CHAPTER 1 FUNDAMENTAL OF SATELLITE POSITIONING

Contents of this chapter

1.1 Introduction to GNSS

Global Navigation Satellite Systems (GNSS) include constellations of Earth-orbiting satellites that broadcast their locations in space and time, of networks of ground control stations, and of receivers that calculate ground positions by trilateration. At present GNSS include two fully operational global systems, the United States' Global Positioning System (GPS) and the Russian Federation's GLObal NAvigation Satellite System (GLONASS), as well as the developing global and regional systems, namely Europe's European Satellite Navigation System (GALILEO) and China's COMPASS/BeiDou, India' s Regional Navigation Satellite System (IRNSS) and Japan's Quasi-Zenith Satellite System (QZSS). Once all these global and regional systems become fully operational, the user will have access to positioning, navigation and timing signals from more than 100 satellites (United Nations, 2012).

In addition to these, there are satellite-based augmentation systems, such as the United States' Wide-area Augmentation System (WAAS), the European Geostationary Navigation Overlay Service (EGNOS), the Russian System of Differential Correction and Monitoring (SDCM), the Indian GPS Aided Geo Augmented Navigation (GAGAN) and Japanese Multi-functional Transport Satellite (MTSAT) Satellite-based Augmentation Systems (MSAS). The successful completion of the work of the International Committee on Global Navigation Systems (ICG), particularly in establishing interoperability among the global systems, will allow a GNSS user to utilize one instrument to receive signals from multiple systems of satellites. This will provide additional data, particularly in urban and mountainous regions, and greater accuracy in timing or position measurements. To benefit from these achievements, GNSS users need to stay abreast of the latest developments in GNSS-related areas and build the capacity to use the GNSS signal (United Nations, 2012).

1.2 Basic Concept of GNSS

The satellite broadcasts two codes – the **coarse acquisition (C/A)** code, unique to the satellite, and the **navigation data message**. The codes bear information the receiver needs to work out its latitude, longitude and altitude and to synchronise its quartz clock with GPS time, common to the GPS system. The information includes almanac data – the predicted orbital parameters of the satellites beamed up to each satellite from the ground stations – and the more accurate ephemeris[1] tracking data divulged by each satellite. The coarse acquisition code is transmitted in binary form – a series of zeros and ones – and

---

[1] Parameters, such as Keplerian coefficients, that can be used to compute a satellite's position at a specified time (United Nations, 2012).

is superimposed on the carrier wave through a method called phase modulation (CASA, 2006).

GNSS uses the difference in the time of travel of radiowaves from four satellites to fix the position of the receiver and get an accurate value for time. The unit's processor computes the distance from a satellite from the time it takes the signal, travelling at 300,000 km/sec – the speed of light – to reach it. The computer deduces the value for time from the degree to which the pattern of zeros and ones in the coarse acquisition code is out of sync with the same pattern retrieved from its own memory and replayed at the same time.

The distance to the receiver is the product of velocity (300,000 km/sec) and time, and the unit's computer plugs these values into equations, which it solves simultaneously to get the **navigation solution**. The radiowaves enter the strange realm of relativity in which time slows down, and this is factored in to the receiver calculations (CASA, 2006).

The GPS unit displays the coordinates as latitude and longitude or as bearing and distance information relative to a known point. (Current approvals for the use of GPS equipment in IFR operations need GPS-derived data to be to WGS-84, or worldwide geodetic datum standard 1984) (CASA, 2006).

The basic GNSS concept is shown in **Figure 1-1**, which illustrates the steps involved in using GNSS to determine time and position then applying this information (Jeffrey, 2010).

Step 1 – Satellites

GNSS satellites orbit the earth well above the atmosphere (GPS and GLONASS satellites orbit at altitudes close to 20,000 km and Galileo satellites orbit around 23,000 km above the earth's surface). The satellites know their orbit ephemerides[2] (the parameter that define their orbit) and the time very, very accurately. Satellites use atomic clocks to determine time, the latest generation of GPS satellites uses rubidium clocks that are accurate to within ± 5 parts in 10^{11} Ground-based control stations adjust satellites' ephemerides and time, when necessary (Jeffrey, 2010).

Time is such a big deal in GNSS systems because the time it takes a GNSS signal to travel from satellites to receivers is used to determine distances (ranges) to satellites. Accuracy is essential because radio waves travel at the speed of light. In one microsecond (a millionth of a second), light travels 300 m. In a nanosecond (a billionth of a second), light travels 30 cm. Small errors in time can result in large errors in position (Jeffrey, 2010).

GPS was the first GNSS constellation to be launched. GLONASS constellation has also been launched and operational. Full operational capability of Galileo is likely sometimes after 2013. The benefit to end users of having access to multiple constellations is redundancy and availability. If one system fails, for any reason, GNSS receivers, if they are equipped to so, can receive and use signals from satellites in other systems (Jeffrey, 2010).

[2] Plural of ephemeris.

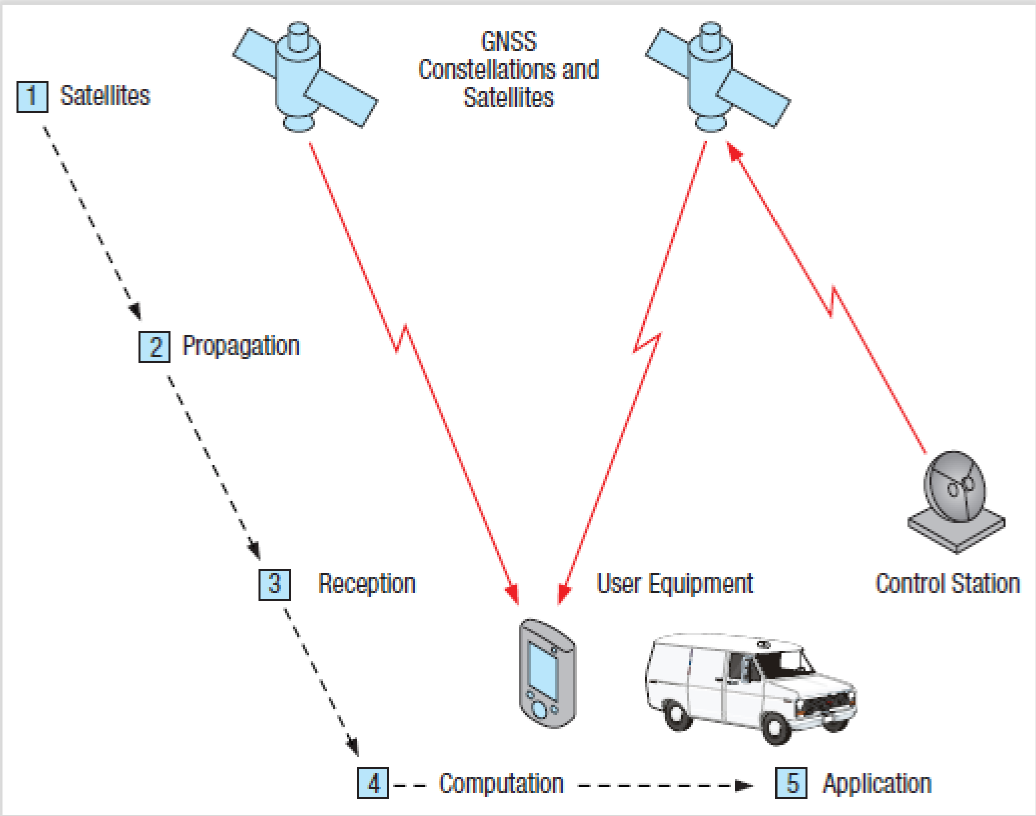


Figure 1-1: GNSS Basic Concept **(Jeffrey, 2010)**

Step 2 – Propagation

GNSS satellites regularly broadcast their ephemerides and time, as well as their status. GNSS radio signals pass through near-vacuum of space, then through the various layers of the atmosphere to the earth, as illustrated in **Figure 1-2**.

In order to obtain accurate positioning and timing, we need to know the length of the direct path from the satellite to the user equipment (which is referred to as the range to the satellite). As shown in the **Figure 1-2**, radio wave do not travel in a straight path, rather they are bent as they pass through the earth's atmosphere. Due to this phenomenon, the time the signal takes to travel from the satellite to the receiver is increased. The distance to the satellite is calculated by multiplying the time of propagation by the speed of light. Errors in the propagation time increase or decrease the computed range to the satellite. Incidentally, since the computed range contains errors and is not exactly equal to the actual range, we refer to it as a pseudorange[3] (Jeffrey, 2010).

---

[3] A measure of the apparent signal propagation time from the satellite to the receiver antenna, scaled into distance by the speed of light. Pseudo-range differs from the actual range by the influence of satellite and user clock (United Nations, 2012).

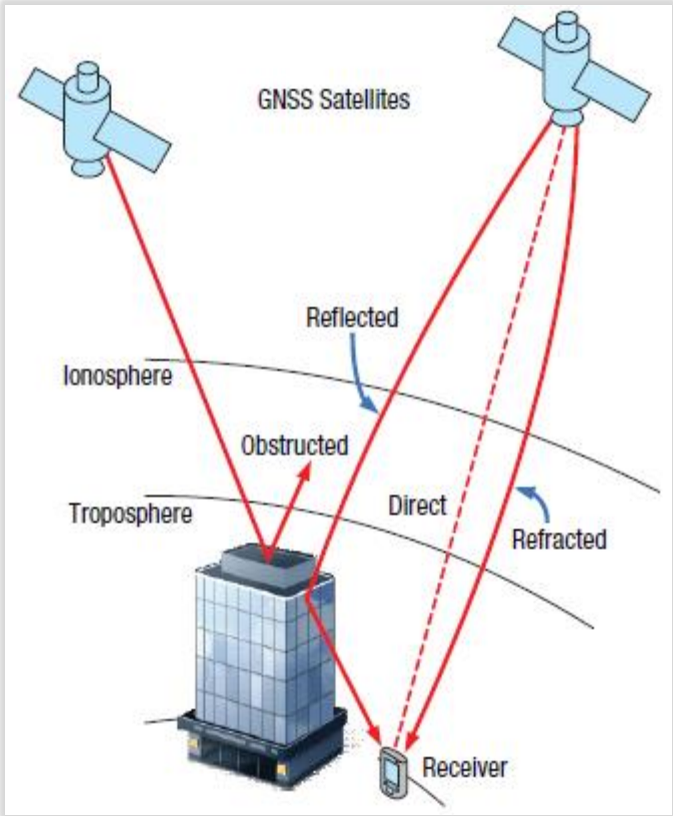


Figure 1-2: GNSS Signal Propagation

The layer of the atmosphere that most influences the transmission of GNSS signals is the ionosphere, the layer 70 to 1000 km above the earth's surface. Ultraviolet rays from the sun ionize gas molecules in this layer, releasing free electrons. These electrons influence electromagnetic wave propagation, including GPS satellite signal broadcasts. Ionospheric delays are frequency dependent so by calculating the range using both L1 and L2, the effect of the ionosphere cane be virtually eliminated by the receiver (Jeffrey, 2010).

The other layer of the atmosphere that influences the transmission of GPS signals is the troposphere, the lowest layer of the Earth's atmosphere. The thickness of the troposphere varies, about 17 km in the middle latitudes, up to 20 km near the equator, and thinner at the poles. Tropospheric delay is a function of local temperature, pressure and relative humidity. L1 and L2 are equally delayed, so the effect of tropospheric delay cannot be eliminated the way ionospheric delay can be. It is possible, however, to model the troposphere then predict and compensate for much of the delay (Jeffrey, 2010).

Some of the signal energy transmitted by the satellite is reflected on the way to the receiver. This phenomenon is referred to as multipath propagation. These reflected signals are delayed from the direct signal and, if they are strong enough, can interfere with the desired signal. Techniques have been developed whereby the receiver only considers the earliest-arriving signals and ignores multipath signals, which arrive later. In the early days of GPS, most errors came from ionospheric and tropospheric delays, but now more attention is being made to multipath effects, in the interests of continually improving GNSS performance (Jeffrey, 2010).

Step 3 – Reception

GNSS user equipment receives the signals from multiple GNSS satellites then, for each satellite, recovers the information that was transmitted and determines the time of propagation, the time it takes the signals to travel from the satellite to the receiver (Jeffrey, 2010). Principally three satellite are good enough, but receivers need at least four satellites to obtain a position. The use of more satellites, if they are available, will improve the position solution, however, the receiver's ability to make use of additional satellites may be limited by its computational power (Jeffrey, 2010).

Receivers vary in terms of which constellation of constellations they track, and how many satellites they track simultaneously. For each satellite being tracked, the receiver determines the propagation time. It can do this because of the pseudorandom nature of the signals. Since the receiver knows the pseudorandom code (a series of zeros and ones) for each satellite, it can determine the time it received the code from a particular satellite (Jeffrey, 2010). In this way, it can determine the time of propagation as shown in the **Figure 1-3**.

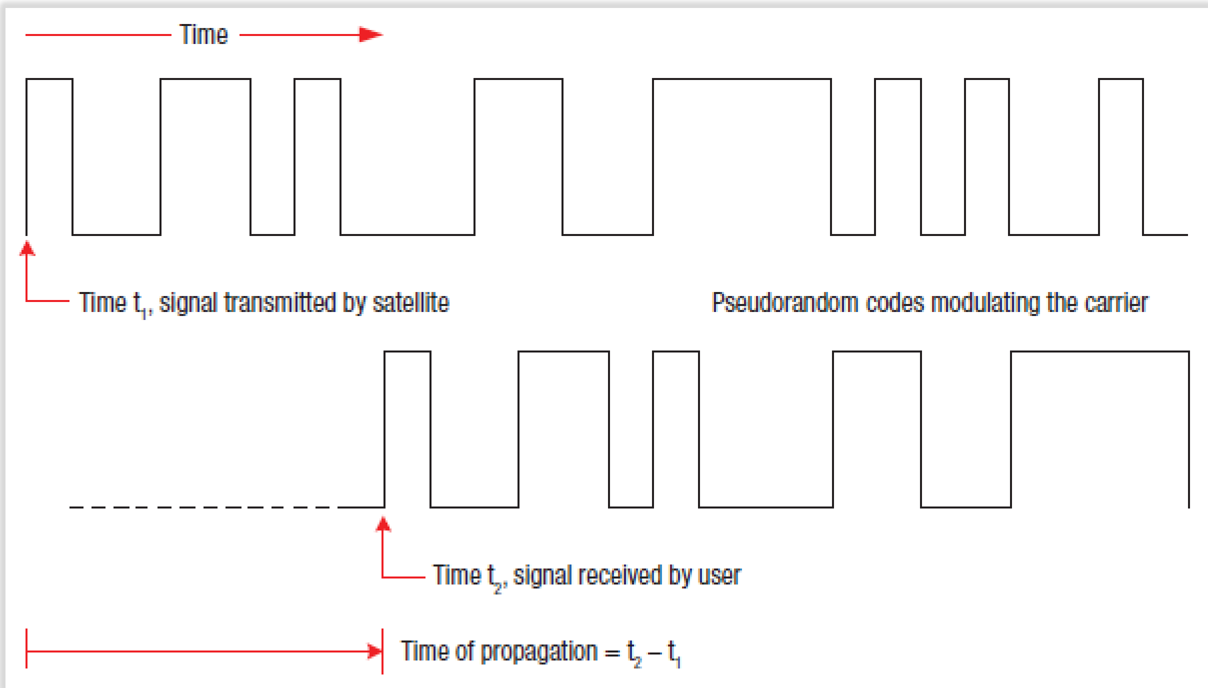


Figure 1-3: Time of Propagation

Step 4 – Computation

GNSS user equipment uses the recovered information to compute time and position. If we know the exact position of three satellites and the exact range to each of them, we would geometrically be able to determine our location. For each satellite being tracked, receiver calculates how long the satellite signal took to reach it, as follows (refer to **Figure 1-3**):

$$\Delta t = t_{sa,j} - t_{st,j}$$

where

Δt = Time of propagation

$t_{sa,j}$ = Time signal left satellite

$t_{st,j}$ = Time signal reached receiver

j = Satellite Number

Multiplying this propagation time by the speed of light gives the distance to the satellite. For each satellite being tracked, the receiver knows where the satellite was at the time of transmission (because the satellite broadcasts its orbit ephemerides) and it has determined the distance to the satellite when it was there. Using trilateration, a method of geometrically determining the position of an object, in a manner similar to triangulation, the receiver calculates its position (Jeffrey, 2010).

A GNSS position solution is determined by passive ranging in three dimensions [4]. The time of signal arrival, t_{sa}, is determined from the receiver clock, while the time of transmission, t_{st}, of each signal is obtained from its ranging code and data message. Where the receiver and satellite clocks are synchronized, the range, ρ, from a satellite to the user, measured by GNSS user equipment, is obtained by differencing the times of arrival and transmission and then multiplying by the speed of light, c. Thus,

$$\rho_j = (t_{sa,j} - t_{st,j})\,c \tag{6.1}$$

where the index j is used to denote the satellite number or receiver tracking channel and error sources have been neglected.

Where a ranging measurement from a single satellite is used, the user position can be anywhere on the surface of a sphere of radius ρ centered on that satellite. Where signals from two satellites are used, the locus of the user position is the circle of intersection of two spheres of radii ρ_1 and ρ_2. Adding a third ranging measurement limits the user position to two points on that circle as illustrated by Figure 6.2. For most applications, only one position solution will be viable in practice; the other may be in space, inside the Earth or simply outside the user's area of operation. Where both solutions are viable, a fourth ranging measurement can be used to resolve the ambiguity.

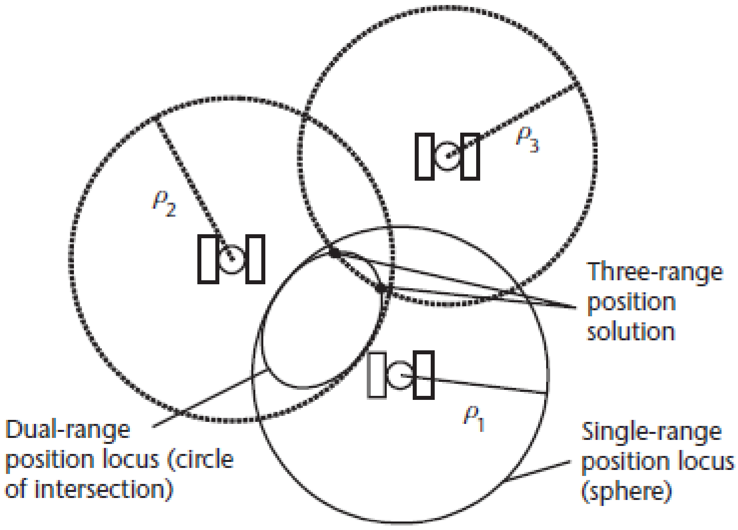


Figure 6.2 Position loci from single, dual, and triple ranging measurements.

In practice, however, the receiver and satellite clocks are not synchronized. If the receiver clock is running ahead of system time, the measured time of arrival, $\tilde{t}_{sa,j}$ will be later than the actual time of arrival, $t_{sa,j}$, resulting in an overestimated range measurement. If the satellite clock is running ahead of system time, the actual time of transmission, $t_{st,j}$, will be earlier than the intended time of transmission, $\tilde{t}_{st,j}$, which is that deduced by the user equipment from the ranging code. This will result in an underestimated range measurement. If the receiver clock is ahead by δt_{rc} and the clock of satellite j ahead by δt_{sj}, the range measurement error, neglecting other error sources, is

$$\begin{aligned} \delta\rho_j &= \tilde{\rho}_{Rj} - \rho_j \\ &= (\tilde{t}_{sa,j} - \tilde{t}_{st,j})c - (t_{sa,j} - t_{st,j})c \\ &= (\tilde{t}_{sa,j} - t_{sa,j})c - (\tilde{t}_{st,j} - t_{st,j})c \\ &= (\delta t_{rc} - \delta t_{sj})c \end{aligned} \tag{6.2}$$

where $\tilde{\rho}_{Rj}$ is the measured range, which is known as the *pseudo-range* to distinguish it from the range in the absence of clock errors. Figure 6.3 illustrates this.

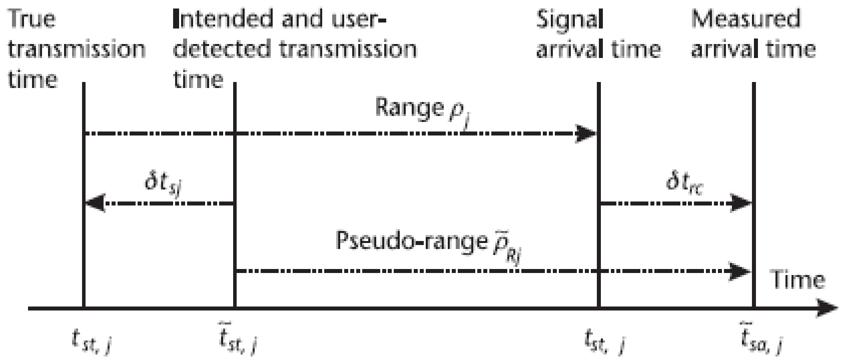


Figure 6.3 Effect of unsynchronized satellite and receiver clocks on range measurement.

The satellite clock errors are measured by the control segment and transmitted in the navigation data message. Therefore, the navigation processor is able to correct for them. The receiver clock offset from system time is unknown. However, as it is common to all simultaneous pseudo-range measurements made using a given receiver, it is treated as a fourth parameter of the position solution to be determined. Therefore, determination of a navigation solution using GNSS requires signals from at least four different satellites to be measured (unless the navigation solution is constrained).

Each pseudo-range measurement, corrected for the satellite clock error (and other known errors), $\tilde{\rho}_{Cj}$, may be expressed in terms of the satellite position, $\mathbf{r}^i_{isj}$, user antenna position, $\mathbf{r}^i_{ia}$, and the range error due to the receiver clock error, $\delta\rho_{rc}$, by

$$\tilde{\rho}_{Cj} = \sqrt{\left(\mathbf{r}^i_{isj}(t_{st,j}) - \mathbf{r}^i_{ia}(t_{sa})\right)^{\mathrm{T}}\left(\mathbf{r}^i_{isj}(t_{st,j}) - \mathbf{r}^i_{ia}(t_{sa})\right)} + \delta\rho_{rc}(t_{sa}) \tag{6.3}$$

noting that a is used to denote the user-antenna body frame and $\delta\rho_{rc} = \delta t_{rc} c$. The satellite position is obtained from the set of parameters broadcast in the navigation data message describing the satellite orbit, known as the *ephemeris* (see Section 7.1.1), together with the corrected measurement of the time of signal transmission. The four unknowns, comprising the antenna position and receiver clock error, are common to the pseudo-range equations for each of the satellites, assuming a common time of signal arrival. Therefore, they may be obtained by solving simultaneous equations for four pseudo-range measurements. Figure 6.4 illustrates the solution geometry. Similarly, the velocity of the user antenna may be obtained from a set of measurements of pseudo-range rate, the rate of change of the pseudo-range. Calculation of the GNSS navigation solution is described in detail in Section 7.5.

Sources of error in the GNSS navigation solution include differences between the true and broadcast ephemeris and satellite clock errors, signal propagation delays through the ionosphere and troposphere, and receiver measurement errors due to delays in responding to dynamics, receiver noise, radio frequency (RF) interference, and signal multipath. These are all discussed in Section 7.4. The ionosphere and troposphere delays may be partially calibrated using models;

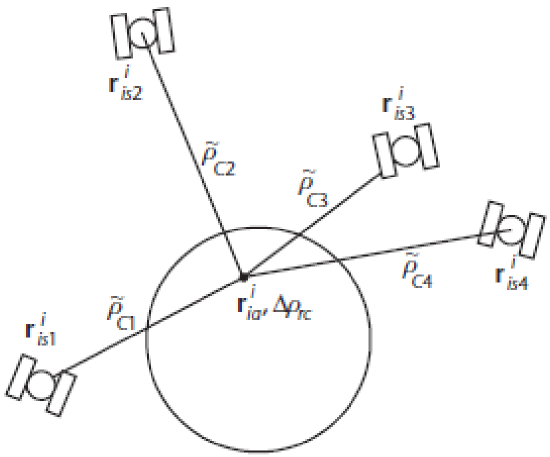


Figure 6.4 Determination of a position solution using four satellite navigation signals.

(Groves, 2008)

To help us understand trilateration, we'll present the technique in two dimensions. The receiver calculates its range to Satellite A. As we mentioned, it does this by determining the amount of time it took for the signal from Satellite A to arrive at the receiver, and multiplying this time by the speed of light. Satellite A communicated its location (determined from the satellite orbit ephemerides and time) to the receiver, so the receiver knows it is somewhere on a circle with radius equal to the range and center at the location of Satellite A, as illustrated in Figure 18. In three dimensions, we would show ranges as spheres, not circles, but stay with us for now.

The receiver also determines its range to a second satellite, Satellite B. Now the receiver knows it is at the intersection of two circles, at either Position 1 or 2, as shown in Figure 19.

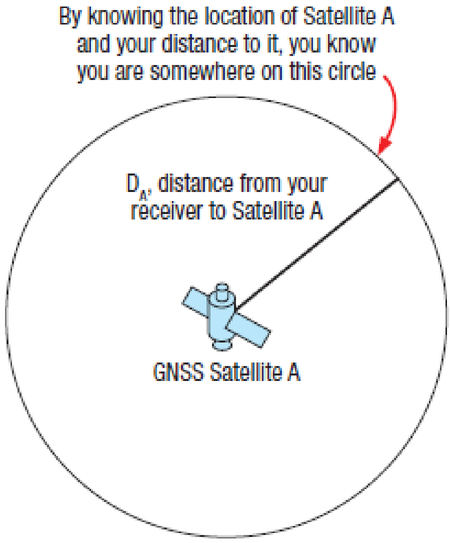


Figure 18 Ranging to First Satellite

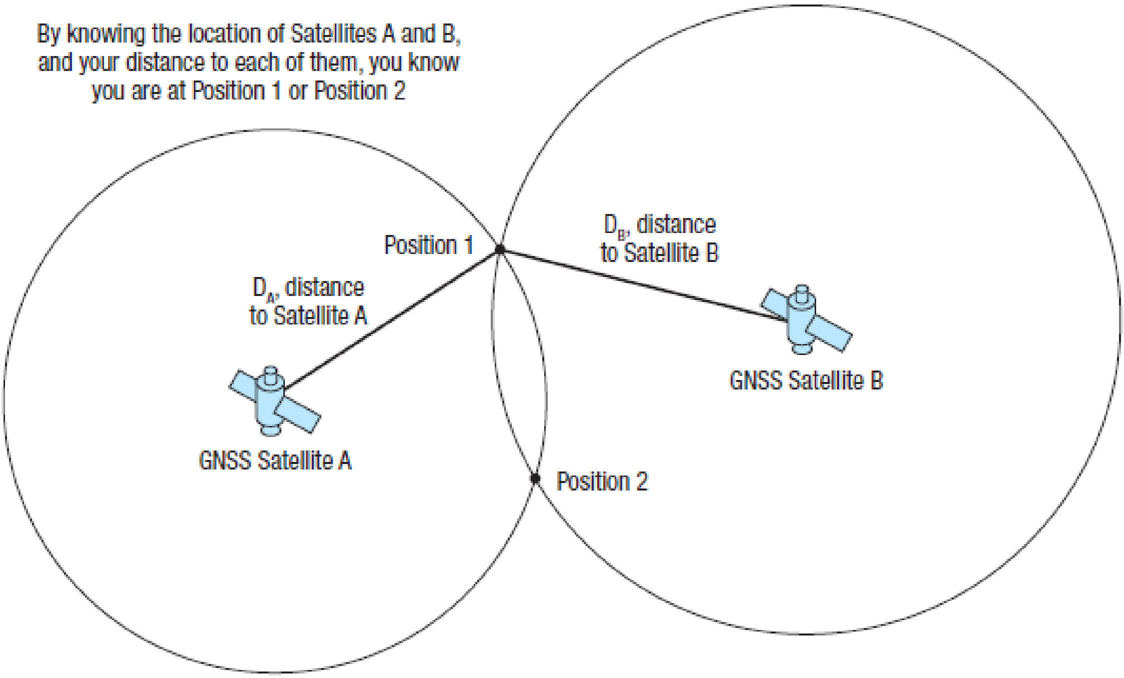


Figure 19 Ranging to Second Satellite

You may be tempted to conclude that ranging to a third satellite would be required to resolve your location to Position 1 or Position 2. But one of the positions can most often be eliminated as not feasible because, for example, it is in space or in the middle of the Earth. You might be tempted to extend our illustration to three dimensions and suggest that only three ranges are needed for positioning.

It turns out that receiver clocks are not nearly as accurate as the clocks on board the satellites. Most are based on quartz crystals. Remember, we said these clocks were accurate to only about 5 parts per million. If we multiply this by the speed of light, it will result in an accuracy of ±1500 metres. When we determine the range to two satellites, our computed position will be out by an amount proportional to the inaccuracy in our receiver clock, as illustrated in Figure 20.

We want to determine our actual position but, as shown in Figure 20, the receiver time inaccuracy causes range errors that result in position errors. The receiver knows there is an error, it just does not know the size of the error. If we now compute the range to a third satellite, it will not intersect the computed position, as shown in Figure 21.

Now for one of the ingenious techniques used in GNSS positioning.

The receiver knows that the reason the pseudoranges to the three satellites are not intersecting is because its clock is not very good. The receiver is programmed to advance or delay its clock until the pseudoranges to the three satellites converge at a single point, as shown in Figure 22.

The incredible accuracy of the satellite clock has now been "transferred" to the receiver clock, eliminating the receiver clock error in the position determination. The receiver now has both an accurate position and a very, very accurate time. This presents opportunities for a broad range of applications, as we shall discuss.

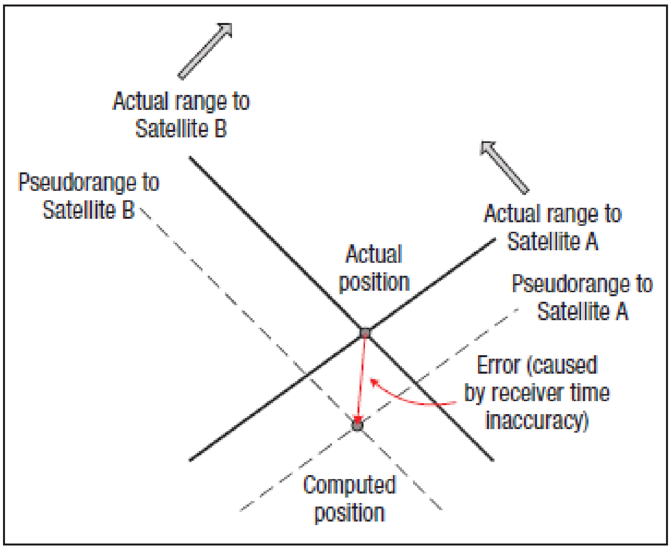


Figure 20 Position Error

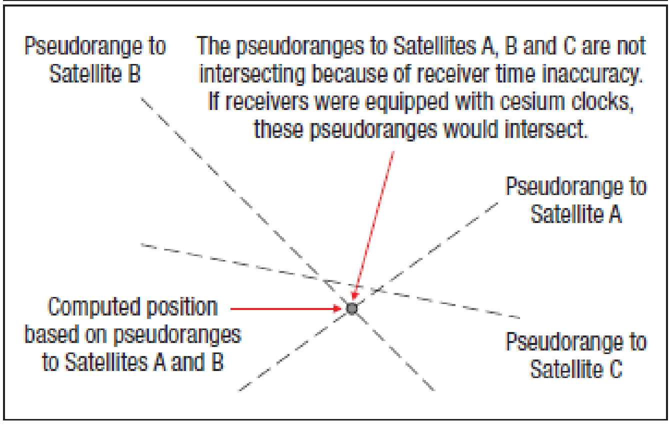


Figure 21 Detecting Position Error

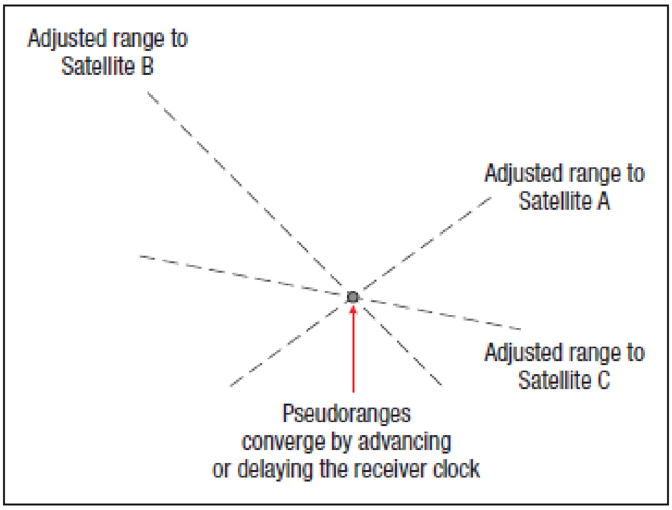


Figure 22 Convergence

The above technique shows how, in a two-dimensional representation, receiver time inaccuracy can be eliminated and position determined using ranges to three satellites. When we extend this technique to three dimensions, we need to add a range to a fourth satellite. This is the reason why line-of-sight to a minimum of four GNSS satellites is needed to determine position.

Step 5 – Application

GNSS user equipment utilizes the position and time information in their applications, for example, navigation, surveying or mapping.

Once the receiver has determined its position and time, this information is passed to and used by the user application. In a handheld receiver, for example, position may be superimposed on a map to graphically show the user's location. Events in electrical substations can be tagged with time that is exactly the same for two remote substations (Jeffrey, 2010).

1.3 GNSS Architecture

GNSS satellite systems consist of three major components or segments: space segment, control segment and user segment (Jeffrey, 2010). These are illustrated in **Figure 1-4**.

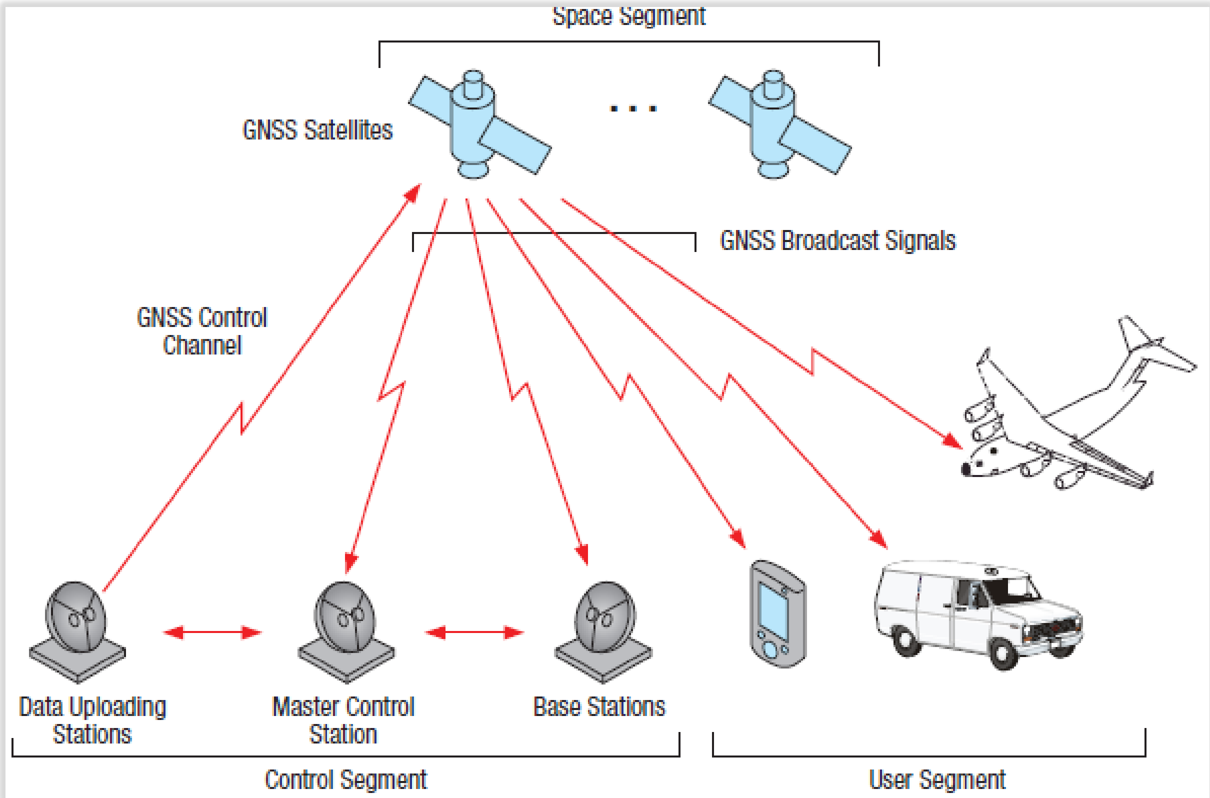


Figure 1-4: GNSS Segments

1.3.1 Space Segment

The space segment comprises of satellites, collectively known as a constellation, which broadcasts signals to both the control segment and the users. Some authors use the term space vehicle (SV) instead of satellite. GPS, GLONASS, and Galileo satellites are distributed between a number of medium earth orbits (MEOs), inclined at roughly 60° to the equator with around two orbits per day. Compared to geostationary orbits, these orbits give better signal geometry for positioning and better coverage in Polar Regions (Groves, 2008).

The space segment consists of GNSS satellites, orbiting about 20,000 km above the earth. Each GNSS has its own constellation of satellites, arranges in orbits to provide the desired coverage, as illustrated in **Figure 1-5** (Jeffrey, 2010).

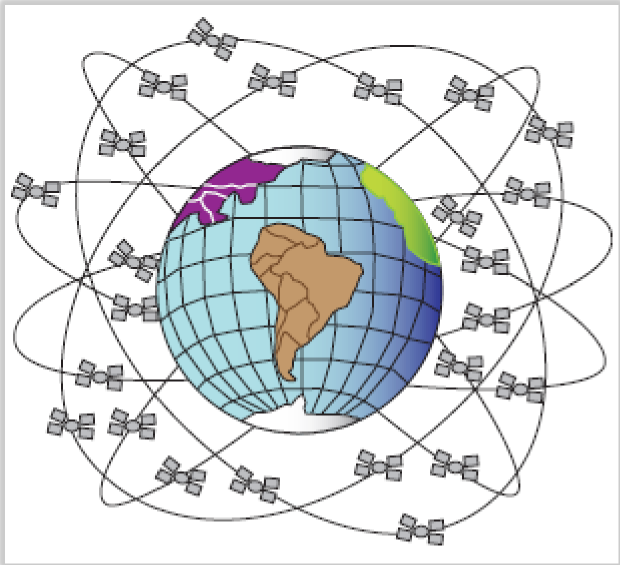


Figure 1-5: Satellites in their Orbits **(Jeffrey, 2010)**

1.3.2 Control Segment

The control segment comprises a ground-based network of master control stations, data uploading stations, and monitor stations; in the case of GPS, two master control stations (one primary and one backup), four data uploading stations and ten monitor stations, located throughout the world. In each GNSS system, the master control station adjusts the satellites' orbit parameter and onboard high-precision clocks when necessary to maintain accuracy (Jeffrey, 2010).

Monitor stations, usually installed over a broad geographic area, monitor the satellites' signals and status, and relay this information to the master control station. The master control station analyses the signals then transmits orbit and time corrections to the satellites through data uploading stations (Jeffrey, 2010).

The control segment, or ground segment, consists of a network of monitor stations, one or more control stations and a number of uplink stations. The monitor stations obtain ranging measurements from the satellites and send these to the control station(s). The monitor stations are at precisely surveyed locations and have synchronized clocks, enabling their ranging measurements to be used to determine the satellite orbits and calibrate the satellite clocks. Radar and laser tracking measurements may also be used (Groves, 2008).

The control stations calculate the navigation data message for each satellite and determine whether any manoeuvres must be performed. This information is then transmitted to the space segment by the uplink stations. Most satellite manoeuvres are small infrequent corrections, known as station keeping, which are used to maintain their satellites in their correct orbits. However, major relocations are performed in the event of satellite failure, with the failed satellite moved to a different orbit and a new satellite moved to take its place. Satellites are not moved from one orbital plane to another. GPS, GLONASS, and Galileo each maintain an independent control segment. Details of each system's space and control segment are given later in this chapter (Groves, 2008).

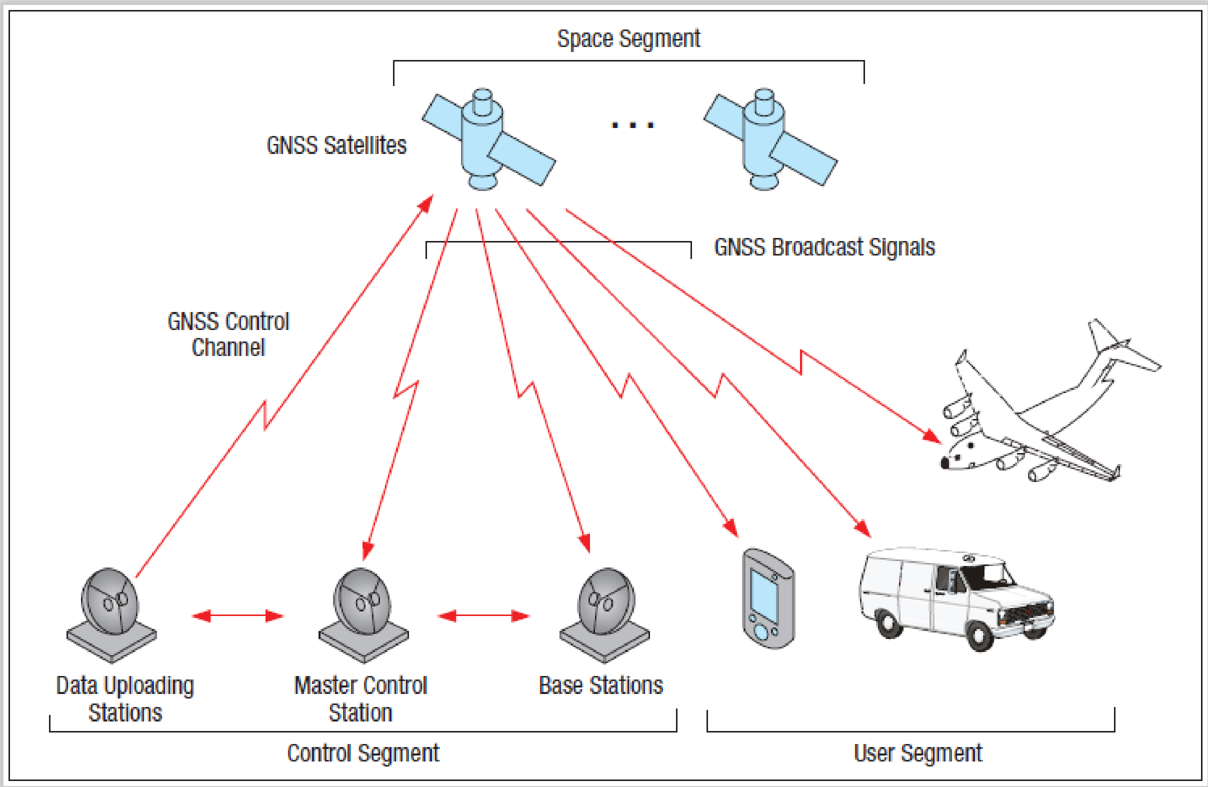


Figure 1-6: GNSS Segments

1.3.3 User Segment

The user segment consists of equipment that processes the received signals from the GNSS satellites and uses them to derive and apply location and time information. The equipment ranges from handheld receivers used by hikers, to sophisticated, specialized receivers used for high-end survey and mapping applications (Jeffrey, 2010).

GNSS user equipment is commonly described as GPS, GLONASS, Galileo, and GNSS receivers, as appropriate. However, as Figure 1-4 shows, the receiver forms only part of each set of user equipment. The antenna converts the incoming GNSS radio signals to electrical signals. These are input to the receiver, which demodulates the signals using a clock to provide a time reference. The ranging processor uses acquisition and tracking algorithms to determine the range from the antenna to each of the satellites used from the receiver outputs. It also controls the receiver and decodes the navigation messages. Finally, the navigation processor uses the ranging measurements to compute a position, velocity, and time (PVT) solution (Groves, 2008).

CHAPTER 2 REFERENCE SYSTEMS, SIGNAL, STRUCTURE AND USER EQUIPMENT

Contents of this chapter

2.1 Reference Coordinate Systems

In determining the positions of points on Earth from satellite observations, three different reference coordinate systems are important. First of all, satellite positions at the instant they are observed are specified in the "space-related" satellite reference coordinate systems. These are three-dimensional rectangular systems defined by the satellite orbits. Satellite positions are then transformed into a three-dimensional rectangular geocentric coordinate system, which is physically related to the Earth. As a result of satellite positioning observations, the positions of new points on Earth are determined in this coordinate system. Finally, the geocentric coordinates are transformed into the more commonly used and locally oriented geodetic coordinate system. The following subsections describe these three coordinate systems (Ghilani and Wolf , 2012: 337).

2.1.1 Satellite Reference Coordinate System

Once a satellite is launched into orbit, its movement thereafter within that orbit is governed primarily by the Earth's gravitational force. However, there are a number of other lesser factors involved including the gravitational forces exerted by the sun and moon, as well as forces due to solar radiation. Because of movements of the Earth, sun, and moon with respect to each other, and because of variations in solar radiation, these forces are not uniform and hence satellite movements vary somewhat from their ideal paths. As shown in **Figure 2-1**, ignoring all forces except the Earth's gravitational pull, a satellite's idealized orbit is elliptical, and has one of its two foci at G, the Earth's mass center. The figure also illustrates a satellite reference coordinate system, X_s, Y_s, Z_s. The perigee and apogee points are where the satellite is closest to, and farthest away from G, respectively, in its orbit. The line of apsides joins these two points, passes through the two foci, and is the reference axis X_s. The origin of the X_s, Y_s, Z_s coordinate system is at G; the Y_s axis is in the mean orbital plane; and Z_s is perpendicular to this plane. Values of Z_s coordinates represent departures of the satellite from its mean orbital plane, and normally are very small. A satellite at position S_1 would have coordinates X_{s1}, Y_{s1}, and Z_{s1}, as shown in Figure 2-1. For any instant of time, the satellite's position in its orbit can be calculated from its orbital parameters, which are part of the broadcast ephemeris (Ghilani and Wolf , 2012: 337).

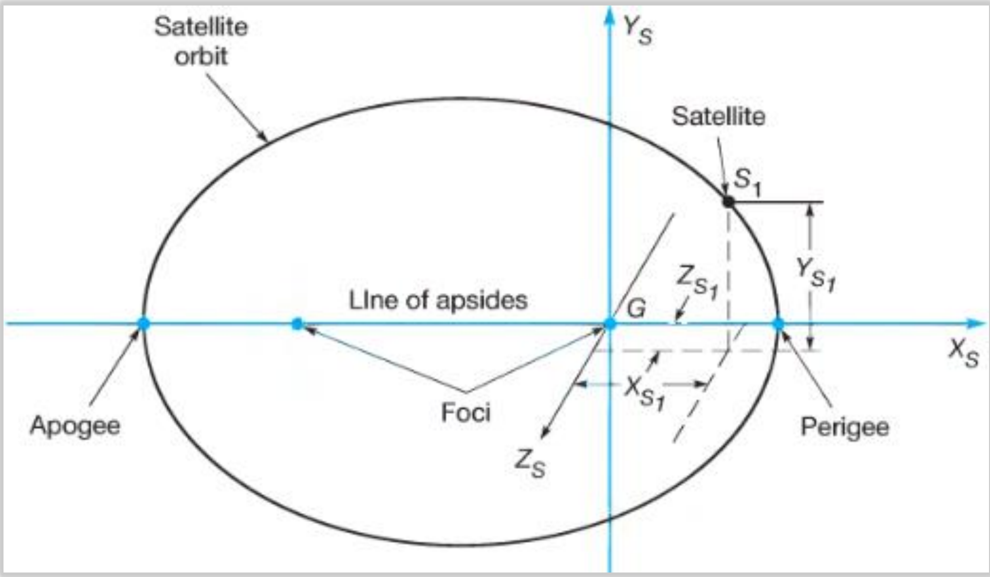


Figure 2-1: Satellite Coordinate Reference System (Ghilani and Wolf , 2012)

2.1.2 Geocentric Coordinate System

Because the objective of satellite surveys is to locate points on the surface of the Earth, it is necessary to have a so-called terrestrial frame of reference, which enables relating points physically to the Earth. The frame of reference used for this is the geocentric coordinate system. Figure 2-2 illustrates a quadrant of a reference ellipsoid[4], with a geocentric coordinate system (X_e,Y_e,Z_e) superimposed. This three-dimensional rectangular coordinate system has its origin at the mass center of the Earth. Its X_e axis passes through the Greenwich meridian in the plane of the equator, and its Z_e axis coincides with the Conventional Terrestrial Pole (CTP) (Ghilani and Wolf , 2012: 338).

To make the conversion from the satellite reference coordinate system to the geocentric system, four angular parameters are required which define the relationship between the satellite's orbital coordinate system and key reference planes and lines on the Earth. As shown in Figure 2-2, these parameters are (1) the inclination angle, i (angle between the orbital plane and the Earth's equatorial plane), (2) the argument of perigee, ω (angle in the orbital plane from the equator to the line of apsides), (3) the right ascension of the ascending node, Ω (angle in the plane of the Earth's equator from the vernal equinox to the line of intersection between the orbital and equatorial planes), and (4) the Greenwich hour angle of the vernal equinox, GHA_γ (angle in the equatorial plane from the Greenwich meridian to the vernal equinox). These parameters are known in real time for each satellite based upon predictive mathematical modelling of the orbits. Where higher accuracy is needed, satellite coordinates in the geocentric system for specific epochs of time are determined from observations at the tracking stations and distributed in precise ephemerides (Ghilani and Wolf , 2012: 339).

---

[4] The reference ellipsoid used for most GPS work is the World Geodetic System of 1984 (WGS84) ellipsoid. As explained in Section 19.1, any ellipsoid can be defined by two parameters, for example the semi-major axis (a), and the flattening ratio (f). For the WGS84 ellipsoid these values are a = 6,378,137 m (exactly), and/ = 1/298.257223563.

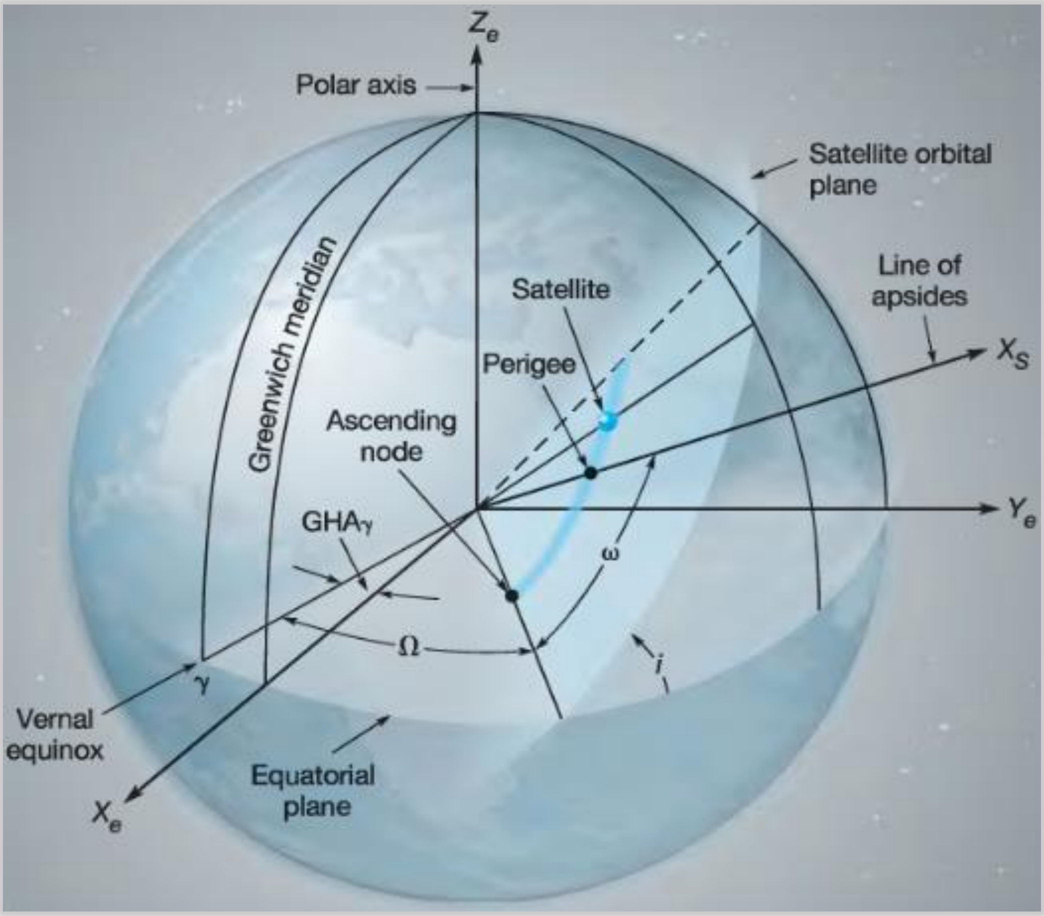


Figure 2-2: Transformation parameters (Satellite Coordinate System to Geocentric Coordinate System) (Ghilani and Wolf , 2012)

2.1.3 Geodetic Coordinate System

Although the positions of points in a satellite survey are computed in the geocentric coordinate system described in the preceding subsection, in that form they are inconvenient for use by surveyors (geomatics engineers). This is the case for three reasons: (1) with their origin at the Earth's center, geocentric coordinates are typically extremely large values, (2) with the X-Y plane in the plane of the equator, the axes are unrelated to the conventional directions of north-south or east-west on the surface of the Earth, and (3) geocentric coordinates give no indication about relative elevations between points. For these reasons, the geocentric coordinates are converted to geodetic coordinates of latitude (ϕ), longitude (λ), and height (h) so that reported point positions become more meaningful and convenient for users (Ghilani and Wolf , 2012: 339).

Figure 2-3 also illustrates a quadrant of the reference ellipsoid, and shows both the geocentric coordinate system (X,Y,Z), and the geodetic coordinate system (ϕ,λ, h). Conversions from geocentric to geodetic coordinates, and vice versa are readily made. From the figure it can be shown that geocentric coordinates of point P can be computed from its geodetic coordinates using the following equations (Ghilani and Wolf , 2012: 339):

$$X_P = (R_{N_P} + h_P)\cos\phi_P \cos\lambda_P$$
$$Y_P = (R_{N_P} + h_P)\cos\phi_P \sin\lambda_P$$
$$Z_P = [R_{N_P}(1 - e^2) + h_P]\sin\phi_P \qquad \textbf{(2-1)}$$

where

$$R_{N_P} = \frac{a}{\sqrt{1 - e^2 \sin^2 \phi_P}} \qquad \textbf{(2-2)}$$

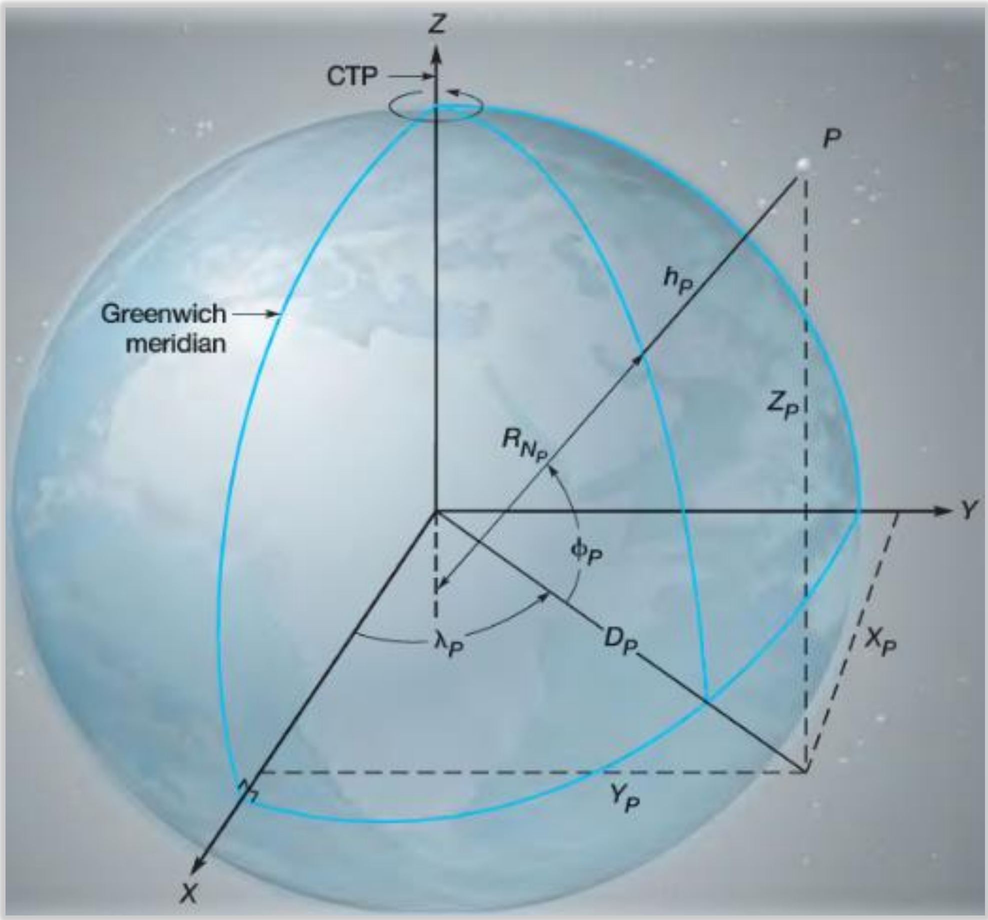


Figure 2-3: Geodetic and Geocentric Coordinate Systems (Ghilani and Wolf , 2012)

In Equations (2.1), X_P, Y_P, and Z_P are the geocentric coordinates of any point P, and the term e, which appears in both Equations (2.1) and (2.2), is the eccentricity of the WGS84 reference ellipsoid. Its value is 0.08181919084. In Equation (2.2), R_{Np} is the radius in the prime vertical of the ellipsoid at point P, and a, as noted earlier, is the semimajor axis of the ellipsoid. In Equations (2.1) and (2.2), north latitudes are considered positive and south latitudes negative. Similarly, east longitudes are considered positive and west longitudes negative (Ghilani and Wolf , 2012: 340).

Example 2.1: The geodetic latitude, longitude, and height of a point A are 41°15'18.2106"N, 75°00'58.6127" W, and 312.391 m, respectively. Using WGS84 values, what are the geocentric coordinates of the point?

Solution:

Substituting the appropriate values into Equations (13.1) and (13.2) yields

$$R_{N_A} = \frac{6{,}378{,}137}{\sqrt{1 - 0.0066943799 \sin^2(41°15'18.2106'')}} = 6{,}387{,}440.3113 \text{ m}$$

$$X_A = (6{,}387{,}440.3113 + 312.391) \cos 41°15'18.2106'' \cos(-75°00'58.6127'')$$
$$= 1{,}241{,}581.343 \text{ m}$$

$$Y_A = (6{,}387{,}440.3113 + 312.391) \cos 41°15'18.2106'' \sin(-75°00'58.6127'')$$
$$= -4{,}638{,}917.074 \text{ m}$$

$$Z_A = [6{,}387{,}440.3113(1 - 0.00669437999) + 312.391)] \sin(41°15'18.2106'')$$
$$= 4{,}183{,}965.568 \text{ m}$$

Conversion of geocentric coordinates of any point P to its geodetic values is accomplished using the following steps (refer again to Figure 2-3).

Step 1: Compute D_P as

$$D_P = \sqrt{X_P^2 + Y_P^2}$$

Step 2: Compute the Longitude as

$$\lambda_P = 2 \tan^{-1}\left(\frac{D_P - X_P}{Y_P}\right)$$

Step 3: Calculate approximate latitude.

$$\phi_0 = \tan^{-1}\left[\frac{Z_P}{D_P(1 - e^2)}\right]$$

Step 4: Calculate the approximate radius of the prime vertical, R_N, using ϕ_o from step 3, and Equation (2.2).

Step 5: Calculate an improved value for the latitude from

$$\phi = \tan^{-1}\left(\frac{Z_P + e^2 R_{N_P} \sin(\phi_0)}{D_P}\right)$$

Step 6: Repeat the computations of steps 4 and 5 until the change in ϕ_P between iterations becomes negligible. This final value, ϕ_P, is the latitude of the station.

Step 7: Use the following formulas to compute the geodetic height of the station. For latitudes less than 45°, use

$$h_P = \frac{D_P}{\cos(\phi_P)} - R_{N_P}$$

For latitudes greater than 45° use the formula

$$h_P = \left[\frac{Z_P}{\sin(\phi_P)}\right] - R_{N_P}(1 - e^2)$$

2.2 WGS-84, Datum Transformations and GPS Time System

2.2.1 World Geodetic System 1984 (WGS 84)

The advent of satellite navigation has enabled the position of points across the whole of the Earth's surface to be measured with respect to a common reference, the satellite constellation, leading to the development of global ellipsoid models. The two main standards are the World Geodetic System 1984 (WGS 84) and the International Terrestrial Reference Frame (ITRF). Both of these datums have their origin at the Earth's center of mass and define rotation using the IRP/CTP (Groves, 2008).

An ellipsoid designed to fit the shape of the entire earth as well as possible with a single ellipsoid. It is often used as a reference on a worldwide basis, while other ellipsoids are used locally to provide a better fid to the earth in a local region. GNSS uses the center of the WGS84 ellipsoid as the center of the GNSS ECEF reference frame (Jeffrey, 2010).

WGS 84 was developed by the Defense Mapping Agency, now the National Geospatial-Intelligence Agency (NGA), as a standard for the U.S. military and is a refinement of predecessors WGS 60, WGS 66, and WGS 72. Its use for GPS and in most INSs led to its adoption as a global standard for navigation systems. WGS 84 was originally realized with 1691 Transit position fixes, each accurate to 1–2m and was revised in the 1990s using GPS measurements and ITRF data. As well as defining an ECEF coordinate frame and an ellipsoid, WGS 84 provides models of the Earth's geoid and gravity field and a set of fundamental constants. WGS84 defines the ellipsoid in terms of the equatorial radius and the flattening. The polar radius and eccentricity may be derived from this. The values are (Groves, 2008)

- $R_0 = 6{,}378{,}137.0$ m, $f = 1 / 298.257223563$
- $R_P = 6{,}356{,}752.3142$ m, $e = 0.0818191908425$

The ITRF is maintained by the IERS and is the datum of choice for the scientific community, particularly geodesists. It is based on a mixture of measurements from satellite laser ranging, lunar laser ranging, very long baseline interferometry (VLBI), and GPS. ITRF is more precise than WGS 84, though the revision of the latter in the 1990s brought the two into closer alignment and WGS 84 is now considered to be a realization of the ITRF. Galileo will use a realization of the ITRF known as the Galileo terrestrial reference frame (GTRF), while GLONASS plans to switch to an ITRF-based datum from the PZ-90 datum. All datums must be regularly updated to account for plate tectonic motion, which causes the position of all points on the surface to move by a few centimeters each year with respect to the center of the Earth (Groves, 2008).

The ellipsoid in WGS84 is defined through four parameters, with specified variances (Strang and Borre, 1997):

1. the semi major axis a = 6 378 137 m (σ_a= 2 m)
2. the Earth's gravitational constant (including the mass of the Earth's atmosphere) kM = 3 986 005 x 10^8 m^3/s^3 (σ_{kM}= 0.6 x 10^8 m^3/s^3)
3. the normalized second degree zonal coefficient of the gravity potential $\bar{C}_{20}$
4. the Earth's rotational rate ω = 7 292 115 x 10^{-11} rad/s. (σ_ω = 15 x 10^{-11} rad/s).

The International Astronomical Union uses σ_ω = 7 292 115.1467 x 10^{-11} rad/s, with four extra digits, together with a new definition of time. In order to maintain consistency with GPS it is necessary to use ω_e instead of ω. The speed of light in vacuum is taken as

C = 299 792 458 m/s with σ_c = 1.2 m/s

Conceptually WGS 84 is a very special datum as it includes a model for the gravity field. The description is given by spherical harmonics up to degree and order 180. This adds 32 755 more coefficients to WGS 84 allowing for determination of the global features of the geoid. In North America the transformation from NAD 27 to WGS 84 is given as

$$\begin{bmatrix} X_{WGS\,84} \\ Y_{WGS\,84} \\ Z_{WGS\,84} \end{bmatrix} = \begin{bmatrix} X_{NAD\,27} + 9\,m \\ Y_{NAD\,27} - 161\,m \\ Z_{NAD\,27} - 179\,m \end{bmatrix}$$

A typical datum transformation into WGS 84 only includes changes in the semi major axis of the ellipsoid and its flattening and three translations of the origin of the ellipsoid.

WGS 84 is a global datum, allowing us to transform between regions by means of GPS. Yet, we finish by a warning about such transformation formulas. They shall only be used with caution. They are defined to an accuracy of a few meters, though the mathematical relationship is exact. The importance of WGS 84 is undoubtedly to provide a unified global datum (Strang and Borre, 1997).

2.2.2 Datum Transformations

Although most navigation systems now use the WGS 84 datum, many maps are based on older datums. Consequently, it may be necessary to transform curvilinear position from one datum to another. The datums may use different origins, axis alignments, and scalings, as well as different radii curvature. No conversion between WGS 84 and ITRF position is needed, as the differences between the two datums are less than the uncertainty bounds (Groves, 2008).

Positions with respect to horizontal and vertical datums have been determined independent of each other. In addition, horizontal datums were nongeocentric and were selected to best fit certain regions of the world. As such, those datums were commonly called local datums. More than 150 local datums have been used by different countries of the world. An example of the local datums is the North American datum of 1927 (NAD 27). With the advent of space geodetic positioning systems such as GPS, it is now possible to determine global 3-D geocentric datums (El-Rabbany, 2006).

Old maps were produced with the local datums, while new maps are mostly produced with the geocentric datums. Therefore, to ensure consistency, it is necessary to establish the relationships between the local datums and the geocentric datums, such as WGS 84. Such a relationship is known as the datum transformation. NIMA has published the transformation parameters between WGS 84 and the various local datums used in many countries. Many GPS manufacturers currently use these parameters within their processing software packages. It should be clear, however, that these transformation parameters are only approximate and should not be used for precise GPS applications. In Toronto, for example, a difference as large as several meters in the horizontal coordinates is obtained when applying NIMA.s parameters (WGS 84 to NAD 27) as compared with the more precise National Transformation software (NTv2) produced by Geomatics Canada. Such a difference could be even larger in other regions. The best way to obtain the transformation parameters is by comparing the coordinates of well-distributed common points in both datums (El-Rabbany, 2006).

2.3 GPS Time Systems

Time plays a very important role in positioning with GPS. As explained in Chapter 1, the GPS signal is controlled by accurate timing devices, the atomic satellite clocks. In addition, measuring the ranges (distances) from the receiver to the satellites is based on both the receiver and the satellite clocks. GPS is also a timing system, that is, it can be used for time synchronization (El-Rabbany, 2006).

A number of time systems are used worldwide for various purposes. Of these, the Coordinated Universal Time (UTC) and the GPS Time are the most important to GPS users. UTC is an atomic time scale based on the International Atomic Time (TAI). TAI is a uniform time scale, which is computed based on independent time scales generated by atomic clocks located at various timing laboratories throughout the world. In surveying and navigation, however, a time system with relation to the rotation of the Earth, not the atomic time, is desired. This is achieved by occasionally adjusting the UTC time scale by 1-second increments, known as leap seconds, to keep it within 0.9 second of another time scale called the Universal Time 1 (UT1), where UT1 is a universal time that gives a measure of the rotation of the Earth. Leap seconds are introduced occasionally, on either June 30 or December 31. As of July 2001, the last leap second was introduced on January 1, 1999, which made the difference between TAI and UTC time scales to be exactly 32 seconds (TAI is ahead of UTC). Information about the leap seconds can be found at the U.S. Naval Observatory web site, http://maia.usno.navy.mil (El-Rabbany, 2006).

GPS Time is the time scale used for referencing, or time tagging, the GPS signals. It is computed based on the time scales generated by the atomic clocks at the monitor stations and onboard GPS satellites. There are no leap seconds introduced into GPS Time, which means that GPS Time is a continuous time scale. GPS Time scale was set equal to that of the UTC on January 6, 1980 [8]. However, due to the leap seconds introduced into the UTC time scale, GPS Time moved ahead of the UTC by 13 seconds on January 1, 1999. The difference between GPS and UTC time scales is given in the GPS navigation message. It is worth mentioning that, as shown in Chapter 3, both GPS satellite and receiver clocks are offset from the GPS Time, as a result of satellite and receiver clock errors (El-Rabbany, 2006).

GPS time is used as the primary time reference for all GPS operation. GPS time is referenced to a universal coordinated time (UTC). The acronym UTC is an English-French mixture for Coordinated Universal Time (CUT) in English or Temps Universal Coordonne (TUC) in French. It was internationally agreed to write Universal Coordinated Time as UTC, rather than CUT or TUC, making it language-independent. The GPS zero time is defined as midnight on the night of January 5/morning of January 6, 1980. The largest unit used in stating GPS time is one week, defined as 604,800 seconds (7 × 24 × 3600). The GPS time may differ from UTC because GPS time is a continuous time scale, while UTC is corrected periodically with an integer number of leap seconds. The GPS time scale is maintained to be within one μs of UTC (modulo of one second). This means the two times can be different by an integer number of seconds (Tsui, 2005).

In each satellite, an internally derived 1.5-second epoch, the Z count, provides a convenient unit for precise counting and communication time. The Z count has 29 bits consisting of two parts: the 19 least-significant bits (LSBs) referred to as the time of the week (TOW) and the 10 most-significant bits (MSBs) as the week number. In the actual data transmitted by the satellite, there are only 27 Z count bits. The 10-bit week number is in the third word of subframe 1. The 17-bit TOW is in the HOW in every subframe. The two LSBs are implied through multiplication of the truncated Z count (Tsui, 2005).

Table 3.1 indicates that the satellite rotates around the earth twice in a sidereal day. The sidereal day is slightly different from an apparent solar day. The apparent day has 24 hours and it is the time used daily. The apparent solar day is measured by the time between two successive transits of the sun across our local meridian, because we use the sun as our reference. A sidereal day is the time for the earth to turn one revolution. Figure 3.2 shows the difference between the apparent solar day and a sidereal day. In this figure, the effect is exaggerated and it is obvious that a sidereal day is slightly shorter than a solar day. The difference should be approximately equal to one day in one year which corresponds to about 4 min ($24 \times 60/365$) per day. The mean sidereal day is 23 hrs, 56 min, 4.09 sec.

The difference from an apparent day is 3 min, 55.91 sec. Half a sidereal day is 11 hrs, 58 min, 2.05 sec. This is the time for the satellite to rotate once around the earth. From this arrangement one can see that from one day to the next a certain satellite will be at approximately the same position at the same time. The location of the satellite will be presented in the next section.

(Tsui, 2005)

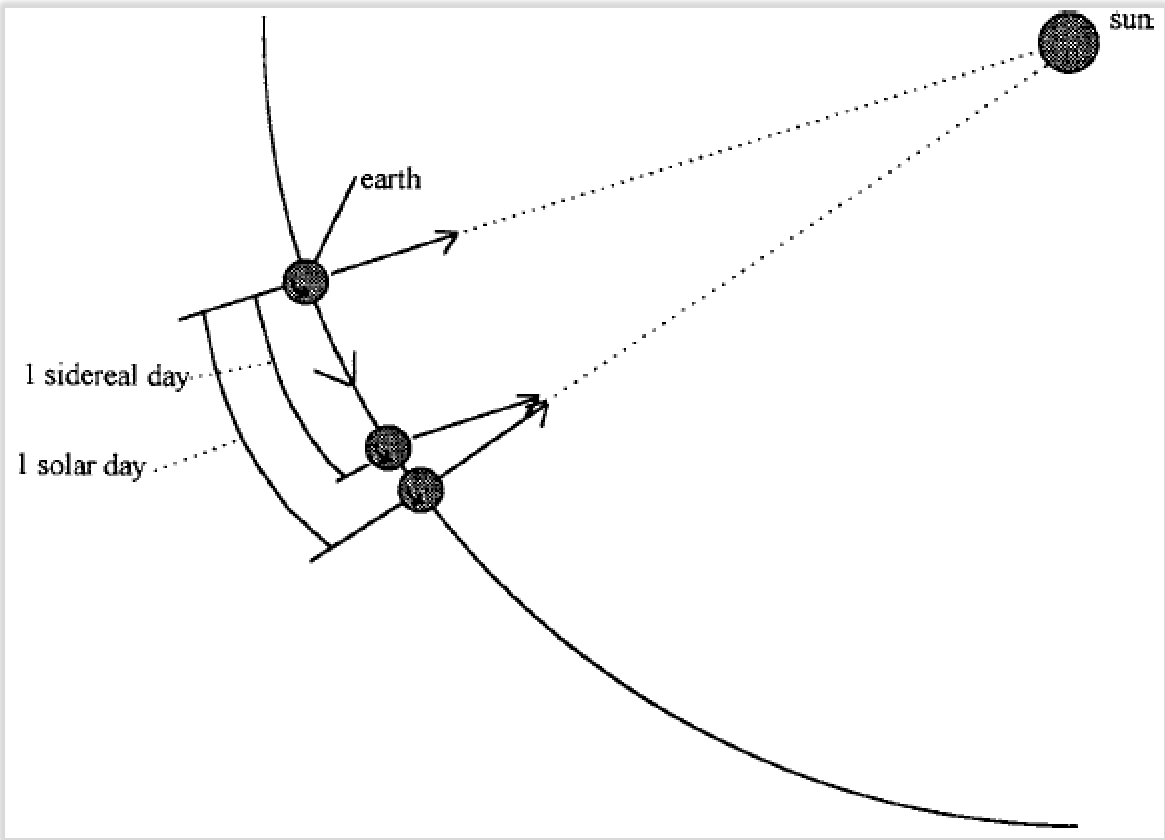


Figure 2-4: Configuration of Apparent Solar Day and Sidereal Day (Tsui, 2005)

Three basic groups of time scales are of importance in satellite geodesy (Seeber, 2003: 31):

(1) The time-dependent orientation of Earth with respect to the inertial space is required in order to relate the Earth-based observations to a space-fixed reference frame. The appropriate time scale is connected with the diurnal rotation of Earth, and is called *Sidereal Time* or *Universal Time*.

(2) For the description of the satellite motion we need a strictly uniform time measure which can be used as the independent variable in the equations of motion. An appropriate time scale can be derived from the orbital motion of celestial bodies around the Sun. It is called *Ephemeris Time*, *Dynamical Time*, or *Terrestrial Time*.

(3) The precise measurement of signal travel times, e.g. in satellite laser ranging, requires a uniform and easily accessible time scale with high resolution. The appropriate measure is related to phenomena in nuclear physics and is called *Atomic Time*.

All these time scales are based on the observation of uniform and repetitive astronomical or physical phenomena. The time interval between two consecutive phenomena forms the scale measure of the particular time scale. A certain multiple or fraction of the scale measure is called the *time unit*. In general, the *second* (s) is used as the basic time unit. Larger time units, such as days or years, are derived from the second (Seeber, 2003).

Within the time scale a starting point or origin has to be fixed. This may be achieved through a certain astronomical event, such as the particular position of a star, or the meridian transit of a particular celestial object (Seeber, 2003).

The instant of the occurrence of some phenomena or observations can be related to a certain reading of the particular time scale, and gives the *datation* of the event. In astronomy such an event is called the *epoch* of the observation. With respect to the particular time scale the epoch determination reflects an *absolute time measurement*. For many purposes, e.g. for the determination of signal travel times, a *relative time measurement*, i.e. the determination of the time interval between two epochs, is sufficient. In many cases the relative time measurement can be done much more accurately than the absolute time measurement. In satellite geodesy the datation of an event is often called *time-tag* or *time-tagging*, e.g. when the instant of transmission or reception of a signal is considered (Seeber, 2003).

Strictly speaking, we have to distinguish between the ideal conception of a time scale and the practical realization through observations. This becomes particularly evident with the atomic time, when we compare the definition of the atomic time second with its practical realization through a group of individual atomic clocks. A time scale may be regarded as an approximation to the particular time concept. In the following we will not use this distinction (Seeber, 2003).

In order to meet the various requirements, stemming from science and technology, the relationship between the different time scales have to be established with the highest possible accuracy. Fig. 2.12 illustrates how timing errors in satellite geodesy are related to a position error of 1 cm:

1 cm motion of a point on the equator caused by Earth's rotation corresponds to about 2×10^{-5} s,
1 cm motion of a near-Earth satellite in the orbit corresponds to about 1×10^{-6} s,
1 cm in the satellite range derived from signal travel time (e.g. laser ranging) corresponds to about 1×10^{-10} s.

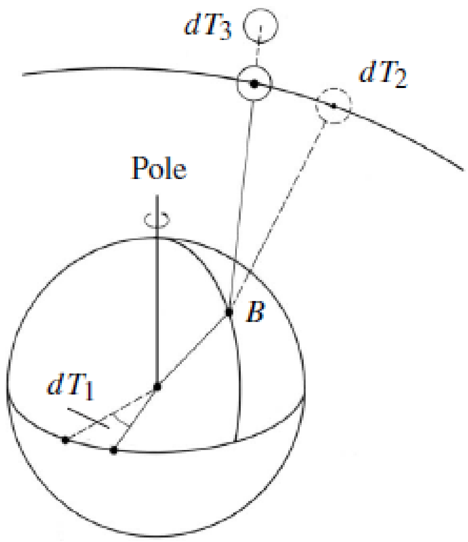


Figure 2.12. Effect of timing errors in satellite geodesy

The related requirements for the accuracy of time determination dT_i are as follows:

$$
\begin{aligned}
&dT_1[\text{s}] \le 2 \times 10^{-5} && \text{for Earth rotation,} \\
&dT_2[\text{s}] \le 1 \times 10^{-6} && \text{for orbital motion, and} \\
&dT_3[\text{s}] \le 1 \times 10^{-10} && \text{for signal travel time.}
\end{aligned}
\qquad (2.50)
$$

(Seeber, 2003)

2.3.1 Sidereal Time and Universal Time

Sidereal time and universal time are directly related to the rotation of Earth, and they are thus equivalent time scales. Sidereal time equals the hour angle of the vernal equinox ϒ, and consequently depends on the geographical longitude of the particular observation station. From Figure 2-5 we may easily derive the following relations. The Local Apparent (or True) Sidereal Time (LAST), referred to the true vernal equinox, is

LAST = Local hour angle of the true vernal equinox.

For Greenwich we obtain the *Greenwich Apparent Sidereal Time* (GAST)

GAST = Greenwich hour angle of the true vernal equinox.

The vernal equinox is subject to the nutation in longitude. Removing the nutation term, we obtain the *Local Mean Sidereal Time* (LMST), and the *Greenwich Mean Sidereal Time (GMST), respectively*

LMST = Local hour angle of the mean vernal equinox

GMST = Greenwich hour angle of the mean vernal equinox.

The difference between the apparent and mean sidereal times is termed Equation of Equinoxes (Eq.E)

$$GMST - GAST = \Delta \Psi \cos \varepsilon \qquad (2.51)$$

with $\Delta \Psi$ *(2.21) the nutation in longitude. For the east longitude* Λ *of the local meridian*

$$LMST - GMST = LAST - GAST = \Lambda \qquad (2.52)$$

The apparent sidereal time is used for the evaluation of astronomical observations. However, for the construction of a time scale, only the mean sidereal time is used. The fundamental unit is the *Mean Sidereal Day*, defined as the interval between two consecutive upper transits of the mean vernal equinox across the meridian. The mean sidereal day does not correspond exactly to a complete revolution of Earth on its spin axis with respect to inertial space, because the position of the vernal equinox is affected by precession. The daily difference is 0.s0084, with the sidereal day being shorter (Seeber, 2003).

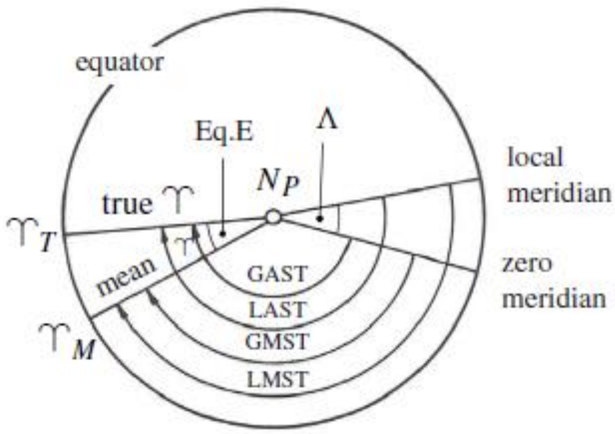


Figure 2-5: Definition of Sidereal Time

For practical purposes a time scale is required which corresponds to the apparent diurnal motion of the Sun. The hour angle of the true Sun experiences rather large variations during the year, caused by the changing declination of the Sun and the ellipticity of Earth's orbit. Consequently, this measure is not suitable for a uniform time scale. It is therefore substituted by a fictitious *Mean Sun*, which moves in the plane of the equator with constant velocity. The *Mean Solar Day* is thus defined as the interval between two successive transits of the mean fictitious Sun across the meridian (Seeber, 2003).

Mean Solar Time is measured by the hour angle of the mean Sun. The Greenwich hour angle of the mean Sun is called *Universal Time* (UT). For practical purposes the day starts at midnight, hence

$$\text{UT} = 12\text{h} + \text{Greenwich hour angle of the mean Sun} \qquad (2.53)$$

Both concepts of time are based on Earth's rotation and they are closely connected to each other. Universal time can be considered as a special form of sidereal time. The difference in the length of the day for both definitions is around 4 minutes, because the diurnal motion of Earth in its orbit amounts to $360^\circ/365 \approx 1^\circ$. The approximate relation is

$$1 \text{ mean sidereal day} = 1 \text{ mean solar day} - 3^{m}55.^{s}909 \qquad (2.54)$$

The raw universal time UT0_B, which is obtained from observations at a particular station *B*, is still affected by the location-dependent influences of the actual true pole position. The reduction to the conventional terrestrial pole (CTP) causes a change, $\Delta\Lambda_P$, in longitude, and hence in time. The universal time, which is referred to CTP, is termed (Seeber, 2003)

$$\text{UT1} = \text{UT0}_B + \Delta\Lambda_P. \qquad (2.55)$$

UT1 is the fundamental time scale in geodetic astronomy and satellite geodesy, because it defines the actual orientation of the conventional terrestrial system with respect to space. UT1 is also the basic time scale for navigation. UT1 contains, however, all variations of Earth's rotation rate, and is thus not a uniform time scale.

2.3.2 Atomic Time

The international atomic time scale TAI (Temps Atomique International) was introduced to meet the requirements for an easily accessible and strictly uniform time scale. The unit of the atomic time was selected in such a way that it equals the duration of the ephemeris second. The definition of the second of the atomic time scale has been worded by the 13th Conference of the International Committee of Weights and Measures in Paris, 1967, as follows (Seeber, 2003):

The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the Cesium 133 atom.

This is also the definition of the unit of time of the *International System of Units* (SI).

The international atomic time scale is maintained by the Time Section of the International Bureau of Weights and Measures (*Bureau International des Poids et Mesures*, BIPM) in Paris, based on the readings of a large number of the most accurate atomic clocks in various laboratories. The Bureau International de l'Heure (BIH) was responsible for maintaining the atomic time scales until the 31st of December 1987 (Seeber, 2003).

In practice, atomic time scales are derived from groups of commercial and laboratory cesium standards which generate time intervals, based on the definition of the SI second. The readings refer to non-moving clocks at sea level. TAI is computed as the weighted mean of individual clocks (about 250 clocks in 2002). TAI is hence a statistically formed common time scale for international use. Each laboratory time scale can be regarded as a particular realization of the atomic time scale. The differences between TAI and the time scales of the participating laboratories are distributed on a monthly basis in the Circular T of the BIPM Time Section (Seeber, 2003).

The epoch of TAI agreed with the epoch of UT1 on January 1, 1958. Due to the deceleration of Earth's rotation the difference between the time scales is increasing. The difference, for some selected dates, amounts to (Seeber, 2003)

$$
\begin{aligned}
\mathrm{TAI} - \mathrm{UT1} &= +\ 6.^{s}1 \quad \text{on January 1, 1968} \\
&= +16.^{s}4 \quad \text{on January 1, 1978} \\
&= +23.^{s}6 \quad \text{on January 1, 1988} \\
&= +30.^{s}8 \quad \text{on January 1, 1998} \\
&= +31.^{s}9 \quad \text{on January 1, 2001} \\
&= +32.^{s}3 \quad \text{on January 1, 2003.}
\end{aligned}
$$

The rather large size of the differences stems from the fact that the unit of the SI-second was adopted from the length of the ephemeris second; and the ephemeris second was derived from the mean duration of the solar day between 1756 and 1895, when Earth's rotation was faster than today (Seeber, 2003).

For many applications, navigation in particular, a time scale is required which provides both a highly uniform time unit and the best possible adaptation to UT1, and hence to Earth rotation. This is why, in 1972, a compromise time scale, *Universal Time Coordinated* (UTC), was introduced. UTC and TAI differ by an integer number *n* of seconds (Seeber, 2003)

$$\mathrm{UTC} = \mathrm{TAI} - n \cdot (1\ \mathrm{s}) \tag{2.57}$$

Depending on the prevailing situation, *n* can be changed at given dates, namely on January 1 and/or July 1. Thus the epoch of UTC is adapted to UT1 by inserting or removing so-called *leap seconds*. The unit of UTC remains the SI second. The difference, DUT1, between both times should not exceed 0.9 s in absolute value

$$|\mathrm{DUT1}| = |\mathrm{UT1} - \mathrm{UTC}| \overset{!}{\leq} 0.9\,\mathrm{s}. \tag{2.58}$$

DUT1 is distributed through the bulletins of the IERS, and it must be taken into account with all calculations related to Earth-fixed reference systems. In most countries the disseminated time signals refer to UTC. On January 1, 2003 the difference between TAI and UTC was

$$\mathrm{TAI} - \mathrm{UTC2003} = +32\ \mathrm{s} \tag{2.59}$$

The Global Positioning System (GPS) uses its own particular time scale *GPS time*. It differs from UTC by a nearly integer number of seconds. Both time scales had identical epochs on January 5, 1980. Because GPS time is not incremented by leap seconds the difference between UTC and GPS time is increasing. The unit of GPS time is the SI second. However, GPS time is only derived from atomic clocks which form part of the GPS system. It is hence a "free" atomic time scale and may show slight differences when compared to TAI. The relation between UTC and GPS time is included in time bulletins of the USNO and the BIPM, and it is also disseminated within the "GPS satellite message". In 2003 the difference was approximately (Seeber, 2003)

$$\text{GPS time} - \mathrm{UTC}_{2003} = +13\ \mathrm{s} \tag{2.60}$$

The exact relation is (e.g. BIPM (2002))

GPS time − UTC = n s − C_0,

where n is an integer number, and the correction term C_0 is in the order of several nanoseconds. Thus the reception of GPS signals provides real-time access to TAI and UTC with uncertainties below 1 microsecond (Seeber, 2003).

A similar relationship holds for GLONASS time and UTC. Note that UTC and GPS time, as well as GLONASS time are atomic time scales (Seeber, 2003).

2.3.3 Ephemeris Time, Dynamical Time, Terrestrial Time

In 1952 the IAU introduced *Ephemeris Time* (ET) as a theoretically uniform time scale for use with ephemeris. The *Ephemeris Second* was defined as a certain fraction of the *Tropical Year* 1900, and hence it was strictly uniform. In practice, the ephemeris time was derived from lunar observations, and it depended on a theory of the Sun and the system of astronomical constants. Its reading accuracy was only about 0.1 s on yearly averages. ET has never been disseminated by time signals. It was made available only through the publication of differences with respect to UT1, and later to TAI (Guinot, 1995 in (Seeber, 2003)).

In 1977 the IAU adopted the so-called *Dynamical Time Scales* in order to meet the arising requirements for a relativistic formulation of orbital motion. *Barycentric Dynamical Time* (TDB) was defined to be the time-like argument for the barycentre of the solar system, and *Terrestrial Dynamic Time* (TDT) was referred to geocentric ephemerides (Seeber, 2003).

In the concept of *General Relativity* a clock, moving with Earth, experiences periodic variations up to 1.6 milliseconds, caused by the annual motion within the gravity field of the Sun. This effect, however, must not be considered in the computation of near-Earth satellite orbits, because the satellites move together with Earth. This is why *Terrestrial Dynamical Time* (TDT) was the appropriate time scale for geocentric calculations in satellite geodesy. A further advantage is, that compared with the *Barycentric Dynamical Time* (TDB), TDT is independent of various forms of relativistic theories (Seidelmann et al., 1992 in (Seeber, 2003)).

Dynamical time has been used as the argument for astronomical ephemerides since January 1, 1984. The SI second was formally introduced as the fundamental time unit in the TDT scale. It corresponds to the time which an atomic clock would measure on the rotating geoid. For the sake of continuity TDT was set equal to ET at the beginning of January 1, 1984. This is why a constant difference of 32.s184 exists between the TAI time scale, and the TDT (or ET) time scale (Seeber, 2003).

In 1991 the IAU has defined new time scales in the framework of the general theory of relativity to clarify the relationships between *space-time coordinates*. In this concept a time scale is regarded as one of the coordinate axes of a space-time reference frame (Guinot, 1995). The new time scales are the *Barycentric Coordinate Time* (TCB), the *Geocentric Coordinate Time* (TCG), and the *Terrestrial Time* (TT). Explicit formulas relating TDB, TDT, TCB, and TCG are given in Seidelmann et al. (1992) (Seeber, 2003).

Terrestrial Time (TT) in essence is a new denomination for TDT. The word "dynamical" has been omitted because TT as an idealized time scale is no longer based on dynamical theories. TT is the time reference for geocentric ephemerides and hence the primary time scale for the relativistic treatment of near-Earth satellite orbits. TT differs from TCG only by a constant rate (Seeber, 2003)

$$dTT/dTCG = 1 - LG \quad (2.61)$$

where LG = 6.969290134 × 10–10 is a defining constant (resolution B1.9, IAU 24th General Assembly 2000). The unit of TT is the SI second, hence TT is realized through the atomic time scale TAI with a constant offset of $32.^s184$ between both scales (Seeber, 2003)

$$TT \equiv TDT \equiv ET = TAI + 32.^s184 \quad (2.62)$$

Consequently, there is only a conceptional, not a practical, difference between both time scales (Seeber, 2003).

2.4 Satellite Orbit Computation and Dissemination

GNSS satellites orbit well above the earth's atmosphere. GPS and GLONASS satellites orbit at altitudes close to 20,000 km. Galileo satellites orbit a bit higher, around 23,000 km. GNSS orbits, which are more or less circular, and highly stable and predictable, fall into the category of MEO (medium earth orbit) (Jeffrey, 2010).

There is not much drag at 20,000 km, but gravitational effects and the pressure of solar radiation do affect GNSS orbits a bit and the orbits have to be occasionally corrected. While its orbit is being adjusted, a GNSS satellite's status is changed to out of service so user equipment knows not us use their signals (Jeffrey, 2010).

GPS satellite orbits are nearly circular (an elliptical shape with a maximum eccentricity is about 0.01), with an inclination of about 55° to the equator. The semi-major axis of a GPS orbit is about 26,560 km (i.e., the satellite altitude of about 20,200 km above the Earth's surface) [4]. The corresponding GPS orbital period is about 12 sidereal hours (~11 hours, 58 minutes). The GPS system was officially declared to have achieved full operational capability (FOC) on July 17, 1995, ensuring the availability of at least 24 operational, non-experimental, GPS satellites. In fact, as shown in Section 1.4, since GPS achieved its FOC, the number of satellites in the GPS constellation has always been more than 24 operational satellites (El-Rabbany, 2006).

The orbit describes the position of a satellite in space. Satellites used for navigation move around the Earth in endless circular or elliptical orbits. The spatial orientation (e.g. orbital inclination, excentricity, length, altitude above the ground) and the parameters of motion (e.g. orbital period) have a significant impact on the usability and performance of these satellites as shown in the Figure 2-6 (Zogg, 2009).

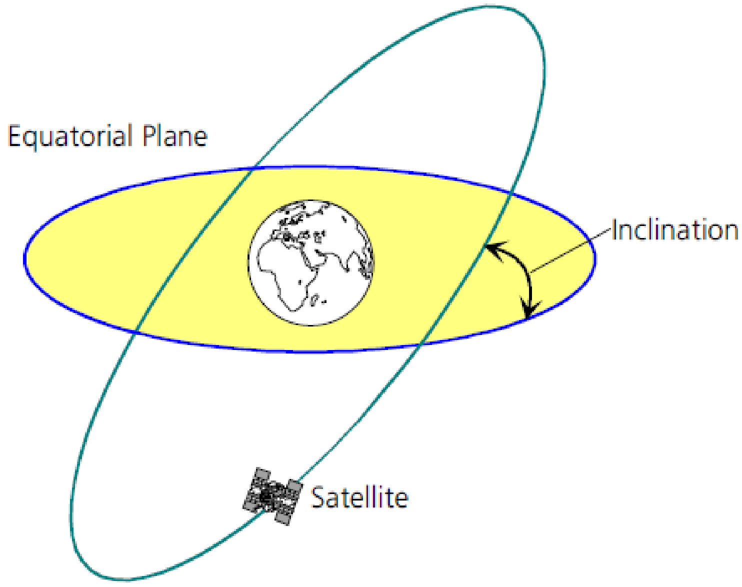


Figure 2-6: Satellite Orbits (Zogg, 2009)

The **inclination**, also referred to as the angle of inclination or the axial tilt, expresses the tilt of the circular or elliptical orbit of the satellite around the Earth relative to the equatorial plane. For example, with an inclination of 90° an orbit would pass directly over the polar caps. All satellite orbits that do not lie along the equatorial plane are referred to as "inclined orbits".

The **Ephemeris** of a satellite is a mathematical description of its orbit. The high precision satellite orbital data is necessary for a receiver to calculate the satellite's exact position in space at any given time. Orbital data with reduced exactness is referred to as an **Almanac**. With the help of the Almanac the receiver can calculate which satellites are visible over the horizon from an approximate position and time. Each satellite transmits its own Ephemeris as well as the Almanacs of all existing satellites. The current Almanac Data can also be viewed over the internet.

The **Elevation** describes the angle of a satellite relative to the horizontal plane. If a satellite is directly above the point of observation on the ground, then the elevation is 90°. If the satellite is at the horizon, then the elevation is 0°.

The **Azimuth** is the angle between a reference plane and a point. In the case of satellites the reference plane is the plane of the horizon based on true North. The Azimuth is the angle between the satellite and true North (North = 0°, East = 90°, South = 180°, West = 270°).

Excentricity defines the so-called Numerical Excentricity "e", which is the deviation of an elliptical satellite orbit (excentric orbit) from a geometrically exact circular orbit. Numerical Excentricity is defined by the equation:

$$e = \sqrt{\frac{a^2 - b^2}{a^2}}$$

where a is the semi-major axis and b is the semi-minor axis of the elliptical orbit. For completely circular orbits the value of e = 0, and approaches 1 the more the length (i.e. the semi-major axis) of the ellipse is stretched relative to the semi-minor axis (Zogg, 2009).

The orbital altitude gives the elevation above the Earth's surface of a point on a circular or elliptical satellite orbit. Originally, commercial communications satellites were preferentially brought into circular equatorial (Inclination 0°) orbits with an altitude of about 36,000 km above the ground. Satellites on this orbit rotate around the earth in 24 hours (orbital period: 24 hours), so that there is no relative movement with respect to the Earth. For this reason such satellites are also referred to as Geosynchronous (**GEO**) satellites, with an orbit referred to as Geostationary. GEO satellites are used by communications satellite systems such as Inmarsat and Thuraya as well as SBAS systems such as WAAS and EGNOS (Zogg, 2009).

In addition to the relatively high altitude GEO satellites, which can provide coverage to large areas of the Earth's surface, other satellite systems (e.g. Iridium, Globalstar, GPS und GALILEO) employ satellites with much lower orbital altitudes. These lower altitude satellites must orbit the Earth with increased speed in order to provide the necessary centrifugal force to compensate for the increased gravitational pull experienced at lower altitudes. In contrast to the GEO satellites, these satellites move relative to the Earth and rotate in so-called Non-Geostationary Satellite Orbits (NGSO) (Zogg, 2009).

Generally, six different categories of orbits are classified (Zogg, 2009):

- Geosynchronous Earth Orbit (**GEO**): geostationary orbit with an altitude of approximately 36,000km
- Medium Earth Orbit (**MEO**): inclined orbit with medium altitude of about 10,000 km
- Low Earth Orbit (**LEO**): low altitude orbit up to approximately 1'000 km
- Highly (Inclined) Elliptical Earth Orbit (**HEO**)
- Inclined Geosynchronous Orbit (**IGSO**)
- Polar Earth Orbit (**PEO**): LEO orbit over the polar caps

Example 1: Determining the altitude of a GEO satellite:

Satellites with a geostationary orbit have a very exact altitude which can be calculated.

The mean siderial[10] Earth day has a duration of 23 hours, 56 minutes, 4.099 seconds = 86164.099 s and represents a geometrically complete rotation of the Earth of 360° in a system with fixed stars.

From Section 3.1.3 we know the formula:

$$h = \sqrt[3]{3.9860042 \bullet 10^{14}\,\frac{\mathrm{m}^3}{\mathrm{s}^2} \bullet \left(\frac{T}{2\pi}\right)^2} - R_e = \sqrt[3]{3.9860042 \bullet 10^{14}\,\frac{\mathrm{m}^3}{\mathrm{s}^2} \bullet \left(\frac{86164.099\,\mathrm{s}}{2\pi}\right)^2} - 6378137\mathrm{m} =$$

$$35786035\mathrm{m} = 35{,}786.035\mathrm{km}$$

Example 2: Determining the orbital period of a GPS satellite.

GPS satellites have a medium level altitude of 20,184.5 km above the Earth. The mean orbital period T of a GPS satellite is determinded by:

$$T = 2 \bullet \pi \bullet \sqrt[2]{\frac{(h+R_e)^3}{3{,}9860042 \bullet 10^{14}\,\frac{m^3}{s^2}}} = 2 \bullet \pi \bullet \sqrt[2]{\frac{(20184500m+6378137m)^3}{3{,}9860042 \bullet 10^{14}\,\frac{m^3}{s^2}}} =$$

$$43084s = 11h\ 58min$$

This represents a half siderial day. Since the Earth also rotates in this time, after two orbits the GPS satellite will find itself over the same point on the Earth's surface.

2.5 Signal Structure

There are 10 different GPS navigation signals, broadcast across three bands, known as link 1 (L1), link 2 (L2), and link 5 (L5). The career frequencies are 1575.42 MHz for L1, 1227.60 MHz for L2 and 1176.45 MHz for L5, while the declared double-sided signal bandwidth is 30.69 MHz in each band. The signals are summarized in *Table 2-1* and their PSDs illustrated by figure 6.11. However, many of these signals are being introduced under the GPS modernization program and are broadcast by all satellites while L1C signals will not be broadcast before 2011. GPS satellites can also transmit a signal on 1381.05 MHz (L3), but this is used by Nuclear Detonation Detection System; it is not designed for navigation [6 in (Groves, 2008)].

The nominal signal powers listed in *Table 2-1* are minimum values. Satellites initially transmit at higher powers, but the power drops as the satellite ages. The Block III satellites will transmit at higher power with equal power in the L1 and L2 bands (Groves, 2008).

Table 2-1: GPS Signal Properties

Signal	*Band and Carrier Frequency (MHz)*	*Service*	*Modulation and Chipping Rate* ($\times$ *1.023 Mchip* s^{-1})	*Navigation Message Rate* (*symbol* s^{-1})	*Minimum Received Signal Power (dBW)*	*Satellite Blocks*
C/A	L1, 1575.42	SPS/PPS	BPSK 1	50	−158.5	All
P(Y)	L1, 1575.42	PPS	BPSK 10	50	−161.5	All
M code	L1, 1575.42	PPS	$BOC_s(10,5)$	Note 2	Note 2	From IIR-M
L1C-d	L1, 1575.42	PPS	$BOC_s(1,1)$ Note 1	100	−163	From III
L1C-p	L1, 1575.42	PPS	$BOC_s(1,1)$ Note 1	None	−158.3	From III
L2C	L2, 1227.60	SPS	BPSK 1	50	−160	From IIR-M
P(Y)	L2, 1227.60	PPS	BPSK 10	50	−164.5	All
M code	L2, 1227.60	PPS	$BOC_s(10,5)$	Note 2	Note 2	From IIR-M
L5I	L5, 1176.45	SPS	BPSK 10	100	−158	From IIF
L5Q	L5, 1176.45	SPS	BPSK 10	None	−158	From IIF

Note 1: This provisional L1C modulation scheme was recently replaced by a modified binary offset carrier (MBOC), comprising a 10/11 power $BOC_s(1,1)$ modulation and 1/11 power $BOC_s(6,1)$ modulation [18].
Note 2: Some details of the M-code signal are not in the public domain. PPS users are directed to the relevant interface control document (ICD) [19].

(Groves, 2008)

Positioning, or finding the user's location, with GPS requires some understanding of the GPS signal structure and how the measurements can be made. Likewise, as the GPS signal is received through a GPS receiver, understanding the capabilities and limitations of the various types of GPS receivers is essential. Furthermore, the GPS measurements, like all measurable quantities, contain errors and biases, which can be removed or reduced by combining the various GPS observables (El-Rabbany, 2006).

Each GPS satellite transmits a microwave radio signal composed of two carrier frequencies (or sine waves) modulated by two digital codes and a navigation message (see Figure 2-7). The two carrier frequencies are generated at 1,575.42 MHz (referred to as the L1 carrier) and 1,227.60 MHz (referred to as the L2 carrier). The corresponding carrier wavelengths are approximately 19 cm and 24.4 cm, respectively, which result from the relation between the carrier frequency and the speed of light in space. The availability of the two carrier frequencies allows for correcting a major GPS error, known as the ionospheric delay. All of the GPS satellites transmit the same L1 and L2 carrier frequencies.

The code modulation, however, is different for each satellite, which significantly minimizes the signal interference (El-Rabbany, 2006).

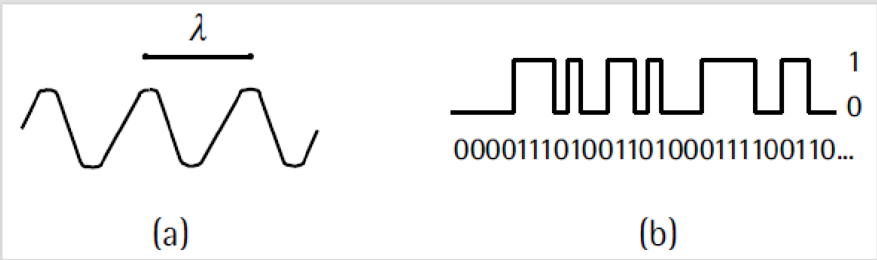


Figure 2-7: (a) Sinusoildal Wave and (b) Digital Code

The two GPS codes are called coarse acquisition (or C/A-code) and precision (or P-code). Each code consists of a stream of binary digits, zeros and ones, known as bits or chips. The codes are commonly known as PRN[5] codes because they look like random signals (i.e., they are noise-like signals). But in reality, the codes are generated using a mathematical algorithm. Presently, the C/A-code is modulated onto the L1 carrier only, while the P-code is modulated onto both the L1 and the L2 carriers. This modulation is called biphase modulation, because the carrier phase is shifted by 180° when the code value changes from zero to one or from one to zero (El-Rabbany, 2006).

The C/A-code is a stream of 1,023 binary digits (i.e., 1,023 zeros and ones) that repeats itself every millisecond. This means that the chipping rate of the C/A-code is 1.023 Mbps. In other words, the duration of one bit is approximately 1ms, or equivalently 300 m. Each satellite is assigned a unique C/A-code, which enables the GPS receivers to identify which satellite is transmitting a particular code. The C/A-code range measurement is relatively less precise compared with that of the P-code. It is, however, less complex and is available to all users (El-Rabbany, 2006).

The P-code is a very long sequence of binary digits that repeats itself after 266 days. It is also 10 times faster than the C/A-code (i.e., its rate is 10.23 Mbps). Multiplying the time it takes the P-code to repeat itself, 266 days, by its rate, 10.23 Mbps, tells us that the P-code is a stream of about 2.35 × 1014 chips! The 266-day-long code is divided into 38 segments; each is 1 week long. Of these, 32 segments are assigned to the various GPS satellites. That is, each satellite transmits a unique 1-week segment of the P-code, which is initialized every Saturday/Sunday midnight crossing. The remaining six segments are reserved for other uses. It is worth mentioning that a GPS satellite is usually identified by its unique 1-week segment of the P-code. For example, a GPS satellite with an ID of PRN 20 refers to a GPS satellite that is assigned the twentieth-week segment of the PRN P-code. The P-code is designed primarily for military purposes. It was available to all users until January 31, 1994. At that time, the P-code was encrypted by adding to it an unknown W-code. The resulting encrypted code is called the Y-code, which has the same chipping rate as the P-code. This encryption is known as the antispoofing (AS) (El-Rabbany, 2006).

The P code is bi-phase modulated at 10.23 MHz; therefore, the main lobe of the spectrum is 20.46 MHz wide from null to null. The chip length is about 97.8 ns (1/10.23 MHz). The code is generated from two pseudorandom noise (PRN)

[5] Pseudorandom Noise

codes with the same chip rate. One PRN sequence has 15,345,000 chips, which has a period of 1.5 seconds, the other one has 15,345,037 chips, and the difference is 37 chips. The two numbers, 15,345,000 and 15,345,037, are relative prime, which means there are no common factors between them. Therefore, the code length generated by these two codes is 23,017,555.5 (1.5 $\times$ 15, 345, 037) seconds, which is slightly longer than 38 weeks. However, the actual length of the P code is 1 week as the code is reset every week. This 38-week-long code can be divided into 37 different P codes and each satellite can use a different portion of the code. There are a total of 32 satellite identification numbers although only 24 of them are in the orbit. Five of the P code signals (33–37) are reserved for other uses such as ground transmission. In order to perform acquisition on the signal, the time of the week must be known very accurately. Usually this time is found from the C/A code signal that will be discussed in the next section. The navigation data rate carried by the P code through phase modulation is at a 50 Hz rate (Tsui, 2005).

The GPS navigation message is a data stream added to both the L1 and the L2 carriers as binary biphase modulation at a low rate of 50 kbps. It consists of 25 frames of 1,500 bits each, or 37,500 bits in total. This means that the transmission of the complete navigation message takes 750 seconds, or 12.5 minutes. The navigation message contains, along with other information, the coordinates of the GPS satellites as a function of time, the satellite health status, the satellite clock correction, the satellite almanac, and atmospheric data. Each satellite transmits its own navigation message with information on the other satellites, such as the approximate location and health status (El-Rabbany, 2006).

The actual modulated signal is divided into two independent components. They are modulated by different bit sequences. The slower Coarse/Acquisition (C/A) code has a chipping rate of 1.023 MHz, and the Precision (P) code is 10 times faster. The C/A code is available to all users. The encrypted P code is referred to as the Y code which is reserved for the military (although it is partly useful to others also) (Strang and Borre, 1997).

All GPS satellites use the same carrier phase frequencies. Each satellite has its own pseudorandom sequence of 1023 bits (its periodic C/A code which is a Gold Code). The repeat time is 1.5 s for the C/A code and one week for the P code. It is most important to know that there are (at present) two coherent transmission frequencies for the P code (and its secret encryption, the Y code) (Strang and Borre, 1997).

L_1 is 1575.42 MHz and L_2 is 1227.60 MHz

The two frequencies are differently delayed by the ionosphere. A receiver that accepts both frequencies can compute this delay (because it is known to be proportional to $1/f^2$). This correction to the speed of light through the ionosphere is essential. A single frequency receiver has to make do with estimates of ionospheric correction whose parameters are broadcast by the satellites.

At this time there is wide discussion of the proposal for a new civilian frequency L_5 (then the military might remove L_2 from civilian use). The difficulties of agreeing on a new frequency before launching expensive satellites seem to have frustrated everyone (Strang and Borre, 1997).

2.6 Antenna Characteristics

A receiving antenna converts an electromagnetic signal into an electrical signal so that it may be processed by a radio receiver. A transmitting antenna performs the reverse operation. The gain of an antenna varies with frequency. Therefore, GNSS user equipment must incorporate an antenna that has peak sensitivity near to the carrier frequency of the signals processed by the receiver and sufficient bandwidth to pass those signals. Where the receiver processes signals in more than one frequency band, either the antenna must be sensitive in all of the bands required or a separate antenna for each band must be used. The antenna bandwidth should match or exceed the pre-correlation bandwidth of the receiver (Groves, 2008).

A GNSS antenna should generally be sensitive to signals from all directions. A typical GNSS antenna has a gain of 2 to 4 dB for signals at normal incidence. This drops as the angle of incidence increases and is generally negative (in decibel terms) for angles of incidence greater than 75°. For a horizontally mounted antenna, a 75° incidence angle corresponds to a satellite signal at a 15° elevation angle (Groves, 2008).

GNSS signals are transmitted with right-handed circular polarization (RHCP). On surface reflection, this is reversed to left-handed circular polarization (LHCP) or to elliptical polarization, which mixes RHCP and LHCP. Therefore, to minimize multipath problems, the antenna should be sensitive only to RHCP signals (Groves, 2008).

For high-precision applications, it is important to know where the electrical phase centre of the antenna is. This is the point in space for which the GNSS user equipment determines the navigation solution and does not necessarily coincide with the physical centre of the antenna. For a given antenna, the phase centre can vary with elevation, azimuth, and frequency (Groves, 2008).

Basic GNSS antennas come in a number of shapes and sizes. Patch, or microstrip, antennas have the advantage of being low cost, flat, and rugged, but their polarization varies with the angle of incidence. Better performance can be obtained from a dome, blade, or helical (volute) antenna. More advanced antenna technology may be used to limit the effects of RF interference sources and/or multipath. For hand-held applications, the antenna is usually included with the user equipment, whereas for vehicles, a separate antenna is generally mounted on the vehicle body (Groves, 2008).

The cable between the antenna and the receiver imposes a common-mode lag on the incoming signal. However, the effects of antenna cable lag and receiver clock offset are indistinguishable, so the navigation processor simply accounts for the lag as part of its clock offset estimate. Signal attenuation in the antenna cable may be mitigated by including an amplifier in the antenna (Groves, 2008).

2.7 Receiver Characteristics

The receivers are the most important hardware in a GPS surveying operation. Their characteristics and capabilities influence the techniques available to the user throughout the work. There are many different GPS receivers on the market. Some of them are appropriate for surveying and they share some fundamental elements. Though no level of accuracy is ever guaranteed, with proper procedures and data handling they are generally capable of accuracies from sub meter to centimetres. Most are also capable of performing differential

GPS, real-time GPS, static GPS, etc., and are usually accompanied by processing and network adjustment software and so on.

GPS receivers come in a variety of shapes and sizes. Some have external batteries, data collectors. Some are tripod mounted. Some are hand-held and have all components built in and some can be used in both ways, with externals and without. Nevertheless most have similar characteristics. Here is a schematic drawing of a GPS receiver. It includes some of the common components.

https://www.e-education.psu.edu/geog862/node/1781

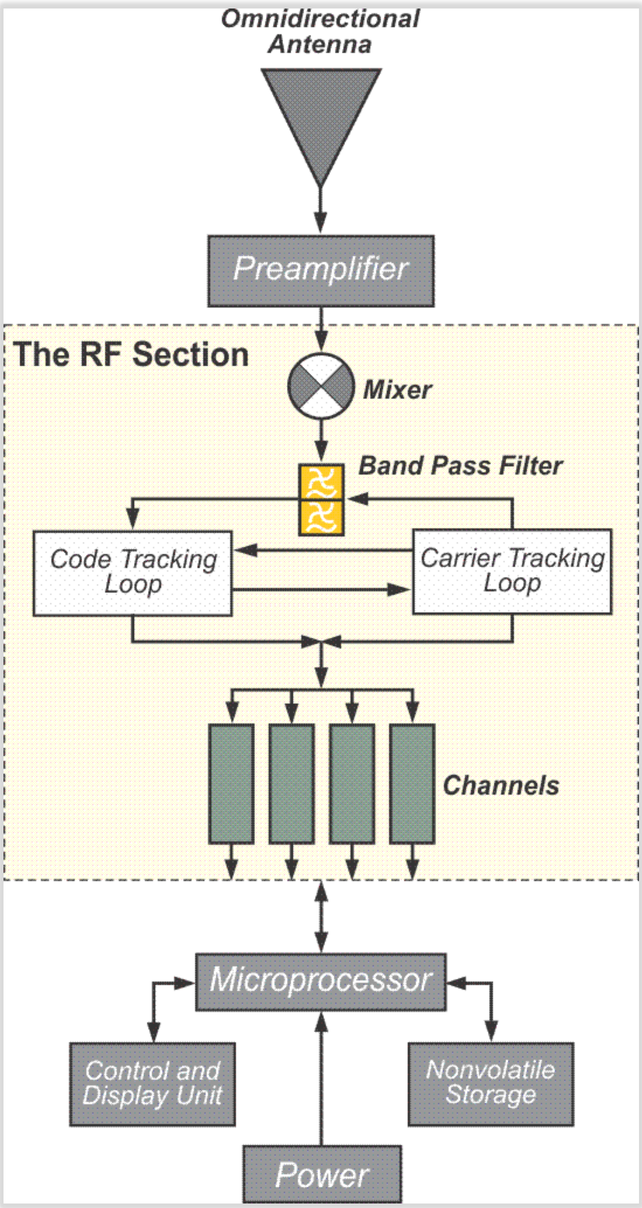


Figure 2-8: GPS Receiver Schematic (Source: *GPS for Land Surveyors)*

A GPS receiver must collect and then convert signals from GPS satellites into measurements of position, velocity and time. There is a challenge in that the GPS signal has low power. An orbiting GPS satellite broadcasts its signal across a cone of approximately 28º of arc. From the satellite's point of view, about 11,000 miles up, that cone covers a substantial portion of the whole planet. It is instructive to contrast this arrangement with a typical communication satellite that not only has much more power, but also broadcasts a very directional signal. Its signals are usually collected by a large dish antenna, but the typical GPS receiver has a small, relatively non-directional antenna. Stated another way, a GPS satellite spreads a low power signal over a large area rather than directing a high power signal at a very specific area. Fortunately, antennas used for GPS receivers do not have to be pointed directly at the signal source. The GPS signal also intentionally occupies a broader bandwidth than it must to carry its information. This characteristic is used to prevent jamming and mitigate multipath but most importantly the GPS signal itself would be completely obscured by the variety of electromagnetic noise that surrounds us if it were not a spread spectrum coded signal. In fact, when a GPS signal reaches a receiver its power is actually less than the receivers natural noise level, fortunately the receiver can still extract the signal and achieve unambiguous satellite tracking using the correlation techniques described earlier. To do this job the elements of a GPS receiver function cooperatively and iteratively. That means that the data stream is repeatedly refined by the several components of the device working together as it makes its way through the receiver.

From the point of view of a GPS satellite, the earth presents a disk that really, from 11,000 miles up, has a spread of approximately 28 degrees. The GPS signal not at high power in the beginning then spreads over this very large area. It is easy for the GPS signal to get lost in overhead cover before it reaches the antenna of the GPS receivers. It can also be disrupted by interference from electromagnetic noise. However, since it is a spread spectrum signal it has a certain robustness against jamming.

So even though it's a weak signal, the design of the signal comes to our rescue to some degree and allows it to be sorted by the receiver from the surrounding noise interference and obstructions-- up to a point, of course. GPS is not going to penetrate substantial cover.

https://www.e-education.psu.edu/geog862/node/1781

Receivers are generally categorized by their physical characteristics, the elements of the GPS signal they can use with advantage, and by the claims about their accuracy. But the effects of these features on a receiver's actual productivity are not always obvious. There are receivers that use only the C/A code on the L1 frequency and receivers that cross-correlate with the P code, or encrypted Y code, on L1 and L2. There are L1 carrier phase tracking receivers, dual-frequency and multi-frequency carrier phase tracking receivers, receivers that track all in view and GPS/GNSS receivers. The more aspects of the GPS signal a receiver can employ, the greater its flexibility, but so too the greater its cost. It is important to understand receiver capabilities and limitations to ensure that the systematic capability of a receiver is matched to the required outcome of a project.

As shown in the illustration it is possible to divide receivers into three categories. They are; recreation, mapping and surveying. These categories can be further divided by the observables they are capable of tracking. Again the contribution of these capabilities to the levels of possible systematic precision and accuracy are given from lower (L1 Code alone) to higher (GPS + GLONASS), that is, from left to right in the illustration.

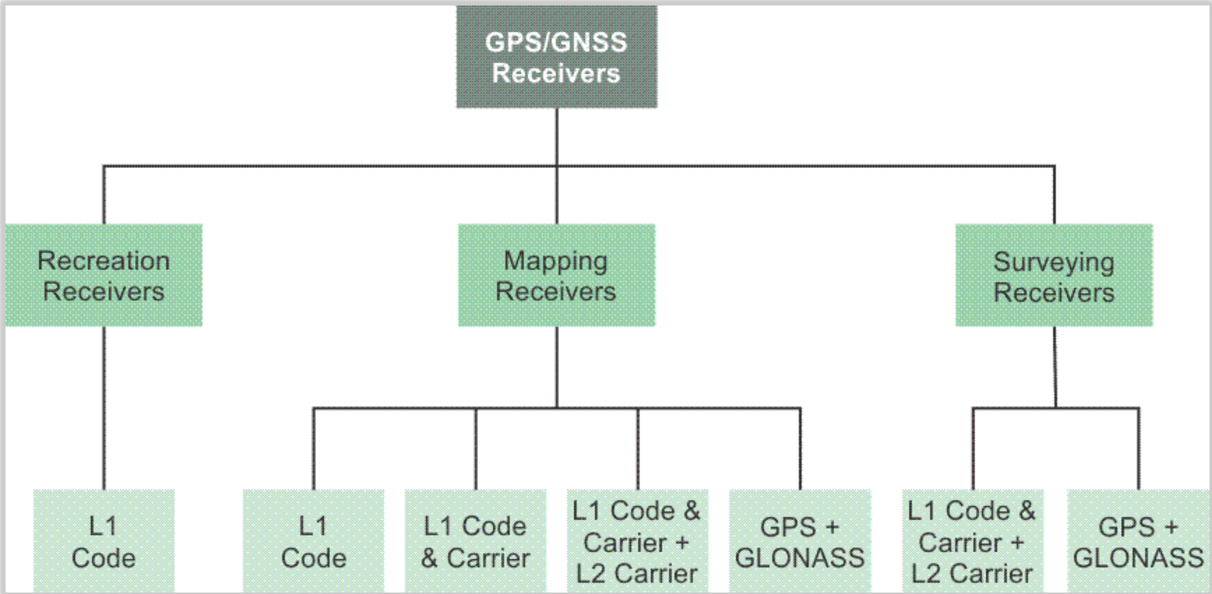


Figure 2-9: Receiver Categories (Source: GPS for Land Surveyors)

2.7.1 Recreational Receiver

These receivers are generally defined as L1 Code receivers which are typically not user configurable for settings such as mask angle, PDOP, the rate at which measurements are downloaded, the *logging rate*, also known as the epoch interval, and *signal to noise ratio SNR*. As you might expect SNR is the ratio of the received signal power to the noise floor of a GPS observation. It is typical for the antenna, receiver and CDU to be integrated into the device in these receivers.

Generally speaking receivers that track the C/A code only provide relatively low accuracy. Most are not capable of tracking the carrier phase observable. These receivers were typically developed with basic navigation in mind. Most are designed for autonomous (stand-alone) operation to navigate, record tracks, waypoints and routes aided by the display of onboard maps. They are sometimes categorized by the number of waypoints they can store. *Waypoint* is a term that grew out of military usage. It means the coordinate of an intermediate position a person, vehicle, or airplane must pass to reach a desired destination. With such a receiver, a user may call up a distance and direction from his present location to the next waypoint.

CHAPTER 3 OBSERVABLE

Contents of this Chapter

In general GPS provides three types of measurements: pseudorange, carrier phase, and Doppler. The pseudorange measurements, at a specific moment from different satellites, have a common clock bias which results in the name of this observation. The carrier phase gives more precise measurements than pseudoranges, by estimating its instantaneous rate, or Doppler measurement (beat frequency) over time. Hence it is also known as the integrated Doppler, or the accumulated phase. The main differences and characteristics between pseudorange and carrier phase measurements are summarized in Table 2.1 (Hofmann Wellenhof et al., 1997). Additional signals are planned to enhance the ability of GPS to support users, provide a new military code and improve the overall performance. The first new signal will be a C code on the L2 frequency (1227.60 MHz). This new feature will enable dual channel civil receivers to correct for ionospheric errors. The L2C signal contains two codes of different length, CM-L2C which is the moderate length code contains 10,230 chips, repeats every 20 milliseconds. CL-L2C is the long code contains 767,250 chips, repeats every 1.5 seconds (Fontana et al., 2001). A third civil signal will be added on the L5 frequency (1176.45 MHz) for use in safety-of-life applications. L5 can serve as a redundant signal to the GPS L1 frequency (1575.42 MHz) with a goal of assurance of continuity of service potentially to provide precision approach capability for aviation users. The initial transmission of L5 signal is planned for 2008, with a full operational availability in 2012 (Qiu, 2007). (Marji, 2008).

Table 2-1: Main Characteristics of Pseudorange and Carrier Phase Data (Ho Wellenhof et al., 1997)

	Code	Carrier
Wavelength	P-code 29.3 m C/A-code 293 m	L1 19.03 cm L2 24.42 cm
Observation noise Classical receiver New development	P-code 0.1-0.3 m C/A-code 0.1-3 m P-code 2-5 cm	 1-3 mm <0.2 mm
Propagation effect	Ionospheric delay $+\Delta T_{ION}$	Ionospheric advance $-\Delta T_{ION}$
Ambiguity	Non-ambiguous	Ambiguous

(Strang and Borre, 1997)

3.1 Code Pseudoranges

The pseudorange is a measure of the range, or distance, between the GPS receiver and the GPS satellite (more precisely, it is the distance between the GPS receiver's antenna and the GPS satellite's antenna). As stated before, the ranges from the receiver to the satellites are needed for the position computation. Either the P-code or the C/A-code can be used for measuring the pseudorange (El-Rabbany, 2006).

The procedure of the GPS range determination, or pseudoranging, can be described as follows. Let us assume for a moment that both the satellite and the receiver clocks, which control the signal generation, are perfectly synchronized with each other. When the PRN code is transmitted from the satellite, the receiver generates an exact replica of that code [3]. After some time, equivalent to the signal travel time in space, the transmitted code will be picked up by the receiver. By comparing the transmitted code and its replica, the receiver can compute the signal travel time. Multiplying the travel time by the speed of light (299,729,458 m/s) gives the range between the satellite and the receiver. Figure 2.3 explains the pseudorange measurements (El-Rabbany, 2006).

Unfortunately, the assumption that the receiver and satellite clocks are synchronized is not exactly true. In fact, the measured range is contaminated, along with other errors and biases, by the synchronization error between the satellite and receiver clocks. For this reason, this quantity is referred to as the pseudorange, not the range [4] (El-Rabbany, 2006).

GPS was designed so that the range determined by the civilian C/A-code would be less precise than that of military P-code. This is based on the fact that the resolution of the C/A-code, 300 m, is 10 times lower than the P-code. Surprisingly, due to the improvements in the receiver technology, the obtained accuracy was almost the same from both codes [4] (El-Rabbany, 2006).

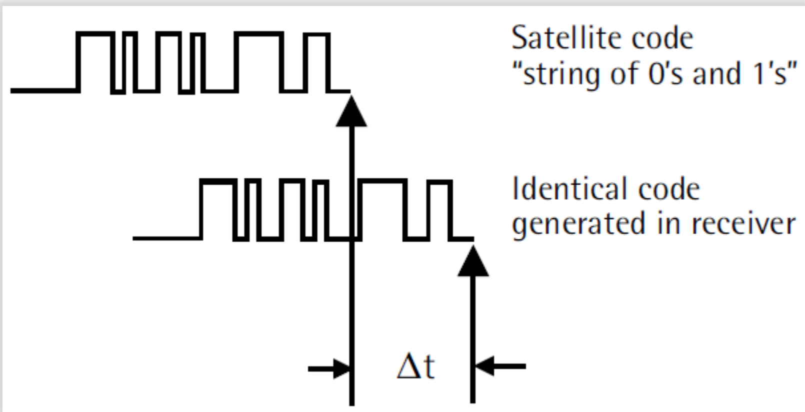


Figure 3-1: Pseudorange Measurement **(El-Rabbany, 2006)**

Refer to (Tsui, 2005) Chapter 5 **GPS C/A Code Signal Structure**

12.540 Principles of the Global Positioning System Lecture 08, Prof. Thomas Herring, http://geoweb.mit.edu/~tah/12.540

a. Pseudo-random noise. The modulated C/A-code is referred to as pseudo-random noise (PRN). This pseudo-random noise is actually a 1023 bit code with a clock rate of 1.023 MHz that repeats every 1 millisecond. The 10.23 MHz P(Y)-code PRN has a coded sequence of 267 days. This sequence of very precise time marks permits the ground receivers to compare and compute the time of transmission between the satellite and ground station. From this transmission time, the range to the satellite can be derived. This is the basis behind GPS range measurements. Each satellite has a different PRN. The C/A-code pulse intervals are approximately every 293 m in range and the more accurate P-code every 29 m--see Table 2-2.

Table 2-2. NAVSTAR GPS Signal Codes and Carrier Frequencies (Block IIR)

Carrier (L-Band)	Codes: Civilian C/A-Code	Codes: Military P(Y)-Code	Satellite Messages
L1 1575.42 MHz 19 cm wavelength	Present 293 m wavelength	Present 29.3 m wavelength	user messages satellite constants satellite positions
L2 1227.60 MHz 24 cm wavelength	Not Present	Present 29.3 m wavelength	

The word observable is used throughout GPS literature to indicate the signals whose measurement yields the range or distance between the satellite and the receiver. The word is used to draw a distinction between the thing being measured, the observable and the measurement, the observation. In GPS there are two types of observables: the *pseudorange* and the *carrier phase*. The latter, also known as the *carrier beat phase*, is the basis of the techniques used for high-precision GPS surveys. On the other hand, the pseudorange can serve applications when virtually instantaneous point positions are required or relatively low accuracy will suffice. These basic observables can also be combined in various ways to generate additional measurements that have certain advantages. It is in this latter context that pseudoranges are used in many GPS receivers as a preliminary step toward the final determination of position by a carrier phase measurement. The foundation of pseudoranges is the correlation of code carried on a modulated carrier wave received from a GPS satellite with a replica of that same code generated in the receiver. Most of the GPS receivers used for surveying applications are capable of code correlation. That is, they can determine pseudoranges from the C/A code or the P(Y) code. These same receivers are usually capable of determining ranges using the unmodulated carrier as well. However, first let us concentrate on the pseudorange.

There are two observables that are of interest right at the moment, one is the code. We've talked about how that is modulated onto the carrier. The codes are one type of observable. There's another type of observable. That's the carrier itself without the codes.

In the upper image, you see the pseudorange code observable illustrated as square waves of code states, and in the lower image, you see the carrier wave observable, which is just a constant sine wave, not modulated. The pseudorange in many GPS receivers is used as the preliminary step to the final position done by the carrier phase measurement. Many GPS receivers use the pseudorange code observable as sort of the front door, a way to begin the determination of a position, and then frequently they switch to the carrier to refine that position.

Carrier wave positions are more accurate than are the positions available from the code phase or the pseudo range code positions alone.

b. Pseudoranges. A pseudorange is the time delay between the satellite clock and the receiver clock, as determined from C/A- or P-code pulses. This time difference equates to the range measurement but is called a "pseudorange" since at the time of the measurement, the receiver clock is not synchronized to the satellite clock. In most cases, an absolute 3-D real-time navigation position can be obtained by observing at least four simultaneous pseudoranges. The Standard Positioning Service (SPS) uses the less precise L1 C/A-code pseudoranges for real-time GPS navigation. The L2 signal is not used in SPS positioning. The Precise Positioning Service (PPS) is the fundamental military real-time navigation use of GPS. Pseudoranges are obtained using the higher pulse rate (i.e. higher accuracy) P-code on both frequencies (L1 and L2). P-codes are encrypted to prevent unauthorized civil or foreign use. This encryption requires a special key.

(US Army Corps of Engineers, 2003)

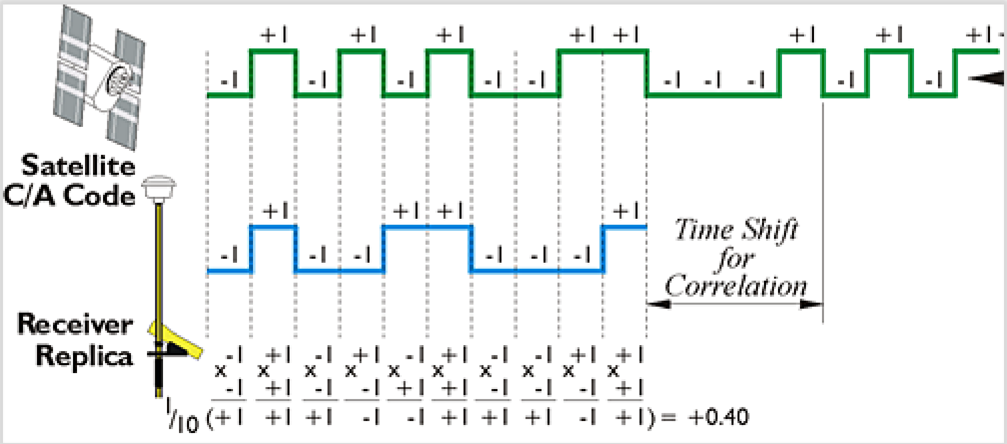


Figure 3-2: Pseudorange Code Observable (*Source: GPS for Land Surveyors*)

3.2 Carrier Phase

Carrier frequency tracking measures the phase differences between the Doppler shifted satellite and receiver frequencies. Phase measurements are resolved over the relatively short L1 and L2 carrier wavelengths (19 cm and 24 cm respectively). This allows phase resolution at the mm level. The phase differences are continuously changing due to the changing satellite earth geometry. However, such effects are resolved in the receiver and subsequent data post-processing. When carrier phase measurements are observed and compared between two stations (i.e. relative or differential mode), baseline vector accuracy between the stations below the centimetre level is attainable in three dimensions. Various receiver technologies and processing techniques allow carrier phase measurements to be used in real-time centimetre positioning (US Army Corps of Engineers, 2003).

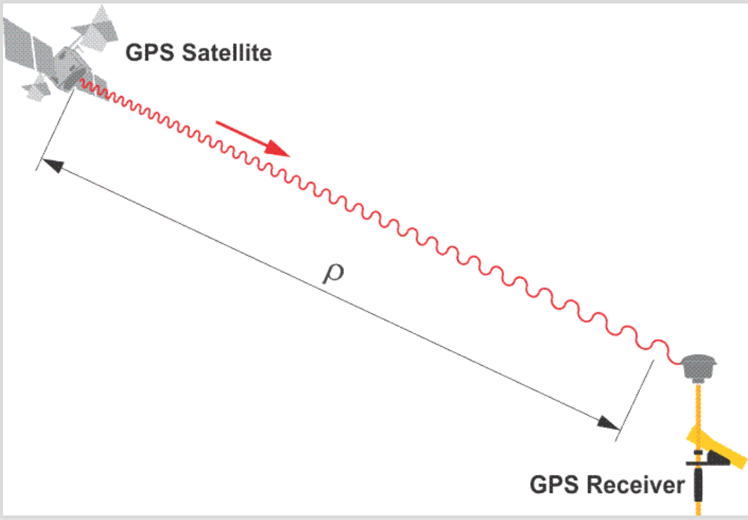
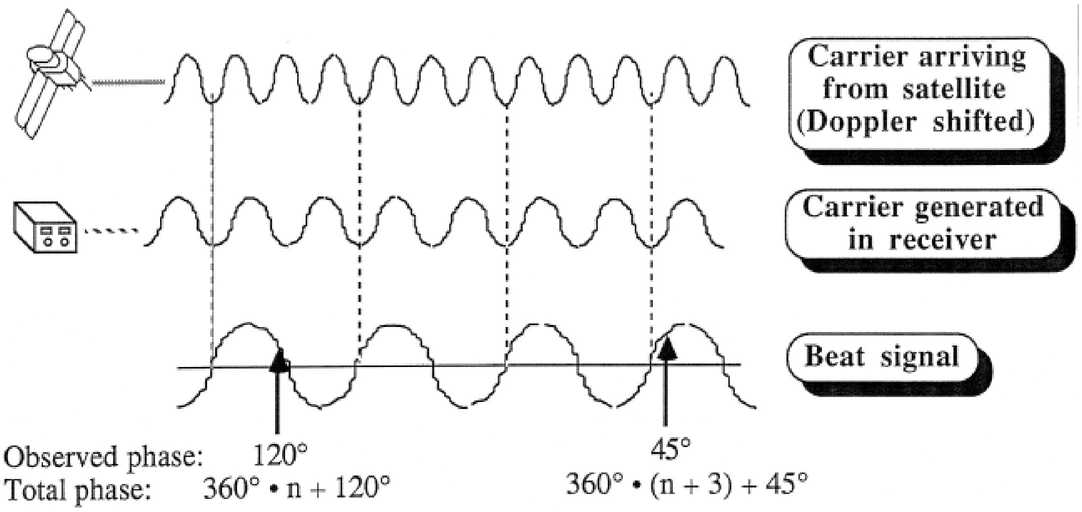


Figure 3-3: Carrier Phase Observable (Source: GPS for Land Surveyors)



Integer cycle count n is

- **not observed, but counted inside the receiver**
- **loss of lock leads to loss of n-count (called cycle slip)**
- **initial value for n-count must be determined (cycle ambiguity problem)**

$$\varphi_a^j(t) = \frac{1}{\lambda}\rho_a^j(t) + N_a^j + f\,\delta^j(t) - f\,\delta_a(t) - \frac{1}{\lambda}\Delta_a^j Ion(t) + \frac{1}{\lambda}\Delta_a^j Trop(t) + \varepsilon \quad (1.5)$$

Where

$\varphi_a^j(t)$	The phase measurements.
N_a^j	The unknown integer ambiguity.
λ	The signal wave length.
f	The signal frequency.
ε	The noise of the phase measurements.

3.3 Doppler Frequency

As the satellite passes overhead, the range between the receiver and the satellite changes, that steady change is reflected in a smooth and continuous movement of the phase of the signal coming into the receiver. The rate of that change is reflected in the constant variation of the signal's Doppler shift. But if the receiver's oscillator frequency is matching these variations exactly, as they are happening, it will duplicate the incoming signal's Doppler shift and phase. This strategy of making measurements using the carrier beat phase observable is a matter of counting the elapsed cycles and adding the fractional phase of the receiver's own oscillator.

Doppler information has broad applications in signal processing. It can be used to discriminate between the signals from various GPS satellites, to determine integer ambiguities in kinematic surveying, as a help in the detection of cycle slips, and as an additional independent observable for autonomous point positioning. But perhaps the most important application of Doppler data is the determination of the *range rate* between a receiver and a satellite. *Range rate* is a term used to mean the rate at which the range between a satellite and a receiver changes over a particular period of time.

We've talked about the Doppler shift in several different contexts. One was the original transit system, NNSS system that operated on the Doppler shift. And I was mentioned earlier that the GPS system uses the Doppler shift as an observable. It is useful to have a concept of how much shift is typical with a GPS satellite. This graphic is intended to indicate that. As you see on the left, with the satellite rising or moving toward the receiver, the Doppler shift is approximately 4 1/2 to 5 cycles per millisecond. At zenith or at its closest approach the shift is nominally zero. It then goes from the positive to negative returning again to approximately 4 1/2 to 5 cycles per millisecond as it's moving away and about to set relative to the receiver. This steady shift is caused by the continuous movement of the satellite relative to the receiver. It is very predictable. That predictability, the constant variation of the signal's Doppler shift, makes it a good observable. If the receiver's oscillator frequency is adjusted to match these variations exactly, as they're happening, it will duplicate the incoming signal's shift and phase. This strategy of making measurements using the carrier beat phase observable is a matter of counting the elapsed cycles and adding the fractional phase of the receiver's own oscillator. This is one way that the phase lock loop maintains its lock on the signal as the Doppler shift occurs with each of the satellites that it is tracking.

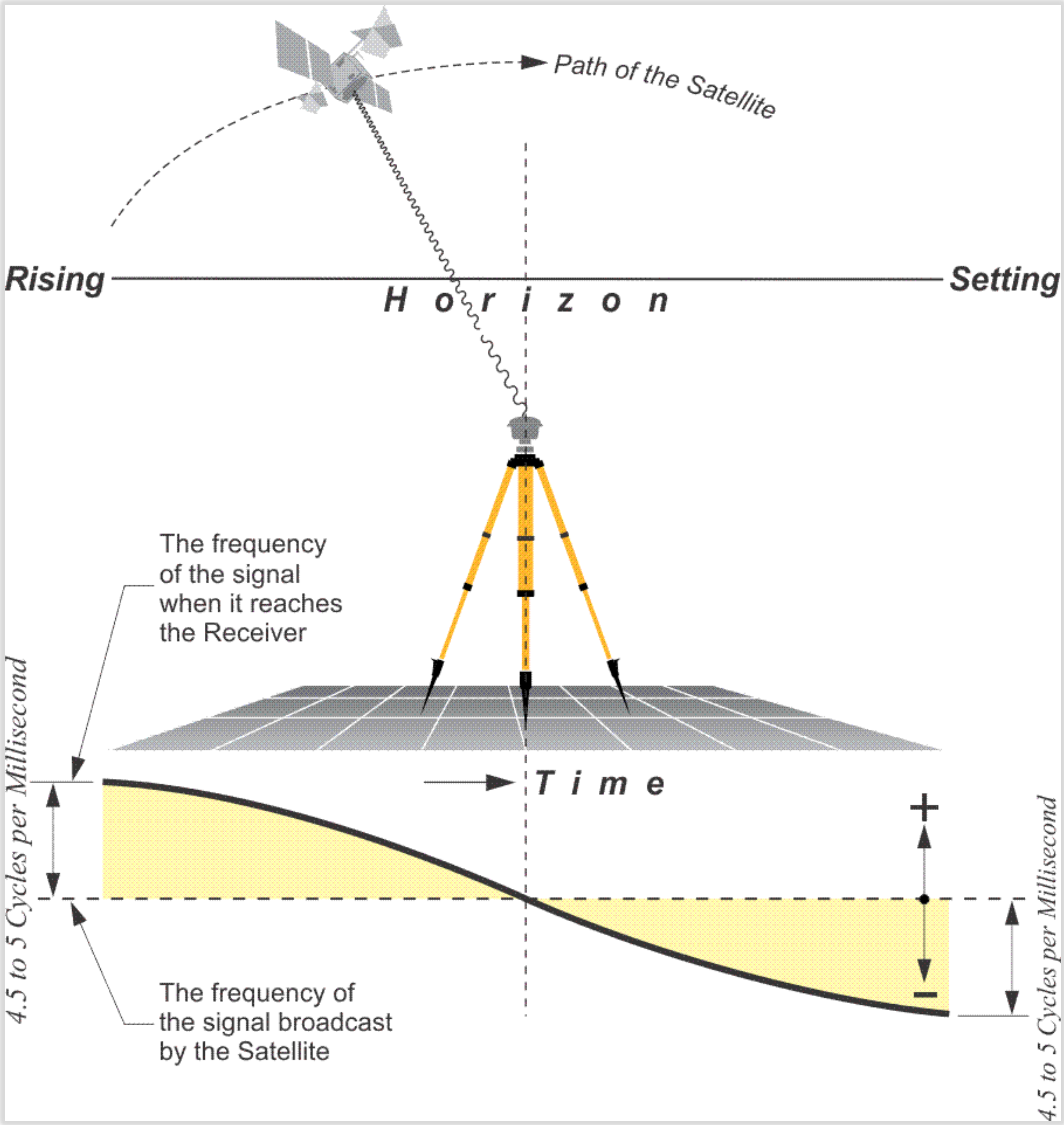


Figure 3-4: Typical Doppler Shift (Source: GPS for Land Surveyors)

3.4 Linear Carrier Phase Combinations

GPS measurements are corrupted by a number of errors and biases (discussed in detail in Chapter 3), which are difficult to model fully. The unmodeled errors and biases limit the positioning accuracy of the stand-alone GPS receiver. Fortunately, GPS receivers in close proximity will share to a high degree of similarity the same errors and biases. As such, for those receivers, a major part of the GPS error budget can simply be removed by combining their GPS observables.

In principle, there are three groups of GPS errors and biases: satellite-related, receiver-related, and atmospheric errors and biases [3]. The measurements of two GPS receivers simultaneously tracking a particular satellite contain more or less the same satellite-related errors and atmospheric errors. The shorter the separation between the two receivers, the more similar the errors and biases. Therefore, if we take the difference between the measurements collected at these two GPS receivers, the satellite-related errors and the atmospheric errors will be reduced significantly. In fact, as shown in Chapter 3, the satellite clock error is effectively removed with this linear combination. This linear combination is known as between-receiver single difference (Figure 2.6).

Similarly, the two measurements of a single receiver tracking two satellites contain the same receiver clock errors. Therefore, taking the difference between these two measurements removes the receiver clock errors. This difference is known as between-satellite single difference (Figure 2.6).

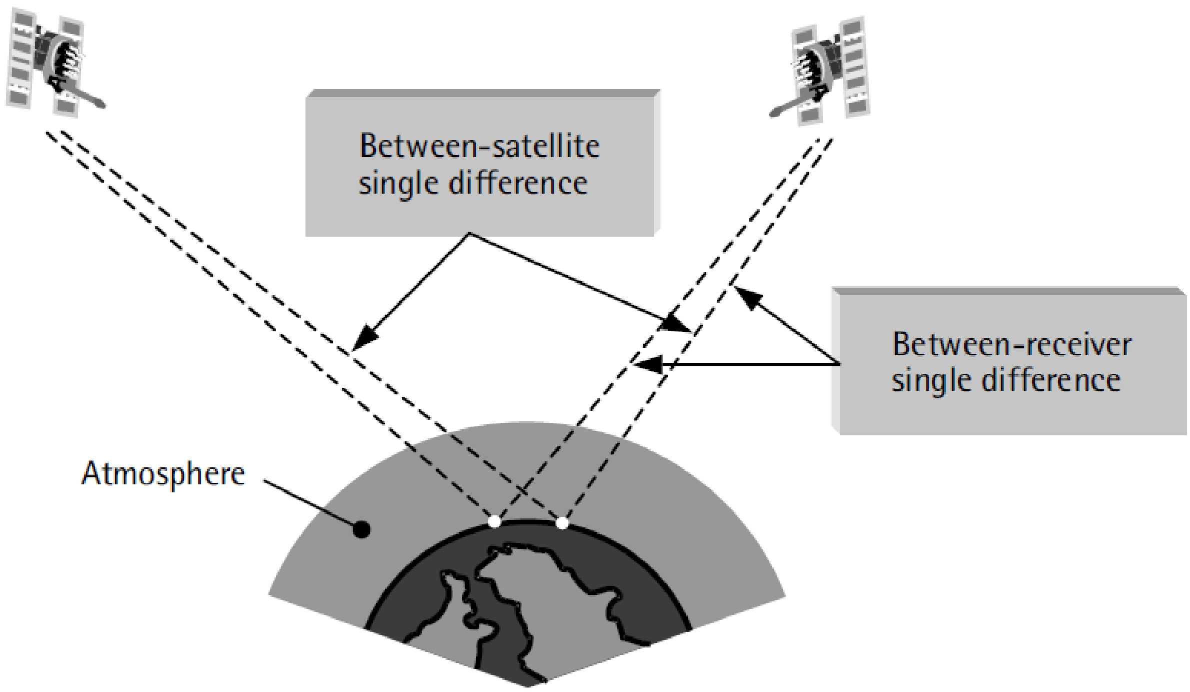


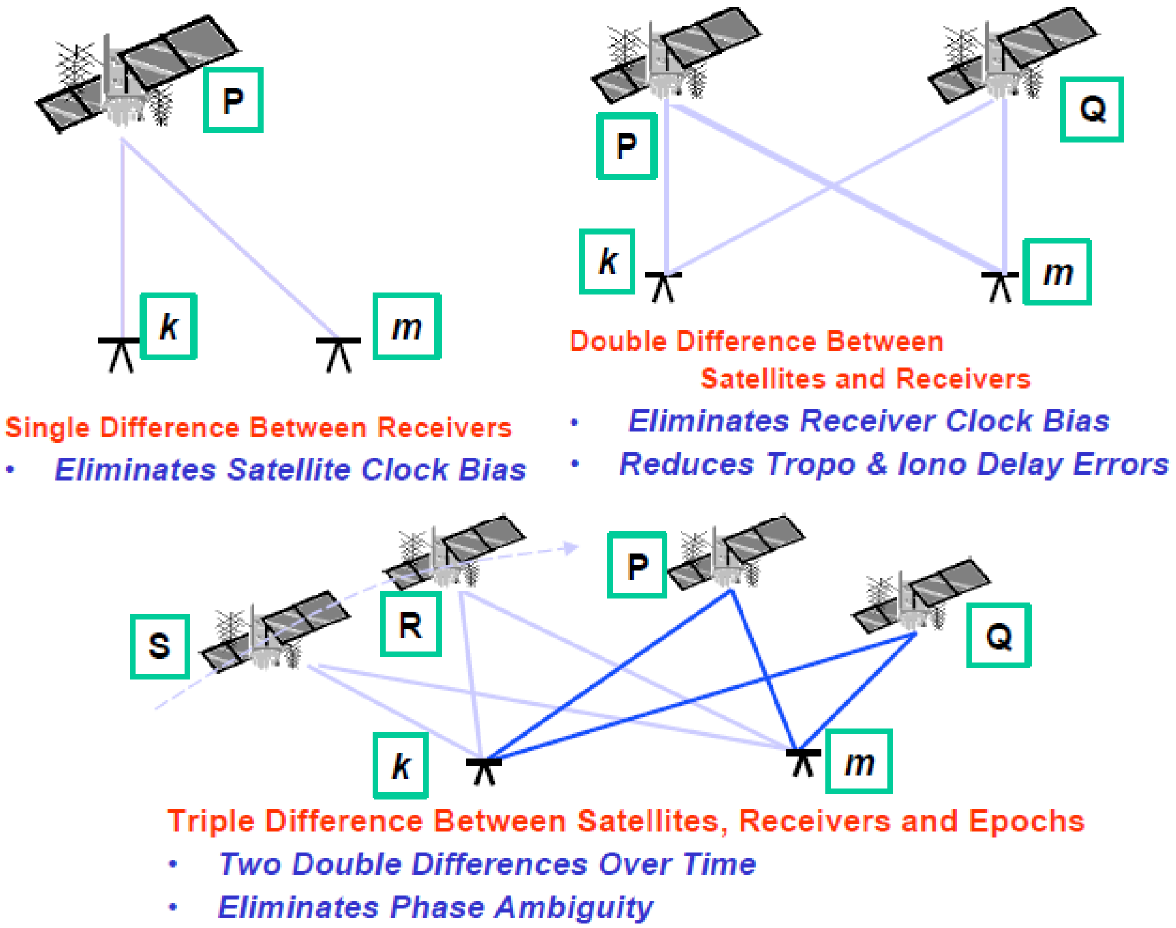
Figure 2.6 Some GPS linear combinations.

When two receivers track two satellites simultaneously, two between-receiver single difference observables could be formed. Subtracting these two single difference observables from each other generates the so-called double difference [3]. This linear combination removes the satellite and receiver clock errors. The other errors are greatly reduced. In addition, this observable preserves the integer nature of the ambiguity parameters. It is therefore used for precise carrier-phase-based GPS positioning.

Another important linear combination in known as the "triple difference," which results from differencing two double-difference observables over two epochs of time [3]. As explained in the previous section, the ambiguity parameters remain constant over time, as long as there are no cycle slips. As such, when forming the triple difference, the constant ambiguity parameters disappear. If, however, there is a cycle slip in the data, it will affect one triple-difference observable only, and therefore will appear as a spike in the triple-difference data series. It is for this reason that the triple-difference linear combination is used for detecting the cycle slips.

All of these linear combinations can be formed with a single frequency data, whether it is the carrier phase or the pseudorange observables. If dual-frequency data is available, other useful linear combinations could be formed. One such linear combination is known as the ionosphere-free linear combination. As shown in Chapter 3, ionospheric delay is inversely proportional to the square of the carrier frequency. Based on this characteristic, the ionosphere-free observable combines the L1 and L2 measurements to essentially eliminate the ionospheric effect. The L1 and L2 carrier-phase measurements could also be combined to form the so-called wide-lane observable, an artificial signal with an effective wavelength of about 86 cm. This long wavelength helps in resolving the integer ambiguity parameters [1].

The accuracy achievable by pseudoranging and carrier phase measurement in both absolute and relative positioning surveys can be improved through processing that incorporates differencing of the mathematical models of the observables. Processing by differencing takes advantage of correlation of error (e.g., GPS signal, satellite ephemeris, receiver clock, and atmospheric propagation errors) between receivers, satellites, and epochs, or combinations thereof, in order to improve GPS processing. Through differencing, the effects of the errors that are common to the observations being processed are eliminated or at least greatly reduced. Basically, there are three broad processing techniques that incorporate differencing: single differencing, double differencing, and triple differencing. Differenced solutions generally proceed in the following order: differencing between receivers takes place first, between satellites second, and between epochs third (Figure 10-1).



3.5 Carrier Smoothing of the Code

The smoothing process uses the carrier phase measurements to smooth the pseudorange data to reduce the measurement noise. The process provides smoothed pseudorange and sampled carrier phase measurements for use by the CS Kalman filter. The smoothing process consists of data editing to remove outliers and cycle slips, converting raw dual-frequency measurements to ionosphere-free observables, and generating smoothed measurements once a sufficient number of validated measurements are available. Figure 3.18 shows a representative data smoothing interval consisting of 600 pseudorange and carrier phase observations, with 595 observations used to form a smoothed pseudorange minus carrier phase offset and the 5 remaining observations used to form a carrier phase polynomial (Kaplan and Hegarty, 2006).

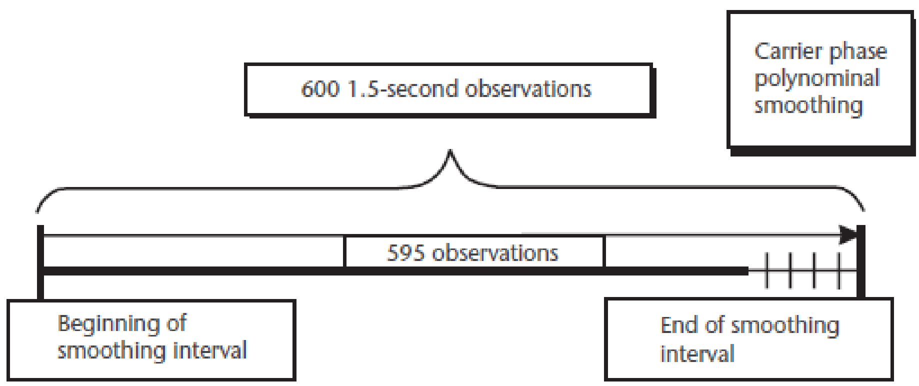


Figure 3.18 Representative MCS data-smoothing interval.

MCS = Master Control Station, CS = Control Segment

CHAPTER 4 SYSTEM BIASES AND ERRORS

Contents of this Chapter

GPS pseudorange and carrier-phase measurements are both affected by several types of random errors and biases (systematic errors). These errors may be classified as those originating at the satellites, those originating at the receiver, and those that are due to signal propagation (atmospheric refraction). Figure 4-1 shows the various errors and biases (El-Rabbany, 2006).

The errors originating at the satellites include ephemeris, or orbital, errors, satellite clock errors, and the effect of selective availability. The latter was intentionally implemented by the U.S. DoD to degrade the autonomous GPS accuracy for security reasons. It was, however, terminated at midnight (eastern daylight time) on May 1, 2000. The errors originating at the receiver include receiver clock errors, multipath error, receiver noise, and antenna phase center variations. The signal propagation errors include the delays of the GPS signal as it passes through the ionospheric and tropospheric layers of the atmosphere. In fact, it is only in a vacuum (free space) that the GPS signal travels, or propagates, at the speed of light (El-Rabbany, 2006).

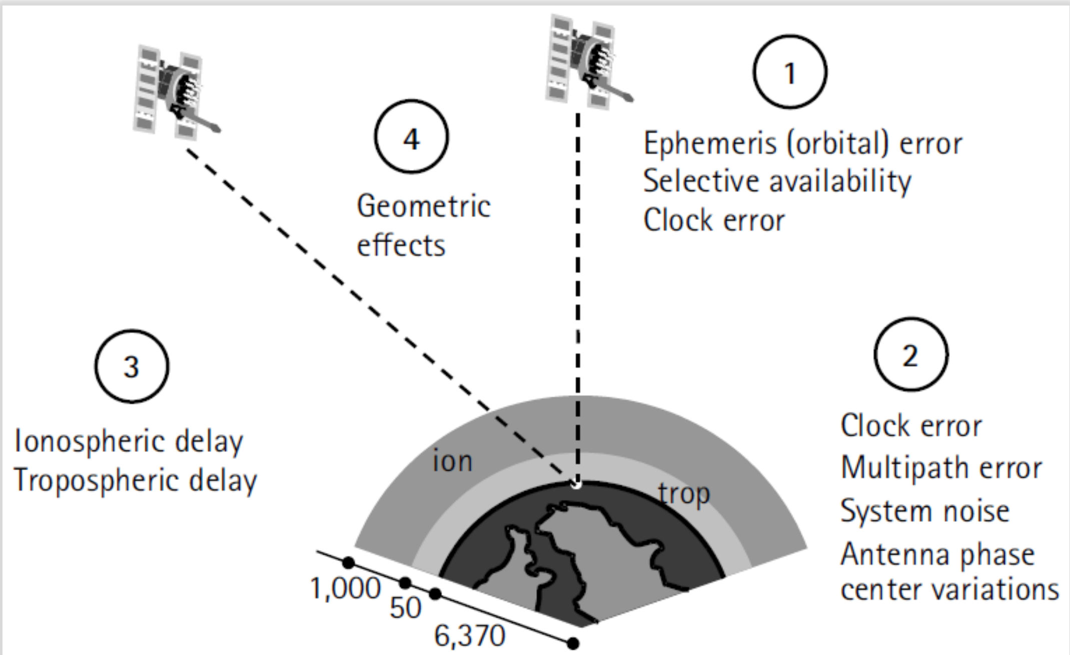


Figure 4-1: GPS Errors and Biases

In addition to the effect of these errors, the accuracy of the computed GPS position is also affected by the geometric locations of the GPS satellites as seen by the receiver. The more spread out the satellites are in the sky, the better the obtained accuracy (El-Rabbany, 2006).

GNSS Error Sources

Before we move on, let's summarize the errors that can affect the accuracy of standard GNSS pseudorange determination, that is, the determination of the pseudorange to a single satellite. These are shown in Table 1:

Contributing Source	Error Range
Satellite clocks	±2 m
Orbit errors	±2.5 m
Ionospheric delays	±5 m
Tropospheric delays	±0.5 m
Receiver noise	±0.3 m
Multipath	±1 m

Table 1 GNSS System Errors

The degree with which the above pseudorange errors affect positioning accuracy depends largely on the geometry of the satellites being used. In Chapter 4, we'll discuss techniques for reducing these errors further.

(Jeffrey, 2010)

4.1 Multipath Errors

Some of the signal energy transmitted by the satellite is reflected on the way to the receiver. This phenomenon is referred to as multipath propagation. These reflected signals are delayed from the direct signal and, if they are strong enough, can interfere with the desired signal. Techniques have been developed whereby the receiver only considers the earliest-arriving signals and ignores multipath signals, which arrive later. In the early days of GPS, most errors came from ionospheric and tropospheric delays, but now more attention is being made to multipath effects, in the interests of continually improving GNSS performance (Jeffrey, 2010).

Multipath interference can occur where the user equipment receives reflected signals from a given satellite in addition to the direct signals. For land applications, signals are generally reflected off the ground, buildings, or trees, as shown in **Figure 4-2**, while for aircraft and ships, reflections off the host-vehicle body are more common. Interference can also occur from diffracted signals. The reflected and diffracted signals are always delayed with respect to the direct signals and have a lower amplitude unless the direct signals are attenuated (e.g., by a building or foliage). Low-elevation-angle signals are usually subject to the greatest multipath interference (Groves, 2008).

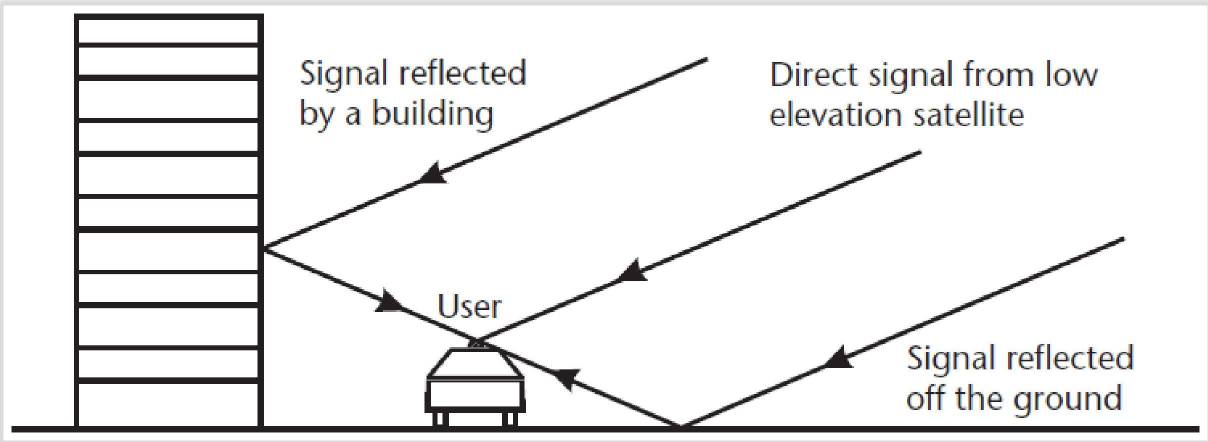


Figure 4-2: Multipath Interference (Groves, 2008)

A GPS signal might follow several paths to a receiver's antenna. The same signal arrives at different times and interferes with itself. This produces ghost images on TV and corresponds to echos for our voice. In GPS, the signal can be reflected from buildings or the ground and create a range of error of several meters or more. Some authors would allow 10 meters for multipath error in C/A code measurements (Strang and Borre, 1997).

Multipath is a serious problem because it is so difficult to model. Sometimes we can improve the site for receiver. The design of the antenna is also critical. Large ground planes, with various antenna elements (diapoles, microstrip), are the most common antidote for multipath. The receiver cam be built with a narrow correlator to block the reflection, or with multiple correlators to allow estimation on several paths. And for a given satellite/static receiver pair, at a given time of day, we could try to estimate repeatable paths (Strang and Borre, 1997).

The multipath errors in the phase observations Φ are much smaller, at the centimetre level. We receive the sum of two signals. The reflected signal has a phase shift $\Delta\Phi$ and its magnitude is attenuated by a factor α (Strang and Borre, 1997):

$$\text{received signal} = A\cos\Phi + \alpha A\cos(\Phi + \Delta\Phi).$$

The multipath error, comparing the phase of this sum to the correct value Φ, is

$$d\Phi = \arctan\left(\frac{\sin\Phi}{\alpha^{-1} + \cos\Phi}\right).$$

The worst case has no attenuation ($\alpha = 1$) and $d\Phi = 90°$. This means only a quarter-wavelength error (5 cm) from multipath.

(Strang and Borre, 1997)

4.2 Timing and Orbital Biases

An atomic clock, with a rubidium or cesium oscillator, is correct to about 1 part in 10^{12}. In a day the offset could reach 10^{-7} seconds; multiplied by 'c' this represents 26 meters. With clock corrections every 12 hours, an average error of 1 meter is reasonably conservative (Strang and Borre, 1997)

The satellite transmits its Keplerian elements, almost exactly but with a small error. This grows from the time of upload by a control station until the next upload. The error growth is slow and smooth, and only the projection of the ephemeris error along the line of sight produces and error in the range. Parkinson and Spikler Jr. (1996) estimate the rms ranging error as 2.1 m (and the estimate now might be smaller) (Strang and Borre, 1997).

Satellite positions as a function of time, which are included in the broadcast satellite navigation message, are predicted from previous GPS observations at the ground control stations. Typically, overlapping 4-hour GPS data spans are used by the operational control system to predict fresh satellite orbital elements for each 1-hour period. As might be expected, modelling the forces acting on the GPS satellites will not in general be perfect, which causes some errors in the estimated satellite positions, known as ephemeris errors. Nominally, an ephemeris error is usually in the order of 2 m to 5 m, and can reach up to 50 m under selective availability. According to, the range error due to the combined effect of the ephemeris and the satellite clock errors is of the order of 2.3 m [1s-level; s is the standard deviation (see Appendix B)] (El-Rabbany, 2006).

An ephemeris error for a particular satellite is identical to all GPS users worldwide. However, as different users see the same satellite at different view angles, the effect of the ephemeris error on the range measurement, and consequently on the computed position, is different. This means that combining (differencing) the measurements of two receivers simultaneously tracking a particular satellite cannot totally remove the ephemeris error. Users of short separations, however, will have an almost identical range error due to the ephemeris error, which can essentially be removed through differencing the observations. For relative positioning, the following rule of thumb gives a rough estimate of the effect of the ephemeris error on the baseline solution: the baseline error / the baseline length = the satellite position error / the range satellite. This means that if the satellite position error is 5m and the baseline length is 10 km, then the expected baseline line error due to ephemeris error is approximately 2.5 mm (El-Rabbany, 2006).

Some applications, such as studies of the crustal dynamics of the earth, require more precise ephemeris data than the broadcast ephemeris. To support these applications, several institutions [e.g., the International GPS Service for Geodynamics (IGS), the U.S. National Geodetic Survey (NGS), and Geomatics Canada] have developed post mission precise orbital service. Precise ephemeris data is based on GPS data collected at a global GPS network coordinated by the IGS. At the present time, precise ephemeris data is available to users with some delay, which varies from 12 hours for the IGS ultra rapid orbit to about 12 days for the most precise IGS precise orbit. The corresponding accuracies for the two precise orbits are in the order of a few decimeters to 1 decimeter, respectively. Users can download the precise ephemeris data free of charge from the IGS center, at ftp://igscb.jpl.nasa.gov/igscb/product/ (El-Rabbany, 2006).

4.3 Ionospheric Delay

Ionosphere Errors GPS signals are delayed as they pass through the ionosphere, which starts 50 km above the Earth and extends to 1000 km or more. The delay is proportional to the number of electrons (integrated density along the signal path) and inversely proportional to f^2. Thus the effect is dispersive; it depends on the frequency f. The density of free electrons varies strongly with the time of day and the latitude. The variations from solar cycles and seasons and especially short-term effects are less strong but less predictable. If the delay were not accounted for at all, the ranging errors on the L_1 frequency in the zenith direction could reach 30 meters. *The effects on the pseudorange P and phase Φ are opposite in sign*; the carrier phase is advanced.

So we *must* estimate the ionospheric delay. A dual-frequency receiver can measure the pseudoranges P_1 and P_2 on both frequencies L_1 and L_2, and solve for the delay:

$$dP_{\text{ion}} = \frac{f_2^2}{f_2^2 - f_1^2}\big(P_1 - P_2\big) + \text{random/unmodelled errors.} \tag{14.3}$$

This should be removed from P_1. Similarly the phase correction for ionospheric delay is

$$d\Phi_{\text{ion}} = \frac{f_2^2}{f_2^2 - f_1^2}\big((\lambda_1 N_1 - \lambda_2 N_2) - (\Phi_1 - \Phi_2)\big) + \text{random/unmodelled errors.} \tag{14.4}$$

Equations (14.3) and (14.4) have equal value (in delay units) but opposite signs, so the ionosphere can be unambiguously calibrated by a combination of pseudorange and phase at both frequencies.

The ambiguities N_1 and N_2 remain constant (but possibly unknown) if there are no cycle slips. So at least a differential delay is known. This estimate is good but there is often a better way. If you have a dual frequency phase receiver, the P code observations allow you to estimate the ionospheric correction. Then the improved pseudoranges can help resolve the ambiguities N_1 and N_2, completing the circle.

For measurements at only one frequency, these formulas for dP_{ion} and $d\Phi_{\text{ion}}$ are useless. In DGPS the ionospheric delay at two receivers is cancelled when we compute the (sufficiently short!) baseline between them. The difference in signal paths produces a slight baseline shortening, proportional to electron content and baseline length. One receiver at one frequency can use the prediction model for dP_{ion} and $d\Phi_{\text{ion}}$ contained in the GPS broadcast message. Tests show better results than promised (but not great).

(Strang and Borre, 1997)

4.4 Tropospheric Delay

Troposphere is the lower part of the atmosphere, thickest over the Equator (about 16 km). The temperature and pressure and humidity alter the speed of radio waves. There effects are nearly independent of the radio frequency, but they depend on the time of passage. For a flat Earth we would divide the zenith dalay (the delay at elevation angle El = $\pi/2$) by sin El. There are number of good mapping functions to improve this to a spherical-surface model. The M-file tropo uses a mapping function proposed by Goad and Goodman (1974) to compute the reduction (Strang and Borre, 1997).

In the zenith direction, the total tropospheric delay is estimated as about 2.3 meters. The hydrostatic component (responsible for 90%) is the path integral of the density of moist air. The wet component is a function of water vapour density which is highly variable. It is questionable how descriptive actual measurements can be. The classical example is sitting in a fog bank only 50 m high. The other extreme is sitting in relatively dry air below a dark thundercloud. Both of these conditions are met in GPS surveys.

In fact Dual et al. (1996) proposed and successfully demonstrated that in the reverse direction, the water vapour density could be measured by GPS! This is a beautiful example of an unexpected contribution coming from accurate measurements of time and distance. The electron content of the ionosphere can also be studied by GPS.

The delay from liquid water in clouds and rain is well below 1 cm. But models of the wet delay (water vapour) using surface meteorology are often wrong by more than 1 cm. Again we recommend the discussion by Langley in the Kleusberg-Teunissen book (Strang and Borre, 1997).

CHAPTER 5 MATHEMATICAL MODELS FOR GPS POSITIONING

Contents of this chapter

5.1 Pseudorange Point Positioning

When a GPS user performs a navigation solution, only an approximate range, or "pseudorange", to selected satellites are measured. In order for the GPS user to determine his precise location, the known range to the satellite and the position of those satellites must be known. By pseudoranging, the GPS user measures an approximate distance between the GPS antenna and the satellite by correlation of a satellite transmitted code and a reference code created by the receiver. This measurement does not contain corrections for synchronization errors between the clock of the satellite transmitter and that of the GPS receiver. The distance the signal has travelled is equal to the velocity of the transmission multiplied by the elapsed time of transmission. The signal velocity is affected by tropospheric and ionospheric conditions in the atmosphere. The figure below illustrates the pseudoranging concept.

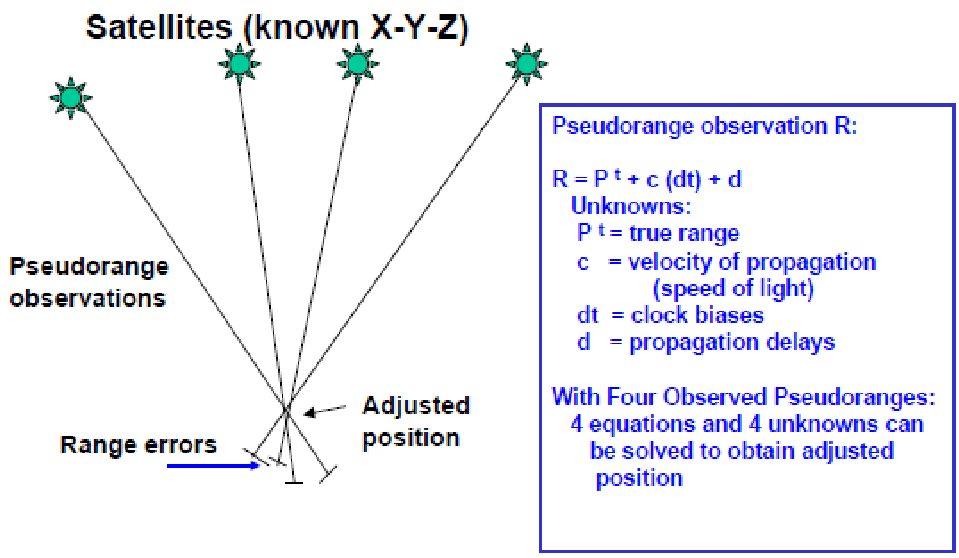


Figure 5-1: Pseudoranging Technique

b. Four pseudorange observations are needed to resolve a GPS 3-D position. (Only three pseudorange observations are needed for a 2-D location.) In practice there are often more than four satellites within view. A minimum of four satellite ranges are needed to resolve the clock biases contained in both the satellite and the ground-based receiver. Thus, in solving for the X-Y-Z coordinates of a point, a fourth unknown (i.e. clock bias--Δt) must also be included in the solution. The solution of the 3-D position of a point is simply the solution of four pseudorange observation equations containing four unknowns, i.e. X, Y, Z, and Δt.

c. A pseudorange observation is equal to the true range from the satellite to the user plus delays due to satellite/receiver clock biases and other effects.

$$R = p^t + c(\Delta t) + d \qquad \text{(Eq 4-1)}$$

where

R	= observed pseudorange
p^t	= true range to satellite (unknown)
c	= velocity of propagation
Δt	= clock biases (receiver and satellite)
d	= propagation delays due to atmospheric conditions

Propagation delays (d) are usually estimated from atmospheric models.

The true range " p^t " is equal to the 3-D coordinate difference between the satellite and user.

$$p^t = [(X^s - X^u)^2 + (Y^s - Y^u)^2 + (Z^s - Z^u)^2]^{1/2} \qquad \text{(Eq 4-2)}$$

where

X^s, Y^s, Z^s = known satellite geocentric coordinates from ephemeris data

X^u, Y^u, Z^u = unknown geocentric coordinates of the user which are to be determined.

When four pseudoranges are observed, four equations are formed from Equations 4-1 and 4-2.

$$(R_1 - c\,\Delta t - d_1)^2 = (X_1^s - X^u)^2 + (Y_1^s - Y^u)^2 + (Z_1^s - Z^u)^2 \qquad \text{(Eq 4-3)}$$
$$(R_2 - c\,\Delta t - d_2)^2 = (X_2^s - X^u)^2 + (Y_2^s - Y^u)^2 + (Z_2^s - Z^u)^2 \qquad \text{(Eq 4-4)}$$
$$(R_3 - c\,\Delta t - d_3)^2 = (X_3^s - X^u)^2 + (Y_3^s - Y^u)^2 + (Z_3^s - Z^u)^2 \qquad \text{(Eq 4-5)}$$
$$(R_4 - c\,\Delta t - d_4)^2 = (X_4^s - X^u)^2 + (Y_4^s - Y^u)^2 + (Z_4^s - Z^u)^2 \qquad \text{(Eq 4-6)}$$

In these equations, the only unknowns are X^u, Y^u, Z^u, and Δt. Solving these four equations for the four unknowns at each GPS update yields the user's 3-D position coordinates--X^u, Y^u, Z^u. These geocentric coordinates can then be transformed to any user reference datum. Adding more pseudorange observations provides redundancy to the solution. For instance, if seven satellites are simultaneously observed, seven equations are derived and still only four unknowns result.

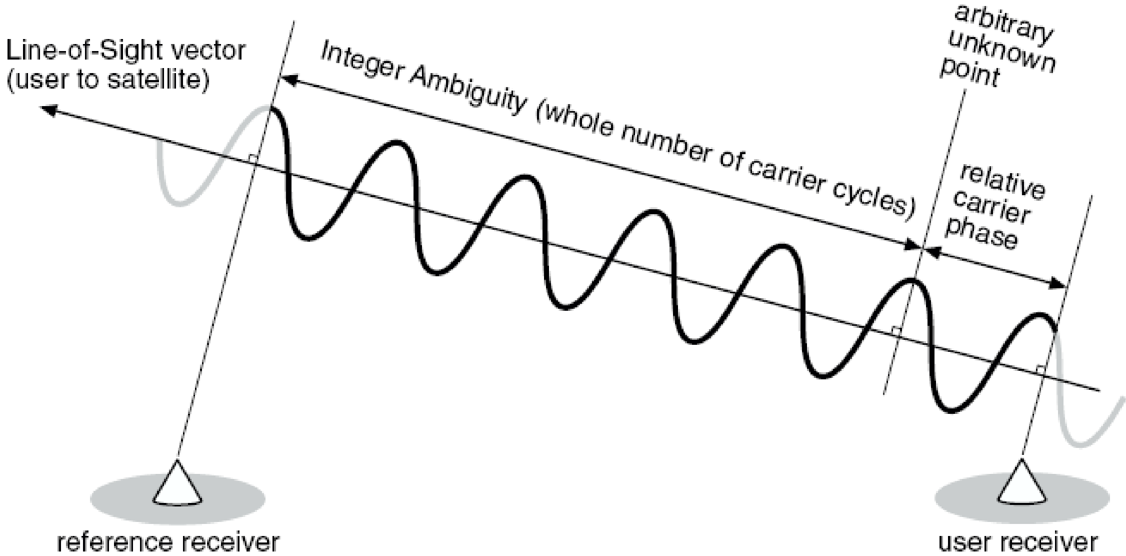
d. This solution quality is highly dependent on the accuracy of the known coordinates of each satellite (i.e. X^s, Y^s, and Z^s), the accuracy with which the atmospheric delays " d " can be estimated through modeling, and the accuracy of the resolution of the actual time measurement process performed in a GPS receiver (clock synchronization, signal processing, signal noise, etc.). As with any measurement process, repeated and long-term observations from a single point will enhance the overall positional reliability.

5.2 Carrier Phase Point Positioning

Carrier phase measurements are similar to pseudorange in that they are the difference in phase between the transmitting and receiving oscillators. Integration of the oscillator frequency gives clock time. Basic notion in carrier phase is: $\Phi = f\Delta t$, where Φ is phase and f is frequency.

Big problem is how to know the number of cycles in the phase measurements.

- GPS pseudo random code has a bit rate of about 1 MHz but its carrier frequency has a cycle rate of over a GHz (which is 1000 times faster!)
- At the speed of light the 1.57 GHz (L1 band), GPS signal has a wavelength of roughly twenty centimeters, so the carrier signal can act as a much more accurate reference than the pseudo random code by itself.
- And if we can get to within one percent of perfect phase like we do with code-phase receivers we would have 3 or 4 millimeter accuracy! This is the principle of carrier-based positioning.
- In essence this method is counting the exact number of carrier cycles between the satellite and the receiver.
- The problem is that the carrier frequency is hard to count because it's so uniform. Every cycle looks like every other.
- The pseudo random code on the other hand is intentionally complex to make it easier to know which cycle you're looking at.
- So the trick with "carrier-phase GPS" is to use code-phase techniques to get close (carrier-phases aided positioning).
- If the code measurement can be made accurate to say, a meter, then we only have a few wavelengths of carrier to consider as we try to determine which cycle really marks the edge of our timing pulse.
- Resolving this "carrier phase ambiguity" (or integer ambiguity) for just a few cycles is a much more tractable problem and as the computers inside the receivers get smarter and smarter



Number of integer (whole) carrier cycles between points N (integer ambiguity). Receiver measures the relative carrier phase up to N. N should be determined separately (e.g., via a reference receiver).

5.3 Pseudorange Relative Positioning

Absolute point positioning, as discussed earlier, will not provide the accuracies needed for most mapping and control projects due to existing and induced errors in the measurement process. In order to minimize these errors and obtain higher accuracies, GPS can be used in a relative or differential positioning mode (US Army Corps of Engineers, 2003).

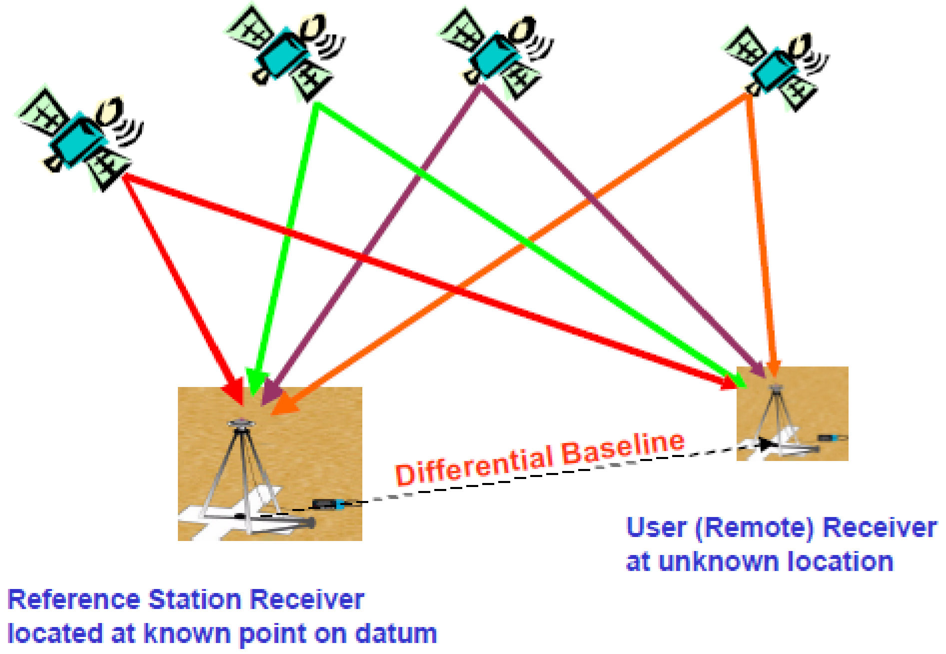


Figure 5-1. Differential or Relative GPS positioning

5-2. Differential Positioning Concepts

As stated in Chapter 2, differential GPS positioning is simply a process of determining the relative differences in coordinates between two receiver points, each of which is simultaneously observing/measuring satellite code ranges and/or carrier phases from the NAVSTAR GPS satellite constellation. These differential observations, in effect, derive a differential baseline vector between the two points, as illustrated in Figure 5-1. This method will position two stations relative to each other--hence the term "relative positioning"--and can provide the higher accuracies required for project control surveys, topographic surveys, and hydrographic surveys. There are basically two general types of differential positioning:

- Code phase pseudorange tracking
- Carrier phase tracking

Both methods, either directly or indirectly, determine the distance, or range, between a NAVSTAR GPS satellite and a ground-based receiver antenna. These measurements are made simultaneously at two different receiver stations. Either the satellite's carrier frequency phase, or the phase of a digital code modulated on the carrier phase, may be tracked--depending on the type of receiver. Through various processing techniques explained below, the distances between the satellites and receivers can be resolved, and the relative positions of the two receiver points are derived. From these relative observations, a baseline vector between the points is generated. The resultant positional accuracy is dependent on the tracking method used--carrier phase tracking being far more accurate than code phase tracking.

5-3. Differential Positioning (Code Pseudorange Tracking)

Code pseudorange tracking is the most widely used differential GPS positioning technique. It can deliver "meter-level" positional accuracies that typically range between 0.5 m to 5 m, depending on the code DGPS reference network and user receiver type. It is the technique used for maritime navigation, including USACE hydrographic surveying and dredge location applications. It is also used for air and land navigation where meter-level accuracy is required. Differential positioning using code pseudoranges is performed similarly to the Absolute Positioning techniques described in Chapter 4; however, some of the major clock error and atmospheric uncertainties are effectively minimized when simultaneous observations are made at two receiver stations. Errors in satellite range measurements are directly reflected in resultant coordinate errors. Differential positioning is not so concerned with the absolute position of the user but with the relative difference between two user positions who are simultaneously observing the same satellites. Since errors in the satellite position (X^s, Y^s, and Z^s) and atmospheric delay estimates (d) are effectively the same (i.e. highly correlated) at both receiving stations, they cancel each other to a large extent. Equation 4-1, which represents a general pseudorange observation, is repeated as Equation 5-1 below.

$$R = p^t + c(\Delta t) + d \qquad \text{(Eq 5-1)}$$

where

R	= observed pseudorange
p^t	= true range to satellite (unknown)
c	= velocity of propagation
Δt	= clock biases (receiver and satellite)
d	= propagation delays due to atmospheric conditions

The clock biases (Δt) and propagation delays (d) in the above equation are significantly minimized when code phase observations are made with two receivers. This allows for a relatively accurate pseudorange correction ($R - p^t$) to be computed at the receiver station set over a known point. This is because the true range (p^t) to the satellite can be determined from inversing between the ground station's coordinates and the broadcast satellite coordinates. If the pseudorange correction ($R - p^t$) is computed for 4 or more satellites, these pseudorange corrections can be transmitted to any number of user receivers to correct the raw pseudoranges originally observed. If 5 or more pseudorange corrections are observed, then a more reliable and redundant position computation is obtained. If more than one "reference station" is used to obtain pseudorange corrections, then the corrections may be further refined using the network of reference stations. Networks of stations transmitting differential GPS code correctors are termed as "augmented" GPS, or a wide area augmented system. Pseudorange corrections are broadcast by standard RF, satellite link, cell phone, or other transmission media. Satellite communications links are typically used for wide area augmentation networks. An alternate differential correction technique computes the position coordinate differences at the reference station and broadcasts these coordinate differences as correctors. This method is not widely used.

(US Army Corps of Engineers, 2003)

5.4 Carrier Phase Relative Positioning

Differential positioning using carrier phase tracking uses a formulation of pseudoranges similar to that done in code or absolute GPS positioning. However, the process becomes somewhat more complex when the carrier signals are tracked such that range changes are measured by phase resolution. The modulated codes are removed from the carrier, and a phase tracking process is used to measure the difference in phase of the received satellite signals between the reference receiver and the user’s receiver at an unknown point. The transmitted satellite signal is shifted in frequency due to the Doppler effect. The phase is not changed. GPS receivers measure what is termed the carrier phase “observable” – usually symbolized by “ϕ”. This observable represents the frequency difference between the satellite carrier and that generated in the receiver, or a so-called “beat” phase difference. This phase measurement observation can be shown in the following expression for the carrier phase observable (Kaplan 1996) (US Army Corps of Engineers, 2003).

$$\boldsymbol{\phi}_k^{\,P}(\mathrm{t}) = \phi_k^{\,P}(t) - \phi^P(t) + N_k^P + S_k + f\tau_P + f\tau_k - \beta_{\mathrm{iono}} + \delta_{\mathrm{tropo}} \qquad \text{(Eq 5-2)}$$

where

$\boldsymbol{\phi}_k^{\,P}(\mathrm{t})$ = length of propagation path between satellite "P" and receiver "k" ... in cycles
$\phi_k^{\,P}(t)$ = received phase of satellite "P" at receiver "k" at time "t"
$\phi^P(t)$ = transmitted phase of satellite "P"
N_k^P = integer ambiguity
S_k = measurement noise (multipath, GPS receiver, etc.)
f = carrier frequency (Hz)
τ_P = satellite clock bias
τ_k = receiver clock bias
β_{iono} = ionospheric advance (cycles)
δ_{tropo} = tropospheric delay (cycles)

For more details on these carrier phase observation models, see also Remondi (1985), Leick (1995), Van Sickle (2001), and other texts listed at Appendix A.

a. Typically, two receivers will be involved in carrier phase observations, and 4 or more satellites will be measured from both receivers. One of the receivers will be placed at a known reference point--the "reference" receiver. The other receiver is usually referred to as the "remote" or "rover" receiver--and is located a point where a map feature or project control point coordinate is required. This "rover" receiver may be stationary over the unknown point--i.e. "static"--or it may be roving from unknown point to unknown point--i.e. "kinematic."

b. Interferomic "differencing" techniques are used to resolve carrier phase observations made at two receivers. Differencing involves forming linear combinations between phase observations. To eliminate clock errors in the satellite, a "single difference" between phase measurements of the reference and remote receivers is performed. Single differencing between receivers eliminates the satellite clock error. This single differencing "between receivers" procedure is performed for all the mutually observed satellites, and the resultant single differences are subsequently differenced "between satellites" (i.e. "double differenced"), thus eliminating the receiver clock error. Double-differenced measurements on three pairs of satellites will yield the difference between the reference and remote locations. "Triple differencing" is the difference of two double differences performed over two different epochs. Triple differencing "between epochs" is used to indirectly resolve the number of whole carrier cycles between the satellite and receiver. There are a number of methods used to determine the integer ambiguity (the number of unknown integer cycles). These range from physical placement of the remote receiver a known distance from the reference receiver to automated Kalman filtering and searching methods. These differencing techniques are more fully described in Chapter 10.

(US Army Corps of Engineers, 2003)

5.5 Cycle Slip Detection and Correction

A loss of count of carrier cycles as they are being measured by a GPS receiver. Loss of signal, ionospheric interference, obstructions, and other forms of interference cause cycle slips to occur. To properly compute a vector between data collected from two GPS receivers, all cycle slips must be corrected (Geog862).

A cycle slip is a discontinuity in a receiver's continuous phase lock on a satellite's signal. A power loss, a very low signal-to-noise ratio, a failure of the receiver software, a malfunctioning satellite oscillator can cause a cycle slip. It can also be caused by severe ionospheric conditions. Most common, however, are obstructions such as buildings, trees and etc., that are so solid they prevent the satellite signal from being tracked by the receiver. Under such circumstances, when the satellite reappears, the tracking resumes (Geog862).

Coded pseudorange measurements are virtually immune from cycle slips, but carrier phase positioning accuracy suffers if cycle slips are not detected and repaired. A cycle slip causes the critical component for successful carrier phase positioning, a resolved integer cycle ambiguity, N, to become instantly unknown again. In other words, lock is lost. When that happens correct positioning requires that N be re-established (Geog862).

There are several methods of handling cycle slips. They are often controlled in post-processing rather than real-time (Geog862).

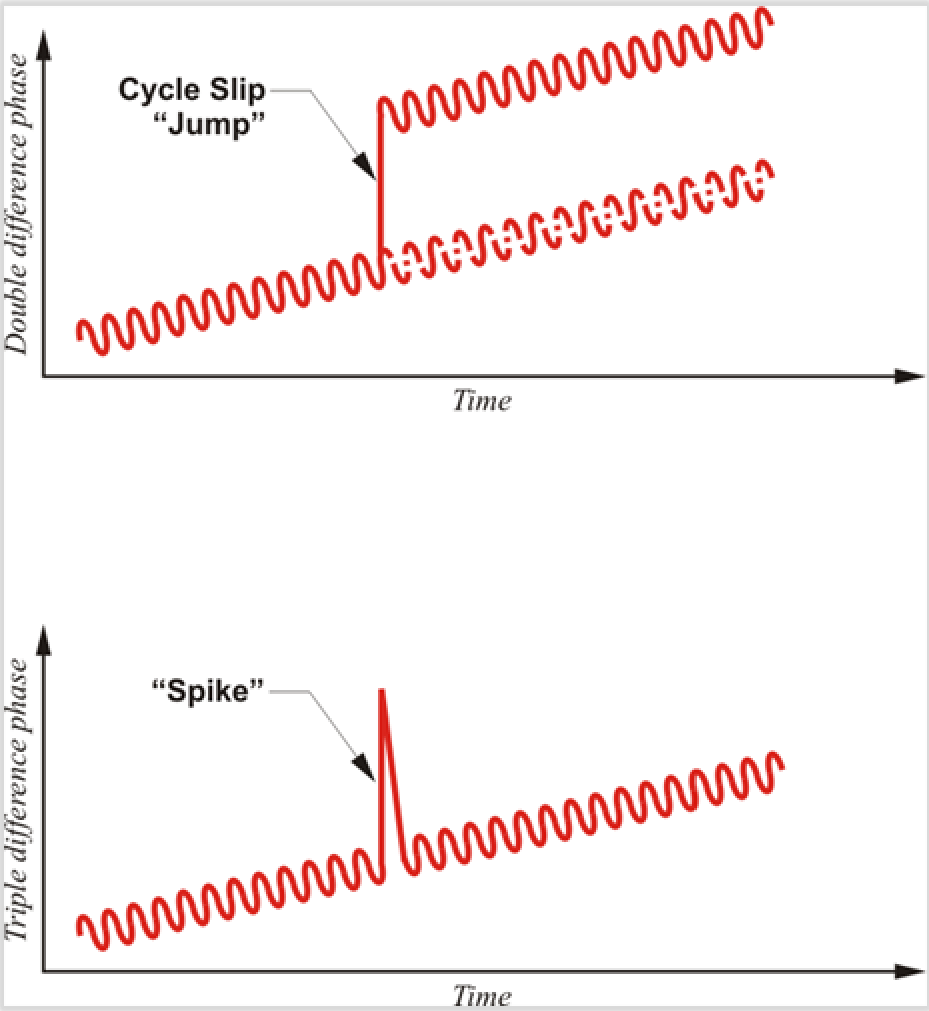


Figure 5-2: Cycle Slip

In post-processing the location and their size of cycle slips must be determined; then the data set can be repaired with the application of a fixed quantity to all the subsequent phase observations. One approach is to hold the initial positions of the stations occupied by the receivers as fixed and edit the data manually. This has proven to work, but would try the patience of Job. Another approach is to model the data on a satellite-dependent basis with continuous polynomials to find the breaks and then manually edit the data set a few cycles at a time. In fact, several methods are available to find the lost integer phase value, but they all involve testing quantities (Geog862).

One of the most convenient of these methods is based on the triple difference. It can provide an automated cycle slip detection system that is not confused by clock drift and, once least-squares convergence has been achieved, it can provide initial station positions even using the unrepaired phase combinations. They may still contain cycle slips but the data can nevertheless be used to process approximate baseline vectors. Then the residuals of these solutions are tested, sometimes through several iterations. Proceeding from its own station solutions, the triple difference can predict how many cycles will occur over a particular time interval. Therefore, by evaluating triple difference residuals over that particular interval, it is not only possible to determine which satellites have integer jumps, but also the number of cycles that have actually been lost. In a sound triple difference solution without cycle slips, the residuals are usually limited to fractions of a cycle. Only those containing cycle slips have residuals close to one cycle or larger. Once cycle slips are discovered, their correction can be systematic (Geog862).

For example, suppose the residuals of one component double difference of a triple-difference solution revealed that the residual of satellite PRN 16 minus the residual of satellite PRN 17 was 8.96 cycles. Further suppose that the residuals from the second component double difference showed that the residual of satellite PRN 17 minus the residual of satellite PRN 20 was 14.04 cycles. Then one might remove 9 cycles from PRN 16 and 14 cycles from PRN 20 for all the subsequent epochs of the observation. However, the process might result in a common integer error for PRNS 16, 17, and 20. Still, small jumps of a couple of cycles can be detected and fixed in the double-difference solutions (Geog862).

In other words, before attempting double difference solutions, the observations should be corrected for cycle slips identified from the triple difference solution. And even though small jumps undiscovered in the triple difference solution might remain in the data sets, the double difference residuals will reveal them at the epoch where they occurred (Geog862).

However, some conditions may prevent the resolution of cycle slips down to the one-cycle level. Inaccurate satellite ephemerides, noisy data, errors in the receiver's initial positions, or severe ionospheric effects all can limit the effectiveness of cycle-slip fixing. In difficult cases, a detailed inspection of the residuals might be the best way to locate the problem.A *cycle slip* is a discontinuity in a receiver's continuous phase lock on a satellite's signal. The coded pseudorange measurement is immune from this difficulty, but the carrier beat phase is not (Geog862).

Here's a graphic to give the idea of a cycle slip. As you see here, the cycle slip is indicated in a double difference, here, as a jump from one place to another. And in a triple difference, it's a spike (Geog862).

5.6 Carrier Phase Ambiguity Resolution

The unknown integer number of cycles of the reconstructed carrier phase contained in an unbroken set of data from a single satellite collected by a single receiver. Also known as integer ambiguity and integer bias.

Cycle ambiguity is the unknown number of whole carrier wavelengths between the satellite and receiver, as was described in Chapter 5. Successful ambiguity resolution is required for baseline formulations. Generally, in static surveying, ambiguity resolution can be achieved through long-term averaging and simple geometrical calibration principles, resulting in solutions to a linear equation that produces a resultant position. Thus, 30 minutes or more of observations may be required to resolve the ambiguities in static surveys. A variety of physical and mathematical techniques have been developed to rapidly resolve the carrier phase ambiguities. The physical methods involve observations over known length baselines or equivalent known points. The most reliable method is to set the base and remote receivers up over known WGS 84 points, and collect data for at least 30 seconds. Initialization can also be accomplished over extremely short baselines, such as those shown in Figure 9-13. Another method that was more commonly used in the past was a reference-rover antenna swapping process. Most GPS systems today can automatically resolve ambiguities mathematically "on-the-fly" (OTF)--the technique used for many real-time kinematic (RTK) applications.

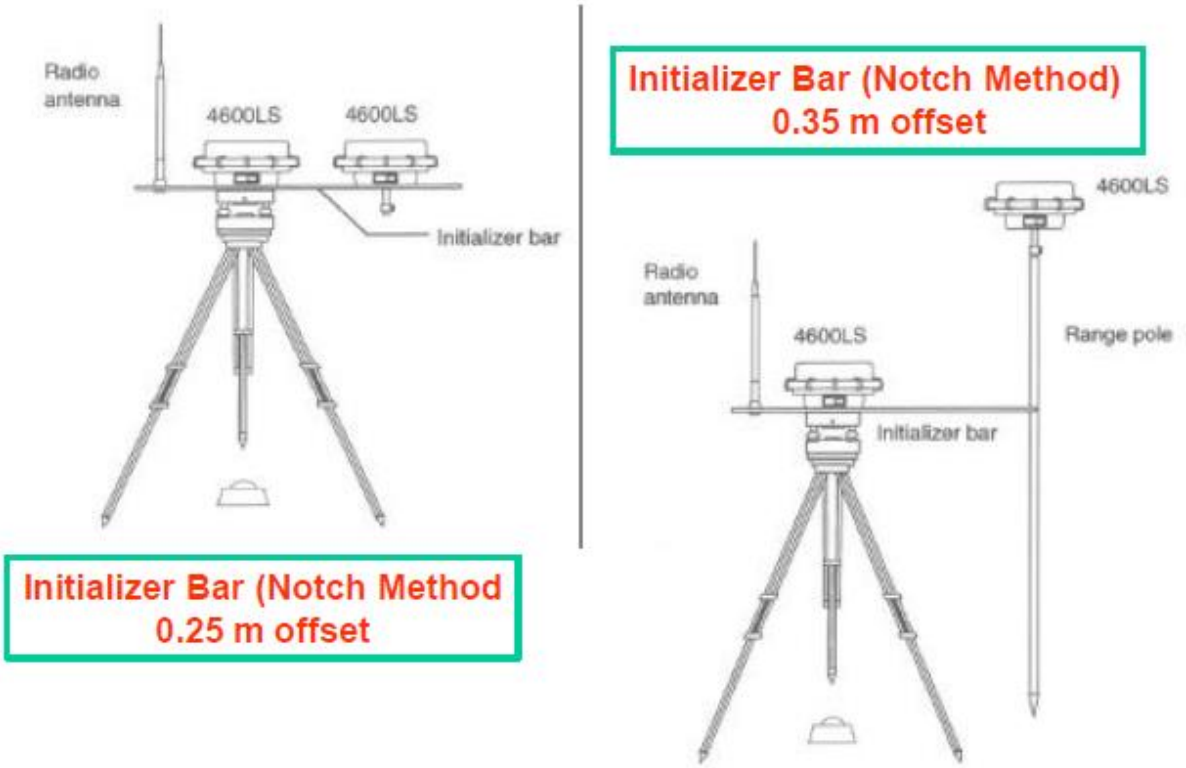


Figure 9-13. Ambiguity resolution of a Trimble 4600LS receiver using an Initializer Bar (Trimble Navigation, LTD)

CHAPTER 6 STATIC AND KINEMATIC POSITIONING

Contents of this Chapter

6.1 Static Positioning Performance and Applications

Static GPS surveying is a relative positioning technique that depends on the carrier-phase measurements. It employs two (or more) stationary receivers simultaneously tracking the same satellites (see Figure 6-1). One receiver, the base receiver, is set up over a point with precisely known coordinates such as a survey monument (sometimes referred to as the known point). The other receiver, the remote receiver, is set up over a point whose coordinates are sought (sometimes referred to as the unknown point). The base receiver can support any number of remote receivers, as long as a minimum of four common satellites is visible at both the base and the remote sites (El-Rabbany, 2006: 72).

In principle, this method is based on collecting simultaneous measurements at both the base and remote receivers for a certain period of time, which, after processing, yield the coordinates of the unknown point. The observation, or occupation, time varies from about 20 minutes to a few hours, depending on the distance between the base and the remote receivers (i.e., the baseline length), the number of visible satellites, and the satellite geometry. The measurements are usually taken at a recording interval of 15 or 20 seconds, or one sample measurement every 15 or 20 seconds (El-Rabbany, 2006: 72).

After completing the field measurements, the collected data is downloaded from the receivers into the PC for processing. Different processing options may be selected depending on the user requirements, the baseline length, and other factors. For example, if the baseline is relatively short, say, 15 or 20 km, resolving the ambiguity parameters would be a key issue to ensure high-precision positioning. As such, in this case the option of fixing the ambiguity parameters should be selected. In contrast, if the baseline is relatively long, a user may select the ionosphere-free linear combination option to remove the majority of the ionospheric error. This is because the ambiguity parameters may not be fixed reliably at the correct integer values (El-Rabbany, 2006: 72).

Static GPS surveying with the carrier-phase measurements is the most accurate positioning technique. This is mainly due to the significant change in satellite geometry over the long observation time span. Although both the single- and dual-frequency receivers can be used for static positioning, the latter is often used, especially for baselines exceeding 20 km. The expected accuracy from a geodetic quality receiver is typically 5 mm +1 ppm (rms), ppm for parts per million and rms for root-mean-square. That is, for a 10-km baseline, for example, the expected accuracy of the static GPS surveying is 1.5 cm (rms). Higher accuracy may be obtained by, for example, applying the precise ephemeris (El-Rabbany, 2006: 74).

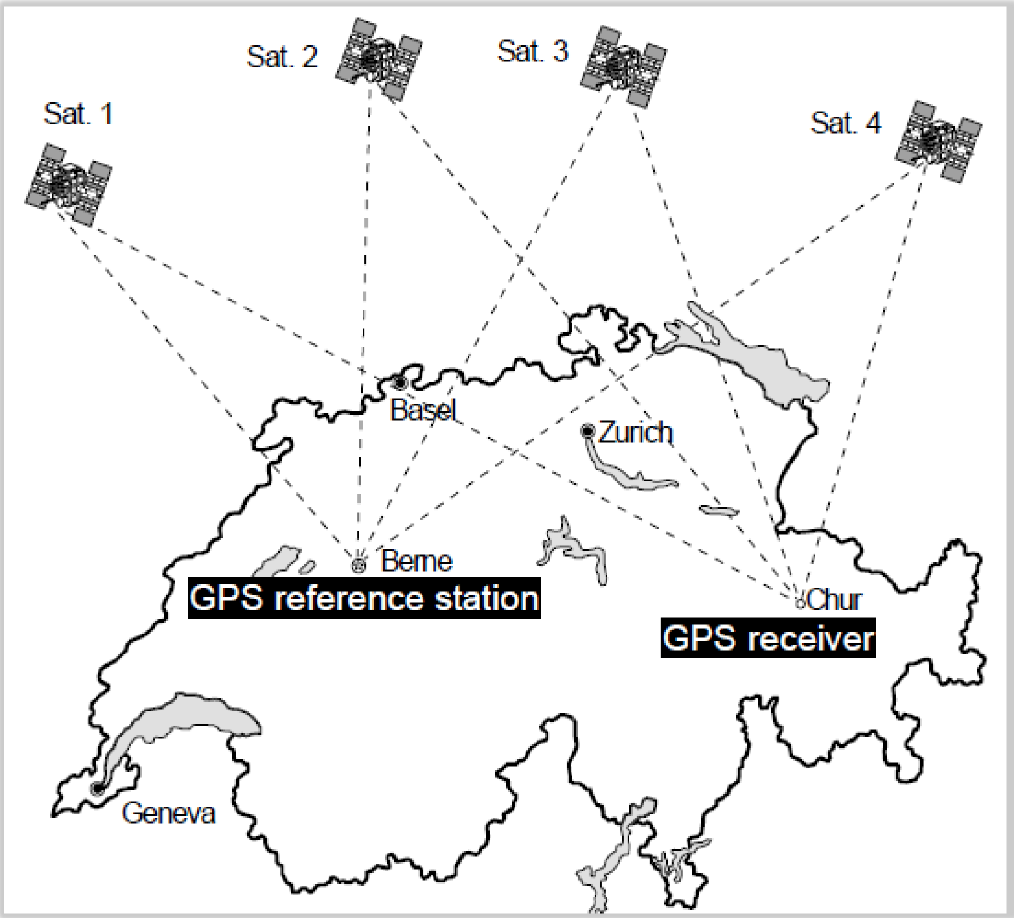


Figure 6-1: Static GNSS Positioning (Zogg, 2009)

6.2 Rapid Static Performance and Applications

Fast, or rapid, static surveying is a carrier-phase based relative positioning technique similar to static GPS surveying. That is, it employs two or more receivers simultaneously tracking the same satellites. However, with rapid static surveying, only the base receiver remains stationary over the known point during the entire observation session (see Figure 6-2). The rover receiver remains stationary over the unknown point for a short period of time only, and then moves to another point whose coordinates are sought. Similar to the static GPS surveying, the base receiver can support any number of rovers (El-Rabbany, 2006: 74).

This method is suitable when the survey involves a number of unknown points located in the vicinity (i.e., within up to about 15 km) of a known point. The survey starts by setting up the base receiver over the known point, while setting up the rover receiver over the first unknown point (Figure 6-2). The base receiver remains stationary and collects data continuously. The rover receiver collects data for a period of about 2 to 10 minutes, depending on the distance to the base as well as the satellite geometry. Once the rover receiver has collected the data, the user moves to the following point with unknown coordinates and repeats the procedures. It should be pointed out that, while moving, the rover receiver may be turned off. Due to the relatively short occupation time for the rover receiver, the recording interval is reduced to 5 seconds (El-Rabbany, 2006: 74).

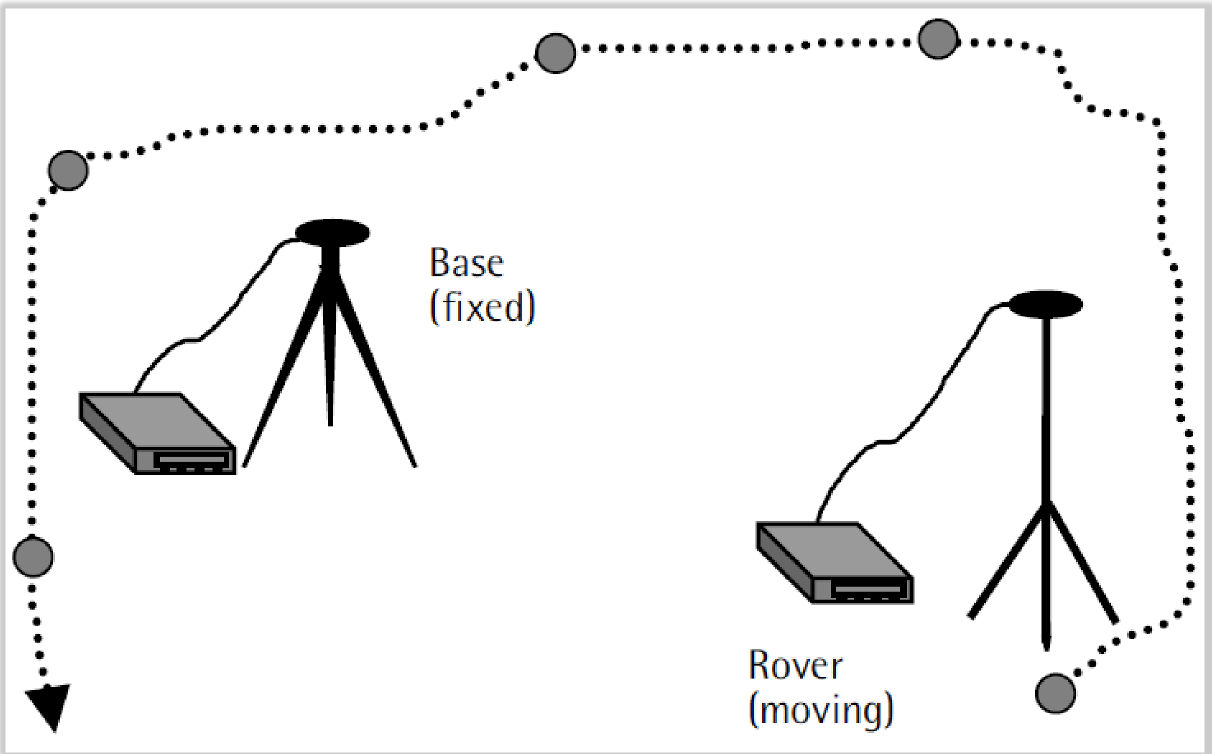


Figure 6-2: Rapid Static GNSS Positioning (El-Rabbany, 2006)

After collecting and downloading the field data from both receivers, the PC software is used for data processing. Depending on whether enough common data was collected, the software may output a fixed solution, which indicates that the ambiguity parameters were fixed at integer values. Otherwise, a float solution is obtained, which means that the software was unable to fix ambiguity parameters at integer values (i.e., only real-valued ambiguity parameters were obtained). This problem occurs mainly when the collected GPS data is insufficient. A fixed solution means that the positioning accuracy is at the centimetre level, while the float solution means that the positioning accuracy is at the decimetre or sub meter level. Although both the single- and dual-frequency receivers can be used for fast static surveying, the probability of getting a fixed solution is higher with the latter (El-Rabbany, 2006: 75).

It is used to measure baselines and determine positions at the centimetre-level with short, static observation times e.g., 5-20 minutes. The observation time is dependent on the length of the baseline and number of visible satellites. Loss of lock, when moving from one station to the next, can also occur since each baseline is processed independent of each other. Unlike Pseudo Kinematic, stations are occupied only once. Dual-frequency receivers are required (US Army Corps of Engineers, 2003: 5-6).

6.3 Kinematic Positioning Performance and Applications

As the name implies, during kinematic surveys one receiver, the rover, can be in continuous motion. This is the most productive of the survey methods but is also the least accurate. The accuracy or a kinematic survey is typically in the range of ±(1 to 2 cm + 2 ppm). This accuracy is sufficient for many types of surveys and thus is the most common method of surveying. Kinematic methods are applicable for any type of survey that requires many points to be located, which makes it very appropriate for most topographic and construction surveys. It is also excellent for dynamic surveying, that is, where the observation station is in motion. The range of a kinematic survey is typically limited to the broadcast range of the base radio. However, real-time networks have made kinematic surveys possible over large regions (Ghilani and Wolf , 2012: 371).

Kinematic surveying is a GPS carrier phase surveying technique that allows the user to rapidly and accurately measure baselines while moving from one point to the next, stopping only briefly at the unknown points, or in dynamic motion such as a survey boat or aircraft. A reference receiver is set up at a known station and a remote, or rover receiver traverses between the unknown points to be positioned. The data is collected and processed (either in real-time or post-time) to obtain accurate positions to the centimeter level. Kinematic survey techniques require some form of initialization to resolve the carrier phase ambiguities. This can be done by setting the remote receiver on a known baseline relative to the reference receiver, by performing an "antenna swap" procedure between the two receivers, and other techniques such as "On-the-Fly" or OTF (US Army Corps of Engineers, 2003).

6.4 Pseudo-Kinematic (Pseudo-Static) Positioning

This procedure is also known as the intermittent or reoccupation method, and like the other static methods requires a minimum of two receivers. In pseudo-kinematic surveying, the base receiver always stays on a control station, while the rover goes to each point of unknown position. Two relatively short observation sessions (around 5 min each in duration) are conducted with the rover on each station. The time lapse between the first session at a station, and the repeat session, should be about an hour. This produces an increase in the geometric strength of the observations due to the change in satellite geometry that occurs over the time period (Ghilani and Wolf , 2012: 371).

A disadvantage of this method, compared to other static methods, is the need to revisit the stations. This procedure requires careful pre-survey planning to ensure that sufficient time is available for site revisitation, and to achieve the most efficient travel plan. Pseudo-kinematic surveys are most appropriately used where the points to be surveyed are along a road, and rapid movement from one site to another can be readily accomplished. During the movement from one site to another, the receiver can be turned off. Some projects for which pseudo-kinematic surveys may be appropriate include alignment surveys, photo-control surveys, lower-order control surveys, and mining surveys. Given the speed and accuracy of kinematic surveys, however, this survey procedure is seldom used in practice (Ghilani and Wolf , 2012: 371).

6.5 Semi (or Stop and Go) Kinematic Positioning

Semi-kinematic or the stop-and-go mode, is useful for mapping and construction surveys where increased accuracy is desired for a specific feature. In the semi-kinematic mode, the antenna is positioned over points of interest and a point identifier is entered into the survey controller for each feature. Since multiple epochs of data are usually recorded at each point, the accuracy of this mode is greater than that obtainable in the true kinematic mode. In both surveys, the rate of data collection at the base station and rover is typically set to 1 sec (Ghilani and Wolf , 2012: 405).

Stop and Go Kinematic involves collecting static data for several minutes (i.e., 10 to 30 minutes) at each station after a period of initialization to gain the integers. This technique does not allow for loss of satellite lock during the survey. If loss of satellite lock does occur, a new period of initialization must take place. This method can be performed with two fixed or known stations in order to provide redundancy and improve accuracy (US Army Corps of Engineers, 2003: 5-6)

6.6 Real-time Positioning

RTK surveying is a carrier phase based relative positioning technique that, like the previous methods, employs two (or more) receivers simultaneously tracking the same satellites (Figure 6-3). This method is suitable when: (1) the survey involves a large number of unknown points located in the vicinity (i.e., within up to about 10.15 km) of a known point; (2) the coordinates of the unknown points are required in real time; and (3) the line of sight, the propagation path, is relatively unobstructed. Because of its ease of use as well as its capability to determine the coordinates in real time, this method is the preferred method by many users (El-Rabbany, 2006: 77).

In this method, the base receiver remains stationary over the known point and is attached to a radio transmitter (Figure 6-3). The rover receiver is normally carried in a backpack and is attached to a radio receiver. Similar to the conventional kinematic GPS method, a data rate as high as 1 Hz (one sample per second) is required. The base receiver measurements and coordinates are transmitted to the rover receiver through the communication (radio) link. The built-in software in a rover receiver combines and processes the GPS measurements collected at both the base and the rover receivers to obtain the rover coordinates (El-Rabbany, 2006: 78).

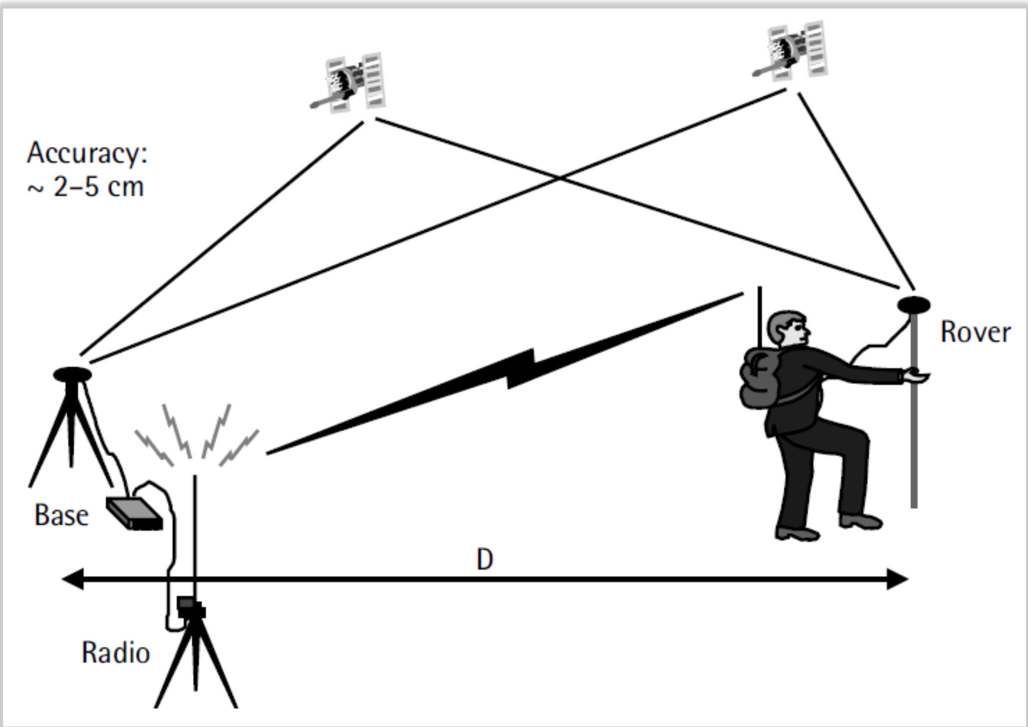


Figure 6-3: Real Time Kinematic Survey (El-Rabbany, 2006)

The initial ambiguity parameters are determined almost instantaneously using a technique called on-the-fly (OTF) ambiguity resolution, to be discussed in the next chapter. Once the ambiguity parameters are fixed to integer values, the receiver (or its handheld computer controller) will display the rover coordinates right in the field. That is, no post-processing is required. The expected positioning accuracy is of the order of 2 to 5 cm (rms). This can be improved by staying over the point for a short period of time, for example, about 30 seconds, to allow for averaging the position. The computed rover coordinates for the entire survey may be stored and downloaded at a later time into CAD software for further analysis. This method is used mainly, but not exclusively, with dual-frequency receivers (El-Rabbany, 2006: 78).

Under the same conditions, the positioning accuracy of the RTK method is slightly degraded compared with that of the conventional kinematic GPS method. This is mainly because the time tags (or time stamps) of the conventional kinematic data from both the base and the rover match perfectly in the processing. With RTK, however, the base receiver data reaches the rover after some delay (or latency). Data latency occurs as a result of formatting, packetizing, transmitting, and decoding the base data. To match the time tag of the rover data, the base data must be extrapolated, which degrades the positioning accuracy (El-Rabbany, 2006: 78).

The RTK positioning methods will yield sub-decimeter accuracies in real-time. This method has become widely used for accurate engineering and construction surveys, including topographic site plan mapping, construction stake out, construction equipment location, and hydrographic surveying. This GPS technique determines the integer number of carrier wavelengths between the GPS antenna to the GPS satellite while the rover receiver is in motion and without static initialization. RTK typically uses an "On-the-Fly" (OTF) integer initialization process whereby initialization can be performed while the roving receiver is moving. Periodic loss of satellite lock can be tolerated and no static initialization is required to regain the integers. This differs from other GPS techniques that require static initialization while the user is stationary. A communication link between the reference and rover receivers is required. A number of techniques have been developed to increase RTK accuracies over local areas, such as placing simulated GPS satellite receivers at fixed ground locations (pseudolites). These have application in obscured areas (underground, tunnels, inside buildings, etc.) or for accurate aircraft landing elevation measurement (US Army Corps of Engineers, 2003: 5-6).

The basic practical concept for real-time kinematic GPS surveying was developed in the early 1980's by Ben Remondi of the National Geodetic Survey. In 1989, the Corps' Topographic Engineering Center (ERDC/TEC) began development of algorithms to enable RTK observation of tides for hydrographic survey and dredge elevation corrections in offshore environments. Today, nearly all GPS receiver manufactures provide RTK survey options for engineering, construction, and boundary survey applications (US Army Corps of Engineers, 2003: 5-7).

6.7 Real-Time Differential GPS

Real-time differential GPS (DGPS) is a code-based relative positioning technique that employs two or more receivers simultaneously tracking the same satellites (**Figure 6-4**). It is used when a real-time meter-level accuracy is enough. The method is based on the fact that the GPS errors in the measured pseudoranges are essentially the same at both the base and the rover, as long as the baseline length is within a few hundred kilometres (El-Rabbany, 2006: 79).

As before, the base receiver remains stationary over the known point. The built-in software in the base receiver uses the precisely known base coordinates as well as the satellite coordinates, derived from the navigation message, to compute the ranges to each satellite in view. The software further takes the difference between the computed ranges and the measured code pseudoranges to obtain the pseudorange errors (or DGPS corrections). These corrections are transmitted in a standard format called Radio Technical Commission for Maritime Service (RTCM) to the rover through a communication link. The rover then applies the DGPS corrections to correct the measured pseudoranges at the rover. Finally, the corrected pseudoranges are used to compute the rover coordinates (El-Rabbany, 2006: 79).

The accuracy obtained with this method varies between a submeter and about 5m, depending on the base-rover distance, the transmission rate of the RTCM DGPS corrections, and the performance of the C/A-code receivers.Higher accuracy is obtained with short base-rover separation, high transmission rate, and carrier-smoothed C/A-code ranges. With the termination of selective availability, the data rate could be reduced to 10 seconds or lower without noticeable accuracy degradation. Further accuracy improvement could be achieved if the receivers are capable of storing the raw pseudorange measurements, which could be used at a later time in the post-processing mode. As the real-time DGPS is widely used, some governmental agencies as well as private firms are providing the RTCM DGPS corrections either at no cost or at certain fees. More about these services will be given in Chapter 7 (El-Rabbany, 2006: 80).

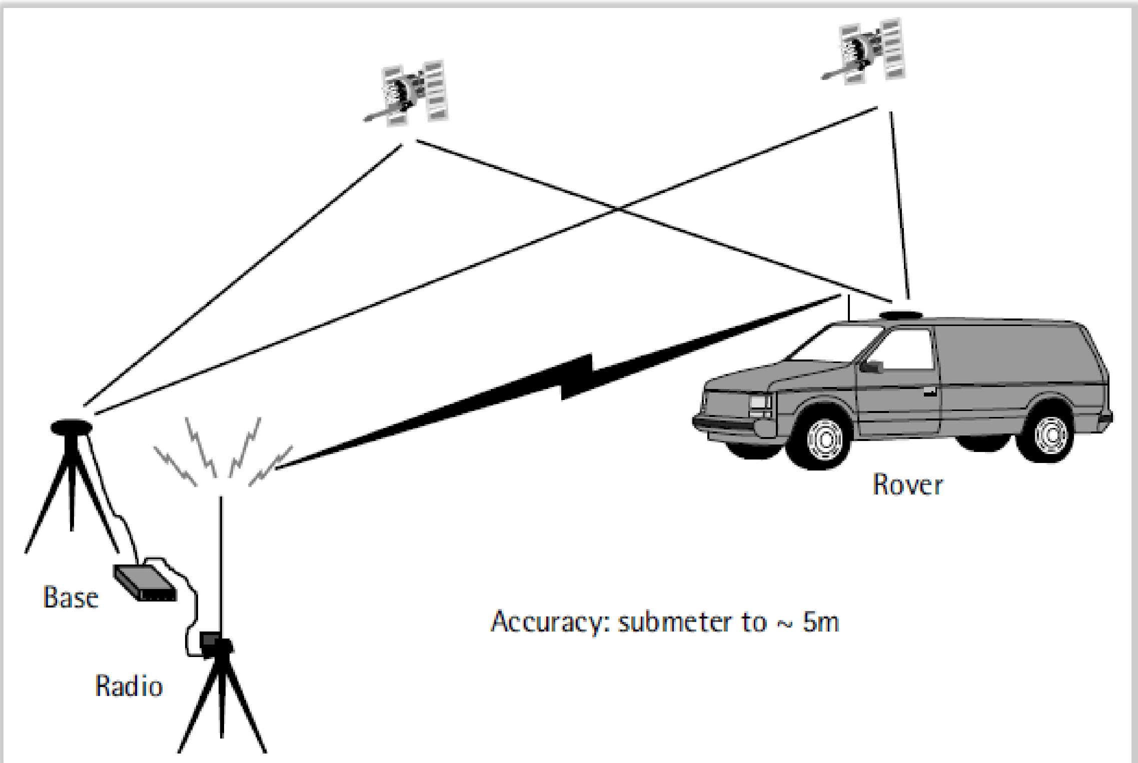


Figure 6-4: Real-time Differential GPS (El-Rabbany, 2006)

CHAPTER 7 SPECIFICATIONS AND FIELD SURVEYS

Contents of this Chapter

7.1 Satellite Survey Planning

Small surveys generally do not require much in the way of project planning. However, for large projects and for higher-accuracy surveys, project planning is a critical component in obtaining successful results. The subsections that follow discuss various aspects of project planning with emphasis on control surveys (Ghilani and Wolf , 2012: 372).

7.1.1 Preliminary Considerations

All new high-accuracy survey projects that employ relative positioning techniques must be tied to nearby existing control points. Thus, one of the first things that must be done in planning a new project is to obtain information on the availability of existing control stations near the project area. For planning purposes, these should be plotted in their correct locations on an existing map or aerial photos of the area (Ghilani and Wolf , 2012: 372).

Another important factor that must be addressed in the preliminary stages of planning for projects is the selection of the new station locations. Of course, they must be chosen so that they meet the overall project objective. But in addition, terrain, vegetation, and other factors must be considered in their selection. If possible, they should be reasonably accessible by either the land vehicles or aircraft that will be used to transport the survey hardware. The stations can be somewhat removed from vehicle access points since hardware components are relatively small and portable. Also, the receiver antenna is the only hardware component that must be accurately centered over the ground station. It is easily hand carried and, when possible, can be separated from the other components by a length of cable, as shown in Figure 14.3. Once the preliminary station locations are selected, they should be plotted on the map or aerial photo of the area (Ghilani and Wolf , 2012).

Another consideration in station selection is the assurance of an overhead view free of obstructions. This is known as canopy restrictions. Canopy restrictions may possibly block satellite signals, thus reducing observations and possibly adversely affecting satellite geometry. At a minimum, it is recommended that visibility be clear in all directions from a mask angle (altitude angle) of 10° to 20° from the horizon. In some cases, careful station placement will enable this visibility criterion to be met without difficulty; in other situations clearing around the stations may be necessary. Furthermore, as discussed in Section 13.6.3, potential sources that can cause interference and multipath errors should also be identified when visiting each site (Ghilani and Wolf , 2012).

A number of factors need to be considered during the planning phase of a proposed GPS data collection. These include:

- Project Application – Purpose of Data Collection Survey
  - Establishing primary control for subsequent location, topographic, hydrographic, or utility survey
  - General site plan, feature mapping, or GIS densification survey
  - Number of horizontal points of benchmarks required or to be occupied
  - Datum – horizontal and vertical
- Accuracy Requirements
  - Horizontal and vertical
  - Will GPS provide the necessary accuracy
- Equipment Resources
  - In-house or contract
  - GPS receiver availability
  - Other auxiliary equipment availability
- GPS Procedure
  - High accuracy – use centimetre-level static of kinematic carrier phase
  - Medium accuracy – use meter-level code phase
  - Low accuracy – use 10 to 30 m level absolute positioning
  - RTK options for topographic mapping
- Network Design and Connections
  - Static baseline connections to local project control
  - Connections to NGRS/CORS points
  - Code phase connections with wide-area commercial, USCG, or FAA WAAS networks
- Data Collection and Adjustment Techniques
  - Feature, attribute, and format requirements
  - Data collection session time
  - GPS initialization and calibration requirements
  - Multiple/repeat baseline requirements
  - Loop requirements
  - Other quality control requirements
  - Adjustment criteria accuracy standards

- Metadata requirements
- Final survey report format

- Site Access Restrictions
  - Reconnaissance survey required
  - Potential visibility restriction or multipath problems

- Funding Considerations (impacts many or the above factors)

The above list is not exhaustive-numerous other project-specific conditions need to be considered. The …

7.2 Initialization

To start a kinematic survey, the receivers must be initialized. This process includes determining the integer ambiguity (see Section 13.5.2) for each pseudo-range observation. Following any of the methods described below can yield initialization of the receivers (Ghilani and Wolf , 2012: 402).

One procedure for initializing the receivers uses a baseline whose ΔX, ΔY, and ΔZ components are known. A very short static observing session is conducted with base and roving receivers occupying two stations with known positions simultaneously. Because the baseline coordinate differences are known, differencing of the observations will yield the unknown integer ambiguities. These differencing computations are performed in a post-processing operation using the data from both receivers. If only one control station is available, a second one can be set using the static or rapid static surveying methods.

An alternative initialization procedure, called antenna swapping, is also suitable if only one control station is available. Here receiver A is placed on the control point and receiver B on a nearby, unknown point. For convenience, the unknown point can be within 30 ft (10 m) of the control station. After a few minutes of data collection with both receivers, their positions are interchanged while keeping them running. In the interchange process, care must be exercised to make certain continuous tracking, or "lock" is maintained on at least four satellites. After a few more minutes of observations, the receivers are interchanged again, returning them to their starting positions. This procedure enables the baseline coordinate differences and the integer ambiguities to be determined, again by differencing techniques.

Finally, the most advanced techniques of initialization are known as on-the-fly (OTF) ambiguity resolution methods. These methods require five usable satellites during the initialization process and dual-frequency receivers. OTF, which involves the solution of a sophisticated mathematical algorithm, has resolved ambiguities to the centimeter level in 2 min for a 20-km line. However, longer sessions are sometimes necessary to resolve the ambiguities since ideal conditions are not always available. The typical period for ambiguity resolution is usually less than five minutes. It is not uncommon with current processing techniques to resolve ambiguities in under a minute. As mentioned in Section 13.11, when four satellites with the L5 frequencies are available, the ambiguities can be mathematically determined in a single epoch of data eliminating the need for the previous methods discussed herein.

7.3 Survey Planning and Dilution of Precision (DOP)

Dilution of Precision (DOP) is an indicator of satellite geometry quality for a unique constellation. Poor satellite geometry leads to poor Dilution of Precision (DOP), triangulation, and location estimation. A low Dilution of Precision (DOP) value represents a better positional precision because of wide angular separation between the satellites used to calculate a terrestrial position (Rydlund Jr. and Densmore, 2012).

The Position Dilution of Precision (PDOP) represents the geometry of the GNSS satellite constellation and its effect on precision. Satellite constellation geometry is the basis for a method, known as trilateration, which provides dimensions of position for the GNSS receiver. Regarding trilateration, the determination of point locations by distance measurements using geometry is much more favorable toward a well-dispersed satellite constellation over the field of view, as opposed to a less-dispersed satellite constellation that biases the field of view; the satellites in the field of view of a GNSS receiver are positionally related to one another to provide a level of precision in each dimension of the receiver measurement. The configuration of these satellites affects horizontal and vertical uncertainties, and represents a unitless positioning value, known as PDOP, simplified and expressed as the ratio of the positioning accuracy to the measurement accuracy (Henning, 2010). A lower PDOP represents a well dispersed satellite constellation, which indicates a favorable ratio of positional accuracy to measurement accuracy. Conversely, geometry of the satellite constellation that is less dispersed produces a higher PDOP value. Additional dimensions of dilution of precision include horizontal, vertical, and time. The Horizontal Dilution of Precision (HDOP) represents horizontal accuracy in two dimensions, and the Vertical Dilution of Precision (VDOP) represents the vertical accuracy in one dimension (height). The relation between these variations and PDOP is expressed as: $PDOP^2=HDOP^2+VDOP^2$ (Henning, 2010). The variations of PDOP and their relation to quality are illustrated in figure 15. The vertical component of the GNSS position is the most likely component to be lacking in quality if the PDOP values are high (Skeen, 2005) (Rydlund Jr. and Densmore, 2012).

A third variation, known as the Time Dilution of Precision (TDOP), represents how the satellite geometry is affecting the ability of the GNSS receiver to determine time. The association of time can be equated to a measure of the overall uncertainty in a GNSS solution, known as the Geometric Dilution of Precision (GDOP). GDOP is defined in a similar manner as PDOP, with the inclusion of time, and the relation between these two is defined as $GDOP^2=PDOP^2+TDOP^2$. General experience in GNSS positioning may broadly classify a PDOP value less than or equal to 3 as generally sufficient, a value greater than 3 but less than or equal to 5 as marginal, and values greater than or equal to 5 as poor. GNSS receivers may be set to mask and cutoff PDOP values of positioning, such that the receiver stops computing position fixes for a satellite orientation that yields a PDOP greater than the mask value (Skeen, 2005). As discussed in the previous section, an appropriate elevation mask needs to be set to remove GNSS signals encountering lengthy travel through the ionosphere and troposphere; however, too much exclusion, such as a mask that is greater than 15 degrees, can introduce undesirable satellite geometry and adversely affect PDOP. Adding a supplemental fixed-height range pole to increase the height of the receiver may improve satellite availability and overall

PDOP; however, the user needs to ensure that the extended height does not compromise the stability of the receiver (Rydlund Jr. and Densmore, 2012).

7.4 Quality Assurance

7.4.1 Dilution of Precision

Dilution of Precision (DOP) is an indicator of satellite geometry quality for a unique constellation. Poor satellite geometry leads to poor Dilution of Precision (DOP), triangulation, and location estimation. A low Dilution of Precision (DOP) value represents a better positional precision because of wide angular separation between the satellites used to calculate a terrestrial position (Rydlund Jr. and Densmore, 2012).

The Position Dilution of Precision (PDOP) represents the geometry of the GNSS satellite constellation and its effect on precision. Satellite constellation geometry is the basis for a method, known as trilateration, which provides dimensions of position for the GNSS receiver. Regarding trilateration, the determination of point locations by distance measurements using geometry is much more favorable toward a well-dispersed satellite constellation over the field of view, as opposed to a less-dispersed satellite constellation that biases the field of view; the satellites in the field of view of a GNSS receiver are positionally related to one another to provide a level of precision in each dimension of the receiver measurement. The configuration of these satellites affects horizontal and vertical uncertainties, and represents a unitless positioning value, known as PDOP, simplified and expressed as the ratio of the positioning accuracy to the measurement accuracy (Henning, 2010). A lower PDOP represents a well dispersed satellite constellation, which indicates a favorable ratio of positional accuracy to measurement accuracy. Conversely, geometry of the satellite constellation that is less dispersed produces a higher PDOP value. Additional dimensions of dilution of precision include horizontal, vertical, and time. The Horizontal Dilution of Precision (HDOP) represents horizontal accuracy in two dimensions, and the Vertical Dilution of Precision (VDOP) represents the vertical accuracy in one dimension (height). The relation between these variations and PDOP is expressed as: $PDOP^2=HDOP^2+VDOP^2$ (Henning, 2010). The variations of PDOP and their relation to quality are illustrated in figure 15. The vertical component of the GNSS position is the most likely component to be lacking in quality if the PDOP values are high (Skeen, 2005) (Rydlund Jr. and Densmore, 2012).

A third variation, known as the Time Dilution of Precision (TDOP), represents how the satellite geometry is affecting the ability of the GNSS receiver to determine time. The association of time can be equated to a measure of the overall uncertainty in a GNSS solution, known as the Geometric Dilution of Precision (GDOP). GDOP is defined in a similar manner as PDOP, with the inclusion of time, and the relation between these two is defined as $GDOP^2=PDOP^2+TDOP^2$. General experience in GNSS positioning may broadly classify a PDOP value less than or equal to 3 as generally sufficient, a value greater than 3 but less than or equal to 5 as marginal, and values greater than or equal to 5 as poor. GNSS receivers may be set to mask and cutoff PDOP values of positioning, such that the receiver stops computing position fixes for a satellite orientation that yields a PDOP greater than the mask value (Skeen, 2005). As discussed in the previous section, an appropriate elevation mask needs to be set to remove GNSS signals encountering lengthy travel through the ionosphere and troposphere; however, too much exclusion, such as a mask that is greater than 15 degrees, can introduce undesirable satellite geometry and adversely affect PDOP. Adding a supplemental fixed-height range pole to increase the height of the receiver may improve satellite availability and overall PDOP; however, the user needs to ensure that the extended height does not compromise the stability of the receiver (Rydlund Jr. and Densmore, 2012).

The precision of positioning with GPS navigation depends on the one hand on the precision of the individual pseudorange measurements and on the other hand on the geometric configuration of the satellites used. This configuration is expressed in terms of a scalar value, which is referred to in navigation literature as DOP (Dilution of Precision).

The DOP value describes the weakening of precision and is therefore a factor or measure of the constellation dependent imprecision. If the DOP values are high (because for example all visible satellites are close to one another), then the anticipated imprecision will be higher.

There are a variety of DOP terms used:

- GDOP (Geometric-DOP): Describes the influence of satellite geometry on the position in 3D space and time measurement.
- PDOP (Positional-DOP): Describes the influence of satellite geometry on the position in 3D space.
- HDOP (Horizontal-DOP): Describes the influence of satellite geometry on the position along upon a plane (2D)
- VDOP (Vertical-DOP): Describes the influence of satellite geometry on height (1D).
- TDOP (Time-DOP): Describes the influence of satellite geometry on time measurement.

The influence of satellite geometry on imprecision is demonstrated in Figure 85. When both satellites are widely separated (figure left) the position error (area in red) is smaller. If the satellites are close to one another (figure right), then the area of error is more spread out. This is valid when the uncertainty for determining the position, known as the Range Error (R-E: yellow and blue areas), is the same for both satellites. R (R1 and R2) refers to the measured distance of the satellites to the user (pseudorange).

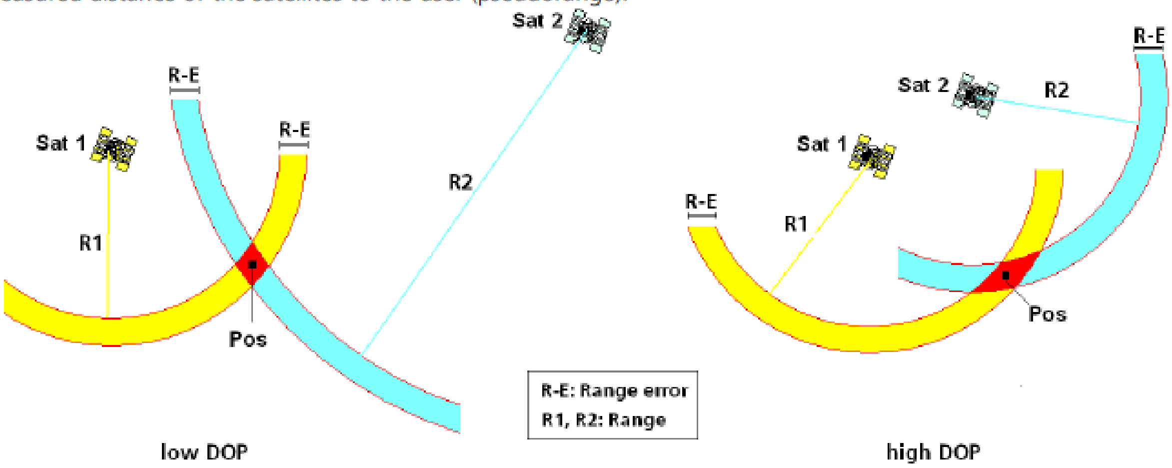


Figure 85: The flatter the angle with which the circles with ranges R1 and R2 intersect, the higher the DOP value

a. The accuracy of the positioned point is a function of the range measurement accuracy and the geometry of the satellites, as reduced to spherical intersections with the earth's surface. A description of the geometrical magnification of uncertainty in a GPS determined point position is termed "Dilution of Precision" (DOP), which is discussed in a later section. Repeated and redundant range observations will generally improve range accuracy. However, the dilution of precision remains the same. In a static mode (meaning the GPS receiver antenna stays stationary), range measurements to each satellite may be continuously remeasured over varying orbital locations of the satellites. The varying satellite orbits cause varying positional intersection geometry. In addition, simultaneous range observations to numerous satellites can be adjusted using weighting techniques based on the elevation and pseudorange measurement reliability.

a. In a more practical sense, GDOP is a scalar quantity of the contribution of the configuration of satellite constellation geometry to the GPS accuracy, in other words, a measure of the "strength" of the geometry of the satellite configuration. In general, the more satellites that can be observed and used in the final solution, the better the solution. Since GDOP can be used as a measure of the geometrical strength, it can also be used to selectively choose four satellites in a particular constellation that will provide the best solution. Satellites spread around the horizon will provide the best horizontal position, but the weakest vertical elevation. Conversely, if all satellites are at high altitudes, then the precision of the horizontal solution drops but the vertical improves. This is illustrated in Figure 4-5. The smaller the GDOP, the more accurate the position.

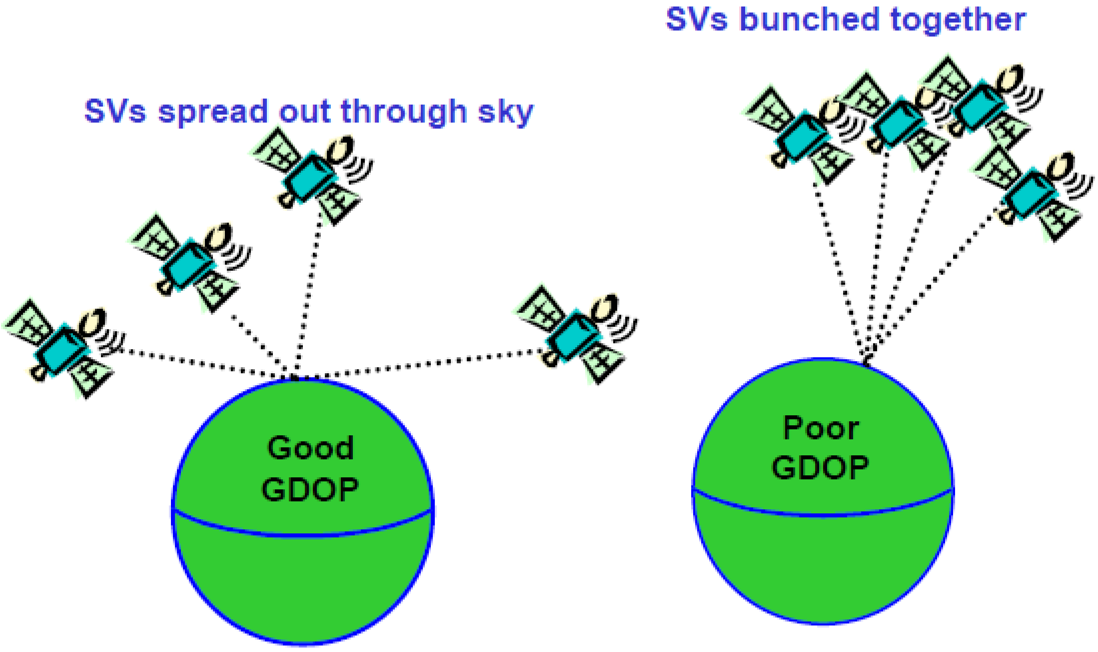


Figure 4-5. Satellite geometry and GDOP--"Good" GDOP and "Poor" GDOP configurations

b. GDOP values used in absolute GPS positioning is a measure of spatial accuracy of a 3-D position and time. The GDOP is constantly changing as the relative orientation and visibility of the satellites change. GDOP can be computed in the GPS receivers in real-time, and can be used as a quality control indicator. GDOP is defined to be the square root of the sum of the variances of the position and time error estimates.

$$GDOP = [\sigma_E^2 + \sigma_N^2 + \sigma_U^2 + \sigma_R^2 + (c * \delta_T)^2]^{0.5} \cdot [1 / \sigma_R] \qquad \text{(Eq 4-9)}$$

where

σ_E = standard deviation in east value, m
σ_N = standard deviation in north value, m
σ_U = standard deviation in up direction, m
c = speed of light (299,338,582.7 m/s)
δ_T = standard deviation in time, seconds
σ_R = overall standard deviation in range in meters, i.e. the UERE at the one-sigma (68%) level

The GDOP value is easily estimated by assuming the UEREs are all unity and then pulling the standard deviations directly from the variance-covariance matrix of the position adjustment. Thus GDOP (and its derivations) can be recomputed at each position update (e.g., every second). Large jumps (increases) in GDOP values are poor performance indicators, and typically occur as satellites are moved in and out of the solution.

c. Positional dilution of precision (PDOP). PDOP is a measure of the accuracy in 3-D position, mathematically defined as:

$$PDOP = [\sigma_E^2 + \sigma_N^2 + \sigma_U^2]^{0.5} \cdot [1 / \sigma_R] \qquad \text{(Eq 4-10)}$$

where all variables are equivalent to those used in Equation 4-9. PDOP is simply GDOP less the time bias.

(1) PDOP values are generally developed from satellite ephemerides prior to conducting a survey. When developed prior to a survey, PDOP can be used to determine the adequacy of a particular survey schedule.

(2) The key to understanding PDOP is to remember that it represents position recovery at an instant in time and is not representative of a whole session of time. When using pseudorange techniques, PDOP values in the range of 4-5 are considered very good, while PDOP values greater than 10 are considered very poor. For static surveys it is generally desirable to obtain GPS observations during a time of rapidly changing GDOP and/or PDOP.

(3) When the values of PDOP or GDOP are viewed over time, peak or high values (>10) can be associated with satellites in a constellation of poor geometry. The higher the PDOP or GDOP, the poorer the solution for that instant in time. This is critical in determining the acceptability of real-time navigation and photogrammetric solutions. Poor geometry can be the result of satellites being in the same plane, orbiting near each other, or at similar elevations.

d. Horizontal dilution of precision (HDOP). HDOP is a measurement of the accuracy in 2-D horizontal position, mathematically defined as:

$$\text{HDOP} = [\sigma_E^{\,2} + \sigma_N^{\,2}]^{0.5} \cdot [1 / \sigma_R] \qquad \text{(Eq 4-11)}$$

This HDOP statistic is most important in evaluating GPS surveys intended for densifying horizontal control in a project. The HDOP is basically the RMS error determined from the final variance-covariance matrix divided by the standard error of the range measurements. HDOP roughly indicates the effects of satellite range geometry on a resultant position.

f. Acceptable DOP values. In general, GDOP and PDOP values should be less than 6 for a reliable solution. Optimally, they should be less than 5. GPS performance for HDOP is normally in the 2 to 3 range. VDOP is typically around 3 to 4. Increases above these levels may indicate less accurate positioning. In most cases, VDOP values will closely resemble PDOP values. It is also desirable to have a GDOP/PDOP that changes during the time of GPS survey session. The lower the GDOP/PDOP, the better the instantaneous point position solution is.

e. Vertical dilution of precision (VDOP). VDOP is a measurement of the accuracy in standard deviation in vertical height, mathematically defined as:

$$\text{VDOP} = [\sigma_U] \cdot [1 / \sigma_R] \qquad \text{(Eq 4-12)}$$

f. Acceptable DOP values. In general, GDOP and PDOP values should be less than 6 for a reliable solution. Optimally, they should be less than 5. GPS performance for HDOP is normally in the 2 to 3 range. VDOP is typically around 3 to 4. Increases above these levels may indicate less accurate positioning. In most cases, VDOP values will closely resemble PDOP values. It is also desirable to have a GDOP/PDOP that changes during the time of GPS survey session. The lower the GDOP/PDOP, the better the instantaneous point position solution is.

The geometric arrangement of satellites, as they are presented to the receiver, affects the accuracy of position and time calculations. Receivers will ideally be designed to use signals from available satellites in a manner that minimizes this so called "dilution of precision."

To illustrate DOP, consider the example shown in Figure 39, where the satellites being tracked are clustered in a small region of the sky.

In the example shown in Figure 39, intentionally a bit extreme to illustrate the effect of DOP, it is difficult to determine where the ranges intersect. Position is "spread" over the area of range intersections, an area which is enlarged by range inaccuracies (which can be viewed as a "thickening" of the range lines).

As shown in Figure 40 below, the addition of a range measurement to a satellite that is angularly separated from the cluster allows us to determine a fix more precisely.

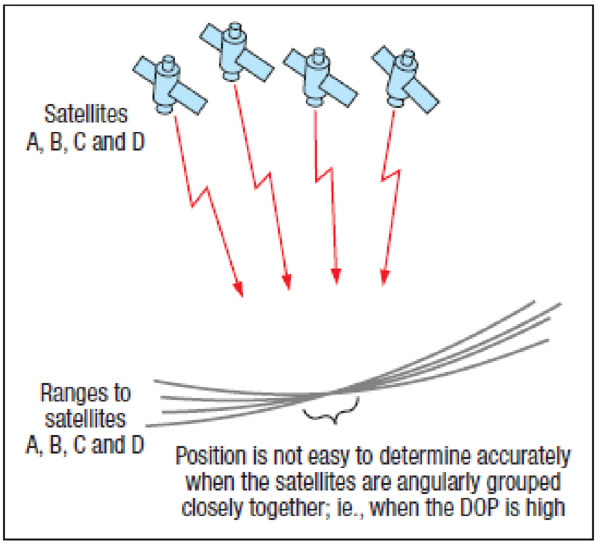


Figure 39 Dilution of Precision (poor satellite geometry)

Although it is calculated using complex statistical methods, we can say the following about DOP:

- DOP is a numerical representation of satellite geometry, and it is dependent on the locations of satellites that are visible to the receiver.
- The smaller the value of DOP, the more precise the result of the time or position calculation. The relationship is shown in the following formula:

  Inaccuracy of Position Measurement = DOP x Inaccuracy of Range Measurement

  So, if DOP is very high, the inaccuracy of the position measurement will be much larger than the inaccuracy of the range measurement.
- DOP can be used as the basis for selecting the satellites on which the position solution will be based; specifically, selecting satellites to minimize DOP for a particular application.
- A DOP above 6 results in generally unacceptable accuracies for DGPS and RTK operations
- DOP varies with time of day and geographic location but, for a fixed position, the geometric presentation of the satellites repeats every day, for GPS.
- DOP can be calculated without determining the range. All that is needed is the satellite positions and the approximate receiver location.

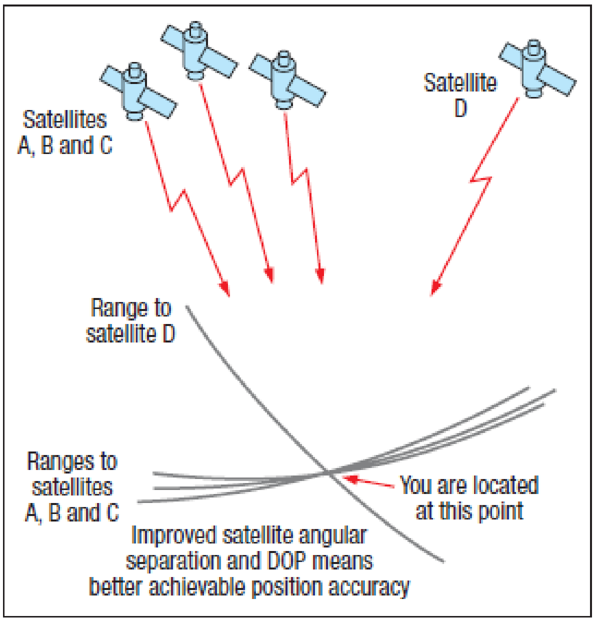


Figure 40 Dilution of Precision (improved geometry)

DOP can be expressed as a number of separate elements that define the dilution of precision for a particular type of measurement, for example, HDOP (Horizontal Dilution of Precision), VDOP (Vertical Dilution of Precision), and PDOP (Position Dilution of Precision). These factors are mathematically related. In some cases, for example when satellites are low in the sky, HDOP is low and it will therefore be possible to get a good to excellent determination of horizontal position (latitude and longitude), but VDOP may only be adequate for a moderate altitude determination. Similarly, when satellites are clustered high in the sky, VDOP is better than HDOP.

When we extend our DOP illustration to three satellites, one way to view dilution of precision is to consider the "tetrahedron" formed by having the satellites at three corners and the receiver at the fourth, as illustrated in Figure 41.

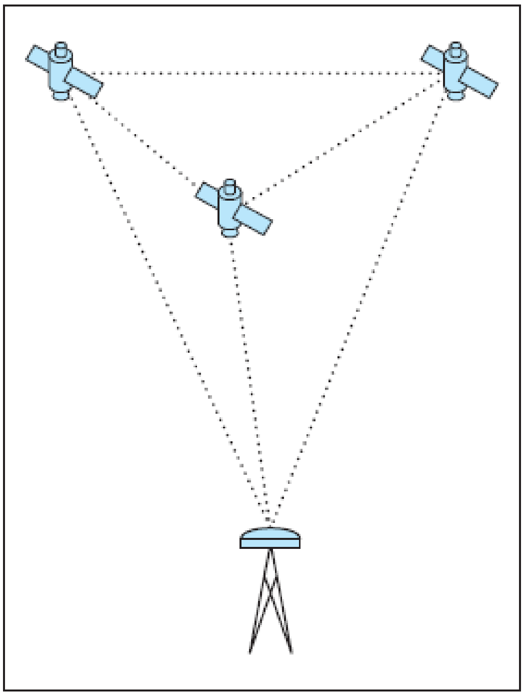


Figure 41 Minimizing Dilution of Precision

Minimizing DOP is not unlike maximizing the
volume of this tetrahedron. When the satellites are tightly clustered and the angle between the satellites is small, the tetrahedron is long and narrow. The volume of the tetrahedron is small and DOP is correspondingly high (undesirable). When the satellites are all located near the horizon, the tetrahedron is flat. Again, the volume of the tetrahedron is small and DOP is high. When the satellites are not tightly clustered in the sky or low in elevation, the volume of the tetrahedron approaches a maximum and DOP is at its lowest (desirable).

In Canada and in other countries at high latitude, GNSS satellites are lower in the sky, and achieving optimal DOP for some applications, particularly where good VDOP is required, is sometimes a challenge.

When there were fewer GNSS satellites, achieving good DOP was sometimes difficult. These difficulties are being reduced with more GNSS constellations and satellites coming on line every year. Applications where the available satellites are low on the horizon or angularly clustered, such as those in urban environments or in deep open-pit mining, may still expose users to the pitfalls of DOP. If you know your application will have obstructed conditions, you may want to use a mission planning tool to determine the ideal time – the time with the ideal DOP – for your survey, as an example.

Further reading in (Tsui, 2005: 25)

CHAPTER 8 OTHER SATELLITE NAVIGATION SYSTEMS

Contents of this Chapter

8.1 Introduction

Currently, the following GNSS systems are operational:

- GPS (United States)
- GLONASS (Russia)

The following GNSS systems are planned and are in varying stages of development:

- Galileo (European Union)
- Compass (China)

The following regional navigation satellite systems are planned and are in varying stages of development:

- IRNSS (India)
- QZSS (Japan)

8.2 Global Positioning System (GPS)

GPS was the first GNSS system. GPS (or NAVSTAR, as it is officially called) satellites were first launched in the late 1970's and early 1980's for the US Department of Defence. Since that time, several generations (referred to as "Block") of GPS satellites have been launched. Initially, GPS was available only for military use but in 1983, a decision was made to extend GPS to civilian use. A GPS satellite is depicted in Figure 24.

Space Segment

The GPS space segment is summarized in Table 2. The orbit period of each satellite is approximately 12 hours, so this provides a GPS receiver with at least six satellites in view from any point on Earth, under open-sky conditions.

Satellites	21 plus 3 spares
Orbital planes	6
Orbit inclination	55 degrees
Orbit radius	26,560 km

Table 2 GPS Satellite Constellation

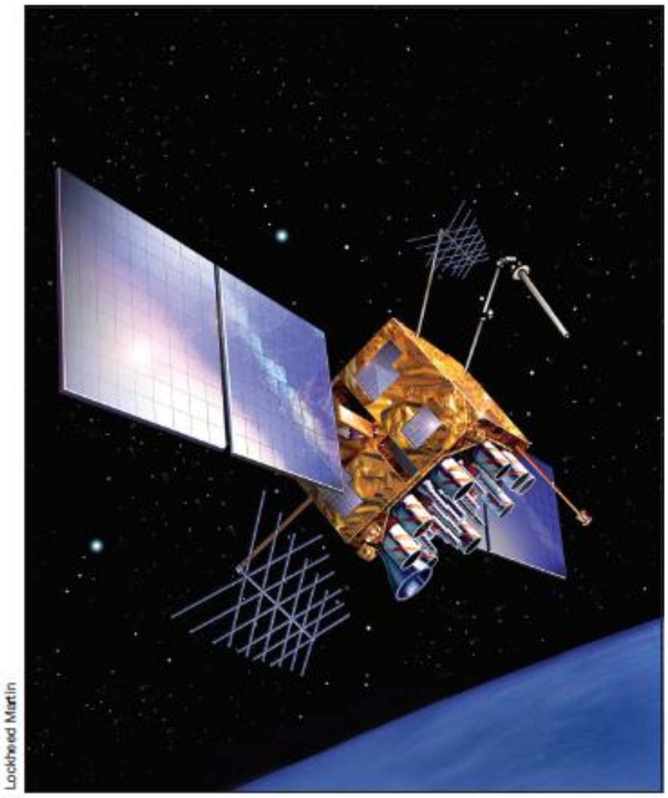


Figure 24 GPS IIRM Satellite (artist's rendition)

A GPS satellite orbit is illustrated in Figure 25.

GPS satellites continually broadcast their identification, ranging signals, satellite status and corrected ephemerides (orbit parameters). The satellites are identified either by their Space Vehicle Number (SVN) or their PseudoRandom code Number (PRN).

Signals

Table 3 provides further information on GPS signals. GPS signals are based on CDMA (Code Division Multiple Access) technology, which we discussed in Chapter 2.

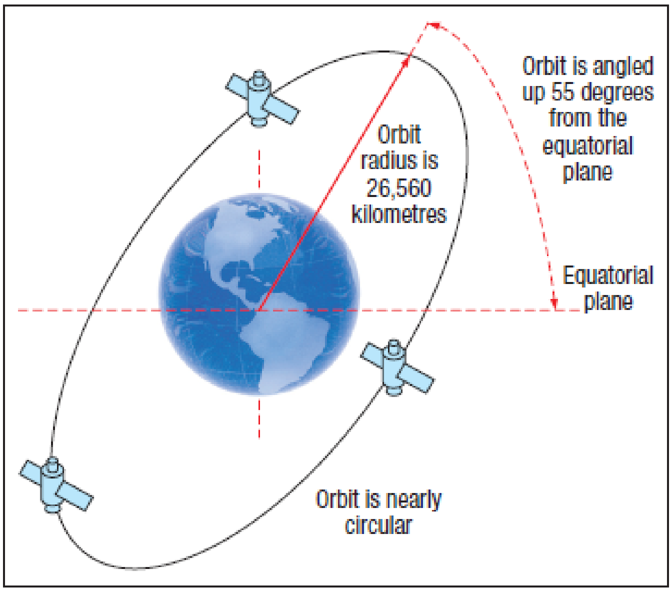


Figure 25 GPS Satellite Orbit

Designation	Frequency	Description
L1	1575.42 MHz	L1 is modulated by the C/A code (Coarse/Acquisition) and the P-code (Precision) which is encrypted for military and other authorized users.
L2	1227.60 MHz	L2 is modulated by the P-code and, beginning with the Block IIR-M satellites, the L2C (civilian) code. L2C, which is considered "under development", is discussed below, under "GPS Modernization".
L5	1176.45 MHz	At the time of writing, L5 is available for demonstration on one GPS satellite. The L5 signal is discussed below, under "GPS Modernization".

Table 3 GPS Signal Characteristics

Control Segment

The GPS control segment consists of a master control station (and a backup master control station) and monitor stations throughout the world, as shown in Figure 26.

Figure 26 GPS Control Segment

Four monitor stations were implemented early in the NAVSTAR program, then six more NGA (National Geospatial Intelligence Agency, also part of the United States Department of Defense) stations were added in 2005.

The monitor stations track the satellites via their broadcast signals, which contain satellite ephemeris data, ranging signals, clock data and almanac data. These signals are passed to the master control station where the ephemerides are recalculated. The resulting ephemeride and timing corrections are transmitted back up to the satellites through data up-loading stations.

8.3 GLONASS Satellite System

GLONASS was developed by the Soviet Union as an experimental military communications system during the 1970s. When the Cold War ended, the Soviet Union recognized that GLONASS had commercial applications, through the system's ability to transmit weather broadcasts, communications, navigation and reconnaissance data.

The first GLONASS satellite was launched in 1982 and the system was declared fully operational in 1993. After a period where GLONASS performance declined, Russia committed to bringing the system up to the required minimum of 18 active satellites. The Russian government set 2011 as the date for full deployment of the 24-satellite constellation and has ensured that the necessary financial support will be there to meet this date.

GLONASS satellites have evolved since the first ones were launched. The latest generation, GLONASS-M, is shown being readied for launch in Figure 27.

JSC "Information Satellite Systems – Reshetnev Company"

Figure 27 GLONASS-M Satellite in Final Manufacturing

The GLONASS constellation provides visibility to a variable number of satellites, depending on your location. A minimum of four satellites in view allows a GLONASS receiver to compute its position in three dimensions and to synchronize with system time.

GLONASS Space Segment

The GLONASS space segment is summarized in Table 4.

Satellites	21 plus 3 spares
Orbital planes	3
Orbital inclination	64.8 degrees
Orbit radius	25,510 km

Table 4 GLONASS Satellite Constellation

When complete, the GLONASS space segment will consist of 24 satellites in three orbital planes, with eight satellites per plane.

The GLONASS constellation geometry repeats about once every eight days. The orbit period of each satellite is approximately 8/17 of a sidereal[1] day so that, after eight sidereal days, the GLONASS satellites have completed exactly 17 orbital revolutions.

Each orbital plane contains eight equally spaced satellites. One of the satellites will be at the same spot in the sky at the same sidereal time each day.
The satellites are placed into nominally circular orbits with target inclinations of 64.8 degrees and an orbital radius of 25,510 km, about 1,050 km lower than GPS satellites.

The GLONASS satellite signal identifies the satellite and includes:

- Positioning, velocity and acceleration information for computing satellite locations.
- Satellite health information.
- Offset of GLONASS time from UTC (SU) [formerly Soviet Union and now Russia].
- Almanac of all other GLONASS satellites.

"The Earth was absolutely round . . . I never knew what the word 'round' meant until I saw Earth from space." Alexei Leonov, Soviet astronaut, talking about his historic 1985 spacewalk.

Figure 28 View of Earth (as seen by Apollo 17 crew)

GLONASS Control Segment

The GLONASS control segment consists of the system control center and a network of command tracking stations across Russia. The GLONASS control segment, similar to that of GPS, monitors the status of satellites, determines the ephemeride corrections, and satellite clock offsets with respect to GLONASS time and UTC (Coordinated Universal Time). Twice a day, it uploads corrections to the satellites.

GLONASS Signals

Table 5 summarizes the GLONASS signals.

Designation	Frequency	Description
L1	1598.0625 - 1609.3125 MHz	L1 is modulated by the HP (high precision) and the SP (standard precision) signals.
L2	1242.9375 - 1251.6875 MHz	L2 is modulated by the HP and SP signals. The SP code is identical to that transmitted on L1.

Table 5 GLONASS Signal Characteristics

GLONASS satellites each transmit on slightly different L1 and L2 frequencies, with the P-code (HP code) on both L1 and L2, and the C/A code (SP code), on L1 (all satellites) and L2 (most satellites). GLONASS satellites transmit the same code at different frequencies, a technique known as FDMA, for frequency division multiple access. Note that this is a different technique from that used by GPS.

GLONASS signals have the same polarization (orientation of the electromagnetic waves) as GPS signals, and have comparable signal strength.

The GLONASS system is based on 24 satellites using 12 frequencies. It achieves this by having antipodal satellites transmitting on the same frequency. Antipodal satellites are in the same orbital plane but are separated by 180 degrees. The paired satellites can transmit on the same frequency because they will never appear at the same time in view of a receiver on the Earth's surface, as shown in Figure 29.

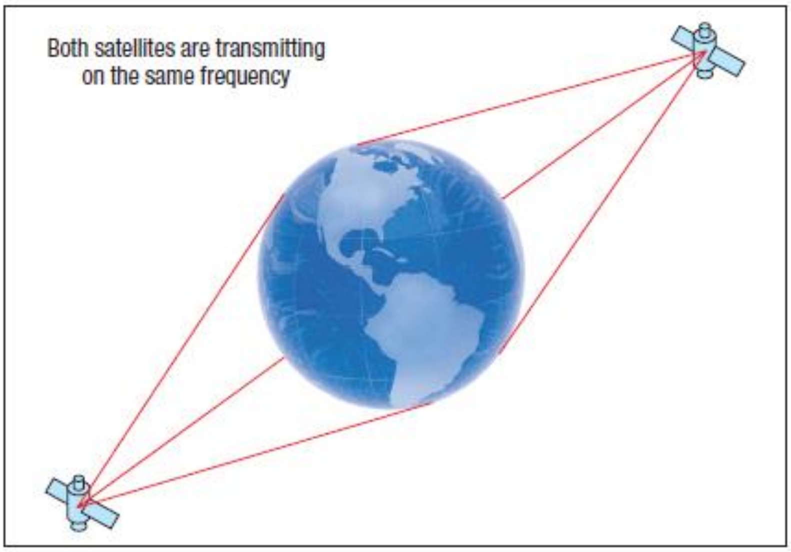


Figure 29 GLONASS Antipodal Satellites

8.4 Chinese Regional Satellite Navigation System (COMPASS)

China has started the implementation of a GNSS system, known as Compass or Beidou-2. The initial system will provide regional coverage. The target is that this be followed after 2015 with the implementation of a constellation of GEO (geostationary orbit) and MEO satellites that will provide global coverage, as shown in Table 8.

Satellites	35, a combination of 5 GEO and 30 MEO
Orbital planes	6
Orbital inclination	55 degrees
Orbit radius	27,528 km

Table 8 Planned Compass Satellite Constellation

Two levels of service will be provided:

- Public service, for civilian use, and free to users in China. The public service will provide location accuracy of 10 m, velocity accuracy within 0.2 m per second and timing accuracy of 50 nanoseconds.
- Licensed military service, more accurate than the public service, and also providing system status information and military communications capability.

8.5 European Global Satellite Navigation System (Galileo)

By 2006, the world's first dedicated civilian GNSS constellation was under development. The European Union's Galileo global navigation satellite system for land, sea and air applications was to have a constellation of 30 satellites divided between three circular orbits inclined at 56 degrees to the equator. Orbiting the Earth at an altitude of around 23,222 km, the satellites would have a bigger footprint than their GPS counterparts, covering the entire surface of the planet. Nine satellites were to be spread evenly around each orbital plane, with each taking about 14 hours to orbit the Earth. Each plane had an extra, dormant satellite able to cover any ailing satellite in its plane (CASA, 2006).

Like GPS, the Galileo system was designed to be supported by a worldwide network of ground stations. In civil aviation, Galileo was designed to lend itself to all phases of flight, for en route navigation, and to airport approach, landing and ground guidance. The system was to broadcast integrity information for some critical applications to assist in assuring the quality of positioning accuracy. The US and EU agreed on interoperability of Galileo and GPS. The different orbital configurations of the systems, together comprising 60 satellites, were to complement each other, boosting integrity, availability and continuity of service. At the time of writing, manufacturers were developing units capable of processing both Galileo and GPS signals to give an integrated navigation solution (CASA, 2006).

Galileo, Europe's planned global navigation satellite system, will provide a highly accurate and guaranteed global positioning service under civilian control. The United States and European Union have been cooperating since 2004 to ensure that GPS and Galileo are compatible and interoperable at the user level. By offering dual frequencies as standard, Galileo will deliver real-time positioning accuracy down to the meter range, previously not achievable by a publicly available system.

Galileo will guarantee availability of service under all but the most extreme circumstances and it will inform users within seconds of a failure of any satellite. This will make it suitable for applications where safety is crucial, such as in air and ground transportation.

The first experimental Galileo satellite, part of the Galileo System Test Bed (GSTB) was launched in December 2005. The purpose of this experimental satellite is to characterize critical Galileo technologies, which are already in development under European Space Agency (ESA) contracts. Four operational satellites are planned to be launched in the 2010-2012 time frame to validate the basic Galileo space and ground segment. Once this In-Orbit Validation (IOV) phase has been completed, the remaining satellites will be launched, with plans to reach Full Operational Capability (FOC) likely sometimes after 2013.

System Design

Satellites	27 operational and three active spares
Orbital planes	3
Orbital inclination	56 degrees
Orbit radius	23,616 km

Table 6 Galileo Satellite Constellation

The Galileo space segment is summarized in Table 6. Once the constellation is operational, Galileo navigation signals will provide coverage at all latitudes. The large number of satellites, together with the optimization of the constellation and the availability of the three active spare satellites, will ensure that the loss of one satellite has no discernable effect on the user segment. Two Galileo Control Centres (GCC), to be located in Europe, will control the satellites. Data recovered by a global network of twenty Galileo Sensor Stations (GSS) will be sent to the GCC through a redundant communications network. The GCCs will use the data from the Sensor Stations to compute integrity information and to synchronize satellite time with ground station clocks. Control Centres will communicate with the satellites through uplink stations, which will be installed around the world.

Galileo will provide a global Search and Rescue (SAR) function, based on the operational search and rescue satellite-aided Cospas-Sarsat[2] system. To do this, each Galileo satellite will be equipped with a transponder that will transfer distress signals to the Rescue Coordination Centre (RCC), which will then initiate the rescue operation. At the same time, the system will provide a signal to the user, informing them that their situation has been detected and that help is underway. This latter feature is new and is considered a major upgrade over existing systems, which do not provide feedback to the user.

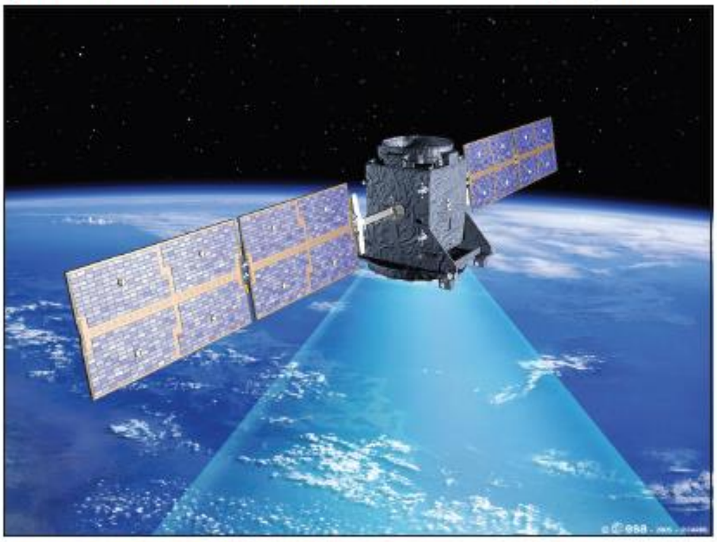


Figure 31 Galileo Satellite in Orbit

Five Galileo services are proposed, as summarized in Table 7.

Service	Description
Free Open Service (OS)	Provides positioning, navigation and precise timing service. It will be available for use by any person with a Galileo receiver. No authorisation will be required to access this service. Galileo is expected to be similar to GPS in this respect.
Highly reliable Commercial Service (CS)	Service providers can provide added-value services, for which they can charge the end customer. The CS signal will contain data relating to these additional commercial services.
Safety-of-Life Service (SOL)	Improves on OS by providing timely warnings to users when it fails to meet certain margins of accuracy. A service guarantee will likely be provided for this service.
Government encrypted Public Regulated Service (PRS)	Highly encrypted restricted-access service offered to government agencies that require a high availability navigation signal.
Search and Rescue Service (SAR)	Public service designed to support search and rescue operations, which will make it possible to locate people and vehicles in distress.

Table 7 Galileo Services

8.6 Indian Regional Navigation Satellite System (IRNSS)

India plans to launch its own regional navigation satellite system to provide regional coverage in the 2011-2012 timeframe. The IRNSS system will consist of seven satellites, three of them in geosynchronous orbits, the other four in geostationary orbits. The system will provide a position accuracy of better than 20 meters throughout India.

8.7 Quasi-Zenith Satellite System (QZSS)

Japan plans to launch demonstration satellites for the QZSS, a three-satellite system that will provide regional communication services and positioning information for the mobile environment. QZSS will provide limited accuracy in standalone mode, so it is viewed as a GNSS augmentation service. QZSS satellites will be placed in a periodic Highly Elliptical Orbit (HEO). These orbits will allow the satellites to 'dwell' for more than 12 hours a day at an elevation above 70° (meaning they appear almost overheat most of the time).

REFERENCES

Unsupported source type (ElectronicSource) for source Sic14.

CASA (2006) *Global Navigation Satellite Systems Overview*, Civil Aviation Safety Authority, Australia.

El-Rabbany, A. (2006) *Introduction to GPS - The Global Positioning System*, 2nd edition, Boston: Artech House.

Erickson, C. (1995) *GPS Positioning Guide*, Ontario: Natural Resources Canada, Available: http://www.geod.nrcan.gc.ca/.

Ghilani , D. and Wolf , P.R. (2012) *Elementary Surveying : An Introduction to Geomatics*, 13th edition, New Jersey: Prentice Hall.

Gleason, S. and Gebre-Egziabher, (2009) 'Global Navigation Satellite Systems: Present and Future', in Gleason, and Gebre-Egziabher, (ed.) *GNSS Applications and Methods*, Boston: Artech House.

GNSS Solutions Reference Manual (2008), Magellan Navigation Inc., Available: www.pro.magellanGPS.com.

Groves, P.D. (2008) *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*, Boston: Artech House.

ICAO (2005) *Global Navigation Satellite System (GNSS) Manual*, 1st edition, Montréal.

Jeffrey, C. (2010) *An Introduction to GNSS - GPS, GLONASS, Galileo and other Global Navitation Satellite Systems*, Calgary: NovAtel Inc., Available: www.novatel.com.

Jin, (ed.) (2012), in *Global Navigation Satellite Systems – Signal, Theory and Applications*, Rijeka: InTech, Available: www.intechopen.com.

Kaplan, D. and Hegarty, J. (ed.) (2006) *Understanding GPS - Principles and Applications*, 2nd edition, Boston: Artech House.

Marji, (2008) *Precise Relative Navigation for Satellite Formation Flying Using GPS*, Calgary: Department of Geomatics Engineering, The University of Calgary.

Rydlund Jr., P.H. and Densmore, B.K. (2012) 'Collection and Delineation of Spatial Data', in *Methods of practice and guidelines for using survey-grade global navigation satellite systems (GNSS) to establish vertical datum in the United States Geological Survey: U.S. Geological Survey Techniques and Methods*.

Seeber, G. (2003) *Satellite Geodesy: Foundations, Methods, and Applications*, 2nd edition, Berlin, New York: Walter de Gruyter.

Strang, G. and Borre, K. (1997) *Linear Algebra, Geodesy, and GPS*, Wellesley: Wellesley-Cambridge Press, Available: http://www-math.mit.edu/-gs.

Tsui, B.-y. (2005) *Fundamentals of Global Position System Receivers - A Software Approach*, 2nd edition, New Jersey: John Wiley & Sons, Inc.

United Nations (2012) *Global Navigation Satellite Systems, Education Curriculum*, New York.

US Army Corps of Engineers (2003) *NAVSTAR Global Positioning System Surveying*, Washington DC: Department of the Army, US Army Corps of Engineers.

Wells, , Beck, , Delikaraoglou, , Kleusberg, , Krakiwsky, , , Lachapelle, , Langley, , , Nakiboglu, , Schwarz, K.-P., Tranquilla, , and Vanicek, (1999) *Guide to GPS Positioning*, Fredericton: Department of Geodesy and Geomatics Engineering, University of New Brunswick.

Zogg, J.-M. (2009) *GPS: Essentials of Satellite Navigation*, u-blox, Available: www.u-blox.com.