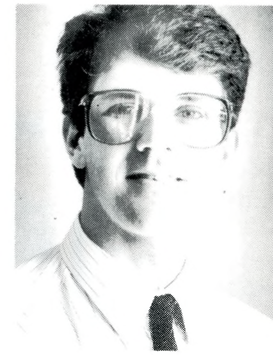
The Geodesy Corner

WHAT IS IT WE OBTAIN FROM GPS?



BY JAMES FERGUSON

For an ever increasing number of AOLS Quarterly readers, the term GPS, acronym for Global Positioning System, is a part of their survey vocabulary. Notwithstanding, many people (surveyors or otherwise) may know the terminology but may still be mystified as to the information we receive (or compute) from GPS. Furthermore, application of the system to solve survey problems is yet another story. This edition of The Geodesy Corner aims to focus on the fundamentals of GPS, and to outline the data and information GPS delivers. A follow up "Corner" will examine some of the ways a GPS user can apply this fantastic technology to traditional and non-traditional survey related problems.

The Basics - Simply Put

The underlying principle of GPS is quite simple - we must accurately measure the time it takes a signal to travel from a GPS satellite orbiting the earth, to a GPS receiver located somewhere on or above the earth's surface. Once we are able to determine what is called the "transit time" of this signal, we can compute the distance between the satellite and the receiver by multiplying the transit time by the speed of light. This principle is based on the theory that the GPS signal travels at the speed of light in a vacuum. Using the distance from satellite to receiver, and knowing the position of the satellite, we are able to determine our position as long as we collect data from a minimum number of satellites at any one time. In general, using a single GPS receiver we are able to determine a horizontal position using three satellites, and a 3-dimensional position if we receive data from at least four satellites.

Throughout the world, there are five precisely located tracking stations which monitor the orbits of all of the GPS satellites. Based on orbital theories, the predicted position of each satellite at any time is calculated through the tracking stations. The position of each of the satellites is "uploaded" to all satellites so that they can in turn "broadcast" this information to the GPS user. Each of the satellites is updated in this fashion several times per day from the "master control station" located at Colorado Springs, USA. In addition to the upload of satellite positions, the master control station tells each of the satellites how accurate their on-board "clocks" are. This is extremely important because the GPS system relies on the timing of the satellite signal (transit time) to accuracies of 0.00000000001 seconds or better. Each satellite has several on-board clocks to ensure that the satellite can maintain this stringent requirement. By the way, each GPS receiver also contains a clock, but a much less accurate one.

A variety of other information also updates the satellites memory via the upload process - items ranging from the health of the satellite to the age of the last information update.

The data that is stored in the satellite is available to the GPS user through transmission of two distinct signals. These two signals, known as L1 and L2, are transmitted on unique frequencies, and one or both may be received by a GPS user depending on the type of GPS receiver he/she has. In general, most GPS receivers acquire only the L1 signal which is sufficient for most survey applications. Receivers that are capable of receiving both signals are called "dual frequency", and are normally used for more precise GPS applications.

Each of the L1 and L2 signals are constructed using a "carrier wave", modulated by the information that the satellite is sending to the receiver. Modulation of the carrier wave simply means that data/information is "piggy-backed" onto the basic carrier radio wave. Information that is sent to the receiver is coded using two predefined code formats, and a receiver must be able to decode the signal if it is to use the information. Most receivers have this capability, and are able to decode what is known as the Standard Positioning Service (SPS) code. Some dual frequency receivers can also decode a second more precise message known as the Precise, or P code. Many GPS receivers have the capability to measure and record the carrier wave information in addition to the coded messages. In doing so, the receiver enhances the GPS users ability to more precisely determine the distance between the satellite and the receiver once it has acquired the coded satellite signal. Subsequent processing techniques of both the recorded code and non-code (carrier wave) raw data ultimately dictate how the two types of measurements produce final results.

Some Errors to Expect

The GPS system is subject to a wide variety of errors that can degrade the obtained results, or in some cases render them totally useless. Several sources of error are virtually beyond the control of the GPS user, while other errors can be reduced or eliminated by proper field/processing techniques.

Some of the more common errors associated with GPS observations are:

* errors in orbital information:
  poor orbit predictions
* errors in the satellite clocks:
  unstable on-board "clocks"
* errors in the receiver clocks:
  possible to have large differences between receivers
* Ionospheric Refraction:
  causes the GPS signal to bend and slow down unpredictably

WHAT IS IT WE OBTAIN FROM GPS cont'd

* Tropospheric Refraction:
  same affect as Ionospheric Refraction
* Deliberate degradation of the Satellite Signal
* Obstructions:
  block the signal from being received at the receiver antenna
* Multipath:
  reflection of the GPS signal off buildings, water, metal surfaces, etc.
* Human Error:
  poor H/I measurements, poor setups, etc.

Errors that descend upon the GPS receiver in terms of incorrect information in the coded messages are often less recoverable than errors that occur from other sources. For instance, the broadcast of a wrong position for a given satellite will be almost impossible to correct, while breaks in the satellite signal caused by trees blocking the signal path may be fixable at the processing stage. These breaks are usually called cycle slips. Satellite and receiver clock errors can generally be eliminated by processing raw data collected simultaneously by two receivers. Refraction errors are more ambiguous and difficult to account for, and severe ionospheric (upper atmosphere) disturbances can play havoc with GPS observations. Since the system is under the control of the United States military establishment, they have the ability to degrade or interrupt parts of the two GPS signals. Most popular of their methods is to degrade the SPS code so that a receiver can only compute a position on the one hundred (100) metre level. Without degradation, it is possible to achieve a thirty (30) metre (or better) positional accuracy using a single receiver. A second method of foiling the GPS user, which to date has not been implemented, is to deny civilian users access to the P code altogether.

Of course, human errors are reduced with experience but can be expected on any project GPS or otherwise.

The Numbers We Get

There are two fundamental ways of obtaining positional information for one or more stations surveyed using GPS. First, we can produce a single absolute, or point position using one receiver which receives the SPS code from several satellites at the same time. Second, we can compute the coordinate differences between any two stations, as long as each receiver acquires the SPS code over the same period of time from the same satellites. Very accurate coordinate differences can be obtained when the receiver records the L1 (for "single frequency" receivers) carrier wave information as well.

When we speak of positions from GPS, we generally refer to them as 3-dimensional quantities. This means that we can obtain a point position in terms of Latitude, Longitude and Height. These 3 values are actually derived from the 3-d Cartesian coordinate system, which defines positions as X,Y and Z referred to the origin of the coordinate system (generally the centre of mass of the earth - ie. WGS84). If we have two stations with X,Y and Z values, then we can obtain three coordinate differences between the stations - dx, dy and dz. The method of computation of the dx, dy, and dz values is termed "differencing", and makes use of two sets of GPS data recorded over the same time interval. This technique is valuable in that it helps to reduce or eliminate some of the major errors associated with the measurements, such as satellite or receiver clock errors.

Once the dx, dy and dz values are computed, it is possible to determine the distance between the stations as well as the orientation of the line between the stations, referred to the same X,Y,Z coordinate system. One other "hidden" quantity enters into the GPS positional computations, and this is the parameter of time. Since we are always trying to determine the most precise time possible for the system, we include it as a fourth dimension to be solved for along with the other three positional parameters. It follows then, that we must observe to a sufficient number of satellites simultaneously, in order to determine our four (Lat, Long, Height, Time) unknown values. If we have only three satellites visible, then we forfeit one of the four parameters. In this case, and at the users discretion, the height component might be held fixed to some value and the Lat, Long and Time parameters would be resolved. This would be sufficient knowledge for a user at sea for example, because the height of vessel is not of great interest. For a land based user who requires all three values, it is of the utmost importance that at least four satellites be observed at all times.

The errors that are left in the system after the raw data has been processed to the coordinate difference stage show up as errors in those same differences. Statistical models are formed throughout the processing phase, and are applied to the data at hand. In general, the errors are expressed in terms of standard deviations for each of dx, dy and dz, along with a correlation matrix forming the relationships between the three quantities. The effects of some of the errors that I spoke about earlier, such as ionospheric refraction,multipath, and signal interruptions, are difficult to assess completely, while others such as tropospheric refraction can be more realistically determined. Depending upon the receiver/software combination used, final error models (the actual errors associated with a measurement) can vary greatly. The most effective way to assess the errors associated with a measurement (dx, dy, dz), and to assess the measurement itself, is to integrate the data into an adjustment process. The adjustment involves inputting all GPS measurements for a particular project so that they may be compensated in an overall network sense. A future Geodesy Corner article will look at the adjustment of GPS data in more detail.

This is only a very brief introduction to the GPS system and what it provides. There are several excellent texts written on GPS positioning, as well as thousands of papers and articles. Two good references are:

- Guide to GPS Positioning
  Canadian GPS Associates
- GPS Satellite Surveying
  Alfred Leick

Write to me care of the Association if you would like further information about these or other publications.

Next time in "The Geodesy Corner" I will elaborate on various methods of using GPS, and touch on a number of unique applications for GPS technology. Thanks for reading.