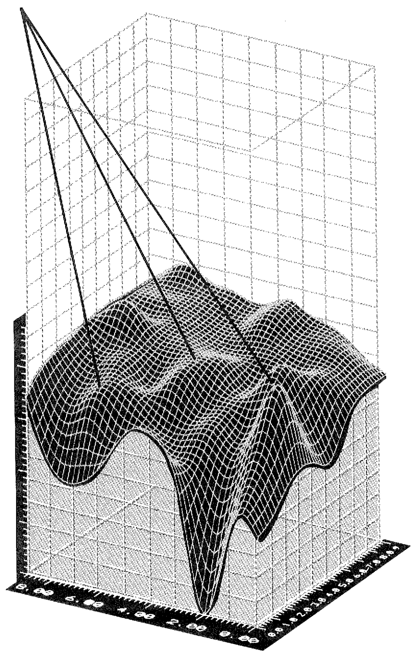
Reprinted from
GPS WORLD
April 1991



Ionospheric Effects on GPS

John A. Klobuchar

Air Force Geophysics Laboratory

"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. In this month's column John Klobuchar, a physical scientist at the Air Force Geophysics Laboratory and the author of the model on which the GPS broadcast ionospheric correction algorithm is based, tells us how the earth's ionosphere perturbs GPS signals and what can be done about it.*

This column is coordinated by Richard Langley and Alfred Kleusberg of the Department of Surveying Engineering at the University of New Brunswick. As always, we welcome your comments and suggestions of topics for future columns.

In the absence of selective availability (SA), the ionosphere can be the largest source of error in GPS positioning and navigation. Furthermore, under certain conditions, amplitude fading and phase scintillation effects can cause loss of carrier lock and intermittent GPS receiver operation. In this article we'll examine how the ionosphere affects GPS signals, how big the effects are, and how these effects can be removed, reduced, or, in some cases, avoided in the first place. But first, let's examine the nature of the ionosphere itself.

The ionosphere is a shell of electrons and electrically charged atoms and molecules that surrounds the earth, stretching from a height of about 50 kilometers to more than 1,000 kilometers. It owes its existence primarily to ultraviolet radiation from the sun. The photons making up the radiation possess a certain amount of energy. When the photons impinge on the atoms and molecules in the upper atmosphere, the photons' energy breaks some of the bonds that hold electrons to their parent atoms. The result is a large number of free, negatively charged, electrons and positively charged atoms and molecules called *ions*.

The free electrons in the ionosphere affect the propagation of radio waves. At frequencies below about 30 MHz the ionosphere acts almost like a mirror, bending the path traveled by a radio wave back toward the earth, thereby allowing long-distance communication. At higher frequencies, such as those used by GPS, radio waves pass right through the ionosphere. They are, nevertheless, affected by it.

The speed of propagation of a radio wave at some point in the ionosphere is determined by the density of electrons there. The speed of a carrier, the pure sinusoidal radio wave conveying a signal, is actually increased by the presence of the electrons. The greater the density of electrons, the greater its speed. The net effect on a radio wave is obtained by integrating the electron density along the whole path that a signal follows from a satellite to a receiver. The result is that a particular phase of the carrier arrives at the receiver earlier than it would have had the signal traveled from the satellite in a complete vacuum. This early arrival is termed a *phase advance*.

On the other hand, the signal that is modulating the carrier (the pseudorandom noise codes and navigation message, in the case of a GPS signal) is delayed by the ionosphere. Because the composite signal can be thought of as being formed by the superposition of a large group of pure sinusoids of slightly different frequencies, the delay of the modulation is called the *group delay*. Oddly enough, the magnitude of the group delay is identical to the magnitude of the phase advance — it's just of the opposite sign.

The electron density is quantified by counting the number of electrons in a vertical column with a cross-sectional area of 1 m^2 (one square meter). This number is called the *total electron content* (TEC). The TEC is a function of the amount of incident solar radiation. On the night side of the earth, the free electrons have a tendency to recombine with the ions, thereby reducing the TEC. As a consequence, the TEC above a particular spot on the earth has a strong diurnal variation.

Changes in TEC can also occur on much shorter time scales. One of the phenomena responsible for such changes is the *traveling ionospheric disturbance* (TID). TIDs, which have characteristic periods on the order of 10 minutes, are manifestations of waves in the upper atmosphere believed to be caused in part by severe weather fronts and volcanic eruptions. There are also seasonal variations in TEC and variations that follow the sun's 27-day rotational period and the roughly 11-year cycle of solar activity.

The size of the phase advance or group delay is also a function of the carrier frequency. The higher the frequency, the smaller the effect. This behavior of the ionosphere is called *dispersion* and is related to the splitting of visible light into its constituent colors by a prism. We can put the dispersive nature of the ionosphere to good advantage. By making measurements on two widely spaced frequencies and combining them, we can actually remove almost all of the ionospheric effect. It is for this reason that GPS satellites transmit signals on two carrier frequencies, L1 at 1575.42 MHz and L2 at 1227.60 MHz.

PSEUDORANGE ERROR

The group delay of the GPS signal introduces a range error into the pseudoranges measured with the P-code or the C/A-code. The range error is obtained by multiplying the group delay by the speed of light (about 300,000 kilometers per second). The resulting range error introduced by the earth's ionosphere can vary from less than 1 meter to more than 100 meters and changes with time of day, season, location of the receiver on the earth's surface, viewing direction, solar activity, and the state of the earth's magnetic field.

Over a single GPS satellite pass, if the ionosphere does not change greatly during the pass (as it does during the morning and evening periods), the ionospheric range error is

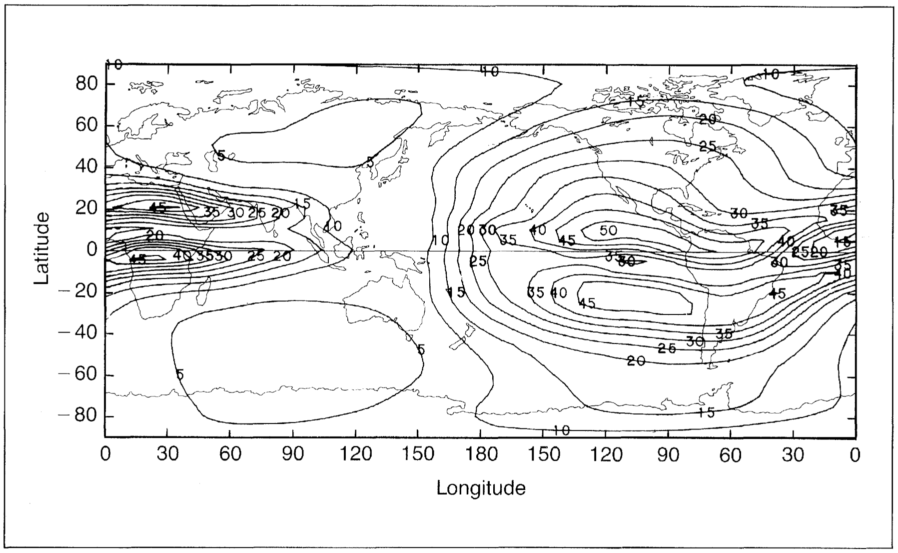


Figure 1. Worldwide contours of average ionospheric group delay at L1 for the month of March 1990, a solar maximum month, at 2000 UTC; units in nanoseconds

proportional to the secant of the satellite zenith angle as observed at the centroid height of the ionosphere, which is approximately 400 kilometers above the earth's surface. The secant of this angle increases from 1 to about 3 as the satellite moves from the zenith to the horizon. Thus, a vertical range error of 10 meters would translate into 30 meters for observations at a 5° elevation angle.

The ionosphere directly over the GPS receiver is important only for satellites viewed near the zenith. Satellite signals received from other directions pass through different regions of the ionosphere. In describing the regional behavior of the ionosphere, a useful concept is the *subionospheric point* — the point on the earth's surface directly below where the signal ray path passes the centroid height of the ionosphere. For satellites at low elevation angles with widely different azimuths, these subionospheric points can be separated by up to 3,000 kilometers. The ionospheric behavior at locations that far apart is generally not the same, giving different range errors along each line of sight.

GPS users with dual-frequency P-code receivers can correct for the ionospheric range error through an appropriate combination of the pseudoranges observed on L1 and L2. Single-frequency users with C/A-code receivers do not have that luxury. They either have to live with the reduced measurement accuracy or employ a model for the correction of ionospheric range errors. The range errors will cause a position error if they remain uncorrected. However, this position error will usually be smaller than the range errors, because the average of the range errors for all signals observed is indistinguishable from the receiver clock error and will be removed along with the clock error when the receiver processes its measurements. A further improvement may be obtained if the receiver's altitude is accurately known, because the major error the ionosphere introduces is to the height component of the receiver's position.

If a receiver uses signals from satellites which are more or less evenly spread out in azimuth, the errors in the horizontal coordinates stemming from errors in the ionospheric propagation delay model for each satellite tend to cancel each other out to a certain degree. But all of the satellites used by the receiver are above the horizon, so that there is no cancelling of the effect of the residual errors on the height coordinate.

Figure 1 shows contours of average vertical ionospheric group delay at L1, in nanoseconds, for the month of March 1990, a year of high solar activity. The regions of highest ionospheric delay are located, on average, approximately ±15° to ±20° either side of the earth's magnetic equator, and delays reach 50 nanoseconds (roughly 15 meters in range error). (The magnetic equator straddles the geographic equator because the axis of the earth's magnetic field is not aligned with the earth's geographic axis.) As discussed earlier, the group delay can be three times the vertical value for elevation angles of 5°. Occasionally there are days during which these monthly averages of delay are exceeded by a factor of two or more, so a "worst case" slant-range group delay due to the earth's ionosphere could be as high as 300 nanoseconds (about 100 meters).

ERROR CORRECTION

The single-frequency user can take advantage of an ionospheric correction model, the coefficients of which are transmitted as part of the GPS satellite message (see "Ionospheric Time-Delay Algorithm for Single-Frequency GPS Users" by J.A. Klobuchar, *IEEE Transactions on Aerospace and Electronic Systems,* vol. AES-23, no. 3, 1987). The correction algorithm was designed to reduce the group delay error by approximately fifty percent in a root-mean-square (rms) sense. Comparisons of the correction model against actual GPS dual-frequency tracking data have shown that this design goal has been exceeded by almost 10 percent.

The ionospheric correction algorithm models the diurnal ionospheric variation with a set of numerical coefficients. These coefficients are updated at 10-day intervals, or more often if necessary, to account for seasonal and solar activity changes. After applying this correction algorithm to single-frequency pseudoranges, the remaining residual range error is due to short-term ionospheric range errors not accounted for by the model.

Because of new developments in code-free GPS receiver design, the single-frequency user is not necessarily stuck with the 50 to 60 percent rms correction provided by the single-frequency ionospheric correction algorithm. Private industry and government research laboratories have developed code-free methods for the direct measurement of GPS range errors due to the ionosphere. Such instrumentation will more fully correct single-frequency measurements for ionospheric range errors.

RANGE-RATE ERRORS

The observed carrier phases of the GPS signals are contaminated by the ionospheric phase advance. Using the wavelength of the GPS carriers, these phase measurements can be converted into carrier ranges. The ionospherically induced error in these carrier ranges, as discussed above, is equal in size to the error of the pseudoranges but opposite in sign.

Changes of the carrier ranges over short time intervals are used to measure the receiver-to-satellite range rate, which in turn determines the receiver velocity. Changes in the ionospheric carrier range error during the range-rate computation interval introduce velocity errors into a single-frequency GPS receiving system. Dual-frequency receivers can automatically correct for any ionospheric range-rate errors in a similar fashion to the correction of range errors.

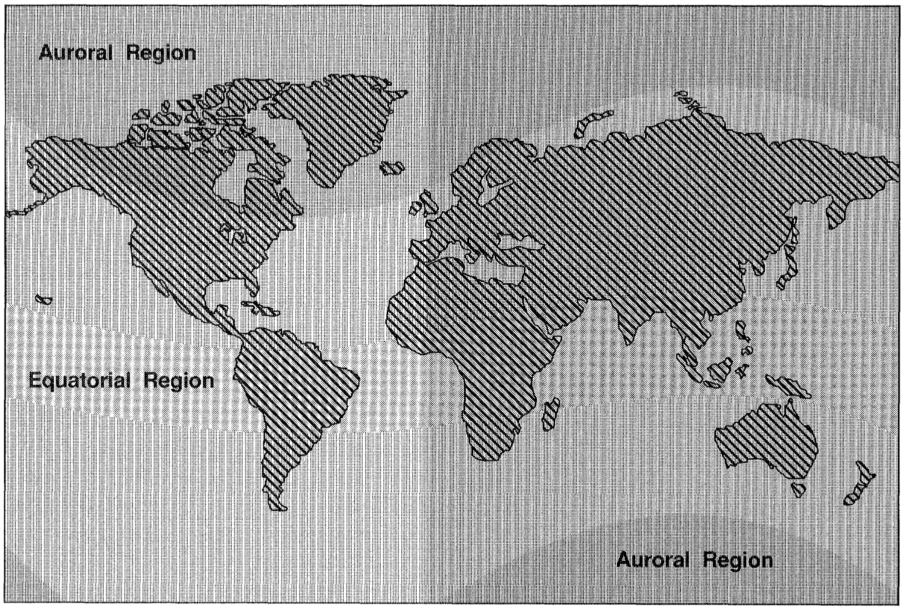


Figure 2. Regions of the world where scintillation fading effects can be especially severe

These range-rate errors are caused by the rate of change of the total number of electrons along the signal path. Because the ionosphere changes in a slightly different manner each day, the use of average models to correct for user range-rate errors generally is not recommended. Therefore, the single-frequency user algorithm transmitted in the GPS message should not be used to correct for ionospheric range-rate errors.

If the ionosphere produces a phase change faster than the receiver frequency bandwidth can handle, receiver phase-lock problems can occur. The receiver bandwidth must be adequate to follow the normal rate of change of geometric Doppler shift, which can be up to 1 Hz per second. Variations in the ionosphere introduce an additional rate of change amounting to one cycle of phase change at L1 for every change of 1.2×10^{16} electrons per square meter (el/m^2) in TEC.

A change of only 1 radian of phase (corresponding to 0.19×10^{16} el/m^2 change in TEC, or only 0.2 percent of a typical 10^{18} el/m^2 TEC) in a time interval equal to the inverse of the receiver bandwidth is enough to cause problems in maintaining receiver loop lock. If the receiver bandwidth is only 1 Hz (just wide enough to accommodate the geometric Doppler shift), there may be times when the rate of change of the ionosphere can cause apparent Doppler phase changes greater than 1 Hz, and loss of lock will result. During those times the signal amplitude is generally fading also, due to ionospheric scintillation effects.

SCINTILLATION EFFECTS

Irregularities in the earth's ionosphere can produce both diffraction and refraction effects causing short-term signal fading that can severely stress the tracking capabilities of a GPS receiver. The fading can be so severe that the signal level will drop completely below the receiver's lock threshold and must be continually reacquired. The geographic regions where scintillation effects normally occur are shown in Figure 2. As shown in the figure, the region of strongest effects extends from approximately $\pm 30°$ either side of the earth's geomagnetic equator.

Equatorial scintillation fading typically can last for several hours in evening-time periods, broken up with intervals of no fading in between. Significant effects also occur in the auroral and polar cap regions, though these regions are much smaller in extent than the equatorial region (despite what the map projection used for the figure appears to indicate).

During times when these fading effects are strong, the refractive effects that produce range-rate errors are also changing, often causing rapid carrier-phase changes. This is the most severe test of a GPS receiver in the natural environment. In fact, the apparent range-rate errors caused by rapid changes in TEC can produce a Doppler shift change of greater than 1 Hz per second, though obviously not all in the same direction. Even without simultaneous deep signal fading, these refractive effects can cause narrow band GPS receiving systems to lose lock due to the rapid frequency changes in the received signal that are greater than the receiver bandwidth.

Fortunately, the times of strong scintillation effects observed in the near-equatorial region are limited to approximately one hour after local sunset to approximately local midnight. There are exceptions, but if one can avoid making precise measurements during the nominal 1900 to 2400 local time period in this region, the chances of encountering strong scintillation effects are small.

Seasonal and solar cycle effects also reduce the chance of encountering scintillation in the near-equatorial regions. From April through August the chances are small of having significant scintillation in the American, African, and Indian longitude regions. In the Pacific region, however, scintillation effects maximize during these months. From September through March the situation is reversed. GPS users planning a campaign to make precise measurements in a near-equatorial region of the world should try to avoid the times of highest scintillation activity.

When severe magnetic storms occur, the auroral effects can move down into the mid-latitudes, and precise positioning with GPS can be affected by the ionosphere over the entire North American landmass for periods lasting up to one or two days.

MAGNETIC STORMS

At auroral and polar cap latitudes, any significant earth magnetic storm activity can produce scintillation effects, but, surprisingly, they are not as severe as those that have been measured in the near-equatorial belt. However, they can last for many hours, even days, and are not limited to the local late evening hours as are the near-equatorial scintillation effects.

Magnetic storms occur due to charged particles from solar flares arriving at the earth and causing changes in the earth's magnetic field. The visible aurora, or "northern and southern lights," is caused by high-energy particles flowing along the earth's magnetic field lines into the high latitudes where they interact with the neutral atmosphere to produce excited ions giving off red and green displays. This process also produces additional electrons that can cause strong scintillation fading effects for GPS receivers operating in the auroral and polar cap latitudes.

The strong electric fields generated in the ionosphere during these magnetically disturbed times push electrons over the polar cap, and large rapid changes in ionospheric group delay can move through a GPS signal ray path, greatly changing the ionospheric range and range-rate errors within time periods on the order of one minute (see "Ranging Errors Due to Disturbances in the Polar Ionosphere" by G.J. Bishop and J.A. Klobuchar, presented at ION GPS '90, Colorado Springs, Colorado, September 1990).

Amplitude fading and phase scintillation effects are associated with these "patches" of enhanced ionization as well, but the amplitude fades recorded have not been as severe as those observed in the near-equatorial region. Also, although these strong, ionospheric storm-induced changes in the polar cap and auroral regions can last many hours, or even days, the geographic area involved, compared with the near-equatorial region, is small. The northern polar cap is less than two percent of the globe. The entire northern auroral and polar cap region comprises less than seven percent of the earth's surface area. The extended equatorial region of potentially strong evening scintillation effects can extend up to $\pm 30°$ of geomagnetic latitude, which is 50 percent of the earth.

During rare, very strong magnetic storms, such as one that occurred in March 1989, these auroral effects can extend well into the mid-latitudes and cause unusual effects on GPS receivers. During those times the ionospheric-delay correction algorithm for single-frequency users would likely do a poor job of correcting for range errors, and the range and range-rate changes would be much larger than during magnetically quiet times. For instance, during the March 1989 magnetic storm during which the aurora extended over most of the continental United States, the range-rate changes produced by rapid variations in TEC exceeded a 1-Hz change in one second. Thus, receivers with only a 1-Hz bandwidth were continually losing lock during the worst part of this storm because of their inability to follow the changes.

Very severe magnetic storms like the one discussed here occur only a few times during each 11-year solar cycle. When they do occur, however, the auroral effects can move down into the mid-latitudes and precise positioning with GPS can be affected by the ionosphere over the entire North American landmass for periods lasting up to one or two days.

These effects try the "souls" of GPS receivers, but they don't occur frequently and, like hurricanes and tornadoes, are part of nature's weather pattern, and so we must learn to live with them. The National Oceanic and Atmospheric Administration operates the Space Environment Services Center in Boulder, Colorado, which gives forecasts of solar and magnetic activity. If you are planning a major field campaign, it would be good to check with the administration for its predictions of ionospheric effects on GPS precise measurements. It can be reached at (303) 497-3171, 24 hours a day, or via modem at (303) 497-5000. The Geophysics Division of the Geological Survey of Canada in Ottawa offers a geomagnetic prediction service. A recorded three-day forecast is available by calling (613) 992-1299. Both services are free.

SOLAR CYCLE

As mentioned earlier, ultraviolet radiation from the sun is primarily responsible for the ionosphere's existence. The ultraviolet output of the sun varies with solar activity, a useful measure of which is the number of sunspots visible on the sun's surface. Sunspots have been observed since the early 1600s, and the record shows a clear variation with a mean period of about 11 years. Figure 3 illustrates the last 50 years of sunspot activity. Of the past 400 years of recorded solar activity, the last 50 years have been the most exciting. Three of the highest four cycles among the 22 cycles systematically observed thus far have occurred within the last 50 years.

We are now in the declining phase of the present solar cycle, with a minimum expected in 1997. The art of predicting long-term solar activity is not very precise; the best guess about the next solar cycle peak, expected to occur at the turn of the century, is that the cycle will be an average one, lower than the last four. From the dashed part of the curve in Figure 3 you can see that we will still be above the average sunspot maximum value for approximately another two years, but we are in the long, slowly declining phase of the present cycle.

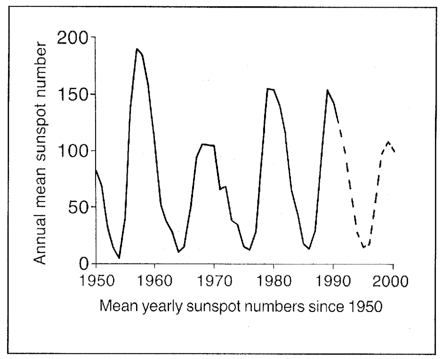


Figure 3. Sunspot activity for the last 40 years, with estimates up to the year 2000

For some unknown reason large magnetic storms, such as the one of March 1989, occur more frequently during the declining phase of the solar cycle, so don't be surprised if a few more periods of interesting ionospheric events occur during the next few years. The ionosphere will be with us forever; we just must learn how to live with it!

CONCLUSION

The ionosphere, in the absence of selective availability, can be the greatest source of range and range-rate errors for GPS users. The dual-frequency automatic correction for these effects is the best solution, and the single frequency user can correct for approximately 50 to 60 percent of the rms range error using the coefficients transmitted as part of the satellite message. As an alternative, a separate, relatively inexpensive, code-free ionospheric monitoring receiver will soon be available to measure ionospheric range error from all visible satellites simultaneously.

All GPS users should be aware of the likely times and regions of strong amplitude fading and phase scintillation effects and should attempt to avoid positioning at those times and in those regions when possible. Finally, the available prediction services should be consulted when you are planning a special observation campaign. ■