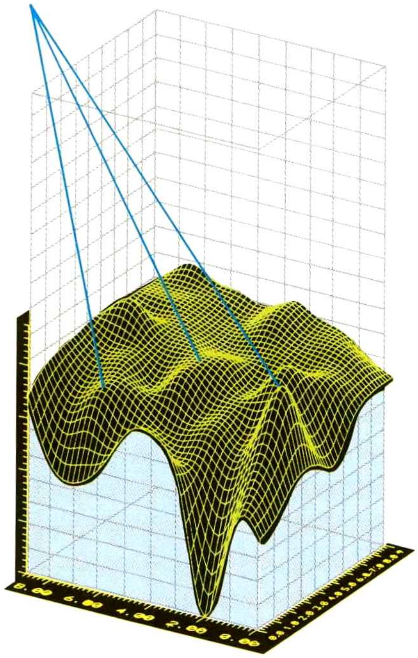

The GPS Observables

Richard B. Langley
University of New Brunswick

In previous columns we have looked at the structure of the signals transmitted by the GPS satellites and the basic operations performed by a GPS receiver in acquiring and processing the signals. In this month's column we'll take a closer look at the nature of the observations themselves, the biases and errors that afflict them, and how these effects can be removed or mitigated through modeling and data-differencing techniques.

"Innovation" is a regular column in GPS World *featuring discussions on recent advances in GPS technology and its applications as well as on the fundamentals of GPS positioning. The column is coordinated by Richard Langley and Alfred Kleusberg of the Department of Surveying Engineering at the University of New Brunswick, and we appreciate receiving your comments as well as suggestions of topics for future columns.*

The word *observable*, used as a noun, was coined by physical scientists to refer to measurable parameters of a system. This seemingly odd word usage enables us to distinguish the thing being measured (the observable) from the measurement itself (the observation). The basic observables of the Global Positioning System — at least those that permit us to determine position, velocity, and time — are the pseudorange and the carrier phase. Combining the basic observables in various ways can generate additional observables that have certain advantages.

THE PSEUDORANGE

Before discussing the pseudorange, let's quickly review the structure of the signals transmitted by the GPS satellites (for a more in-depth examination, see "Why is the GPS Signal So Complex?" in the May/June 1990 issue of *GPS World*). Each GPS satellite transmits two signals for positioning purposes: the L1 signal, centered on a carrier frequency of 1,575.42 MHz, and the L2 signal, centered on 1,227.60 MHz. Modulated onto the L1 carrier are two pseudorandom noise (PRN) ranging codes: the 1-millisecond-long C/A-code with a chipping rate of about 1 MHz, and a week-long segment of the P-code with a chipping rate of about 10 MHz. Also superimposed on the carrier is the navigation message, which, among other items, includes the ephemeris data describing the position of the satellite and predicted satellite clock correction terms. The L2 carrier is modulated by the P-code and the navigation message — the C/A-code is not present.

The PRN codes used by each GPS satellite are unique and have the property that the correlation between any pair of codes is very low. This characteristic allows all satellites to share the same carrier frequencies.

The PRN codes transmitted by a satellite are used to determine the pseudorange — a measure of the range, or distance, between the satellite and the antenna feeding a GPS receiver. The receiver makes this measurement by replicating the code being generated by the satellite and determining the time offset between the arrival of a particular transition in the code and that same transition in the code replica. In principle, the time offset is simply the time the signal takes to propagate from the satellite to the receiver. The pseudorange is this time offset multiplied by the speed of light. The observable is called a pseudorange because it is biased by the lack of time synchronization between the clock in the GPS satellite, which governs the generation of the satellite signal, and the clock in the GPS receiver, which governs the generation of the code replica. The receiver determines this synchronization error along with its position coordinates from the pseudorange measurements. The pseudorange is also biased by several other effects including ionospheric and tropospheric delay, multipath, and receiver noise. We can write an equation for the pseudorange observable that relates the measurement and the various biases:

$$p = \rho + c \times (dt - dT) + d_{ion} + d_{trop} + \varepsilon_p$$

where p is the measured pseudorange, ρ is the geometric range to the satellite, c is the speed of light, dt and dT are the offsets of the satellite and receiver clocks from GPS time, d_{ion} and d_{trop} are the delays imparted by the ionosphere and troposphere, respectively, and ε_p represents the effect of multipath and receiver noise. The receiver coordinates are hidden in the geometric range along with the coordinates of the satellite. The objective in GPS positioning is to mathematically describe all the terms on the right-hand side of

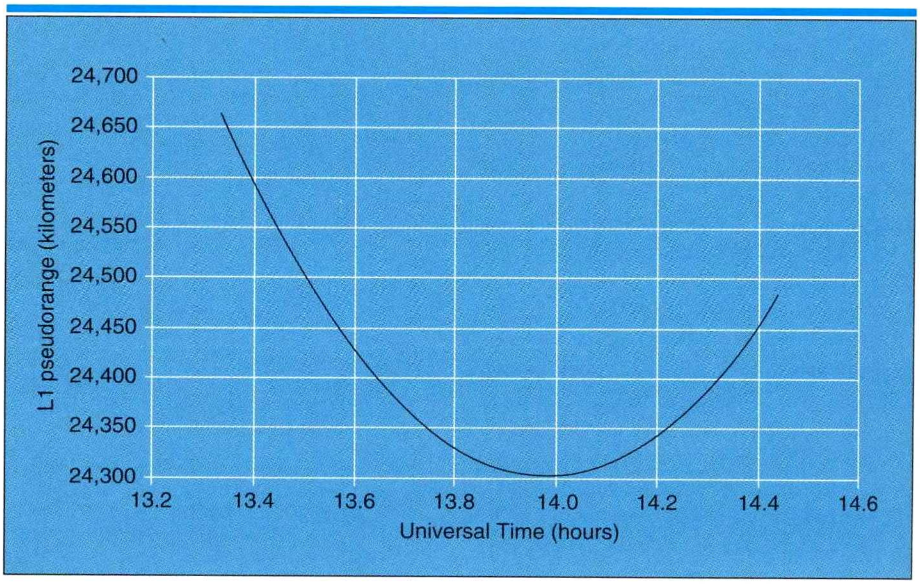

Figure 1. Typical variation in L1 pseudorange measurements over approximately a one-hour period.

the equation — including the initially unknown receiver coordinates in the geometric range term — so that the sum of the terms equals the measurement value on the left-hand side. Any error in the description of the terms will result in errors in the derived receiver coordinates. For example, both the geometric range term and the satellite clock term may include the effects of selective availability (SA), which, if uncompensated, introduce errors into the computed position of the receiver.

Pseudoranges measured by a GPS receiver at a site with accurately known coordinates can be used to monitor the behavior of an atomic clock at the site. By observing the same satellites at the same times and averaging the derived clock offsets, precise time and time interval laboratories on different continents are able to compare the performance of clocks and their respective time scales at accuracy levels of 10 nanoseconds or better.

Figure 1 illustrates the variation in the pseudorange of a particular satellite as measured by a stationary GPS receiver. The large variation is dominated by the change in the geometric range due to the satellite's orbital motion and the rotation of the earth.

Pseudoranges can be measured using either the C/A-code or the P-code. Figure 2 shows typical C/A-code pseudorange noise. This "noise record" was obtained by subtracting the geometric range, clock, and atmospheric contributions from the pseudorange measurements illustrated in Figure 1. What remains is chiefly pseudorange multipath and receiver measurement noise. Because of its higher chipping rate, the P-code generally provides higher-precision observations. However, recent improvements in receiver technologies have resulted in higher-precision C/A-code measurements than were previously achievable.

CARRIER PHASE

Even with the advances in code measurement technology, a far more precise observable than the pseudorange is the phase of the received carrier with respect to the phase of a carrier generated by an oscillator in the GPS receiver. The carrier generated by the receiver has a nominally constant frequency, whereas the received carrier is changing in frequency because of the Doppler shift induced by the relative motion of the satellite and the receiver. The difference between the received carrier and the receiver-generated one is sometimes referred to as the carrier beat phase because the phase difference relates to the difference, or *beat* frequency, of the two carriers. Such a beat frequency is well known to musicians who tune their instruments by listening for the beat note generated when they play two notes slightly different in frequency. The phase of the received carrier is related to the phase of the carrier at the satellite through the time interval required for the signal to propagate from the satellite to the receiver.

So, ideally, the carrier-phase observable would be the total number of full carrier cycles and fractional cycles between the antennas of a satellite and a receiver at any instant. The problem is that a GPS receiver has no way of distinguishing one cycle of a carrier from another. The best it can do, therefore, is measure the fractional phase and then keep track of changes to the phase; the initial phase is undetermined, or ambiguous, by an integer number of cycles. To use the carrier phase as an observable for positioning, this unknown number of cycles or *ambiguity*, *N*, must be estimated along with the other unknowns — the coordinates of the receiver.

If we convert the measured carrier phase in cycles to equivalent distance units by multiplying by the wavelength, λ, of the carrier, we can express the carrier-phase observation equation as

$$\Phi = \rho + c \times (dt - dT) + \lambda \times N - d_{ion} + d_{trop} + \varepsilon_{\Phi}$$

which is very similar to the observation equation for the pseudorange — the major difference being the presence of the ambiguity term. In fact, the carrier phase can be thought of as a biased range measurement just like the pseudorange. Note also that the sign of the ionospheric term in the carrier-phase equation is negative, whereas in the pseudorange equation it is positive. This comes about because the ionosphere, as a dispersive medium, slows down the speed of propagation of signal modulations (the PRN codes and the navigation message) to below the vacuum speed of light, whereas the speed of propagation of the carrier is actually increased beyond the speed of light. Don't worry, Einstein's pronouncement on the sanctity of the speed of light has not been contradicted. The speed of light limit applies only to the transmission of information, and a pure continuous carrier contains no information.

Although all GPS receivers must lock onto and track the carrier of the signal to measure pseudoranges, they may fail to measure or record carrier-phase observations for use in navigation or positioning. Some, however, may internally use carrier-phase measurements to smooth — reduce the high frequency noise on — the pseudorange measurements.

GPS carrier phases are also useful for accurately determining the velocity of a platform. The change in the receiver–satellite range over an interval of time divided by the interval — the *range rate* — is a function of the relative velocity of the receiver and the satellite (and their positions). By combining pseudorange and range rate measurements, a GPS receiver can determine both the position and the velocity of the receiver along with the satellite–receiver clock offset and the time rate of change of this offset. Range rates can be estimated by differencing a pair of carrier-phase measurements over a suitable short time interval (say, 0.2–2 seconds). The range rate is proportional to the Doppler shift of the received carrier frequency and hence is sometimes referred to as the Doppler observable. Further details on velocity determination can be gleaned from "Measuring Velocity Using GPS" in the September 1992 issue of *GPS World.*

Incidentally, in comparison with the carrier phase, pseudoranges when measured in units of the wavelengths of the codes (about 300 meters for the C/A-code and 30 meters for the P-code) are sometimes referred to as code phase measurements.

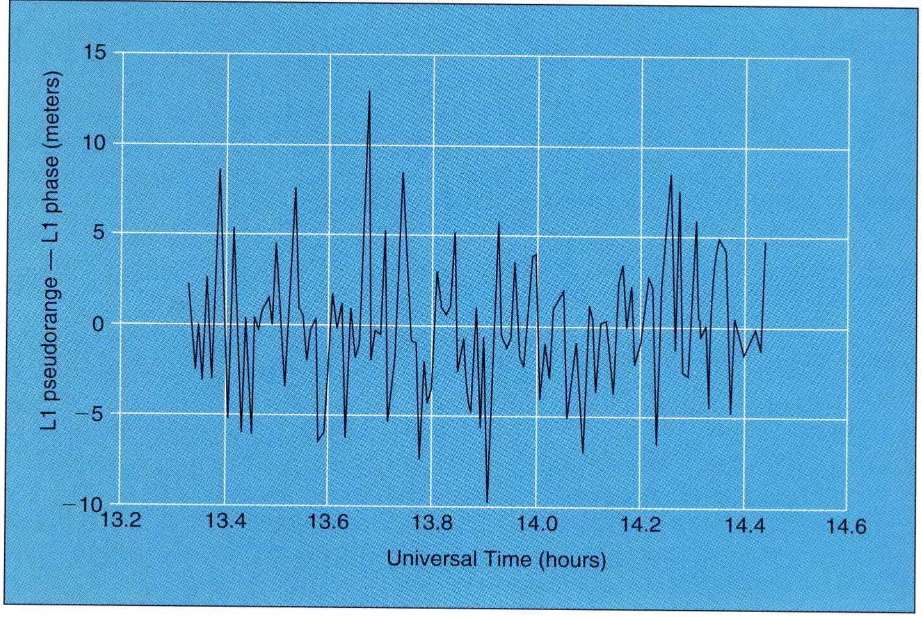

Figure 2. The difference between the L1 pseudorange measurements shown in Figure 1 and the corresponding phase measurements.

POINT POSITIONS

Most civilian receivers intended primarily for navigation exclusively use C/A-code pseudorange measurements to establish the position of a point or the trajectory of a moving platform. The manner in which a GPS receiver solves for its position using pseudoranges was described in "The Mathematics of GPS" in this column in the July/August 1991 issue of *GPS World.* The accuracy of point positions afforded civilians is primarily limited by SA rather than by receiver measurement precision. In fact, before SA was implemented, it was shown that low-cost civilian C/A-code receivers could obtain positions almost as accurate as those provided by P-code military receivers.

RELATIVE POSITIONS

The accuracy of point positions is limited by unmodelable or residual errors in the pseudorange observation equation as well as by measurement noise. To get higher accuracy, we need to take a different approach.

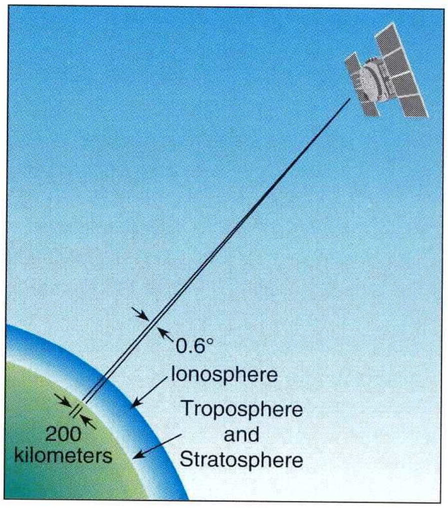

Figure 3. Common-mode error. Distances are approximately to scale.

The pseudoranges and carrier phases measured simultaneously by a pair of GPS receivers tracking a particular satellite will share to a large degree the same satellite ephemeris, satellite clock (including the effects of SA), and atmospheric errors (see Figure 3). The closer together the receivers are, the more similar the errors. If we set up a receiver at a site with a known location, we can monitor the accuracy of the receiver-computed positions. We can attribute any difference between the known and computed positions to pseudorange modeling errors. These errors can be computed and transmitted to another receiver tracking the same set of satellites and used to correct the measured pseudoranges before its coordinates are computed. In this fashion, real-time position accuracies better than 5 meters are possible. The corrections could also be stored in the receiver or an attached computer and applied in postprocessing of the collected data for applications not requiring positions in real time. A similar approach with carrier-phase measurements is possible (difficulties associated with the integer ambiguities notwithstanding), giving real-time position accuracies at the centimeter level.

The Single Differences. Rather than computing and transmitting or storing pseudorange or carrier-phase corrections, we can form what are known as between-receiver single-difference observables — new observables with significantly reduced errors. Although both pseudoranges and carrier phases can be used to form single differences, we will concentrate on the use of the carrier phase. The observation equation for the between-receiver single difference is

$$\Delta\Phi = \Delta\rho - c \times \Delta dT + \lambda \times \Delta N - \Delta d_{ion} + \Delta d_{trop} + \varepsilon_{\Delta\Phi}$$

where Δ denotes the operation of forming differences between receivers. Note that the satellite clock term has disappeared from the equation because the effect of satellite clock errors on the phase measurements of the two receivers is essentially identical. We could use the between-receiver single differences for determining the relative coordinates of the receivers, but in addition to modeling the residual satellite orbit and atmospheric errors and estimating the relative integer ambiguities, we would have to carefully model the relative behavior of the receiver clocks — not an easy task. But, if differencing between receivers is useful in removing satellite clock errors, then it stands to reason that differencing between satellites should remove receiver clock errors. This is just what happens. If we take the carrier phases (or pseudoranges) measured by a single receiver tracking two satellites and difference them, we form the between-satellite single differences:

$$\nabla\Phi = \nabla\rho + c \times \nabla dt + \lambda \times \nabla N - \nabla d_{ion} + \nabla d_{trop} + \varepsilon_{\nabla\Phi}$$

The symbol ∇ denotes the operation of differencing between satellites. Notice that ∇, as an upside down triangle has two vertices on the top (in the sky) whereas Δ had two vertices on the bottom (on the ground).

The Double Difference. The advantages of the between-receiver and between-satellite single differences can be combined by forming the double difference. Consider two receivers tracking two satellites at the same time (see Figure 4). We can form two between-receiver single differences, each involving a different satellite, and two between-satellite single differences, each involving a different receiver. We can difference either the between-receiver or the between-satellite difference pairs to generate a double difference:

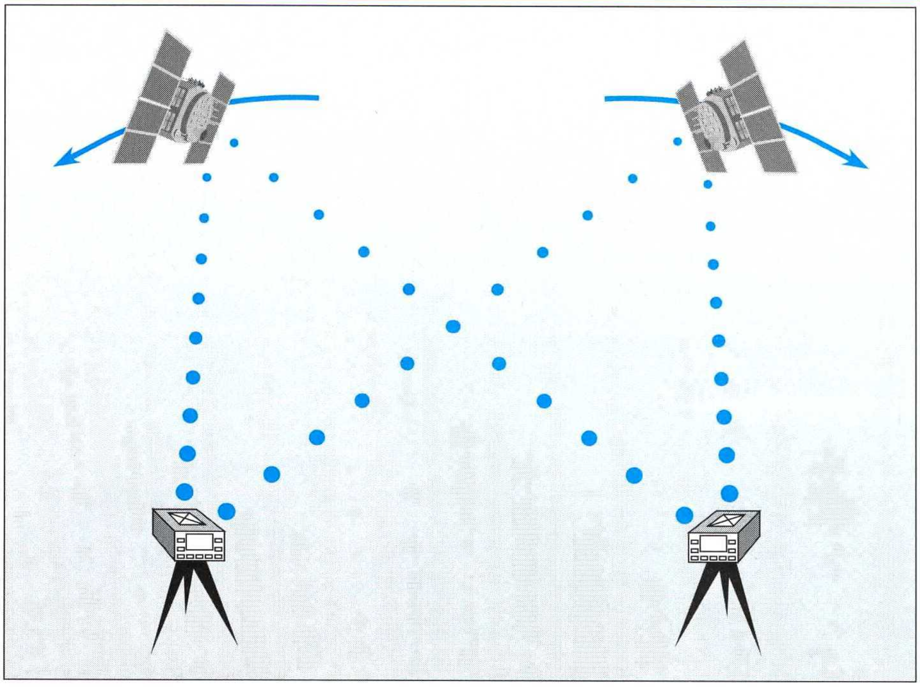

Figure 4. The double difference.

$$\nabla\Delta\Phi = \nabla\Delta\rho + \lambda \times \nabla\Delta N - \nabla\Delta d_{ion} + \nabla\Delta d_{trop} + \varepsilon_{\nabla\Delta\Phi}$$

The great advantage of the double difference is that it removes essentially all satellite and receiver clock errors from the observations. Notice that the ambiguity term in this equation is still an integer number of wavelengths; $\nabla\Delta N$ is simply the difference of integers and is therefore an integer itself. The double-difference observable has become the standard observable in precise differential positioning.

We must estimate the ambiguities for all the satellite pairs forming an observation set in the least-squares adjustment along with the coordinates of one of the receivers (the coordinates of the other receiver are held fixed at a priori known values). Usually these ambiguities are initially estimated as real or floating-point numbers rather than integer numbers. If the data quality is good enough and the model used in the adjustment accurately describes the observations, the real number estimates from the float solution will be very close to integers. The estimates can then be rounded off to the nearest integer and then held fixed at these integer values in a second adjustment of the data. The so-called fixed solution will in general provide more accurate results than those afforded by the float solution.

The Triple Difference. A potential problem with the double-difference observable is associated with the integer ambiguity term. As long as the receivers do not lose carrier lock on the signals, the integer ambiguities remain constant for the whole data set. If, however, one or both receivers loses lock because of a satellite passing behind an obstacle, low signal-to-noise ratio, rapid motion of the receiver, or a severe ionospheric disturbance, then one or more carrier-phase cycles will be lost or slipped. The receiver essentially loses track of the continuous cycle count. This introduces a discontinuity or cycle slip into the data. If any data gap accompanying the cycle slip is short and any noise corrupting the data is minimal, then it is generally possible to determine the correct number of slipped cycles and to correct the phase measurements to produce a continuous phase record. If this cannot be done, then a new set of ambiguities must be adopted for the measurements following the cycle slip.

An observable insensitive to both the initial integer ambiguities and cycle slips is the triple difference. Triple differences are formed by sequentially differencing double differences in time. If we have double differences at epochs one, two, and three, for example, then we can create two triple differences: double difference two minus double difference one, and double difference three minus double difference two. The triple-difference observation equation is written as

$$\delta\nabla\Delta\Phi = \delta\nabla\Delta\rho - \delta\nabla\Delta d_{ion} + \delta\nabla\Delta d_{trop} + \varepsilon_{\delta\nabla\Delta\Phi}$$

where δ is the time-difference operator. The differencing in time results in an observable that has less information content; as a result, receiver coordinates estimated from triple differences tend to be less accurate than those obtained from double differences, especially when the latter are used in fixed ambiguity solutions. Nevertheless, triple-difference results may be accurate enough for certain applications. Also, the triple-difference observable is very useful for obtaining initial estimates of receiver coordinates that can then be used as a priori coordinates in a double-difference solution. Triple differences are also very useful in spotting and correcting cycle slips. In a triple-difference data series, a cycle slip usually appears as an easily identified spike. If the data are particularly noisy — for example, when they are corrupted by severe ionospheric irregularities — identifying and repairing cycle slips can be quite difficult.

OTHER LINEAR COMBINATIONS

The single-, double-, and triple-difference observables are known in mathematical parlance as *linear combinations* of the measured carrier phases (or pseudoranges). Several other linear combinations of the basic GPS observables have also found utility in GPS navigation, positioning, and time transfer. Foremost among these perhaps is the ionosphere-free linear combination of the raw or undifferenced L1 and L2 carrier phases measured by a single receiver. The ionospheric delay term in the carrier-phase observation equation is, with negligible error, inversely proportional to the square of the carrier frequency. Therefore, by combining the equation written for the L2 observation with the equation for the L1 observation, we can create an equation and hence an observable that is essentially free of the ionospheric effect. This observable is sometimes referred to as Lc or L3, although use of the latter term should be discouraged because of possible confusion with the L3 signal associated with the nuclear burst detection package on the GPS satellites. We can remove the ionospheric delay from pseudorange data with a similar operation on the L1 and L2 pseudoranges. Ionosphere-free observations can be combined into single, double, or triple differences and processed in almost the same manner as single-frequency data. We say "almost" because in the ionosphere-free observations, the carrier-phase ambiguities are no longer integers, and ambiguity resolution is more complex than in the single-frequency case. Another distinction is that for baselines shorter than about 20 kilometers or so, ionosphere-free observations are noisier than their single-frequency counterparts. This characteristic arises from the fact that the dominant error source on short baselines is multipath and receiver noise, which in large part is uncorrelated between the L1 and L2 signals. Therefore, for short baselines it may be preferable to use single-frequency observations.

Instead of combining the L1 and L2 phase measurements to remove the ionospheric delay, we can simply difference the L1 and L2 phases measured in distance units and determine changes in the ionospheric delay. Unfortunately, we cannot determine the absolute ionospheric delay from undifferenced phase measurements because of the unknown integer ambiguities. Nevertheless, because orbit, clock, and tropospheric effects are removed, this linear combination of the L1 and L2 phases is particularly useful for locating and potentially removing cycle slips.

Other linear combinations of the L1 and L2 phase measurements have been devised to help with the ambiguity resolution problem. Differencing the L1 and L2 phases measured in cycles results in an observable with an effective wavelength of about 86 centimeters, whereas summing the L1 and L2 phases in cycle units gives an observable with a wavelength of about 10.7 centimeters. Algorithms have been developed to use the so-called wide-lane and narrow-lane ambiguities to help resolve the L1 and L2 ambiguities. Still other linear combinations of both phase and pseudorange observations are being investigated for data cleaning, particularly under ionospherically noisy conditions.

We can also combine carrier-phase and pseudorange measurements. In fact, Figure 2 shows such a combination. Differencing the carrier-phase and C/A-code pseudorange measurements on L1 removes all of the common effects: geometric range, clock terms, and tropospheric delay. What remains is the effect of the ionosphere (doubled because of the sign difference), multipath, receiver noise, and the carrier-phase ambiguity. An estimate of the ambiguity and the constant part of the other effects can be removed by subtracting the arithmetic mean of the carrier phase — pseudorange difference. We are left with an observable usually dominated by the pseudorange multipath and the pseudorange measurement noise. Also remaining is the variation in the ionospheric delay. Over the one-hour period of the measurements shown in Figure 2, the ionospheric delay changed by only about 1 meter in a very smooth fashion. Of course, if we had used dual-frequency data, we could have removed the effect of the ionosphere *ab initio* by using the Lc carrier phase linear combination.

With the advent of receivers providing low-noise pseudorange observations, we can effectively use the pseudoranges to help determine the integer ambiguities of the carrier-phase measurements. This synergistic combination of pseudoranges and carrier phases has spawned the rapid static surveying technique with which we can determine the ambiguities with observation times of mere minutes.

CONCLUSION

In this article we have presented only a brief overview of the various observables used in GPS navigation, positioning, and time transfer. For further details, interested readers can consult one of the available textbooks on GPS, such as *The Guide to GPS Positioning* published by Canadian GPS Associates or *GPS Satellite Surveying,* by Alfred Leick, published by John Wiley and Sons. ■

For reprints (250 minimum), contact Mary Clark, Marketing Services, (503) 343-1200.