The Stochastics of GPS Observables

Christian Tiberius, Niels Jonkman, and Frank Kenselaar

Delft University of Technology

We live in a noisy world. In fact, the laws of physics actually preclude complete silence unless the ambient temperature is absolute zero — the temperature at which molecules have essentially no motion. Consequently, any electrical measurement is affected by noise. Although minimized by GPS receiver designers, noise from a variety of sources both external (picked up by the antenna) and internal (generated within the receiver) contaminates GPS observations. This noise will impact the results we obtain from processing the observations. In this month's column, we investigate possible ways of minimizing this impact by considering the random nature, or stochastics, of GPS noise.

The authors are staff members at Delft University of Technology's Department of Mathematical Geodesy and Positioning, directed by professor Peter Teunissen. Their research concerns precise-positioning modeling and data processing with GPS and GLONASS for surveying, geodesy, and navigation. The main topics currently studied include quality control and integer carrier-phase ambiguity resolution.

"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.

Random noise affects GPS pseudoranges and carrier phases as it does all observations. In distilling the desired information from noise-contaminated measurements, we must account for the noise's random behavior. In processing data, we mathematically formalize a noisy observable as a stochastic quantity, which is just a fancy way of saying the laws of probability govern some component of the obtained observations. (The word *stochastic* comes from the Greek word *stokhazomai*, which means to aim at or guess.) This article concentrates on the stochastic modeling of GPS data and explains the importance of using an adequate stochastic model in data processing. Our primary intent is to foster further investigation and understanding of stochastic modeling as interest in the topic emerges.

We have analyzed data samples collected using dual-frequency, geodetic-quality GPS receivers to demonstrate several stochastic aspects of GPS observables. We consider — as a first step toward developing a better stochastic model — satellite elevation angle dependence, cross correlation between observations, time correlation, and data-noise probability distribution.

It should be noted that we refrain from giving a mathematical description of the full measurement process inside a GPS receiver and deriving the stochastics of the observables by formal propagation of error. Instead, we consider the receiver more or less as a black box providing pseudorange and carrier-phase observations.

MATHEMATICAL BACKGROUND

GPS data processing involves computing unknown parameter values from a set of measurements or observations, preferably incorporating some measures that express the quality of the estimated parameter values. Without a doubt, least-squares estimation constitutes the most popular technique to obtain the desired estimates. One can implement this method in a "batch" algorithm to determine the full vector of unknown parameters in a single mathematical process using all observations simultaneously.

Alternatively, one may employ a Kalman filter, which is a recursive least-squares algorithm. This process allows for time-based sequential estimates derived from observations up to the moment of computation. Consequently, Kalman filtering is particularly suited for (real-time) navigation applications.

Functional Model. If we are to compute parameter values according to the (weighted) least-squares criterion, we must specify a deterministic or functional model that prescribes the relationship between observations (pseudoranges and carrier phases) and the unknown parameters, including coordinates and possibly atmospheric delays, as well as such "nuisance" parameters as clock errors and carrier-phase ambiguities. Although we may not be interested in knowing clock-error or carrier-phase ambiguity values, these parameters must be explicated in the model to obtain accurate values for those parameters we wish to determine.

The functional model may take various forms depending on whether we process pseudoranges or carrier phases that are undifferenced observations or single, double, or triple differences.

Stochastic Model. In addition to the functional model, we need a stochastic model to describe the observations' noise characteristics. The variance–covariance matrix (vc-matrix) serves this purpose. The matrix describes both the precision and correlation of GPS pseudorange and carrier-phase observations. Variances are on the main diagonal (the square root of the variance gives the standard deviation), and the off-diagonal elements are covariances (from which the correlation coefficients are readily obtained). To achieve optimal estimation results, the inverse of the observations' vc-matrix should function as the weight matrix in the least-squares algorithm.

Unfortunately, GPS researchers have yet to concentrate on the stochastic model as much as the functional model. As a result, simple and rudimentary stochastic models are commonly used in practice. If the stochastic model is not completely correct, it will affect, although usually not dramatically, computed parameter estimates such as receiver coordinate values.

As an example, Figure 1 shows the positioning results from a data set processed with two different vc-matrices. We obtained the coordinate values in panel (a) by processing the data with a default (equal-weighting, no-correlation) vc-matrix. Those in (b) were calculated with a "correct" (in other words, more realistic) vc-matrix that accounts for data cross correlation. (We will return to the issue of cross correlation later.) With the correct vc-matrix, the empirical spread in computed positions is clearly smaller than the default vc-matrix, thus demonstrating the role of the stochastic model in parameter estimation.

The stochastic model also affects such important issues as quality description, integrity monitoring (for which a statistical testing procedure is used to spot anomalous data), and ambiguity resolution (for which uncertainty inferences must be made to validate the integer ambiguity values).

Table 1 lists the precisions of the results shown in Figure 1, illustrating the issue of quality description for parameter estimates. The table includes both the empirical standard deviation determined from the scatter in each coordinate and the formal standard deviation resulting from propagation of the assumed observation errors through the data-processing model. With the correct cross-correlation vc-matrix, the agreement between the two value sets is quite good. With the default no-correlation vc-matrix, the formal precision is much too optimistic. In practice, we usually do not know the true position coordinates, so we rely on the formal standard deviation. And if this description is too optimistic, as our example indicates, users may believe their results meet quality requirements when in fact they do not.

Table 1. Position estimator precision in terms of standard deviation, in meters, for north and east coordinates shown in Figure 1

vc-matrix type		empirical	formal
no correlation	N	0.26	0.16
	E	0.21	0.13
cross-correlation	N	0.20	0.23
	E	0.16	0.18

EXPERIMENTS

To more thoroughly investigate the importance of stochastic modeling in GPS data processing, we collected and processed several data sets and examined the impact on the computed results of such factors as satellite elevation angle, observable cross correlation, observation time correlation, and noise probability distribution.

The analysis should be seen in the context of precise relative positioning in, for instance, surveying. By default, we used observations obtained with a 1-second sampling interval in a 10-minute period (which yields 600 epochs). The abbreviations defined in the Receiver-independent Exchange Format (RINEX) designate the observation types:

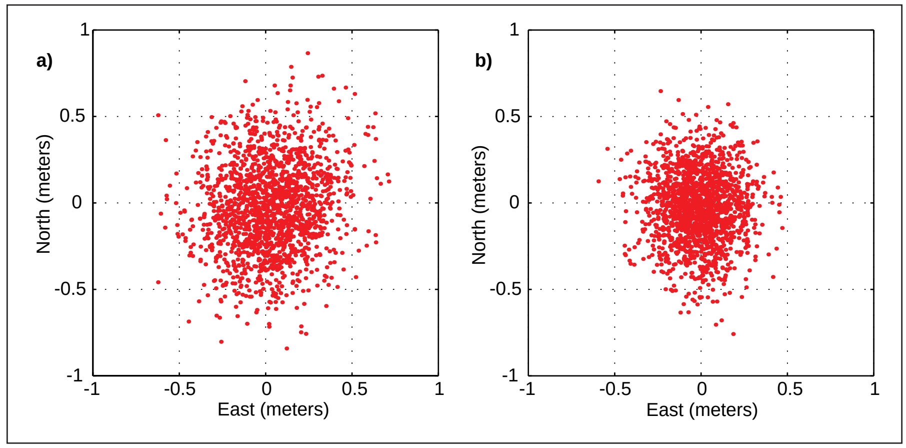

Figure 1. If we ignore stochastic modeling when processing GPS observations, a lower position precision may result. The position scatter from 28 minutes of dual-frequency, pseudorange data collected at 1-second intervals on a short 3-meter baseline is significantly greater when a rudimentary default model is used (a) compared to that obtained using a model that accounts for the correlations between observations (b). The positions of one end of the baseline are given in a local north–east coordinate system, with the origin at the antenna's known position.

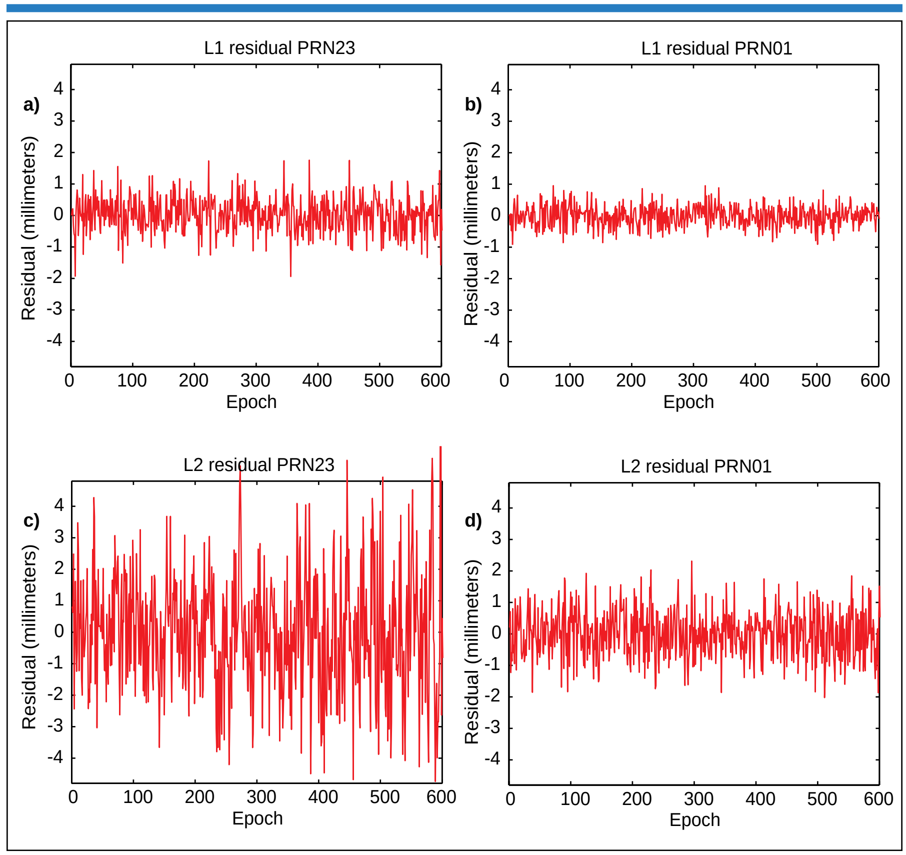

Figure 2. Time series of L1 and L2 least-squares phase residuals for two satellites: (a) L1 residuals for PRN23 at a 17-degree elevation angle; (b) L1 residuals for PRN01 at a 65-degree elevation angle; (c) L2 residuals for PRN23; (d) L2 residuals for PRN01.

■ C1 — pseudorange by C/A-code correlation (in meters);

■ P1, P2 — pseudorange on L1 and L2 respectively (in meters), obtained by a codeless or semicodeless technique to circumvent antispoofing (AS); and

■ L1, L2 — carrier phase on L1 and L2 (in millimeters).

Elevation Angle. We first concentrated on the noise of the observables from a single receiver channel. Figure 2 shows a time-series of least-squares residuals derived from observations of two satellites — one at a low elevation angle (a and c) and one at a high elevation angle (b and d). The least-squares residuals should have a mean of zero, with a random behavior resulting from propagating the observable's noise through the analysis algorithm.

In processing GPS data, we might be inclined to use a constant standard deviation for all receiver channels, assuming equally noisy observations from all satellites and therefore weighting all data the same. Figure 2, however, clearly demonstrates that observations from a low-elevation satellite are subject to substantially more noise.

Figure 2 also shows a distinct difference in noise between the L1 (a and b) and L2 (c and d) phase observations from the same receiver. The use of a codeless or semicodeless measurement technique in making L2 signal observations in the presence of AS causes most of the larger noise in the L2 phase observations.

The observation noise's dependence on elevation angle can be distinctly different for different makes of receivers as well. For example, we estimated the standard deviation of the residuals on a satellite-by-satellite basis using two different receivers, each tracking the same six satellites. Figure 3 shows the results. One receiver (a) shows practically no elevation-angle dependence for L1 observations but a strong dependence for L2. In comparison, the second receiver (b) indicates an elevation-angle dependence for both L1 and L2 observations, with less noise in low–elevation angle, L2 observations than those from receiver (a).

The elevation-angle dependence of a measurement's noise is induced mainly by the receiver antenna's gain pattern, with other factors such as atmospheric signal attenuation contributing to a lesser degree. (We'll discuss multipath and scattering in conjunction with elevation angle–dependent effects later.)

The results shown in Figure 3 also suggest a direct relationship between elevation angle and measurement precision. Some data-processing computer programs use mathematical functions, such as the sine of the elevation angle, to weight observations accordingly. Alternatively, because observation noise directly relates to the receiver's measured signal-to-noise ratio (SNR), we may use SNR values in constructing a stochastic model.

Cross Correlation. Looking at the correlation between different observation types, we found that a significant positive correlation may exist between observations from the

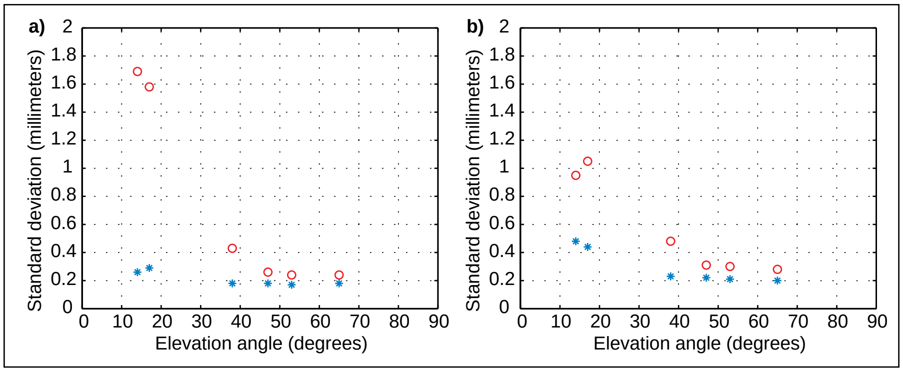

Figure 3. The standard deviations of undifferenced carrier phases depend on elevation angle, carrier frequency, and receiver model. These plots show the L1 (*) and L2 (o) standard deviations from two different receivers.

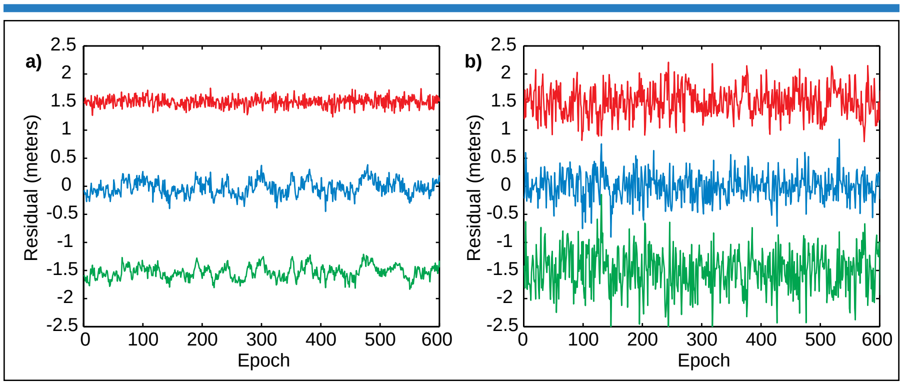

Figure 4. For some receivers, the noise on the L2-frequency observables is correlated with the noise on the L1-frequency observables. Shown here are time series of pseudorange least-squares residuals from receiver (a) and receiver (b) for satellite PRN15 at a 38-degree elevation angle. The three series in each graph are C1 (intentionally offset by 1.5 meters), P2, and P2–C1 (offset by –1.5 meters). The empirical standard deviations for C1, P2, and P2–C1 are 0.09, 0.14, and 0.12 meter for receiver (a) and 0.27, 0.24, and 0.36 meter for receiver (b).

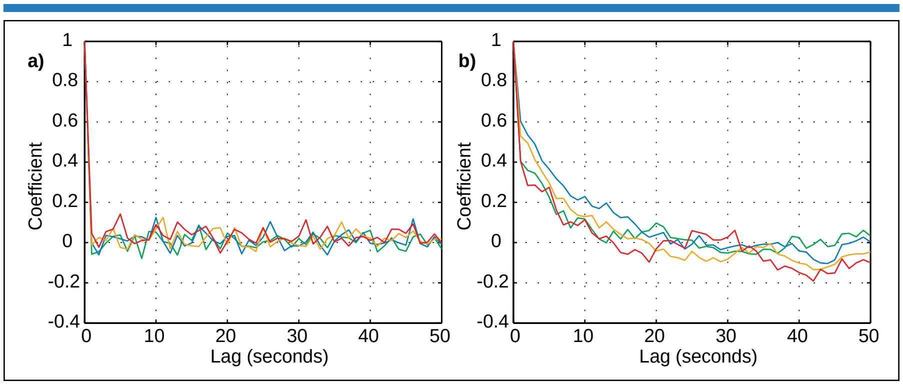

Figure 5. The observations from a particular receiver may exhibit temporal correlation, which may vary between different observables. This figure displays the autocorrelation coefficients for C1 (a) and P2 (b) pseudoranges as derived from the same receiver for four different satellites.

same receiver at the two carrier frequencies. This implies that, although the receiver outputs two apparently distinct observables, the actual "information content" provided by the additional second observable is less than that supplied by the first alone.

Figure 4 illustrates this effect by showing several time series of pseudorange least-squares residuals from two different receivers (a and b) observing the same satellite simultaneously. The top trace shows C1 (a) or P1 (b) residuals and the middle trace portrays the P2 residuals. The plots give an impression of the noise characteristics for each receiver's two pseudorange observables.

In addition, the bottom trace in each panel illustrates the algebraic difference between the two pseudorange time series. Following the error propagation law, the variance (noise) of the difference should be the sum of C1 or P1 and P2 variances in the absence of correlation between the two pseudorange observables. Thus, the noise in the pseudorange difference should be larger than the noise in either of the two pseudoranges. Figure 4a indicates that this is not the case for the first receiver.

For our example, we expect a standard deviation of 0.17 meter for the difference, whereas 0.12 meter is actually observed. The difference noise is even somewhat smaller than the P2 series noise, indicating the presence of a positive correlation between the two pseudorange observables. A different measurement technique for the P2 pseudorange, implemented in the second receiver (Figure 4b), appears to produce uncorrelated P1 and P2 pseudoranges. Applying the error propagation law to the variances yields a 0.36-meter standard deviation for the difference, as is indeed observed.

The effect illustrated in Figure 4a occurs in receivers that use cross-correlation measurement, a so-called codeless technique, for obtaining an L2 pseudorange measurement in the presence of AS, P-code encryption. Such a receiver constructs the L2 pseudorange using the C1 pseudorange, thereby inducing a strong correlation between the two observables. Figure 1 has already shown the consequences of neglecting such correlation.

A significant correlation is similarly present between the first receiver's L1 and L2 carrier-phase observations. Again, we can explain this correspondence as a result of the cross-correlation measurement technique, although the L2 phase construction is slightly different.

Interestingly, we found a positive correlation between the L1 and L2 carrier phases for all the receivers we analyzed, including those

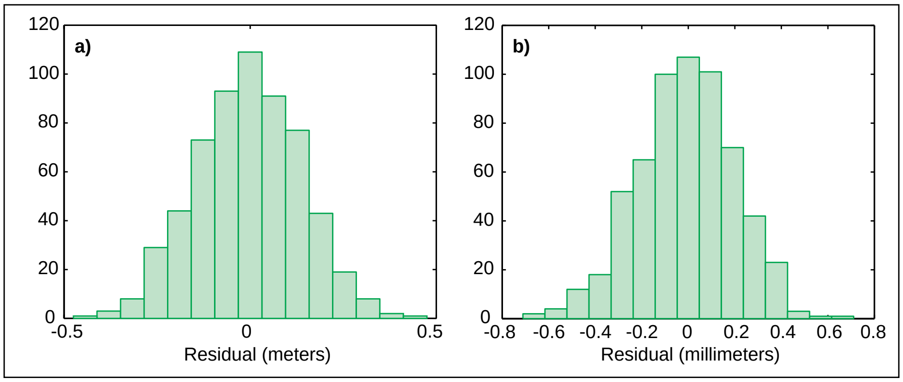

Figure 6. Histograms of the P1 (a) and L1 (b) least-squares residuals for satellite PRN21 (at a 53-degree elevation angle) indicate that the observation noise is essentially distributed normally.

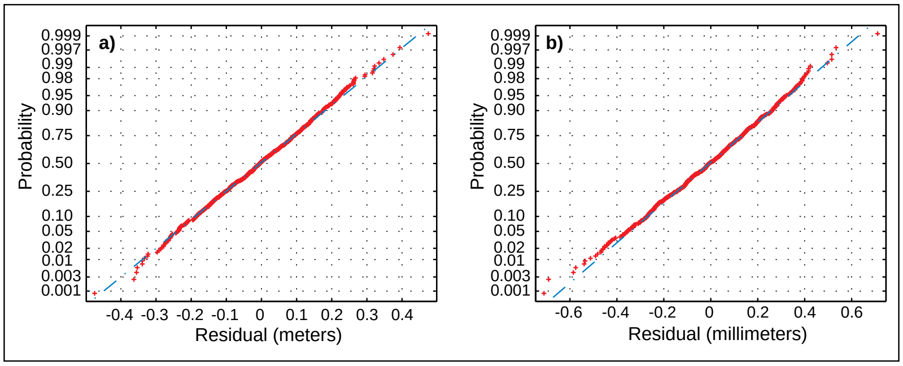

Figure 7. Normal probability plots of the least-squares residuals for those histograms shown in Figure 6 further support the conclusion that P1 (a) and L1 (b) observation noise closely follows a normal distribution.

that do not use cross correlation for L2 pseudorange construction. Numerical values for the different receivers' correlation coefficients range from 0.3 to 0.7.

Correlation between a pseudorange and carrier-phase observable turned out to be negligible. Mutual correlation between channels (that is, between observations to different satellites for one observable type) may exist. Such correlation is often referred to as spatial correlation. And, although several researchers have identified it, how to effectively account for spatial correlation remains unclear.

Time Correlation. To simplify processing, most GPS users assume that observations possess only white noise — that they are not correlated from epoch to epoch. To verify this assumption, we examined time-series correlograms of least-squares residuals. A correlogram portrays the autocorrelation coefficient versus lag (the interval between two samples). The coefficient at lag zero equals one by definition. If a white noise process can describe the residuals, then all other coefficients should be approximately zero because a residual at one epoch is not

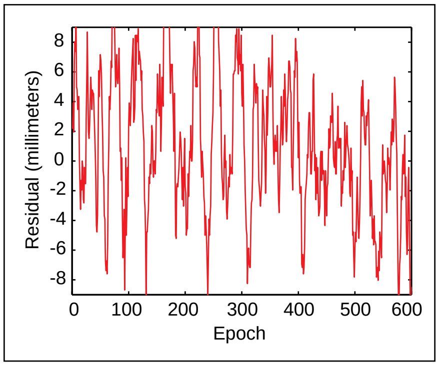

Figure 8. Multipath is evident in these L2, least-squares, phase residuals from PRN29 short-baseline observations at a 17-degree elevation angle.

correlated with a residual at any other epoch. (There will be some small variability because the time series is of finite length.)

As shown in Figure 5, we computed correlograms on a satellite-by-satellite basis for C1 and P2 pseudorange samples. The correlograms shown in Figure 5a indicate that the C1 pseudorange noise is more or less consistent with the assumption of white noise, but this is certainly not the case for the P2 correlograms (Figure 5b). For the P2 observable, the autocorrelation coefficient approaches zero after approximately 10 seconds. Because AS encryption causes quite noisy "raw" P2 pseudoranges, the receiver probably applies some smoothing or filtering that induces a slight time correlation.

We could already see this correlation in the time series of Figure 4a. For this receiver, the C1 pseudorange can thus be modeled as a white noise random variable, but the P2 pseudorange cannot; modeling this P2 pseudorange as white noise would require increasing the sampling rate interval to 15 or 30 seconds. For this receiver's corresponding L1 and L2 carrier phases, the findings are similar. With regard to time correlation, though, various receivers may possess distinctly different characteristics.

Probability Distribution. The least-squares criterion typically used for parameter estimation does not require any assumption about the observables' noise distribution type. Yet, subsequent statistical testing procedures for quality control, such as those used to assess outliers and cycle slips, will necessitate this information. For those procedures, a normal or Gaussian distribution is commonly presumed.

We examined a distribution set of 600 data residuals to test this supposition's validity. We grouped the residuals into intervals, or bins, counting the number of residuals in each bin to create histograms. Figure 6a shows the histogram of the P1 residuals, and Figure 6b displays the histogram for the L1 residuals. Visual inspection of the histograms indicates that the residuals appear consistent with the assumption of normality.

The normal probability plot of Figure 7 shows the cumulative distribution of the residuals. The horizontal axis is linear in the residual value. Note that the vertical axis scale indicating probability is not uniform, as the tickmarks correspond to quantiles of a normal distribution. Sorted in ascending order, normal distribution samples will tend to form a straight line and pass through value 0.0 (indicating an unbiased or zero-mean sample set) at probability 0.50 (indicating a symmetrical distribution).

The data displayed in Figure 7 nicely resemble these normal distribution characteristics. For the most part, deviations at the beginning and end of the line are relatively insignificant as they represent only a few tenths of a percent of the total data set. The standard deviations of the least-squares residuals are 0.14 meter (P1) and 0.21 millimeter (L1). The maximum deviation between the empirical and theoretical cumulative distribution is small for both examples, indicating with a high degree of probability that the assumed theoretical, normal distribution is likely correct.

We initially conclude that the normal distribution is a reasonable model for the observables of several geodetic-quality GPS receivers; however, this may not hold true for all receivers. We recommend further research to verify this conclusion.

Further Considerations. We would like to point out that for our analyses we used (static) zero-baseline data to better assess *receiver* noise (antenna preamplifier noise, for example, was not considered). In other words, we were interested in "nominal" correlation and noise. Such a zero-baseline arrangement does not necessarily allow a full valuation of a receiver's *practical* performance. For a zero (distance) baseline, differential atmospheric delays are completely absent. In practice, environmental effects and especially multipath (and the related phenomenon of scattering) may also affect observations.

Of these two effects, multipath is by far the most difficult to incorporate in an observation model. Figure 8 illustrates the extent to which multipath may prohibit a proper match between actual data and a theoretical model. The figure displays the L2 phase residuals for satellite PRN29 as it rises over the horizon. The residuals' large magnitude and periodicity clearly indicate the occurrence of (phase) multipath. Other satellites on this short (nonzero) baseline show more typical noise characteristics at only the 1–2-millimeter level. Therefore, multipath may significantly increase the observation noise level and induce time correlation between observations, as well as introduce a bias that can propagate into parameter estimates.

CONCLUDING REMARKS

To obtain optimal results, it seems obvious that any observations' vc-matrix should adequately reflect the noise characteristics of the GPS data being processed. Some GPS users, however, commonly use a rather simple and rudimentary vc-matrix or even a scaled-identity matrix. In this article, we attempted to demonstrate the need for a more sophisticated stochastic model.

Nonetheless, from the receiver side, we think it should be possible (technically) to deliver white noise observables even at a 1-second sampling interval. In addition, we are not yet concerned with the probability distribution type for GPS data. Most observable noise seems more or less normally distributed.

We have some reason for concern, though, regarding measurement precision dependence on satellite elevation angle and the cross correlation between observation types. We feel that in developing a new stochastic model it is most important to account for both these aspects. ■

MANUFACTURERS

The authors used data from **Trimble** (Sunnyvale, California) 4000SSi receivers, Z-XII-3 receivers from **Magellan Corporation, Ashtech Precision Products** (Sunnyvale, California), and SR399 units manufactured by **Leica Geosystems Inc., GPS** (Torrance, California).

FURTHER READING

For an overview of GPS noise sources, see

■ "GPS Receiver System Noise," by R.B. Langley, in *GPS World,* Vol. 8, No. 6, June 1997, pp. 40–45.

For a discussion about correlograms and other aspects of time-series analysis, see

■ *The Analysis of Time Series: An Introduction,* 4th edition, by C. Chatfield, published by Chapman & Hall, London, 1989.

For RINEX format details, see

■ "RINEX: Receiver-independent Exchange Format," by W. Gurtner, in *GPS World,* Vol. 5, No. 7, July 1994, pp. 48–52.

■ <ftp://igscb.jpl.nasa.gov/igscb/data/format/rinex2.txt>

For an introduction to Kalman filtering, see

■ "The Kalman Filter: Navigation's Integration Workhorse," by L.J. Levy, in *GPS World,* Vol. 8, No. 9, September 1997, pp. 65–71.

For a discussion about the importance of stochastic modeling in ambiguity resolution, see

■ "Weighting GPS Dual-Frequency Observations: Bearing the Cross of Cross-correlation," by P.J.G. Teunissen, N.F. Jonkman, and C.C.J.M. Tiberius, in *GPS Solutions,* Vol. 2, No. 2, 1998, pp. 28–37.

For related work regarding GPS stochastic modeling, see

■ "Stochastic Modelling for Very High Precision Real-time Kinematic GPS in an Engineering Environment," by J.B. Barnes, N. Ackroyd, and P.A. Cross, published in the *Proceedings of the Fédération Internationale des Géomètres (FIG) XXI International Congress,* Brighton, England, July 19–25, 1998, Commission 6, Engineering Surveys, pp. 61–76.

■ "Stochastic Modeling for Static GPS Baseline Data Processing," by J. Wang, M.P. Stewart, and M. Tsakiri, in *Journal of Surveying Engineering,* Vol. 124, No, 4, November 1998, pp. 171–181.