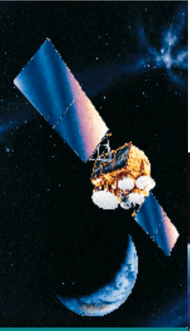
Grains Research & Development Corporation



2

Global Navigation Satellite Systems

Satellite based navigation systems are truly the enabling technology of Precision Agriculture. They provide a relatively simple and robust technique for identifying any location on the earth's surface, or, in the case of aircraft, relative to the surface. This permits any agricultural and environmental operations to be geo-referenced and spatially analysed. A wide range of satellite-based navigation and geo-location tools are available to suit different agronomic situations from point crop/soil sampling to autonomous vehicle guidance.

Introduction

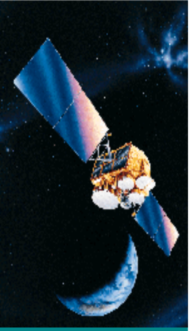
Satellite navigation systems utilise a constellation of satellites orbiting the earth to geo-locate a receivers position on or near the earths surface. Two systems are currently in operation, the NAVSTAR Global Positioning System (GPS), owned by the government of the United States of America, and the Global Navigation Satellite System (GLONASS), which is controlled by a consortium headed by the Russian Government. Two more systems are being planned. The European Space Agency intends to have their network, Gallileo, fully operational in 2008. A Japanese consortium is also planning to launch a satellite navigation system designed for satellite navigation and communication for automobiles. All four existing and proposed systems are basically similar however far more receivers have been developed by commercial enterprises to utilise the information from the GPS satellites so its operation will form the basis of the following review.

How Satellite Based Navigation Systems Work

The GPS, GLONASS and proposed Gallileo and Japanese systems are all designed with three core segments, Space, Control and User.

Space Segment

This consists of the satellite constellation that is orbiting the earth and the Delta rockets used to launch the satellites. In the GPS constellation there are 24 satellites that orbit the earth every 12 hours at an altitude of 20,200km. The satellites are organised into 6 equally spaced orbital planes (60 degrees apart), with 4 satellites per plane. Each satellite is inclined at 55 degrees to the equatorial plane to ensure



coverage of the polar regions. A visible explanation of the satellite constellation is provided in Figure 1. This combination is designed to provide a user anywhere on the earths surface with 5-8 visible satellites. The satellites are powered by solar cells, programmed to follow the sun, and have 4 on-board atomic clocks that are accurate to a nanosecond (a billionth of a second). The satellites also have a variety of antennas to generate, send and receive signals. On-board the satellite signals are generated by a radio transmitter and sent to land-based receivers by L-band antennas

Control Segment

The control segment of the GPS Navigation Systems consists of a Master Control Station which is supported by Monitor Stations and Ground Antennas. The Monitor Stations check the exact altitude, position, speed and overall health of the GPS satellite. A Monitor Station can track up to 11 satellites simultaneously and each satellite is checked twice a day by each Monitor Station. The information collected by the Monitoring Stations is relayed to the Master Control Station to assess the behaviour of each satellites orbit and clock. If any errors are noted then the Master Control Station directs the relevant Ground Antenna to relay the required corrective information to the relevant satellite. The global locations of the Control Segments are shown in Figure 2.

User Segment

The User Segment refers to the civilian and military personnel who use the signals generated by the GPS satellites. There are a wide range of receivers available for civilian use, ranging in price from a few hundred dollars to over fifty thousand dollars per receiver. The price is usually strongly related to the precision and accuracy of the receivers.

How Geo-Location Is Determined

Geo-location using satellite navigation systems is based on the ability to measure the time taken for a signal to travel from a satellite to the receiver. Radio signals travel at the speed of light, which is constant, so if the time of travel is known then the distance between the satellite and the receiver can be determined. Since the position of the satellites is always known, thanks to the work by the Control Segment of the system, an unknown point (the users receiver) can be calculated if the receiver is obtaining signals from at least four satellites.

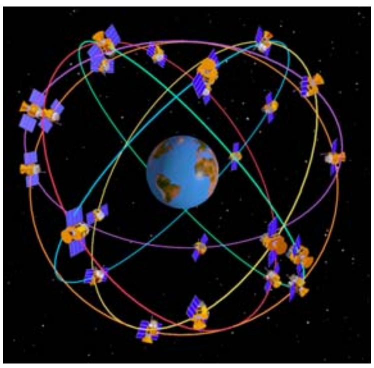
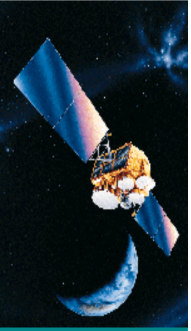


Figure 1: Schematic view of the orbit paths of the GPS satellites. There are 6 orbits with 4 satellites per orbit. (Courtesy of www.aero.org).



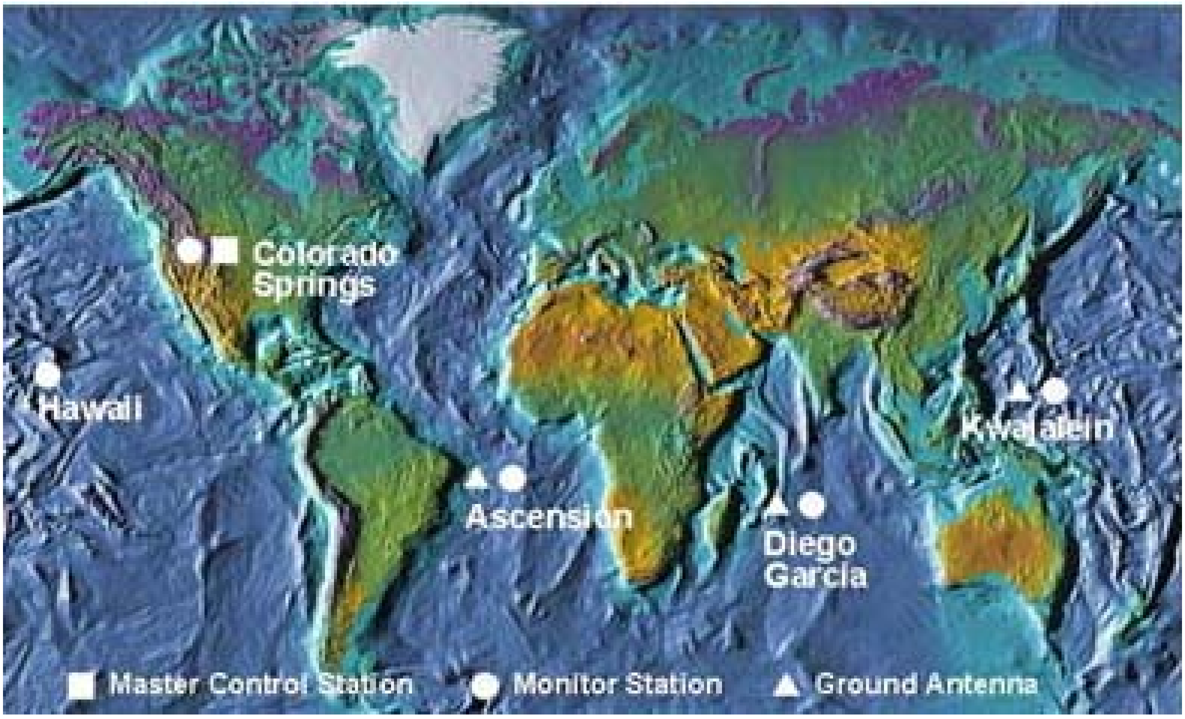


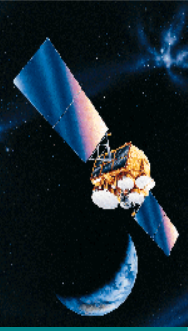
Figure 2: Locations of the Control Segment of the GPS Satellite Navigation System (Courtesy of gps.faa.gov)

Measuring Distance

Each GPS navigation satellite continuously broadcasts its position along with timing data on two frequencies (L1 and L2). The L1 band carries two codes, Coarse Acquisition (C/A) code and Precision (P) code. The L2-band only carries the P code. The C/A signal is also termed "code phase" or Standard Positioning Service (SPS) and is the main signal used in civilian activity. The P signal is also referred to as "Precise Positioning Service" and was designed for US government and military use only. It requires special cryptographic equipment to decode. The C/A and P signals take ~6 milliseconds (6/100ths of a second) to travel from satellite to receiver. The signals require a direct line of sight so receivers will not work indoors or under vegetation canopies/trees. This is a major problem in the use of GPS in horticultural tree crops.

Both the C/A and P signals have a time reference digital code referred to as a pseudo random code. Receivers contain an almanac of the pseudo random codes generated by the satellites and the time they are generated. When a receiver intercepts the digital code from a satellite it can compare the digital signal to its almanac to determine when the signal was generated. The time of travel is simply the difference between the time the signal was intercepted and the time it was generated (Figure 3). The difference between the C/A and P code is in the resolution of the code and thus the accuracy of timing and distance determination.

As well as transmitting in "code phase", satellites also transmit general satellite information in "carrier phase". Carrier phase signals are broadcast on both the L1 and L2-band and at a much higher frequency than the code phase. The higher



frequency permits a more accurate measurement of the range between the satellite and receiver. However the carrier phase is not time referenced like the code phase. This makes the interpretation of the signal susceptible to "cycle slip". To minimise this effect, carrier phase receivers use the C/A code to provide a rough estimation and the carrier phase signal to improve the estimation. The frequency of the different code and carrier signals is illustrated in Figure 4 and a further explanation the difference between code and carrier phase is given in Appendix 1)

Calculating Position

If the distance (d_1) from the receiver to a satellite is known then the receiver must be somewhere on a sphere with a radius of d_1 that is centred on the satellite. If the distance (d_2) to a second satellite is determined then the receiver must also lie somewhere on a sphere of radius d_2 centred on the second satellite. Given this knowledge, the receiver must lie on the ellipse that forms the intersection of the spheres. If a third satellite is located then the receiver position is narrowed down to two points where the spheres of the three satellites intersect. Usually one of these positions can be discarded as it is not near the earth's surface. Thus by locating three satellites, the three unknowns in the receiver's location (latitude, longitude and altitude or X, Y, Z) can be determined.

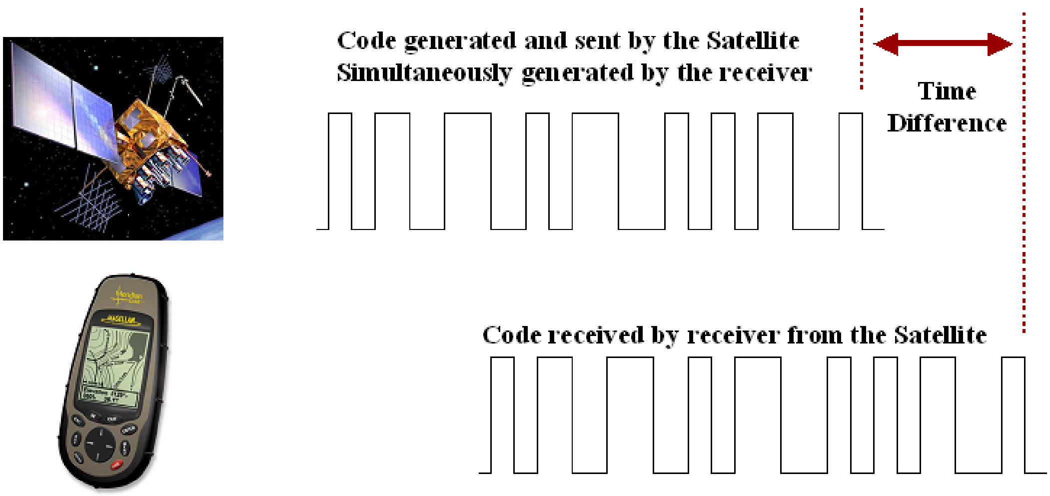


Figure 3 (above): Diagrammatic representation of the C/A code and how it is used to determine time and distance between the satellite and the receiver. (Adapted from of Paul Bolstad, http://bolstad.gis.umn.edu/chapt5figs.)

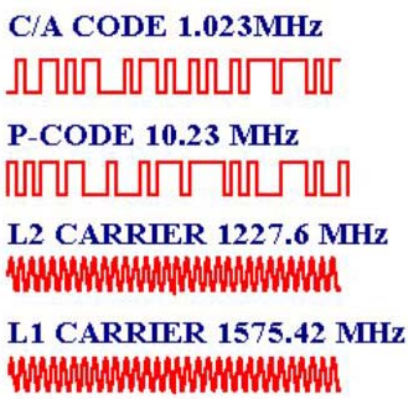


Figure 4 (left): A comparison of the waveforms and frequency of the different signals emitted by the L1 and L2-band antennas from a GPS satellite. (adapted from www.go.ednet.ns.ca)

However, the determination of the distance between the receiver and satellite relies on very accurate timing. Satellites have very accurate timing due to the use of atomic clocks on-board and constant monitoring by the Control Segment. Unfortunately atomic clocks are too heavy (~20kg) and expensive (US$50,000) to mount into GPS receivers. Therefore GPS receivers need to use inferior clocks. This creates a problem as errors in the receiver clock will degrade the estimation of distance by ~300,000m per millisecond. This problem can be overcome by assuming that the receiver clock error is a fourth unknown in the system. By connecting to a fourth satellite the receiver is able to solve the four simultaneous equations to resolve the four variables (X, Y, Z and clock error). In this case there is a trade off between the number of satellites required to calculate the receiver position and the cost of the receiver. Aschematic illustration of positioning is shown in FIgure 5.

Geo-Location Error Sources

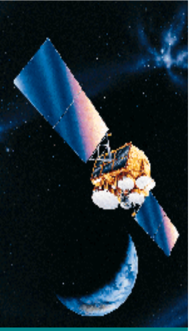
Any error source will affect the ability of a GPS receiver to accurately determine the range to satellites which creates uncertainty in geo-location (Figure 6). Apart from the quality of the signal (C/A vs. P vs Carrier) the error in geo-location calculation by a receiver may be affected by a one or more of the following error sources.

Satellite Errors - These may be due to either errors in the timing of the on-board atomic clocks or an error in the transmitted location of the satellite (ephemeris error). Regular monitoring by the Control segment is aimed at minimising these errors. A special error associated with the Satellite clock time is Selective Availability (SA). This was a man-made error introduced into the satellite time to limit the accuracy of a GPS receiver to *±100m* so that it could not confidently be used by other military organisations outside the USA.. In 2000, the President of the USA (and Commander-in-Chief of the US Armed Forces) turned off the SA error to promote the use of the GPS system over other Satellite-based Navigation Systems. By 2000 the development of differential correction techniques to compensate for the SA error had effectively removed any benefits derived by the military having the SA error turned on.

Receiver Errors - The ability of the GPS receiver and associated software to cope with thermal and electronic noise will affect how accurately the receiver can geo-locate itself.

Atmospheric Errors - To reach a GPS receiver, the satellite signal needs to pass through the Earths atmosphere and, in particular, the Ionosphere and Troposphere which affect the signals.

The Ionosphere contains charged particles that have the effect of slowing the code phase signal and speeding up the carrier phase signal. The speed of a signal through the Ionosphere is related to its frequency thus, by using a receiver's dual frequency capabilities, the Ionospheric errors can be corrected. This is the primary reason why satellites broadcast both L1 and L2-bands. Traditionally only the military have been able to access the L2-band and utilise it for Ionospheric error correction. However some GPS receivers have been developed with sophisticated techniques to take



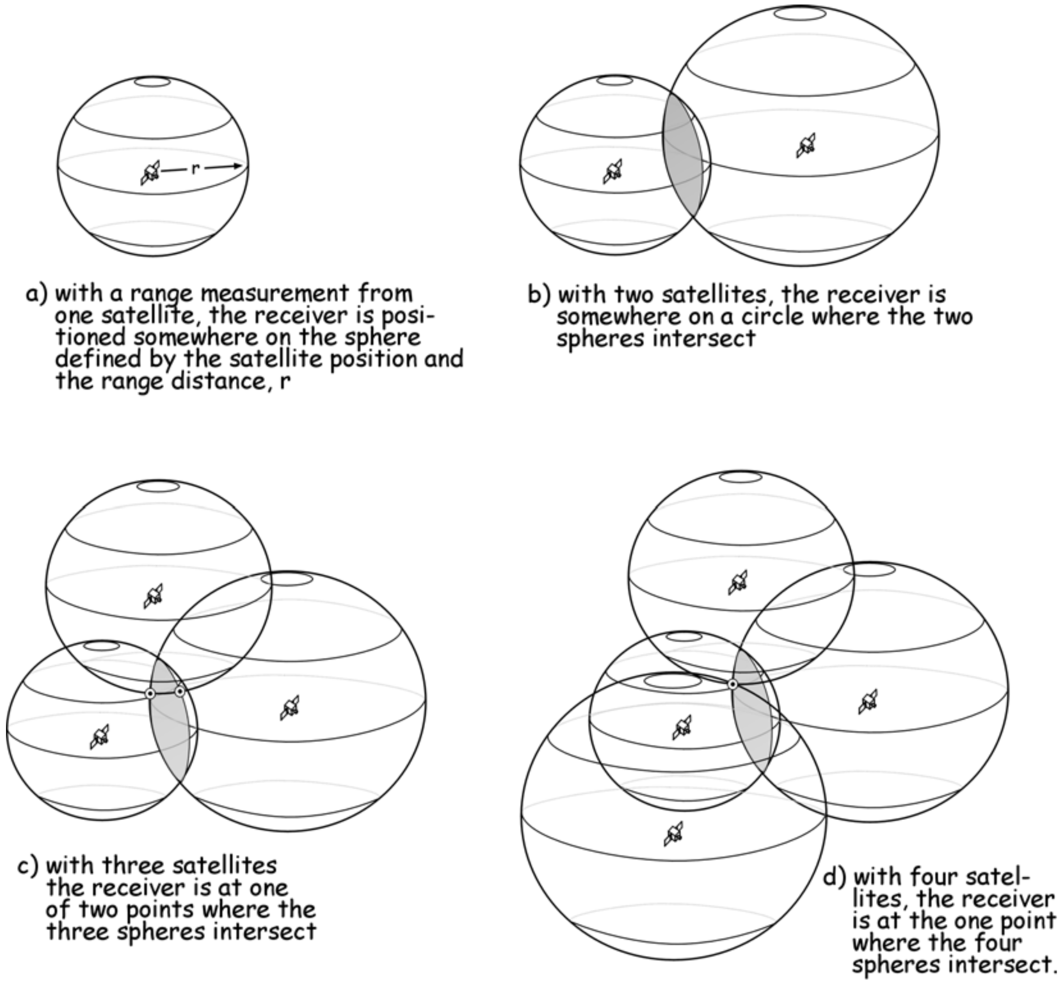


Figure 5: A schematic illustration of how ranging from a receiver to three or more satellites can be used to pinpoint an exact location. (Courtesy of Paul Bolstad, http://bolstad.gis.umn.edu/chapt5figs.)

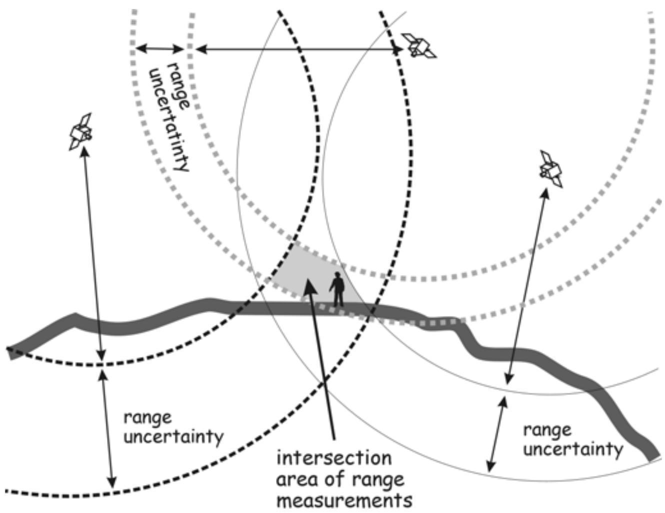


Figure 6: An graphical description of how the inaccuracies in ranging can create some uncertainty in the estimation of actual location. (Courtesy of Paul Bolstad, http://bolstad.gis.umn.edu/chapt5figs.)

advantage of the L2-band without contravening the security objectives of the US Department of Defense. For single frequency receivers mathematical models to reduce the error have been developed however these only remove ~50% of the error.

The troposphere, the lower level of the atmosphere, contains water vapour that slows both the code and carrier phase signals. Dual frequency systems cannot compensate for this error and it must be modeled using measurements of atmospheric moisture, pressure and temperature.

The amount of error introduced into the system by atmospheric effects will be related to the distance that the signal has to travel through the atmosphere. Signals from satellites low on the horizon will travel further through the atmosphere than satellites positioned directly above the receiver. This is illustrated in Figure 7.

Multipath Errors - These are errors caused when the GPS antenna receives signals that have been reflected from a secondary source. This lengthens the travel time and thus creates error in the distance determination.

Satellite Geometry - Apart from errors in determining the distance between the satellites and receiver, the accuracy of geo-location is also a function of the geometry of the satellites used for geo-location. Satellite geometry is measured by the Dilution of Precision (DOP) statistics. The dilution of precision can be determined horizontal (HDOP), vertical (VDOP) or as a timing factor (TDOP). Alternatively the individual DOP statistics can be joined to produce more general estimations of the positional DOP (PDOP = $\sqrt{(\mathrm{HDOP})^2 + (\mathrm{VDOP})^2}$) or geometric DOP (GDOP = $\sqrt{(\mathrm{HDOP})^2 + (\mathrm{VDOP})^2 + (\mathrm{TDOP})^2}$. The DOP statistics are unitless and the lower the value the better the positional accuracy.

In general to have confidence in your GPS receiver GDOP should be < 5 and PDOP < 4.

Geometrically, PDOP is proportional to 1 divided by the volume of the pyramid formed by lines running from the receiver to four observed satellites. Four widely separated satellites reduce the DOP error compared with satellites clustered in one sector of the sky (Figure 8). The optimum geometry is for one satellite to be directly overhead and the other three spread out evenly. As satellites orbit the earth, their geometry relative to a receiver varies and the DOP errors will vary and this is the main cause of daily variation in the accuracy of geo-location. However, since

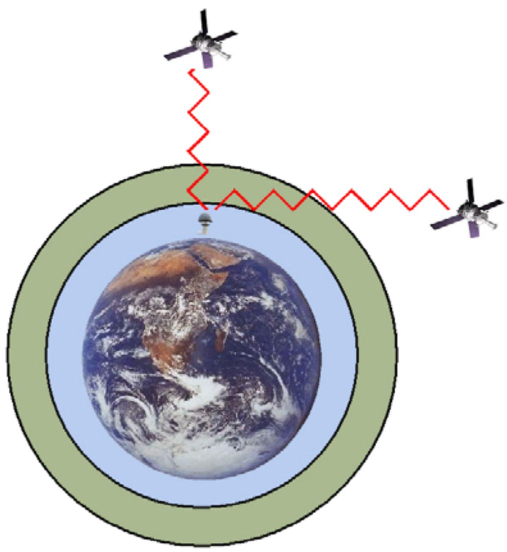


Figure 7: An illustration of the extra atmospheric distance that satellites low on the horizon must travel through.

the path of the satellites is fixed and known, mission planning software is available to determine when these errors can be minimised. Unfortunately, in agricultural situations, it is usually impractical to delay or alter farm management to take advantage of these windows. (Further information on Dilution of Precision errors is provided in Appendix 2).

Types Of Receivers

Code Phase Receivers

Stand Alone GPS receivers

Also known as Standard Position System (SPS) receivers, these receivers operate using only the basic C/A code on the L1-band from the navigation satellites. They are the cheapest GPS receivers available as there is no additional correction signal or complex circuitry to utilise the P code or carrier phase. However SPS receivers have the worst geo-location accuracy (usually ±5m but may be greater) of all the GPS receivers on the market. Most SPS receivers contain filtering algorithms designed to smooth the signal noise when the GPS is moving. This makes SPS receivers more accurate when moving and suitable for wide swathing, low resolution applications.

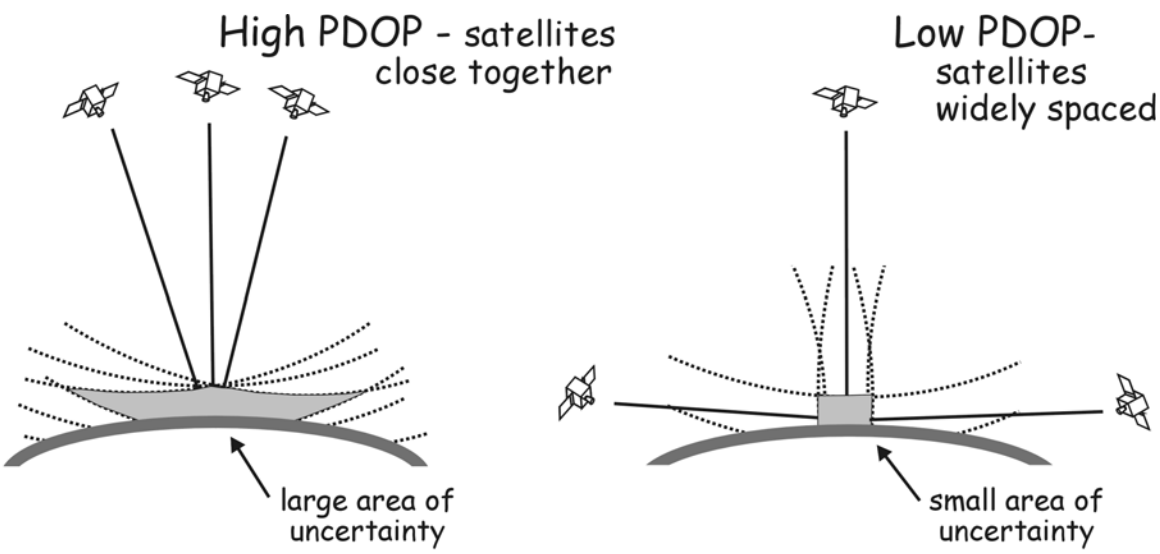


Figure 8: A Diagrammatic representation of how satellite geometry (relative to receiver) can influence the uncertainty and Dilution of Precision (DOP) of a geo-location estimation. The diagram on the left represents poor geometry and an high DOP while the diagram on the left represents good geometry and a low DOP. (Courtesy of Paul Bolstad, http://bolstad.gis.umn.edu/chapt5figs.)

Differential Correction

The error in a GPS signal can be determined by recording the GPS signal at a fixed surveyed location. By comparing the GPS receiver position to the surveyed position the physical error can be determined. Differential GPS (DGPS) take advantage of this known error to correct the SPS geo-location. The correction can be recorded independently and the SPS geo-location corrected later (post processing) or the correction can be applied in real-time. Real-time DGPS systems require two antennas: one to collect the C/A code and determine a geo-location and a second to receiver a correction factor to improve the accuracy of the geo-location. There are a variety of different sources available for the correction signal. These GPS receivers tend to be more expensive then the stand-alone GPS receivers, as they require extra componentry to accept the correction signal and update the geo-location.

Local Base Station

Any user can establish a local base station using a second GPS receiver and a pair of radios to transmit and receive the correction signal. A typical local base station set-up is shown in Figure 9.

Using a single fixed base-station for correction assumes that all errors applying at the reference station apply equally to the mobile receiver. Therefore the effects of ionospheric delay can be compensated for to a degree. As the distance between the two receivers increases, the receivers begin to observe different satellite information errors and receive the satellite signals via different travel paths through the atmosphere.

Initial studies on civilian use of DGPS, with user-controlled base stations and selective availability turned on, suggested an accuracy between 2-4m was attainable if the two receivers were positioned close together. This accuracy would degrade at approximately 1cm per km until the separation distance reached 100 - 200km. Further separation would subject the user to position error up to approximately 15m at 500km. Such degradation limits the usefulness of operating DGPS with a single base station to short-range operation.

Generally a single fixed local base stations are limited to ranges less than 30km by the strength of the radio signals used and the possibly of terrain interference. To overcome this and make the signal more accessible some alternative approaches to signal delivery have been used however issues with correction signal degradation need to be understood when using these wider area signals.

FM Frequency

If the differential signal is broadcast on a FM frequency sideband the system can accommodate multiple users within the effective range. Such differential correction signal coverage was available for many of the major cropping regions in Australia though it is now indefinitely out of service.

Coast Guard Beacon

It possible for some farming areas in Victoria, South Australia and Queensland to receive a free correction signal from a maritime navigation "beacon" system. There are plans for more of these maritime beacons to be installed along the east coast of Australia in the coming years, but their effective range will be dependent on a users location, intervening terrain and the signal strength of each beacon.

WADGPS and WAAS

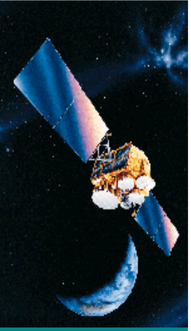
Problems with radio signal limitations and loss of accuracy by moving away from the base station can be overcome by using two or more base stations. The corrections from numerous base-stations can be combined into a correction algorithm that is optimised for any user located within the base station network (Figure 10). This form of correction is term Wide Area DGPS (WADGPS). Two competing companies, OmniSTAR and Thales Geosolutions Australia, offer WADGPS services with submetre accuracy across Australia.

A similar network, Wide Area Augmentation Service or WAAS, has been established in North America to aid flight navigation. The WAAS correction is free and improves the accuracy of a stand alone SPS receiver to ±3m (with a 95% Confidence).

Carrier Phase Receivers

More accurate modes of operation are available whereby the distance to satellites is determined in a codeless manner. This approach uses the phase shift of the information carrier signal between propagation at the satellite and reception by the user. This method offers potentially greater accuracy (centimetre level) but also requires more expensive receivers. Carrier phase systems may be either single frequency, i.e. accessing only the L1-band signals, or dual frequency, i.e. accessing both L1 and L2-band signals. Dual frequency receivers have the advantage of faster acquisition time. Many receivers are also capable of accessing the GLONASS as well as GPS satellites if required. Similar to code phase receivers, carrier phase receivers can be improved by using a local base station or a WADGPS correction.

OmniSTAR (OmniSTAR HP) and John Deere (Starfire) have both produced WADGPS dual frequency carrier phase networks offering sub decimetre accuracy in Europe, North America and Australia. The use of the WADGPS correction negates the need for a local base station. However the trade off is a loss of accuracy. Dual frequency carrier phase GPS receivers, with a local base station, are able to provide sub 2cm accuracy. These systems are often referred to as Real-Time Kinematic GPS (RTK-GPS). The on-farm use of carrier phase receivers is rapidly increasing with the adoption of auto-steer farm machinery.



GRDC
Grains Research & Development Corporation

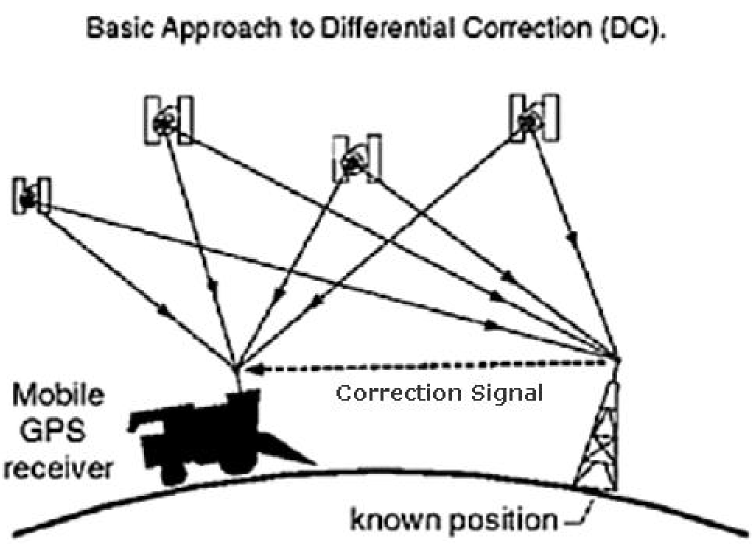


Figure 9: Operation of a DGPS network from a single fixed base station. Both the mobile receiver and the fixed point receive satellite signals. The error at the fixed point is calculated and transmitted to the mobile receiver for correction. (Adapted from http://pasture.ecn.purdue.edu/~abegps/)

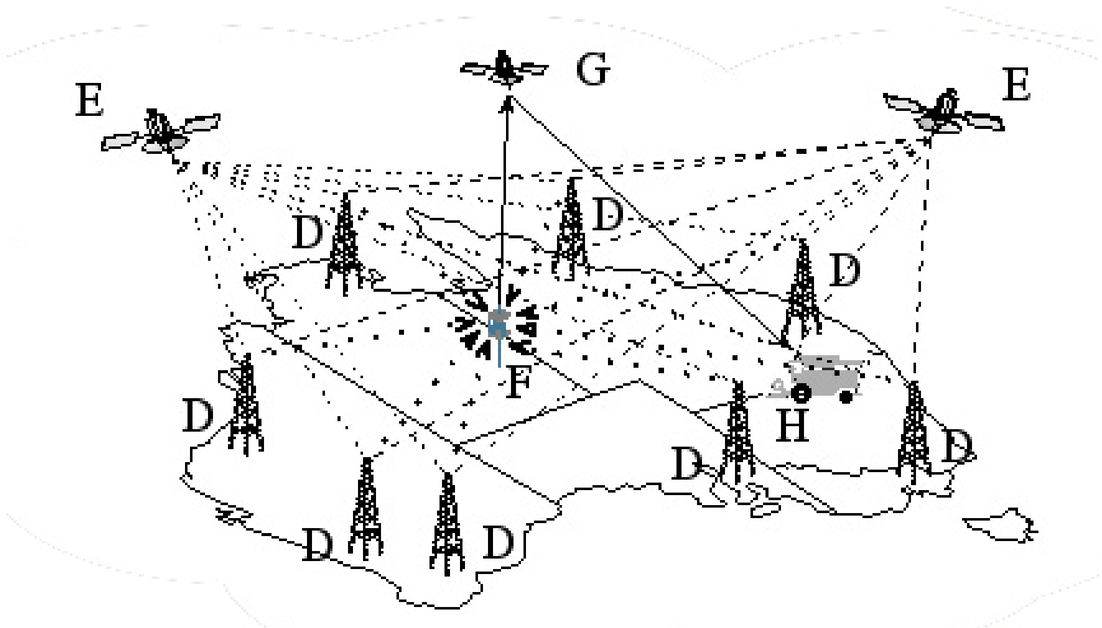


Figure 10: The network of fixed position receivers (D) communicate with the GPS satellites (E) and calculate an individual correction algorithm which is then passed to a master station (F). The master station computes a system-wide correction from all the individual stations and relays this to a general communications satellite (G) that increases the broadcast range to remote users (H). The correction transmission is supplied in a standard format (RTCM-104) defined by the Radio Technical Commission for Maritime Services. (Courtesy of Australian Centre for Precision Agriculture, University of Sydney)

Measuring Precision And Accuracy

Precision and accuracy are two terms that tend to be confused when discussing GPS (or any navigating/targeting system) performance. Precision is most simply defined as repeatability i.e. how close are the data points while accuracy is defined as how close the data points are to the actual point i.e. an estimation of the bias in the system (see Figure 11 for a graphical explanation of this concept).

Accuracy is a measure of how well the logged point approximates the actual point. It is usually determined by finding the radial error of the logged position from the actual point. The mean radial error can be determined if the logged position is averaged over a certain time period.

There are various methods of representing the precision of data. The most common approaches in GPS specifications are the 95% confidence interval (2s) and the Circular Probable Error (CEP). The 95% C.I. describes the radius of the circle within which 95% (i.e. 2 standard deviations) of the data lies. The CEP describes the radius of the circle within which 50% of the dataset resides. The CEP is equivalent to the median radial error of the data. The CEP statistic must always be smaller, or in extreme cases equal to, the 2s statistic thus the precision of the GPS system appears improved when quoting the CEP. For 3D descriptions the Spherical Error Probable (SEP) statistic is used. This is the same as the CEP except the SEP describes the radius of a sphere not a circle. Another useful alternative if multiple measurements are taken is the root mean squared radial error ($RMSE_r$). The difference between the RMSE of accuracy and RMSE of precision is that for accuracy the absolute location of the point is used as a reference while for precision the reference point is the mean location of the logged points.

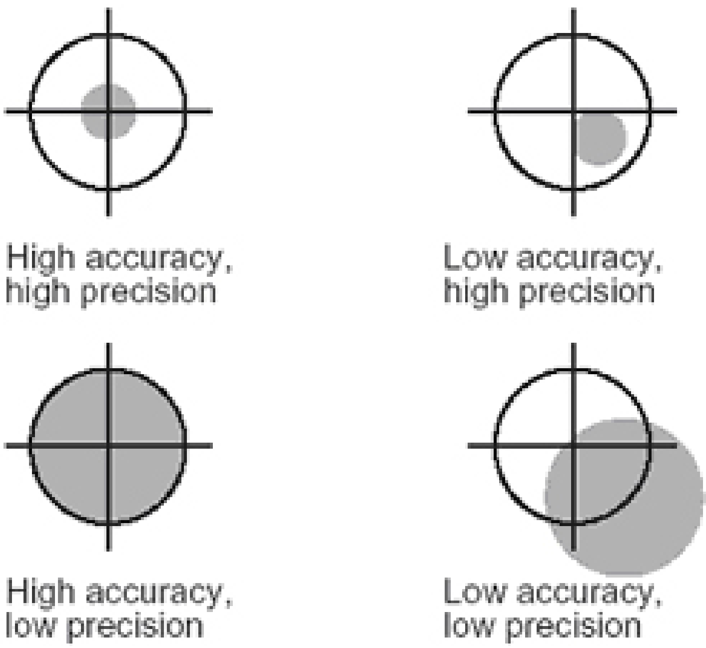
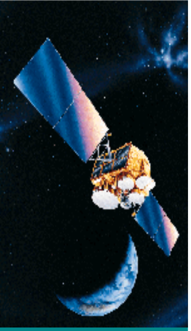


Figure 11: Comparison of the concepts of Precision and Accuracy in geo-location (after Environment Canada, 1993, Guideline for the Application of GPS Positioning)



Choosing a GPS

In an agricultural context, the required location accuracy and precision will depend on the operation being undertaken. As a guide, Table 1 presents some of the common usages of GPS in agricultural and the type of GPS required for each operation. There are many retailers of GPS equipment in Australia. Growers are advised to talk to various manufacturers and distributors to determine which GPS system is right for them and usages for different types of GPS receivers.

	Pegs $50	Standalone GPS ~$500	DGPS (C/A) ~$5000	DGPS High Quality ~$15-25000	DGPS Own Base Station ~$40-60000
Soil tests	*	*	*	*	*
Fertiliser Strips	*	*	*	*	*
Strategic Trials		*	*	*	*
Yield Maps		*	*	*	*
Guidance			*	*	*
AutoSteer				*	*

Table 1: Types of GPS currently commercially available and their potential uses on-farm.

References Used

Ashjaee, J. and Ashjaee, N. (1998) A GPS Tutorial. (http://www.topconps.com/gpstutorial)

Bolstad, P. (2002) GIS Fundamentals: A First Text on Geographical Information Systems (http://bolstad.gis.umn.edu/chapt5figs.html)

Cooksey, D. Understanding the Global Positioning System (GPS) Montana State University. (http://www.montana.edu/places/gps/understd.htm)

Dana, P.H. (1994) Global Positioning System Overview . The Geographer's Craft Project, Department of Geography, The University of Colorado (http://www.colorado.edu/geography/gcraft/notes/gps/gps_f.htm)

Environment Canada (1993) Guideline for the Application of GPS Positioning.

van Hooijdonk, A. (2004) The Practice of GPS Navigation and what else GPS can do for you. (http://www.gps-practice-and-fun.com/gps-tests.html)

Satellite Navigation Product Team. (http://gps.faa.gov/gpsbasics/spacesegment-text.htm)

Other Useful Resources

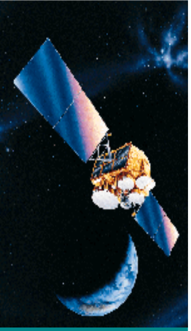
All About GPS
http://www.trimble.com/gps/

US Navel Observatory
http://tycho.usno.navy.mil/gps.html

US Coast Guard Navigation Centre
http://www.navcen.uscg.gov/

European Space Agency
http://www.esa.int/esaNA/index.html

James Taylor & Brett Whelan
Australian Centre for Precision Agriculture
www.usyd.edu.au/su/agric/acpa



Appendix 1: Measuring Distances To Satellites

If you use this material please reference the original source:

www.topconps.com/gpstutorial/Chapter2.html#Measuring%20Distances%20to%20Satellites

Time Is Distance

Have you noticed that during a thunderstorm, you hear the sound sometime after you see the light? The reason is that sound waves travel much slower than light waves. We can estimate our distance to the storm by measuring the delay between the time that we see the thunder and the time that we hear it. Multiplying this time delay by the speed of sound gives us our distance to the storm (assuming that the light reaches us almost instantaneously compared to sound). Sound travels about 344 meters (1,130 feet) per second in air. So if it takes 2 seconds between the time that we see the lightning and the time that we hear it, our distance to the storm is 2 x 344 = 688 meters. We are calculating the distance to an object by measuring the time that it takes for its signal to reach us.

In the above example, the time that we see the lightning is the time that the sound waves are generated in the storm. Then we start to measure the delay until the time that we hear the sound. In this example, the light is our start signal. What about the cases for which we don't have a start signal? Consider the next example.

Codes and Patterns

Assume that your friend at the end of a large field repeatedly shouts numbers from 1 to 10 at the rate of one count per second (10 seconds for a full cycle of 1 to 10 count). And assume that you are doing the exact same thing, synchronized with him, at the other end of the field. Synchronization between you and him could have been achieved by both starting at an exact second and observing your watches to count 1 number per second. We assume that you both have very accurate watches. Because of the sound travel time, you will hear the number patterns of your friend with a delay relative to your patterns. If you hear your friend's count with a delay of one count relative to yours then your friend must be 344 meters away from you (1 sec x 344 meters/sec = 344 m). This is because the counts are one second apart.

Now assume that you and your friend count twice as fast, two counts in one second. Then at the same distance between you and your friend you will hear a two-count delay. This is because now each count takes 0.5 seconds and each count delay measures 172 meters. If you could count 100 times faster then each count would take 0.01 seconds and each count delay between you and your friend would measure the distance of 3.44 meter. Counting faster is like having a ruler with finer graduation. Of course in real world, you need appropriate devices and instruments to generate and receive very fast counts.

Next assume that you and your friend are far apart and counting very fast, say each count in 0.01 second (each delay count is 3.44 meters), and, as before, both are repeatedly counting from 1 to 10. Assume when you say 7 you hear your friend's voice say 5. You hear a delay count of 2 but you know your distance is more than 6.88 meters. This is because the delay is not just only 2 counts, but rather 2 counts plus some multiples of 10 counts (i.e. some multiples of the pattern cycle). This is as if your measuring tape is not long enough and there are some multiples of the full length of measuring tape plus some fraction. We refer to this unknown number of full pattern delays as *unknown integer*. If you and your friend were to count repeatedly from 1 to 1000 (instead of 1 to 10) then you could hear 212 count delays between the numbers that you hear and your numbers, which would produce the distance of 212 count delays x 3.44 meters = 729.28 meters. This is 21 full cycles of the 1-to-10 pattern, plus 2 counts. The number of full cycles, 21, that we were not able to observe with our short pattern is our *unknown integer.*

What we demonstrated above are the concepts of *pattern granularity* (fineness of tape marks) and *pattern length* (tape length).

The concept of measuring distances to satellites is much like what we discussed above, but satellites transmit electronic patterns rather than voice counts. Likewise, our receiver generates similar electronic patterns for comparison with the received patterns from satellites in order to measure the distances to them.

Satellites generate two types of patterns: One has a granularity of about 1-millimeter and a length of about 20 centimeters. The other has a granularity of about 1 meter and effectively an unlimited length. In satellite terminology, the first pattern is called *"carrier"* and the second is called *"code"*. The distance measured by carrier is called *"carrier phase"* and the distance measured by code is called *"code phase"*. Because code pattern is long, the code phase measurements are complete and do not have any unknown integer. We can measure our distance to a satellite as 19,234,763 meters, for example. In contrast, the carrier pattern is short and carrier phase has a large unknown integer. You may think that it is useless to say, for example, that our distance to satellite is 13.2 centimeters plus an unknown number of carrier cycles. The unknown integer is in the order of several tens of millions. You may ask what good will it do to measure the fractional part so accurate when millions of full cycles are missing? We will explain more.

Initial Unknown Integer (Integer Ambiguity)

In the previous counting example with a short pattern, assume that you and your friend are standing next to each other and synchronized together counting fast from 1 to 10. You hear no delay because you are standing next to each other. Then your friend starts to move away. The count delays start to grow from 0 (no delay) to 9. After it reaches 9 it will drop back to 0. This is actually 10 and not zero. You know that this is the case (that the zero count delay actually represents one full cycle count) because you have been following the count delays continuously. You will keep in mind, as your friend moves away, to count the whole number of cycles that are being added to your distance. In this case, there is no unknown integer as long as you keep track of him continuously.

If, instead of starting next to each other, you start at some unknown distance, then you are starting from an unknown integer of cycles. However, if after starting your friend moves away from or towards you, you can account for the number of full cycles that must be added to or subtracted from the *initial unknown integer*. All the distances that you measure every second contains the same initial unknown integer. This is true as long as you keep track of him continuously. If you don't hear him for some period of time, then you don't know how many full cycles he moved and you will have to start with another unknown number of cycles. The point is that as long as you keep track of him you have only one initial unknown integer.

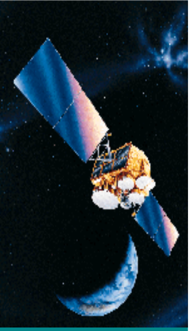
The concepts of code and carrier are very important. Let us use another analogy for better understanding. You may consider that code phase is like a watch that only has an "hours" hand (call it code watch). At any time you can look at this watch and know the time of the day approximately. You may consider carrier phase like a watch that only has a "seconds" hand (call it carrier watch). You can keep track of the elapsed time with this watch with the accuracy of one second as long as you monitor the watch continuously to keep track of the elapsed full minutes. If you somehow can determine the number of full minutes initially (the initial unknown integer when you started looking at this watch) then you can keep track of time very accurately. If you get distracted and lose track of the number of minutes, then you have a new "initial unknown integer" that you somehow must determine again. With code phase watch you always get the time of the day instantly but with the accuracy of not better than 10 minutes by estimating the location of the hour hand. The code watch can narrow the estimate of unknown minutes (integers) of the carrier watch to plus or minus few minutes. You see that there is a gap between the seconds hand and the hours hand. We are missing the minutes hand. GPS manufacturers have developed techniques to narrow the gap such that code phase and carrier phase can make unambiguous and accurate distance measurements as fast as possible. We will explain the reason for the gap later.

The good news is that the integer ambiguity of carrier phase can be determined by tracking satellites for some period of time. This is the fundamental concept in precision applications like geodesy.

With carrier phase, tracking the correct number of full cycles that the distance to satellite is changing is very critical. You will miscalculate this number if you miss a cycle or add an extra cycle. In GPS terminology, this is called a *"cycle slip"*. In our previous example, cycle slips can happen if you don't hear your friend's voice correctly due to noise or other effects, or if he suddenly jumps a very long distance. Cycle slips is like missing the meter marks while you are concentrating on reading the millimeter ticks. It can create large errors. Most GPS systems are able to detect and repair cycle slips.

Note that not all receivers can measure carrier phase. Carrier phases are typically used in high precision receivers.

We can measure the distances to the satellites with the accuracy of 1 meter with code phase and 1 millimeter with carrier phase. This does not mean that we can determine our position with a GPS receiver with the accuracy of one meter or one millimeter. There are several sources that introduce inaccuracies into the GPS measurement



APPENDIX 2: DILUTION OF PRECISION

If you use this material polease reference the original source

www.gps-practice-and-fun.com/gps-tests.html

Here we will describe some practical tests about GPS satellite reception. These tests should not be considered scientific and all exclusive. However, the results will have practical value in the field.

A GPS receiver determines its *Position* (horizontal and vertical), its *Velocity* and the *Time* from the signals of at least four satellites by means of triangulation. The precision of the computations by triangulation depends on the constellation of all satellites of which the signals are taken into account (four or more). As the number and position of satellites will seldom be ideal, the maximum obtainable precision will be *diluted* in practice. Here we present the different terms of dilution of precision.

Dilution of precision (DOP) is a measure of the quality of the GPS data being received from the satellites. DOP is a mathematical representation for the quality of the GPS position solution. The main factors affecting DOP are the number of satellites being tracked and where these satellites are positioned in the sky. The effect of DOP can be resolved into HDOP, VDOP, PDOP and TDOP.

HDOP (Horizontal Dilution Of Precision) is a measure of how well the positions of the satellites, used to generate the Latitude and Longitude solutions, are arranged. PDOP less than 4 gives the best accuracy, between 4 and 8 gives acceptable accuracy and greater than 8 gives unacceptable poor accuracy. Higher HDOP values can be caused if the satellites are at high elevations.

VDOP (Vertical Dilution Of Precision) is a measure of how well the positions of the satellites, used to generate the vertical component of a solution, are arranged. Higher VDOP values mean less certainty in the solutions and can be caused if the satellites are at low elevations.

TDOP (Time Dilution Of Precision) is a measure of how the satellite geometry is affecting the ability of the GPS receiver to determine time.

PDOP (Positional Dilution OF Precision) is a measure of overall uncertainty in a GPS position solution with TDOP not included in the estimated uncertainty. The best PDOP (lowest value) would occur with one satellite directly overhead and three others evenly spaced about the horizon.

The **Position Accuracy = Dilution Of Precision (DOP) X Measurement Precision**. So, if the Measurement Precision = 1m and the DOP = 5, then the best position accuracy will be 5m.