Dilution of Precision

Richard B. Langley

University of New Brunswick

Dilution of precision, or DOP: we've all seen the term, and most of us know that smaller DOP values are better than larger ones. Many of us also know that DOP comes in various flavors, including geometrical (GDOP), positional (PDOP), horizontal (HDOP), vertical (VDOP), and time (TDOP). But just what are these DOPs? In this month's column, we examine GPS dilution of precision and how it affects the accuracy with which our receivers can determine position and time.

"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.

CORRECTION

The time scales of Figs. 3 and 5 in the original article were incorrect. They should run from 1200 to 1200 UTC and have been corrected in this version. – R.B.L.

How accurate is GPS? This is a question that almost every newcomer to GPS asks. And the answer? It depends. It depends on whether we are talking about standalone (single receiver) or differential positioning, single- or dual-frequency receivers, real-time or postprocessed operation, and so on. Even if we confine ourselves to the Standard Positioning Service (SPS), the official, standalone service the United States government provides to all users worldwide, the answer is still — it depends.

The specified SPS accuracy is given in terms of "minimum performance levels;" that is, accuracy will be no worse than a certain level for a certain percentage of time. For any point on the globe, the horizontal accuracy is equal to or better than 100 meters based on the twice-distance–root-mean-square error measure. This means that over a 24-hour period, the horizontal coordinates of a position determined by GPS will be within 100 meters of the true position about 95 percent of the time. The corresponding specified accuracy for heights is 156 meters and 340 nanoseconds for time transfer.

These predicted accuracies are predicated on a 24-satellite constellation (additional satellites are a bonus), a 5-degree satellite elevation mask angle with no obstructions, and at least four satellites in view with a position dilution of precision (PDOP) of six or lower. So, even the basic SPS accuracy is qualified. This means that depending on where we are and the time of day, actual SPS accuracy will vary. In urban canyons, we may in fact not even have four satellites in view, and if we do, the PDOP may be greater than six.

The variability of actual SPS accuracy from place to place and time to time is dominated by the effects of dilution of precision, a geometric factor that, when multiplied by measurement and other input errors, gives the error in position, some component of position, or time. Before we examine how observation geometry affects GPS, let's look at a simple, non-GPS example.

GEOMETRY: A SIMPLE EXAMPLE

Imagine a radio positioning system in which a receiver measures the ranges to two terrestrial transmitters to determine its horizontal coordinates. The receiver lies at the intersection of the circular lines of position that are centered on the transmitters (see Figure 1). There is some uncertainty, however, in the receiver's measurements, and so the location of the range circles will be inexact and result in an error in the computed position. This error depends on the geometry relating the receiver and the transmitters.

In Figure 1a, the transmitters are far apart, giving a relatively small region in which the receiver must lie with some degree of certainty. Transmitter 1 lies in a direction orthogonal to that of transmitter 2, so the receiver's X and Y coordinates are deter-

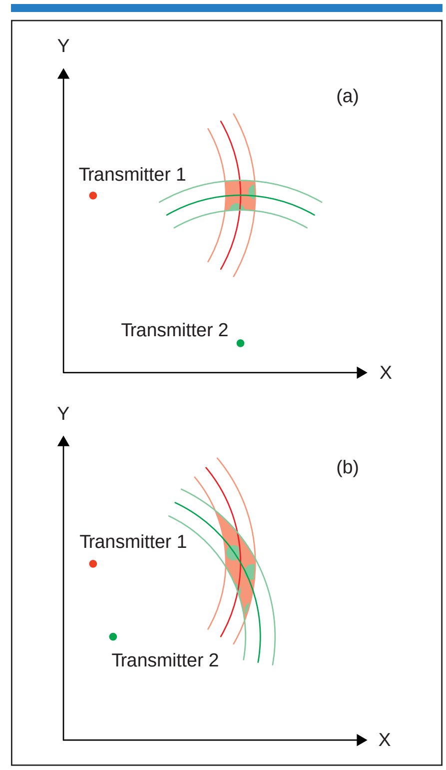

Figure 1. In any ranging system, receiver–transmitter geometry influences position precision. In this figure, the uncertainty in the receiver's position is indicated by the patterned areas. In (a), the position uncertainty is small (low dilution of precision). In (b), transmitter 2 is moved closer to transmitter 1, and, although the measurement uncertainty is the same, the position uncertainty is considerably larger (high dilution of precision).

mined with equal precision. In panel (b), the transmitters are closer together resulting in a considerably larger uncertainty region, with the confidence in the Y coordinate being smaller than the X coordinate. We say that the precision in case (b) is diluted in comparison to that of (a).

Although fictitious, this simple example is not too far removed from the case of Loran-C radionavigation (although in Loran-C, because we typically measure range *differences*, the lines of position are usually hyperbolas, not circles). In fact, the concept of dilution of precision originated with Loran-C users.

With this simple analogy under our collective belt, we can now examine the effect of geometry on GPS accuracy. First, though, let's quickly review the basics of GPS positioning using pseudoranges.

PSEUDORANGE MEASUREMENTS

A GPS receiver computes its three-dimensional coordinates and its clock offset from four or more simultaneous pseudorange measurements**.** These are measurements of the biased range (hence the term *pseudo*range) between the receiver's antenna and the antennas of each of the satellites being tracked. This is derived by cross-correlating the pseudorandom noise code received from a satellite with a replica generated in the receiver. The accuracy of the measured pseudoranges and the fidelity of the model used to process those measurements determine, in part, the overall accuracy of the receiver-derived coordinates.

The basic pseudorange model is given by

$$P = \rho + c(dT - dt) + d_{ion} + d_{trop} + e \qquad [1]$$

in which P denotes the pseudorange measurement; ρ is the geometric range between the receiver's antenna at signal reception time and the satellite's antenna at signal transmission time; dT and dt represent receiver and satellite clock offsets from GPS Time, respectively; d_{ion} and d_{trop} are the ionospheric and tropospheric propagation delays; e accounts for measurement noise as well as unmodeled effects such as multipath; and c stands for the vacuum speed of light.

Assuming the receiver accounts for the satellite clock offset (using the navigation message) and atmospheric delays (from models programmed into its firmware), we can simplify the pseudorange model as follows:

$$P_c = \rho + c \cdot dT + e_c \qquad [2]$$

In this equation, e_c represents the original measurement noise plus model errors and any unmodeled effects (such as selective availability[SA]). There are n such equations that a receiver must solve using the n-simultaneous measurements.

The parameter ρ is a nonlinear function of the receiver and satellite coordinates. To determine the receiver coordinates, we can linearize the pseudorange equations using some initial estimates or guesses for the receiver's position (the linearization point). We can then determine corrections to these initial estimates to obtain the receiver's actual coordinates and clock offset. Grouping our equations together and representing them in matrix form, our model is now

$$\mathbf{\Delta P}_c = \mathbf{A}\,\mathbf{\Delta x} + \mathbf{e}_c \qquad [3]$$

in which $\boldsymbol{\Delta P_c}$ is the n-length vector of differences between the corrected pseudorange measurements and modeled pseudorange values based on the linearization point coordinates; $\boldsymbol{\Delta x}$ designates the four-element vector of unknowns — the receiver position and clock offset (in distance units) — from the linearization point; $\boldsymbol{A}$ is the $n \times 4$ matrix of the partial derivatives of the pseudoranges with respect to the unknowns; and $\boldsymbol{e}_c$ is the n-length vector of measurement and other errors. The first three columns of the $\boldsymbol{A}$ matrix are simply the components of the unit vectors pointing from the linearization point to the satellites; the fourth column is all ones.

The receiver (or postprocessing software) solves the matrix equation using least squares. (The receiver might use a Kalman filter, which is a more general form of conventional least squares.) Equation 4 gives this solution:

$$\mathbf{\Delta x} = -\left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\mathbf{\Delta P}_c \qquad [4]$$

A weight matrix ($\boldsymbol{W}$) characterizes the differences in the errors of the simultaneous measurements as well as any correlations that may exist among them. This weight matrix is also equal to $\sigma_0^2\mathbf{C}^{-1}_{\Delta \mathbf{P}_C}$, in which $\boldsymbol{C}_{\Delta P_C}$ is the covariance matrix of the pseudorange errors and σ_0^2 is a scale factor known as the a priori variance of unit weight. In general, the solution of a nonlinear problem must be iterated to obtain the result. However, if the linearization point is sufficiently close to the true solution, then only one iteration is required.

The Covariance Matrix. So how accurate are the receiver's coordinates and clock offset from such a solution? What we are actually asking is how do the pseudorange measurement and model errors affect the estimated parameters obtained from Equation 4? This is given by the law of propagation of error — also known as the covariance law:

$$\mathbf{C}_{\Delta x} = \left[\left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\right]\mathbf{C}_{\Delta P_c}\left[\left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\right]^T = \left(\mathbf{A}^T\mathbf{C}^{-1}_{\Delta P_c}\mathbf{A}\right)^{-1} \qquad [5]$$

in which $\boldsymbol{C}_{\Delta x}$ is the covariance matrix of the parameter estimates.

Equation 5 represents a fundamental relationship widely used in science and engineering not only for actual measurement analysis but also for experiment and system design studies. It allows a scientist or engineer to examine the effect a particular design or measurement capability will have on specified parameters without actually making any measurements.

In GPS-related studies, for example, we might use the equation to answer a variety of questions: What is the behavior of the estimated parameter covariance matrix as a function of particular satellite configurations? How do various model errors propagate into the receiver coordinates as a function of satellite configurations? What is the tolerance value that a particular model error should not exceed to achieve a desired positioning accuracy? Such questions are not limited to the analysis of pseudoranges; they can also be asked about more precise carrier-phase measurements and differenced observables.

In Equation 5, if we assume that the measurement and model errors are the same for all observations with a particular standard deviation (σ) and that they are uncorrelated, then $\boldsymbol{C}_{\Delta P_C}$ is $\mathbf{I}\sigma^2$ (in which $\boldsymbol{I}$ is the identity matrix). The expression for the covariance of $\boldsymbol{\Delta x}$ thus simplifies to

$$\mathbf{C}_{\Delta x} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\sigma^2 = \mathbf{D}\sigma^2 \qquad [6]$$

(A more accurate error analysis can be carried out using slightly more realistic nonhomogeneous variances and nonzero correlations, but for pedagogical as well as planning and assessment purposes, the equal-variance, zero-correlation assumption is usually more than adequate.)

Because the least-squares estimates of the parameter offsets are simply added to the linearization point values — a linear operation — the parameter estimates and the corrections have the same covariance. The diagonal elements of $\boldsymbol{C}_{\Delta x}$ are the estimated receiver-coordinate and clock-offset variances, and the off-diagonal elements (the covariances) indicate the degree to which these estimates are correlated.

UERE. As we mentioned, σ represents the standard deviation of the pseudorange measurement error plus the residual model error, which we've assumed to be equal for all simultaneous observations. If we further assume that the measurement error and the model error components are all independent, then we can simply root-sum-square these errors to obtain a value for σ. When we combine receiver noise, satellite clock and ephemeris error, atmospheric error, multipath, and SA — all expressed in units of distance — we obtain a quantity known as the total user equivalent range error (UERE), which we can use for σ.

For SPS, the total UERE is typically in the neighborhood of 25 meters. When SA is turned off, total UERE could be less than 5 meters, with the actual value dominated by ionospheric and multipath effects. Dual-frequency Precise Positioning Service users, with the capability to remove almost all of the ionospheric delay from the pseudorange observations, can experience even smaller UEREs. Future users of the proposed new civilian GPS signals will likewise be able to compensate for ionospheric effects and achieve superior UEREs.

THE DOPS

With a value for σ, we can compute the components of $C_{\Delta x}$ using Equation 6. We then can get a measure of the overall quality of the least-squares solution by taking the square root of the sum of the parameter estimate variances:

$$\sigma_G = \sqrt{\sigma_E^2 + \sigma_N^2 + \sigma_U^2 + \sigma_T^2} = \sqrt{D_{11} + D_{22} + D_{33} + D_{44}}\,\sigma \qquad [7]$$

in which σ_E^2, σ_N^2, and σ_U^2 are the variances of the east, north, and up components of the receiver position estimate, and σ_T^2 is the variance of the receiver clock offset estimate. If the solution algorithm is parameterized in terms of geocentric Cartesian coordinates, it is a straightforward procedure to transform the solution covariance matrix to the local coordinate frame. This estimate of solution accuracy — the square root of the trace of the solution covariance matrix — is equal to the pseudorange measurement and modeling error standard deviation (σ) multiplied by a scaling factor equal to the square root of the trace of matrix $\boldsymbol{D}$. The elements of matrix $\boldsymbol{D}$ are a function of the receiver–satellite geometry only. And because the scaling factor is typically greater than one, it amplifies the pseudorange error, or dilutes the precision, of the position determination. This scaling factor is therefore usually called the geometric dilution of precision (GDOP).

Rather than examining the quality of the overall solution, we may prefer to look at specific components such as the three-dimensional receiver position coordinates, the horizontal coordinates, the vertical coordinate, or the clock offset. To do this, we simply take or combine the appropriate $C_{\Delta x}$ variances:

$$\sigma_P = \sqrt{\sigma_E^2 + \sigma_N^2 + \sigma_U^2}$$

$$\sigma_H = \sqrt{\sigma_E^2 + \sigma_N^2}$$

$$\sigma_U = \sqrt{\sigma_U^2}$$

$$\sigma_T = \sqrt{\sigma_T^2} \qquad [8]$$

For each of these error measures, we can determine the corresponding position, horizontal, vertical, and time DOPs:

$$\mathrm{PDOP} = \frac{\sqrt{\sigma_E^2 + \sigma_N^2 + \sigma_U^2}}{\sigma} = \sqrt{D_{11} + D_{22} + D_{33}}$$

$$\mathrm{HDOP} = \frac{\sqrt{\sigma_E^2 + \sigma_N^2}}{\sigma} = \sqrt{D_{11} + D_{22}}$$

$$\mathrm{VDOP} = \frac{\sigma_U}{\sigma} = \sqrt{D_{33}}$$

$$\mathrm{TDOP} = \frac{\sigma_T}{\sigma} = \sqrt{D_{44}} \qquad [9]$$

Note that $\mathrm{PDOP}^2 = \mathrm{HDOP}^2 + \mathrm{VDOP}^2$, and $\mathrm{GDOP}^2 = \mathrm{PDOP}^2 + \mathrm{TDOP}^2$. These relationships may be useful for interrelating the various DOP values. Because the various DOPs are functions only of receiver and satellite coordinates, they may be predicted ahead of time for any given set of satellites in view from a specified location using a satellite almanac.

If the tips of the receiver–satellite unit vectors lie in a plane, the DOP factors are infinitely large. In fact, no position solution is possible with this receiver–satellite geometry, as the matrix $\boldsymbol{A}^T\boldsymbol{A}$ (see Equation 6) is singular: The solution cannot distinguish between an error in the receiver clock and an error in the position of the receiver. DOP values are smaller and hence solution errors are smaller when the satellites used in computing the solution are spread out in the sky.

We can most easily visualize the dependence of solution error on receiver–satellite geometry if we assume a receiver is observing only four satellites. This scenario has no measurement redundancy and makes possible a direct solution of the linearized observation equations (as long as $\boldsymbol{A}$ is not singular). However, the covariance of the solution, again assuming equal uncorrelated errors, has the same form as that of the least-squares solution given in Equation 6.

The Tetrahedron. The tips of the four receiver–satellite unit vectors form a tetrahedron (see Figure 2). The volume of this geometrical figure is related to the DOP values. The larger the tetrahedron's volume, the smaller the DOPs. The largest possible tetrahedron is one for which one satellite is at the zenith and three satellites are below the earth's horizon at an elevation angle of –19.47 degrees and equally spaced in azimuth: GDOP works out to be 1.581. Of course, a GPS receiver on or near the earth's surface cannot see the three below-horizon satellites, so in this case, the lowest possible GDOP (1.732) is obtained with one satellite at the zenith and three satellites equally-spaced on the horizon. Some early GPS receivers could only track a maximum of four satellites simultaneously. Such receivers made use of a satellite selection algorithm to choose the best four satellites of those visible — the four that would produce the lowest DOP values.

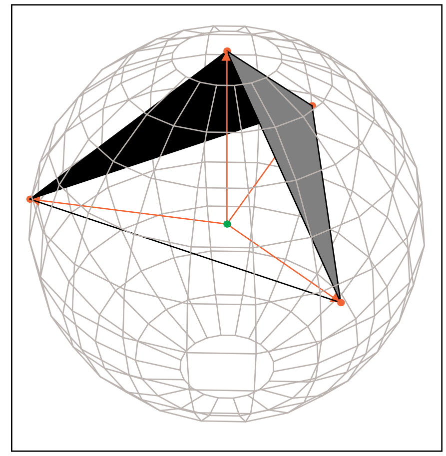

Figure 2. If only four GPS satellites are observed, the tips of the receiver-satellite unit vectors form a tetrahedron circumscribed by a unit sphere. Two faces of the tetrahedron — formed by one satellite at the zenith and three at a 10-degree elevation angle, equally spaced in azimuth — are shaded in this figure. The tetrahedron's volume is highly correlated with GDOP. Maximizing the volume tends to minimize GDOP.

HDOP versus VDOP. Generally, the more satellites used in the solution, the smaller the DOP values and hence the smaller the solution error. Figure 3 shows the DOP values computed for the current GPS constellation viewed from Fredericton, New Brunswick, Canada, with an elevation mask angle of 15 degrees. HDOP values are typically between one and two. VDOP values are larger than the HDOP values indicating that vertical position errors are larger than horizontal errors. We suffer this effect because all of the satellites from which we obtain signals are above the receiver. The horizontal coordinates do not suffer a similar fate as we usually receive signals from all sides.

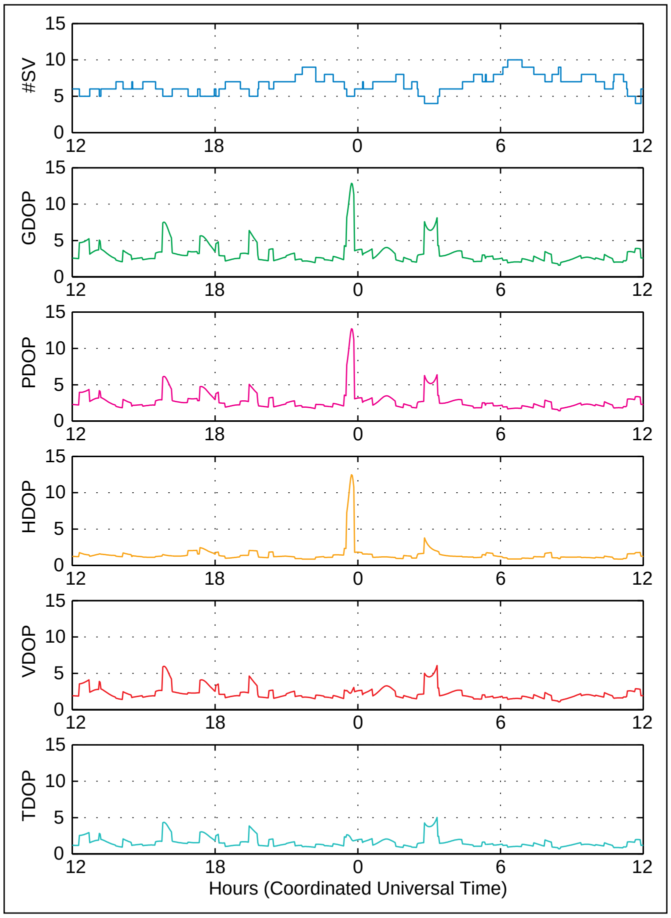

Figure 3. DOP values at Fredericton, New Brunswick, Canada, — computed from broadcast satellite ephemerides using a 15-degree elevation mask angle — are sufficiently small except just before 0000 hours when HDOP, PDOP, and GDOP increase to more than 12 (#SV represents the number of visible satellites [space vehicles]).

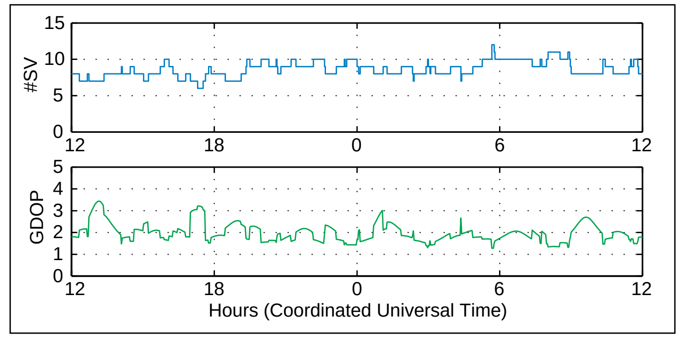

Figure 5. Lowering the elevation mask angle to 5 degrees removes the spikes just before 0000 hours in the DOPs of Figure 3. Now, the GDOP has an average value of about two and never exceeds 3.5 (#SV represents the number of visible satellites [space vehicles]).

If the Earth were transparent to radio waves, we would be able to determine vertical coordinates with about the same accuracy as horizontal coordinates. More realistically, we can also get improved vertical coordinates if we have an accurate receiver clock or one whose offset from GPS Time can be accurately determined so that the receiver only needs to estimate its position. For example, let's assume we are observing one satellite at the zenith and three equally spaced in azimuth at a 15-degree elevation angle. If we estimate the receiver clock offset along with the receiver coordinates, HDOP is 1.195, and VDOP is 1.558. If, however, we assume the clock error to be zero, and we only estimate the receiver coordinates, then HDOP is still 1.195, but VDOP is 0.913 — actually better than HDOP.

Getting back to Figure 3, we notice a large spike in the HDOP values (and hence in the PDOP and GDOP values as well) just before 0000 hours. What's going on here? At this time, the number of visible satellites above the mask angle has dropped to five. While this is not particularly unusual, the arrangement of these five satellites in the sky is (see Figure 4). The satellite positions, projected onto the user's horizon plane, are almost colinear, which makes the elements of matrix ***D*** large. If we accept a lower elevation mask angle of 5 degrees, several more satellites are available and the HDOP value drops to about one, with a correspondingly small GDOP (see Figure 5). In fact, the HDOP value stays close to one for the whole day except for short periods when it grows to about 1.5 or so.

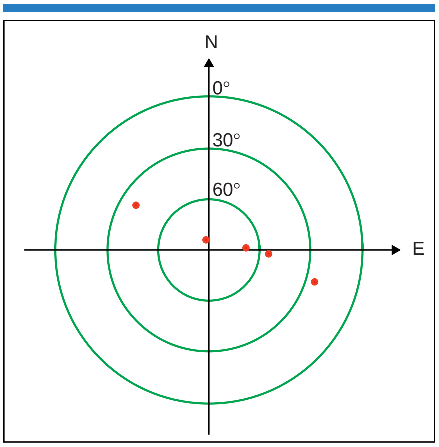

Figure 4. The spike in the DOP values in Figure 3 is caused by the almost perfect alignment of the five satellites above the elevation mask angle.

Latitude. The disparity between HDOP and VDOP values is larger for higher (north or south) latitudes because there are fewer satellites high in the sky. This limitation comes from the fact that the inclination of the GPS satellite orbits is about 55 degrees, which means that you can never have a satellite directly overhead at a latitude north of 55 degrees north (or south of 55 degrees south). At the poles, the highest elevation angle possible is about 45 degrees.

If we use an elevation mask angle of 15 degrees and track only four satellites — the four that produce the lowest DOP values — we find that while HDOP values are always between one and two, VDOP values are almost always above three and sometimes as large as seven. This isn't too surprising because we are only using satellites in an elevation-angle band of about 30 degrees around the sky. How bad is a VDOP of seven? If the UERE is 25 meters, the root-mean-square vertical error would be about 175 meters, and at the 95 percent uncertainty level, this error would increase to 350 meters. Dropping the elevation mask angle to 5 degrees improves the VDOP values to between two and three with an occasional excursion to four.

More Satellites. High DOP values can sometimes occur even for all-in-view receivers operating at midlatitudes. In some environments, such as heavily forested areas or urban canyons, a GPS receiver's antenna may not have a clear view of the whole sky because of obstructions. If it can only receive GPS signals from a small region of the sky, the DOPs will be large, and position accuracy will suffer. Being able to track more satellites can help in such situations, and a combined GPS/GLONASS receiver may provide acceptable accuracies. New receiver technology permitting use of weaker GPS signals, even those present inside buildings, will also be beneficial.

CONCLUSION

In this brief article we have introduced the concept of dilution of precision and examined the important role that receiver–satellite geometry plays in determining GPS position accuracy. While this geometry will always be of some concern in GPS positioning, the new satellite signals, improvements in receiver design, and use of additional signals from GLONASS or the proposed European Galileo constellations of satellites will help to minimize its impact. And in the not too distant future, we might expect real-time, standalone GPS position accuracies even in urban areas of a few meters or perhaps better. Stay tuned. ■

ACKNOWLEDGMENT

Thanks to University of New Brunswick's Paul Collins for generating the DOP plots for Figures 3 and 5.

FURTHER READING

For a more in-depth analysis of GPS dilution of precision (DOP), see

■ "Satellite Constellation and Geometric Dilution of Precision," by J.J. Spilker Jr. and "GPS Error Analysis," by B.W. Parkinson, in *Global Positioning System: Theory and Applications*, Vol. 1, edited by B.W. Parkinson and J.J. Spilker Jr., Progress in Astronautics and Aeronautics, Vol. 163. American Institute of Aeronautics and Astronautics, Washington, D.C., 1996, pp.177–208 and 469–483.

■ "Performance of Standalone GPS," by J.L. Leva, M.U. de Haag, and K. Van Dyke, in *Understanding GPS: Principles and Applications,* edited by E.D. Kaplan, Artech House Publishers, Norwood, Massachusetts, 1996, pp. 237–320.

For a discussion about how DOP and various error sources affect positioning accuracy, see

■ "GPS Performance in Navigation," by P. Misra, B.P. Burke, and M.M. Pratt, *Proceedings of the IEEE* (Special Issue on GPS), Vol. 187, No. 1, January 1999, pp. 65–85.

Several Internet sites provide online computation of DOPs for a specified location and time interval, including the U.S. Naval Air Warfare Center Weapons Division at China Lake, California:

■ <http://sirius.chinalake.navy.mil/satpred/>.

For a Java-based DOP demonstration, see

■ <http://www.ualberta.ca/~norris/gps/DOPdemo.html>.

For a detailed examination of the role geometry plays in high-precision relative GPS positioning, see

■ "Impact of GPS Satellite Sky Distribution," by R. Santerre, in *Manuscripta Geodaetica,* Vol. 16, 1991, pp. 28–53.

For a discussion of GDOP in terrestrial hyperbolic ranging systems, see

■ "Geometric Dilution of Precision," by E.R. Swanson, in *Navigation: Journal of The Institution of Navigation,* Vol. 25, No. 4, 1978–79, pp. 425–429.