EIGHTH PERIOD: Topic 2(a) 2005

Fundamentals of Global Navigation Satellite Systems including the Global Positioning System

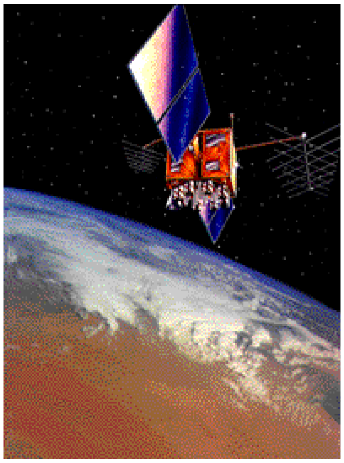
Reading: (Optional) Chapters 1 and 2 of "GPS for Land Surveyors – Second Edition" by Jan Van Sickle

Introduction

The Geodesy concepts covered in Topic 1 are used in many surveying applications today. The main surveying application utilizing those concepts are the Global Navigation Satellite Systems (GNSS), which includes the Global Positioning System (GPS). The current and/or future GNSS are GLONASS, GPS, and Galileo. The GLONASS is a satellite navigation system developed and maintained by the Russian government while GPS was developed and is maintained by the United States. Galileo is a future system conceived by the European Community. All of these systems are/will provide the same type of information to the end user with some distinctions in the system design. For the purposes of this topic, the discussion will focus almost exclusively on GPS.

Background

Before GPS, there were many different types of navigation systems. The immediate satellite navigation system to GPS was the Navy Navigation Satellite System (NNSS), also known as TRANSIT. It had 6 satellites. It was developed by the U.S. military to locate by coordinates land vehicles, seafaring vessels, and aircraft by coordinates. Civilian use was eventually authorized, and TRANSIT became used worldwide for navigation as well as surveying.



GPS was developed to replace TRANSIT because of two shortcomings to the TRANSIT system. The first problem with TRANSIT was that it had only 6 satellites which created large time gaps in the coverage of the system. This meant that a point had to be occupied for several days and then the data had to be processed in order to achieve an accuracy of about one meter. The second problem was that it had a relatively low accuracy level for navigation.

GPS is considerably more accurate than the TRANSIT system was and answers three important questions: what is the position of an object; what is the velocity of an object; and what time is it. This is why GPS is called a PNT system for position, navigation and time. GPS answers these questions quickly, accurately and inexpensively anywhere on the planet at any time.

GPS, formally known as the Navstar Global Positioning System, was initiated in 1973 and is operated and maintained by the Department of Defense (DoD). The Interagency GPS

Executive Board (IGEB) manages GPS, while the U.S. Coast Guard acts as the civil interface to the public for GPS matters.

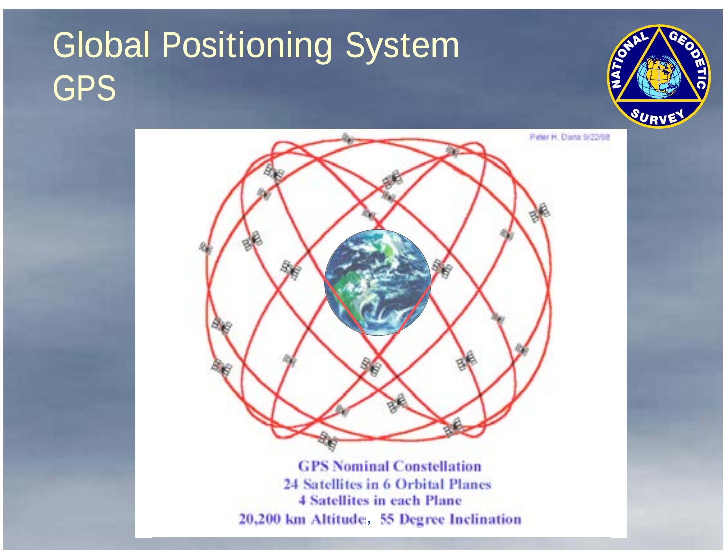
By creating a system that overcame the limitations of the existing navigation systems, GPS became attractive to a wide range of users worldwide. GPS has been successful in virtually all navigation applications, and because it's able to be accessed using small, inexpensive equipment, GPS is being utilized in a wide variety of applications across the globe, in particular surveying applications.

It should be noted that acquiring information from the GPS system is relatively easy. In other words, it is easy to operate a GPS receiver as they are very user friendly. This means that getting coordinates from a GPS receiver is easy; however, knowing what those coordinates are and the statistical indicators on the GPS information is not. This makes it imperative that the GPS surveyor understand the fundamentals of the GPS system, GPS surveying techniques, standards and specifications for GPS surveys, the processing of GPS data, and the analysis of the processing results. These are the concepts that Topic 2 of Period 8 will cover.

GPS Fundamentals – The GPS System

The GPS system is comprised of three segments: the space segment, the control segment, and the user segment.

Space Segment



The space segment of GPS contains 24 evenly spaced active satellites in circular 12 hour orbits that are inclined 55 degrees with respect to the equatorial plane. This scheme for the GPS satellite constellation provides for a minimum of four satellites in view 24 hours a day everywhere and, at times, up to 10 satellites.

Today, there are anywhere from 27 to 29 satellites in orbit where the additional ones above the main 24 are for replacement purposes in case a satellite becomes defective. The picture to the left shows the constellation for the GPS spaced based segment.

The types of satellites and their characteristics are described as "blocks". The current block types are listed below.

- Bloc I: Block I satellites were launched between 1978 and 1985. All of the block I satellites are now non operational although one is still intermittently switched between on and off. These satellites had a design life of 4.5 years. The main difference between these and the later generation satellites is that there was no ability to degrade the transmitted signals.

- Bloc II and IIa: These satellites were first launched in 1985. They have the capability to degrade the signal and have a design life of 7.5 years.
- Bloc IIr: These satellites are designed to have a longer life (10 years) and they are capable of satellite to satellite communications. They were launched from 1996 in order to maintain a full constellation.
- Different blocks will be launched in the future that will upgrade the L1 and L2 signal and add the L5 signal (discussed below).

Control Segment

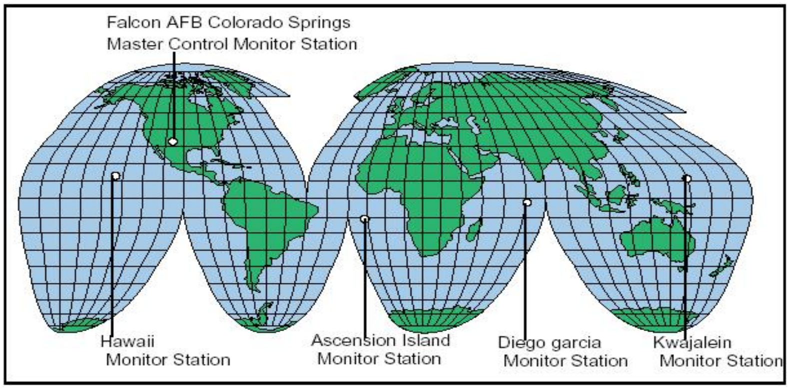
The second segment of the GPS system is the control segment. This segment is comprised of a master control station, monitor stations, and ground antennas.

The main station of the control segment is the master control station, located at Falcon Air Force Base in Colorado Springs, Colorado. It is responsible for overall management of the remote monitoring and transmission sites. As the center for support operations, it calculates any position or clock errors for each individual satellite, based on information received from the monitor stations, and then "orders" the appropriate ground antennas to relay the requisite corrective information back to that satellite.

There are 5 monitor stations that are located at Falcon Air Force Base in Colorado, Hawaii, Ascension Island in the Atlantic Ocean, Diego Garcia Atoll in the Indian Ocean, and Kwajalein Island in the South Pacific Ocean. Each of the monitor stations checks the exact altitude, position, speed, and overall health of the orbiting satellites. The control segment uses measurements collected by the monitor stations to predict the behavior of each satellite's orbit and clock. The prediction data is up-linked, or transmitted, to the satellites for transmission back to the users. The control segment also ensures that the GPS satellite orbits and clocks remain within acceptable limits. A monitor station can track up to 11 satellites at a time. This "check-up" is performed twice a day, by each station, as the satellites complete their journeys around the earth. Noted variations, such as those caused by the gravity of the moon, sun and the pressure of solar radiation, are passed along to the master control station.

Ground antennas monitor and track the satellites from horizon to horizon. They also transmit correction information to individual satellites. The following picture shows the complete control segment.

Control Segment



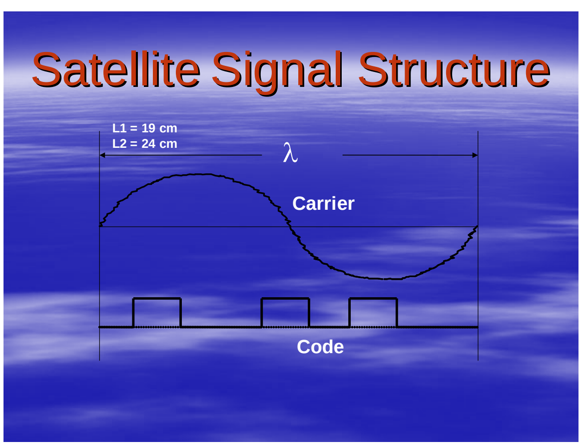
User Segment

The user segment is comprised of anyone or any entity that utilizes the data transmitted from the GPS satellites for any PNT purpose. This includes the military, the aviation industry, the shipping industry, the agriculture industry, the scientific community, and the surveying community to name a few.

GPS Fundamentals – The Satellite Signal

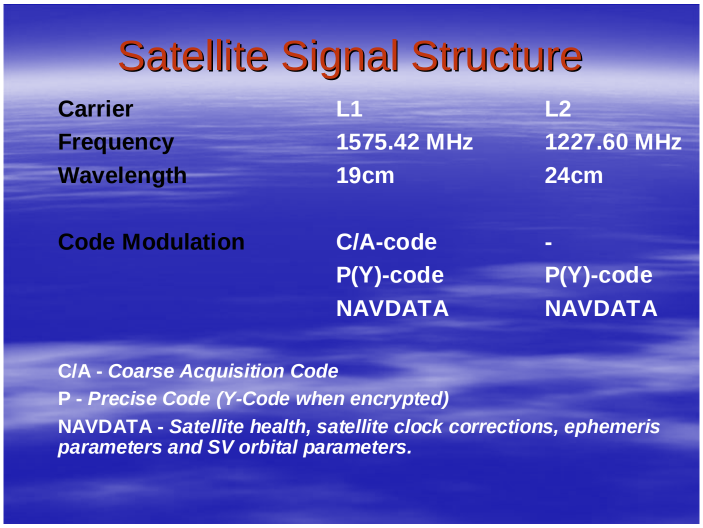
The GPS information is sent from the satellite via electromagnetic waves that are captured by a GPS receiver. These waves, of course, conform to the sinusoidal (Sine) laws of physics. There are, at present, two carrier waves in the L-band denoted L1 and L2. In the future there will also be a L5 carrier wave for civilian use.

The information sent on the L1 and L2 waves contains navigation data for the satellite and the satellite clock information. Two codes are used for transmitting the satellite clock readings. They are the precision code (P-code) and Coarse/Acquisition Code (C/A). The P-code is placed on both signals while the C/A code is only placed on the L1 signal.



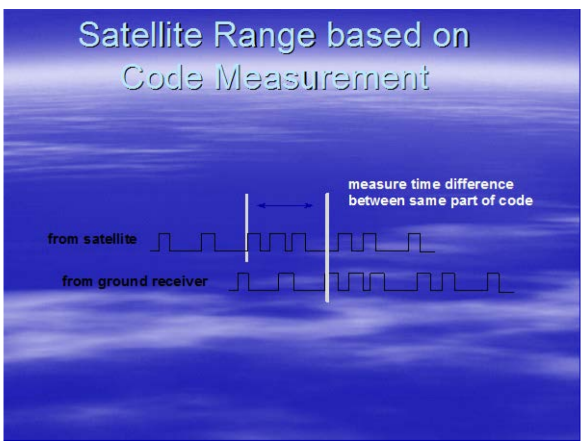
The goal of GPS signal processing by the GPS receiver (or post processing) is to recover the signal components, including the reconstruction of the carrier wave and the extraction of the codes for the satellite clock readings and the navigation data. From that information a distance from the satellite to the GPS receiver can be calculated. The picture above depicts the satellite signal structure. Note: the wave length of L1 is 19 cm and L2 is 24 cm.

The following also illustrates the L1 and L2 signal structure:



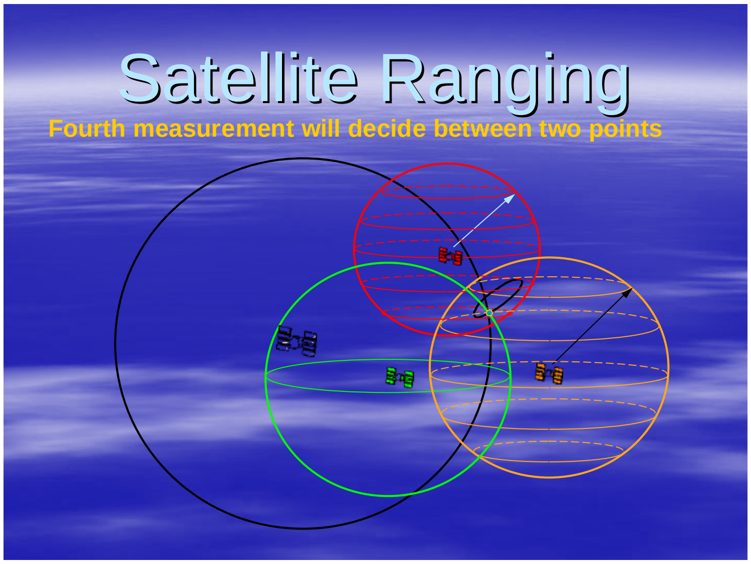
GPS Fundamentals – Pseudorange

The main goal in GPS surveying is to calculate a distance from the GPS receiver to the various satellites in view and then to calculate the GPS receivers location from that information. The distances can be calculated from the information contained on the L1 and L2 signals (and the L5 signal – in the future) or the signals themselves. From the code information on the signals a method called *pseudoranging* calculates this distance. The resultant distance is called a *pseudorange*. Note: some GPS receivers are only capable of receiving the L1 signal.



What is a pseudorange? A pseudorange is a distance from the GPS receiver to a particular satellite that is computed from the code information placed on the satellite signals. The codes are produced on the satellite at a specific point in time and put onto the signal. At the same time the GPS receiver is also generating the same code. Because of delays in the signal from the satellite reaching the GPS receiver, the codes don't match up (correlate). So, a time shift is applied to correlate them. This allows for the pseudorange (distance) to be calculated from the speed of light and the time it took the signal to travel to the receiver. I.e distance = speed x time. Note: the C/A-code based pseudorange is ten times less precise than the P-code psuedorange and has a stand alone (no correction information received fro other sources) accuracy at the GPS receiver of 10 -15 meters.

From these pseudoranges the position of the GPS receiver can be computed. With pseudoranges from three satellites there are two possible solutions. By having a psuedorange from a fourth satellite, one of the solutions can be eliminated. The following picture shows the circles around four satellites that have a radius that is the pseudorange to the GPS receiver.



The resultant positions produced from pseudoranges are used in many different applications. Specifically, they are used for geographic Information Systems (GIS) but are also used for navigation purposes. Additionally, they are used in any other application that doesn't require a high degree of accuracy in the required positions.

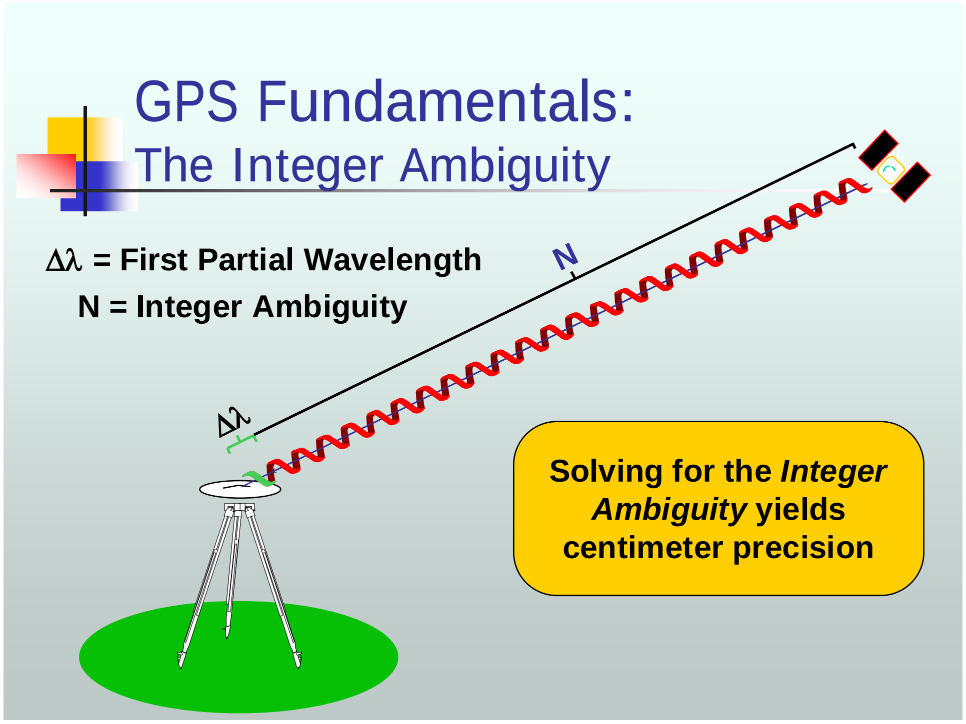
GPS Fundamentals – Carrier Phase Ranging

While a psuedorange is computed from the code information on the L1 and L2 carrier signals, a carrier phase range is derived from the signals themselves along with the code information. The carrier phase observable (signal) is the main information used in high accuracy surveying and scientific applications of GPS.

The process of determining a carrier phase range is analogous to determining a pseudorange. For the pseudorange a correlation is produced between the codes received from the satellites and the corresponding codes produced at the receiver. With code phase ranging a correlation is produced between the carrier phase signals themselves (L1 and L2) and a corresponding replica generated by the receiver.

In carrier phase ranging the game is to determine the number of complete sinusoidal wave lengths from the satellite to the receiver as well as the fractional part of the wavelength. Once the total number of wave lengths and the partial wavelength is known the carrier phase range can be directly computed. This is because we know the wavelength of the L1 and L2 signals (19 cm and 24 cm respectively – see diagram above).

The computing of the partial wavelength is derived from the differences in the signals received from the satellite and the replica produced by the receiver (phase difference). The more difficult problem is computing the total number of complete wavelengths. The computing of the total number of wave lengths is called the *integer ambiguity*. (Note: in essence the production of a pseudorange is also calculating an integer ambiguity). The following picture illustrates carrier phase ranging.



GPS Fundamentals - Error Sources

The processing of the GPS signals, whether to compute a psuedorange or a carrier phase range, is subject to a number of error sources. These include, but are not limited to, Anti-Spoofing and Selective Availability, Positional Dilution of Precision (PDOP), atmospheric delays, multipath, and blunders.

Anti-Spoofing and Selective Availability

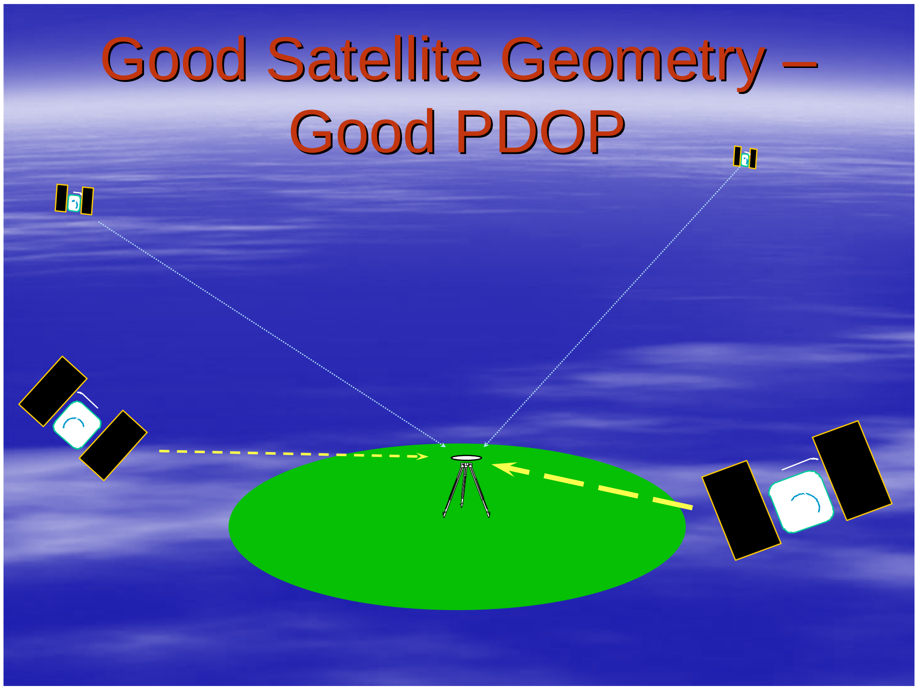
Anti-Spoofing (AS) and Selective Availability (SA) are methods created by the Department of Defense to encrypt or degrade the signals sent from the GPS satellites. AS guards against fake transmissions of satellite data by encrypting the P-code to form another code (Y-code) and is designed to prevent the receiver from making P-code measurements. Most manufactures of GPS receivers have devised techniques to still produce the P-code measurements making it a non issue for GPS surveyors.

SA is the denial of full accuracy of the code information, which is accomplished by manipulating the code information related to the satellite navigation data and/or the satellite clock information. This produced an error of approximately 100 meters in autonomous GPS positions. SA was turned off in May of 2000, however, may be reactivated by the president at anytime.

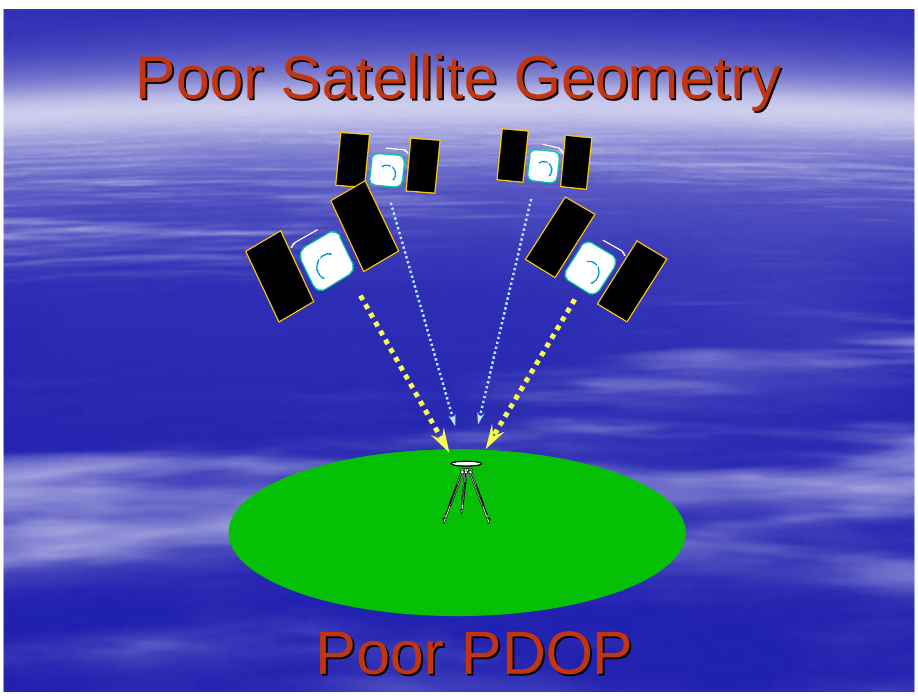
Positional Dilution of Precision

The Positional Dilution of Precision (PDOP) is a number that indicates the geometry of the satellites and the geometrical effect on the calculated position's accuracy. It is derived from the Horizontal Dilution of Precision (HDOP) and the Vertical Dilution of Precision (VDOP). The positions of the GPS satellites in view by a receiver determine the PDOP values for a given site. A value of 1 is the best value for PDOP. The higher the PDOP value the less accurate the position solution at the GPS receiver. The maximum value for the PDOP, before discarding the solution, is dependent on the GPS surveying technique utilized (discussed in the next lesson) and the intended use of the resultant positions.

The following illustration shows good satellite geometry. Notice that by having good satellite geometry the position of the GPS receiver has less room for error.



On the contrary, the following shows poor satellite geometry and hence more room for error at the GPs receiver.



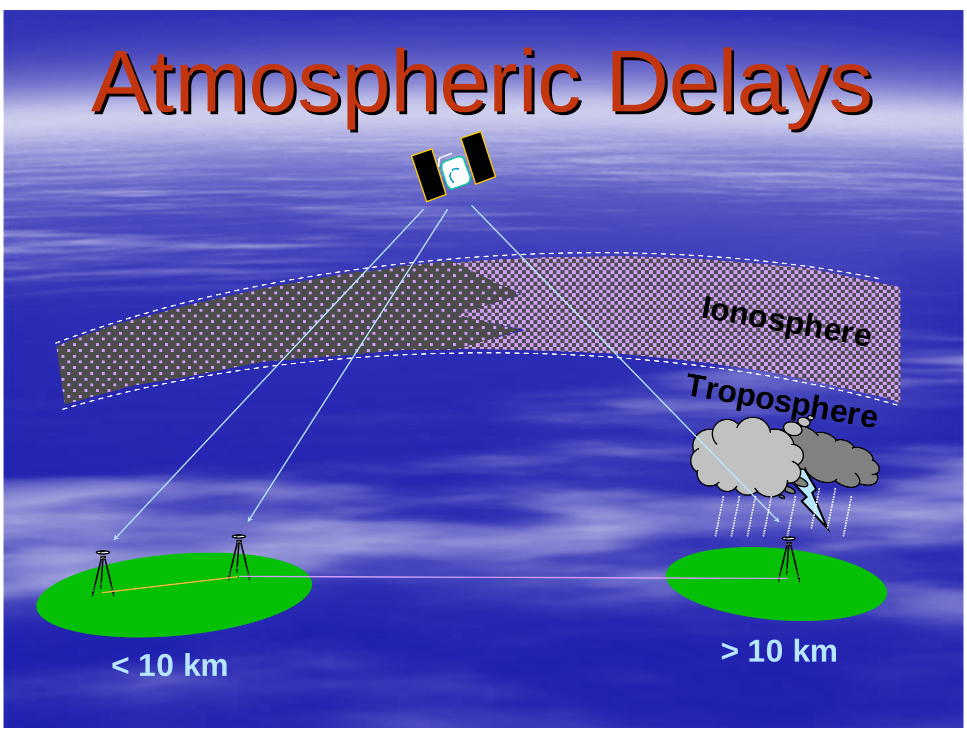
Atmospheric Delays

The L1 and L2 signals when traveling from the satellites to a GPS receiver have to travel through the earth's atmosphere. The earth's atmosphere varies in density and composition as the altitude increases above the surface. The lowest part of the atmosphere is called the troposphere and the higher part is called the ionosphere.

The troposphere is where all of the earth's weather is formed. It contains about 90% of the earth's atmosphere. The ionosphere contains free electrons that are excitable by solar radiation. Both the weather and the excited electrons cause delays in the GPS signals by refraction and diffraction of the signal resulting in an error in the produced GPS position. These errors can be reduced by applying a model of the troposphere and/or a model of the ionosphere. Note: the model of the ionosphere can be improved during post processing by using a GPS receiver that has the capabilities of receiving both the L1 and the L2 signal.

Two or more GPS receivers within 10 kilometers of each other receive GPS signals that travel through virtually the same ionosphere and, therefore, can be receivers that only collect L1 signal information. However, for two or more GPS receivers that are more than 10 kilometers from each other they must be able to collect both the L1 and L2 signal when performing high precision surveying for post processing. This is because the L1 and L2 signals have different wavelengths from which the delay produced by the ionosphere can be deduced.

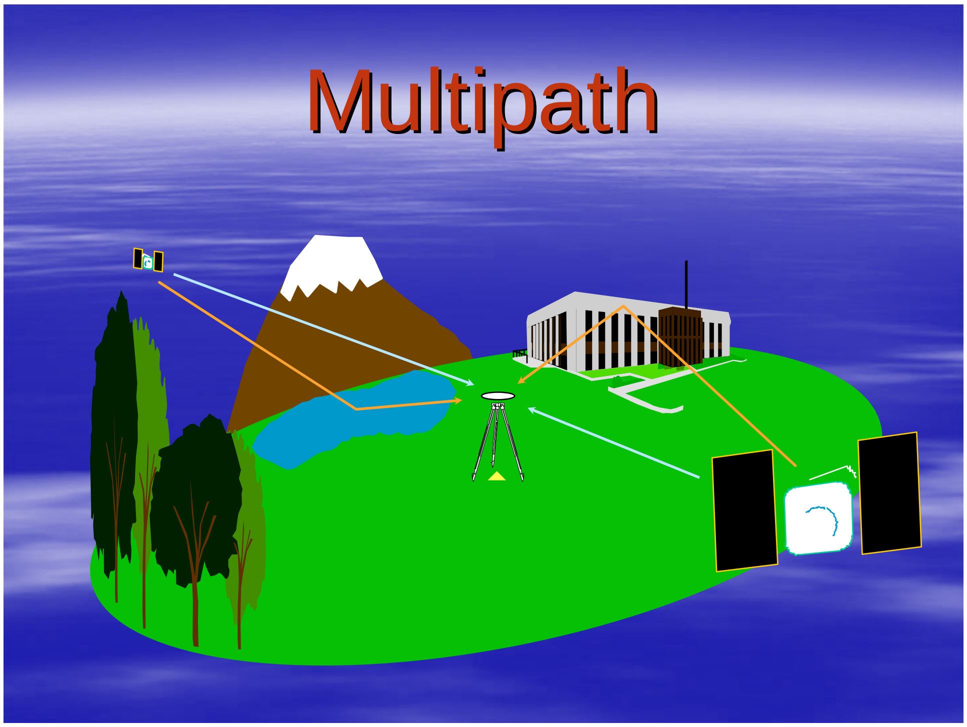
The following picture shows the relative atmospheric delays:



Multipath

The GPS surveyor also needs to be aware of the error source known as multipath. Multipath is the range delay, pseudorange or carrier phase, in the GPS signal by the reflection of the signal off of some object before it is received. The object can be a building, tree, pavement, ground, or anything else. Most GPS receivers can detect when a signal has be reflected.

The following picture illustrates multipath:



Blunders

Blunders are always a potential source for errors in any type of surveying and should be avoided at all cost. This is also true for GPS surveying. The surveyor should be very careful when performing a GPS survey to make sure that the correct station is occupied, the correct station name is recorded, the correct height of the equipment is measured and recorded, and the instrumentation is set up correctly.

The GPs surveyor should always complete a site log that contains the date, time, length of occupation, station name, measured height of the equipment, serial number of the antenna and receiver, and any problems that occurred while collecting data. This allows for some blunders to be removed from the data while post processing.

Summary

The following is a bulleted list of the key points contained in this lesson:

- There are two current and one proposed GNSS system today. GLONASS (Russian), GPS (United States), and Galileo (European Community).
- The GPS system has 24 active orbiting satellites.
- The GPS system is comprised of three segments: the space segment, the control segment, and the user segment.
- The satellites currently send out two carrier phase signals, L1 and L2.
- The L1 and L2 signals have “codes” placed on them. L1 has both the C/A code and the P-code, while the L2 signal has the P-code.
- A pseudorange is a distance from the satellite to a GPS receiver that is based on the C/A code. Pseudorange measurements are used to produce positions for applications such as GIS.
- A carrier phase range is a distance from the satellite to a GPS receiver and is based on the wavelength of the L1 and/or L2 signal. It is used to solve the integer ambiguity for the signals. It is more precise than a pseudorange and is used for high precision surveying and scientific applications.
- Some of the error sources associated with GPS surveying are AS and SA, PDOP, atmospheric delays, multipath, and blunders.

The next lessons will described in detail the different types of GPS surveying methods.