Field Techniques Manual:
GIS, GPS and Remote Sensing

- <u>Section B: Data</u>

Chapter 6: The Global Positioning System (GPS): Principles & Concepts

6 The Global Positioning System (GPS): Principles & Concepts

Over the last five years Global Positioning Systems (GPS) have changed the way fieldwork is conducted. There are two principal reasons for using GPS in the field; these are navigation and determining co-ordinates for points in the GIS. This manual will not deal in depth with navigation, as this topic is described well elsewhere (for a good introduction see Simmonds (2004), which is included on the CD accompanying this book). Navigation is touched on briefly in Section 6.1 and the reader should note that, even though GPS are excellent tools for field navigation, their very nature as electrical equipment means they are fallible. As such, a more traditional backup including a map and compass are essential.

This chapter discusses GPS in detail. It does not aim to describe all manufacturers' units and does not replace the unit's user manual. Even though many aspects of GPS are described, not all of them will be relevant to any given expedition. There are many different types of GPS and different methods for using them. These differences give accuracies ranging from several centimetres to tens of metres. Chapter 11 discusses the appropriate use of GPS for various expeditions and teams should not always be concerned with obtaining the most accurate sets with the most features if this is not appropriate for their studies. An informed decision cannot be made without a thorough understanding of all the aspects of GPS so this chapter describes as much relevant GPS information as possible. Some of the techniques will be too involved for smaller expeditions and expeditions should study this chapter in conjunction with Chapter 11 to select the most practical and appropriate methodologies. Expeditions should not select expensive, time consuming and difficult to use navigation solutions if they are not required. Although there is always a push towards more accurate and precise methods, they should not be used if not required. Studying this chapter should help you to make an informed choice.

6.1 GPS and field navigation

Navigation is vital to the safety of any field expedition. When combined with the necessity of fixing a location's co-ordinates for scientific research, the need for accurate, rapid and cost-effective navigation tools becomes paramount. Increasingly GPS receivers are becoming a standard – some would say essential – item of expedition equipment. Determining the co-ordinates of a point in the field can be achieved in a number of ways. The most common traditional approach involves triangulation with a map and magnetic compass. Triangulation (see Chapter 10) is often very accurate but relies on accurate maps and navigable objects. The Ordnance Survey of Great Britain produces very reliable maps but even they admit:

"*On top of the nationwide errors in OSGB36, individual features on the map may only have been surveyed to a local accuracy of 7 m (for 1:25,000 scale maps) and some features such as boulders may only be shown schematically.*"

Ordnance Survey Website

The result is that any triangulation achieved is relative to the map, which may in fact be quite inaccurate. Lines on navigation charts have accuracy on paper of ±1.5 mm. On a 1:10,000 chart that could be an error of 75 m. In addition, when drafting, the tools used may introduce additional errors. Triangulation is also time consuming and of limited use outside of areas of human influence i.e. those areas with man made objects surveyed to an acceptable accuracy. Other methods have been employed to determine location but they are either difficult in the field or rely on expensive equipment, examples include sextants for astronomical positioning and various types of theodolites for astronomical triangulation. There has for some time been a move to establish Global Navigation Systems (GNS) that are quick, cost effective and reliable. GPS has been the most successful of these systems.

6.2 Introduction to GPS functions/features

GPS use satellite data to calculate an accurate position on the earth. These calculations can relate the user's position to almost any map projection within milli-seconds. All GPS work in a similar manner but they often look very different and have different software. The most significant difference between GPS receivers is the number of satellites they can simultaneously communicate with. Most receivers are described as 12 channel meaning they can communicate with 12 satellites. Older models may be 8 or even 5 channel with more modern receivers capable of communicating with 14 – 20. Given the current (2005) makeup of the GPS satellite's constellation 12 channel is more than adequate.

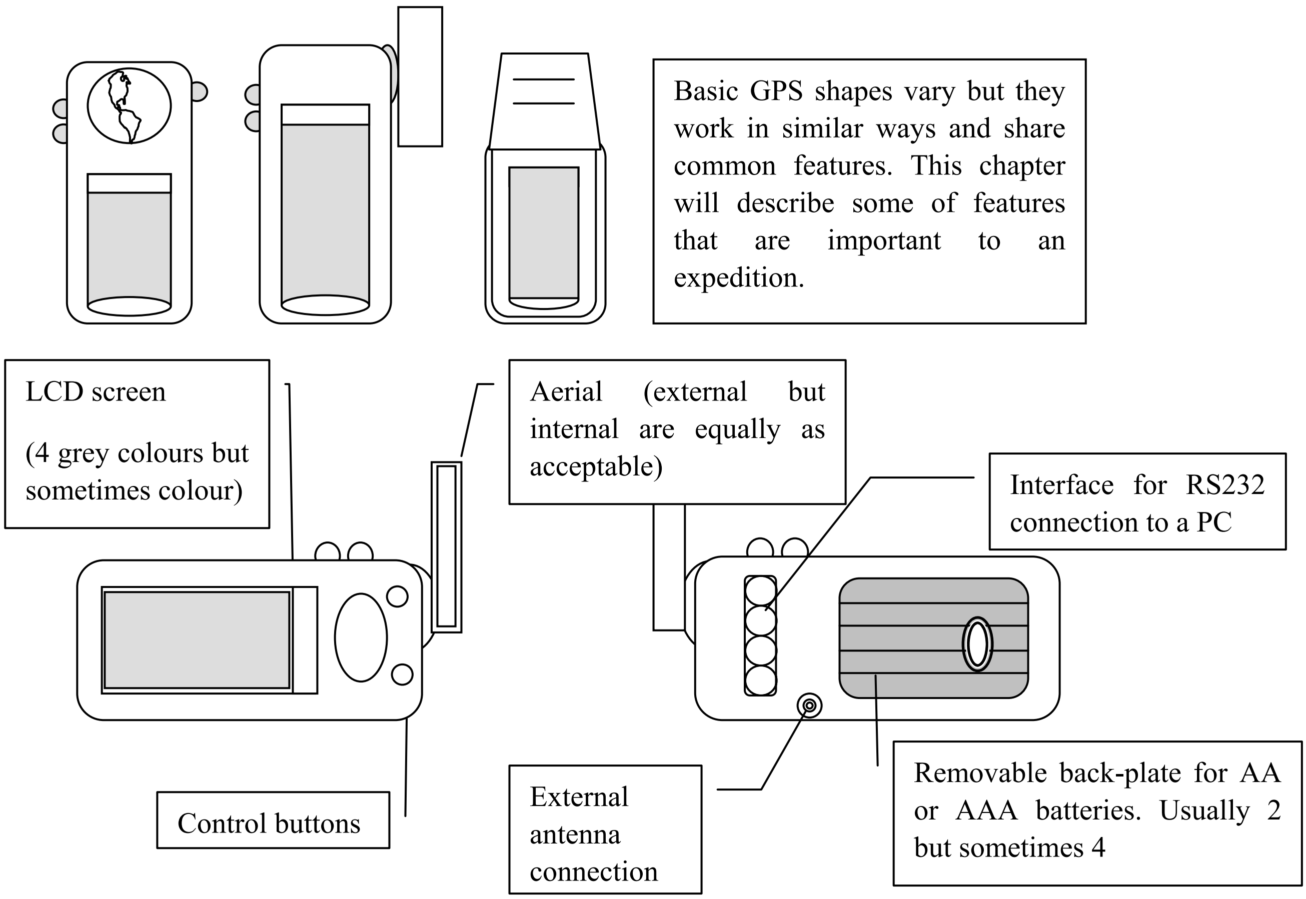


Figure 6-1 A basic guide to GPS models.

Almost all units have an LCD screen or at least software that links to a PC/PDA with an output screen. The unit might have several different pages that can be displayed on screen but usually the default page is very similar. Commonly on starting a receiver you will be presented with a map of the satellites in view. The GPS receiver shows a view of the sky split into four quadrants. These represent the NE, SE, SW, NW parts of the sky, with the concentric circles representing the horizon at 90° from the zenith, with the inner circles representing 60° and 30°. The cross at the centre represents the zenith. The dots/circles represent the satellites and the bars at the bottom represent satellite signal strength. The higher the bar the stronger the signal. This display is typical of a 12 channel set. The dots and bars will commonly be labelled with a number to represent the identity of the satellite. The bars are commonly either hollow or solid (usually white or black on a monochrome display). Hollow lines represent a satellite for which the Ephemeris data is not known. It is therefore not being used to calculate a position. Black bars represent “Fixed” satellites whose ephemeris data has been collected successfully. These satellites are thus available for calculating a position. This is not consistent across all models and some may use grey bars as well as hollow bars to represent satellites not yet fixed.

The number, position and strength of signal from the satellites allows the GPS to calculate a rough estimate of the error in its reported position. This error or dilution of precision is a good guide to how accurate any reading would be. It should be closely monitored and readings should only be taken when this is below 10 m (ideally below 5 m).

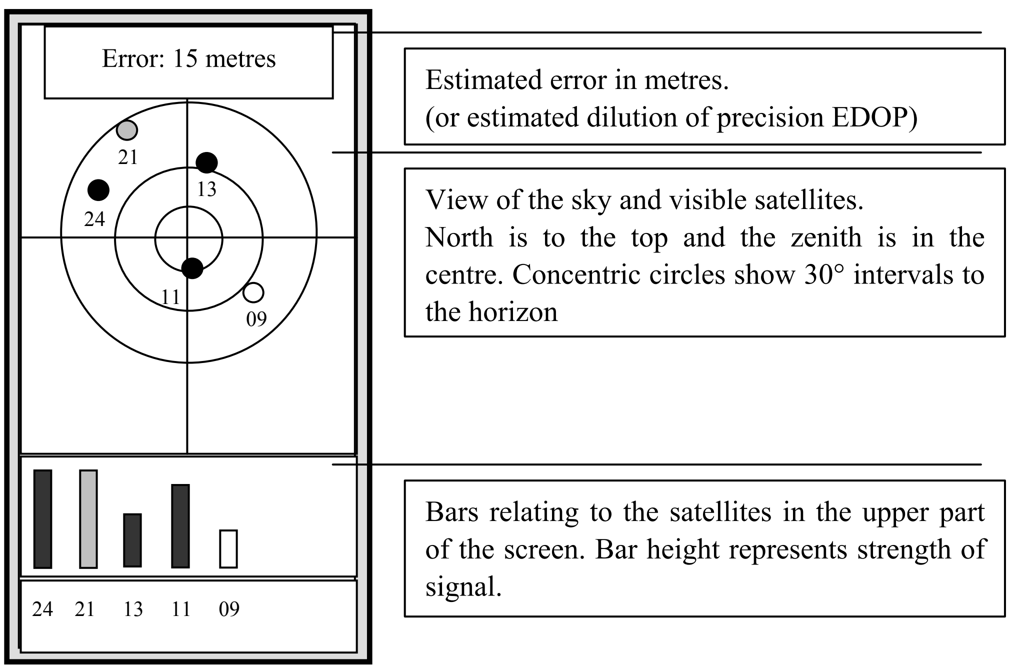


Figure 6-2 Basic GPS screen layout.

The way the GPS records data is generally the same across all units. GPS receivers automatically records data into their memory according to elapsed time or distance moved. These points are called trackpoints. The device can be forced to record additional data, generally with additional information, at user discretion. These user recorded points are called waypoints. Some of the common pages used for viewing this data are shown below. The more expensive sets have more detailed screens.

The GPS receiver can display the entire tracklog ‘all trackpoints collected’ or a selected route ‘specific associated waypoints’ on its display. This can show the direction travelled or plot a course to follow. GPS receivers can receive data uploads, either through a COM port or from user input. These uploaded co-ordinates might describe a route to follow or mark locations of scientific interest. The GPS set can then be used to navigate to these areas. In the diagram opposite, two symbols are used for the start and end of a route and waypoints along the route are marked with the locational Ids 1-4. In between these a route has been drawn from automated trackpoints.

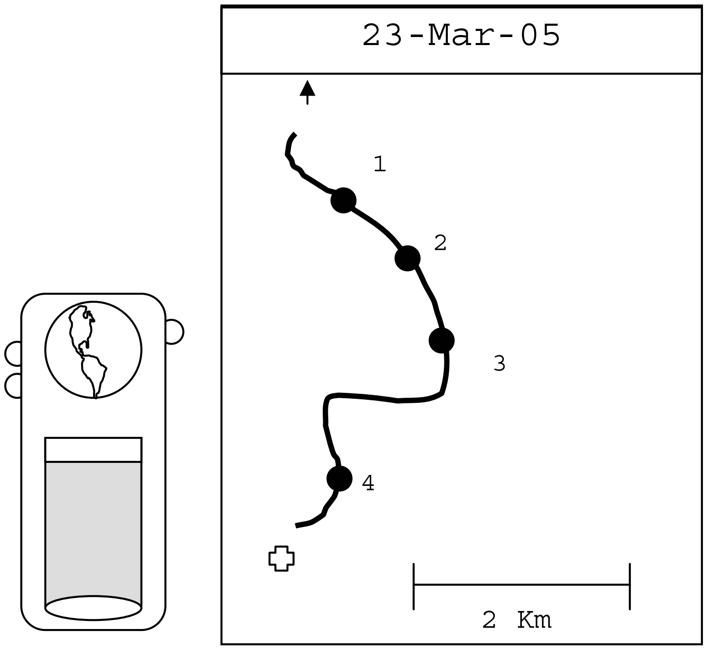


Figure 6-3 Basic GPS Tracklog screen showing track and way points.

In some newer models trackpoints can be used for another purpose. Those sets with inbuilt ‘*area calculations’* can use tracklogs to calculate the area of an object. By selecting the beginning and end of a log, the GPS receiver will attempt to calculate the area enclosed.

The trackpoints on the diagram to the right have been used to create an area calculation tracklog of a lake. This method can be quickly used to calculate areas of larger objects. However, the ~15 m GPS positional errors mean that anything smaller than around 2500 m^2 will be subject to very large errors:

E.g.: Area of a 50 m wide object

50m ±15m x 50m ±15m =

$\underline{1225m^2 – 4225m^2}$

A large percentage difference (*345%*)
Whereas: Area of a 1km wide object
1000m ±15m x 1000 ± 15m =
$\underline{970225m^2 - 1030225m^2}$
Which is a significantly smaller percentage difference (*6%*)

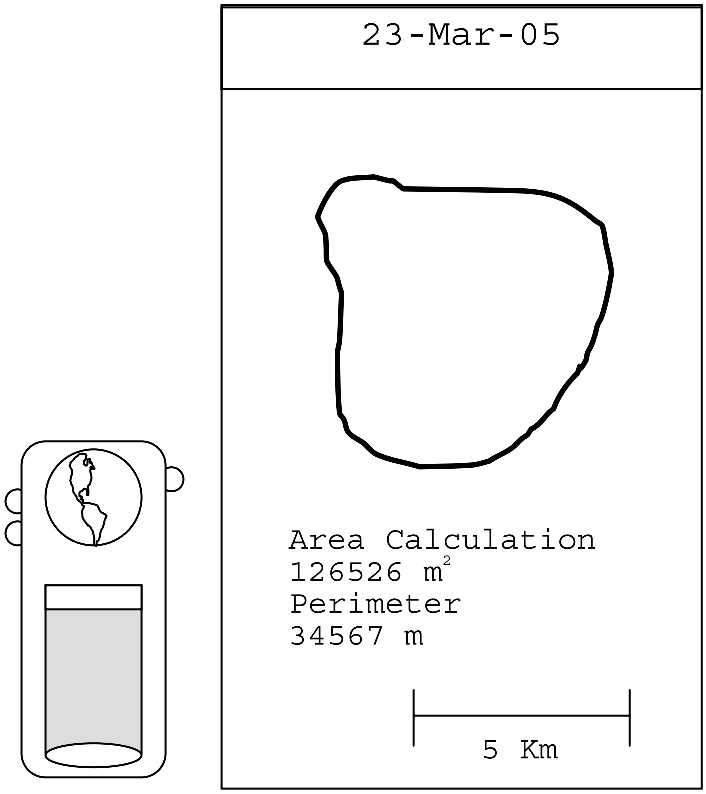


Figure 6-4 GPS area calculations showing when they should and should not be used.

Care must always be exercised before leaving for the field that the characteristics of a given GPS receiver are understood. For example, some receivers will not record elevation within trackpoints while others will not record elevation below sea level. Either of these might be crucial to the aims and objectives of a team. Other common GPS features include support for a ‘Man Over Board’ (MOB) alert. MOB automatically records details of the current location and immediately instructs the user how to return to that point. This is commonly used at sea to return to a lost crewmember but can equally be used to return to any expedition location or the expedition vehicle.

GPS receivers can often be used as a complete navigation tool, not only offering directions and location details but also navigation tools to move between locations. Most receivers come with an inbuilt digital compass. The digital compass is based on data from satellites and is not a magnetic compass. The digital compass will only work when moving and will not re-orientate if the set is rotated. More expensive models can come equipped with a magnetic compass and an inbuilt barometric altitude calculator.

The navigation pages often include a digital compass, an odometer showing distance travelled since the counter was reset, a current speed indicator, a maximum speed and average speed for a trip. The navigation page can also be tailored to show a variety of other statistics e.g. to show how far is remaining until an objective (waypoint, MOB etc) is reached. Some high-end sets include a magnetic compass and a barometric altimeter. These are generally more accurate than the satellite determined heading and elevation. GPS receivers can also display the time of sunrise and sunset at a given location.

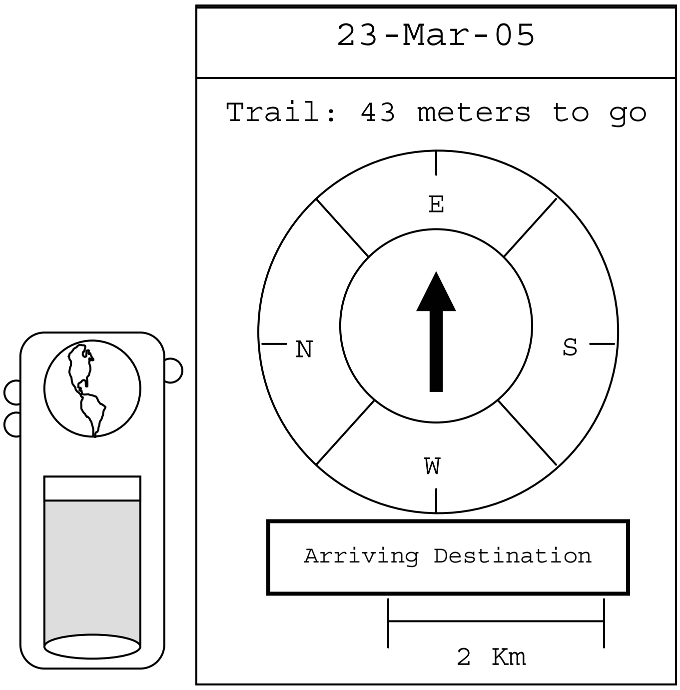


Figure 6-5 Basic GPS Man Over Board or Track Back features.

Standard GPS receivers range in price from under £100 to over £400. There are specialised high accuracy models that can cost up to £40,000 but the common models that an expedition might use are typically under £200. The difference between the models is not related to their accuracy but the number of additional screens and levels of data in the receiver’s database as shown in Figure 6-6. This database can usually be expanded by purchasing CDs of street-level data. These might cost around £100 on a per country or area basis but are of little use to an expedition. Most commonly fieldwork will take place in areas where road maps for GPS may not be available and printed maps would probably be better.

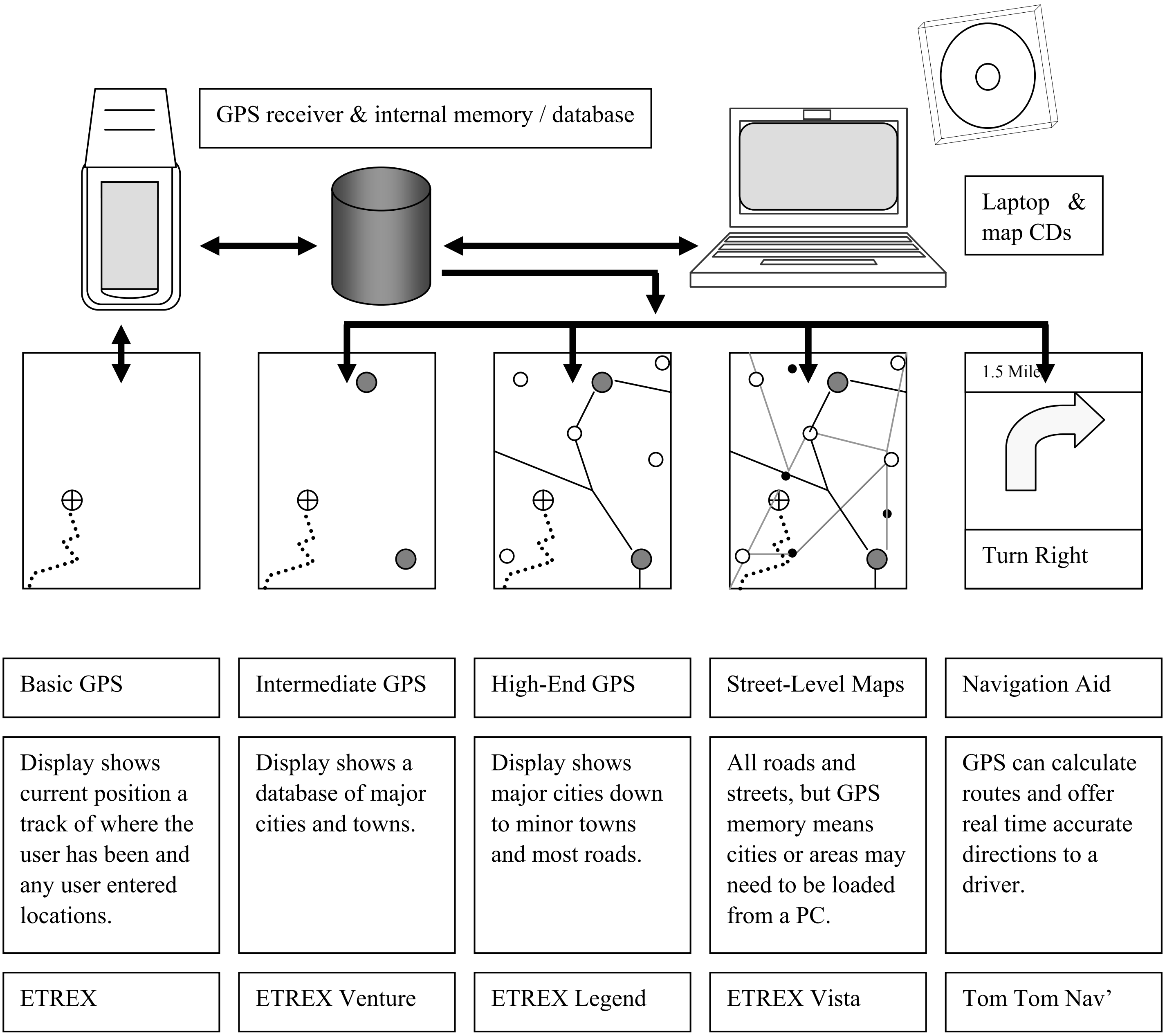


Figure 6-6 The difference between GPS models and their databases.

For most expedition work, a basic or intermediate GPS would be sufficient. The rest of this chapter describes in more detail how GPS actually calculates its position and how the data can be manipulated in a GISci context. Street level mapping and in car navigation aids are rarely useful for GISci fieldwork. It is better to concentrate on basic units that have features that can benefit the team. The buyer’s guide in Appendix 4 describes these features in more detail.

6.3 GNS history

Though there are various land based navigational services such as DECA and Loran, this chapter looks purely at satellite GNS. In the 1950s the US Navy began a programme to study navigation from artificial satellites. The first satellite navigational aid, TRANSIT, was accurate to approximately 160 m for stationary receivers. Moving receivers introduced additional errors of around 1 km per 1 m per second speed. These initial tests were generally accurate for a ship at sea but were of limited use for navigating into ports or shallow waters. The TIMATION I satellites launched in 1967 allowed comparatively slow moving receivers to calculate positions via embedded atomic clocks. This was much more accurate for ships but soon the aviation industry became interested and a system was developed for faster objects. The current Global Position System suitable for aircraft and high-speed navigation, NAVSTAR, was initiated by the US Air Force in 1978. Further details about this can be found at: *www.britannica.com* using a search for GPS.

NAVSTAR theoretically gives global coverage with accurate positional information down to sub metre levels with elevation at sub 10 m accuracy. Four years later the USSR launched a similar system, GLONASS. GLONASS offers global coverage like the NAVSTAR system but has less than half the satellites (9 as of 2004 but with more launches planned to take this back to around 14 by late 2005). The system has a strong bias towards the Northern Hemisphere. GLONASS results obtained below the equator are less accurate and the system has suffered from poor maintenance in recent years. Northern hemisphere results are considered slightly more accurate than standard NAVSTAR results. In December 2004 the US and Russian Federation agreed to co-operate on the development of their systems.

6.3.1 The NAVSTAR system

The NAVSTAR system is managed by the Interagency GPS Executive Board (IGEB) of the US government. Details of this can be found on their website at *www.igeb.gov*. A statement by President G.W. Bush in December 2004 indicates that this may change in the future but the IGEB website remains an excellent source of up to date information. The current specification requires satellites to orbit the Earth in one of six orbits inclined at different degrees to the equator, between –55° +55° at an altitude of 20,200 km. The DoD maintains 4 satellites in each orbital plane, giving a total constellation of 24 satellites, currently supported by up to 5 spares. Satellites are being replaced over time and the newest satellites are referred to as GPS IIR SVs. The design of the satellites allows navigation at all latitudes during all weather conditions. The L-band radio wave used to communicate to Earth from the satellites is effectively immune to local atmospheric conditions such as rain, storms etc. The satellites broadcast two L-band signals (L1 and L2) operating at the following frequencies, L1 = 1575.42 MHz and L2 = 1227.6 MHz. The NAVSTAR system operates two services, standard and precise. The Standard Positioning System (SPS) is available worldwide at no charge and operates the L1 frequency. The Precise Positioning Service (PPS) broadcasts on the L2 band only and accurate data are attained by correlating the two bands. PPS receivers are used solely by the U.S. Military and allies, as well as by the U.S. Federal Government. Applications for access to PPS by non-Federal Government organisations, both domestic US and foreign, can be made and are considered on a case-by-case basis. As a result, all expedition GPS use will probably involve only the SPS signal. Further details about the SPS signal used by most expeditions can be found at *www.igeb.gov/SPS-2001-final.pdf*.

6.4 How GPS works

GPS signals do not contain positional data. The position reported by the receiver on the ground is a calculated position based on range-finding triangulation. GPS positioning is achieved by measuring the time taken for a signal to reach a receiver. Almost one million times a second the satellite transmits a one or a zero in a complex string of digits that appears random. In actuality this code is not random and repeats every 266 days. The receiver knows that the portion of the signal received from the satellite matches exactly with a portion it generated a set number of seconds ago. When the receiver has determined this time, the distance to the satellite can be calculated using simple trigonometry where:

Distance to the satellite = speed x *(t_r - t_{to})* (where speed is c, the speed of light, in a vacuum (299792.5 x 10^3 ms^{-1}). t_{to} is the time at the origin and t_r is the time at the receiver).

The DoD maintains very accurate telemetry data on the satellites and their positions are known to a high level of precision. This simple operation allows the distance to a satellite to be calculated accurately. When the distance to three satellites is known then there is only one point at which the user can be standing. This principle is demonstrated in the diagrams on the following pages.

From one measurement we know the receiver can be anywhere at a uniform distance from the satellite with a radius equal to $r = c \times (t_r - t_{to})$. This defines the outer surface of a sphere of radius r.

Where:
r = radius
c = speed of light
t_{to} is the time at the origin
t_r is the time at the receiver

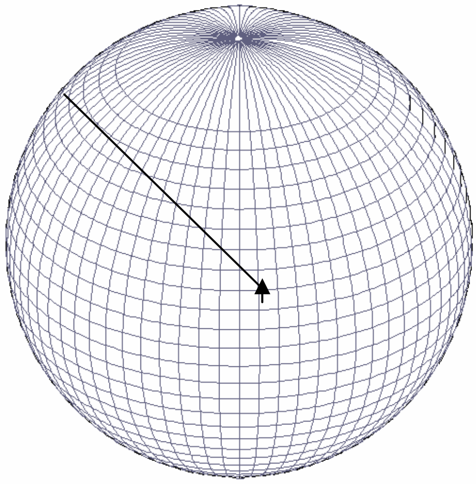


Figure 6-7 Basic Trigonometry - Single Satellite.

From two measurements we know the receiver must be anywhere on the line of the outer edge of a circle of intersection between the two spheres shown as a shaded ellipse below:

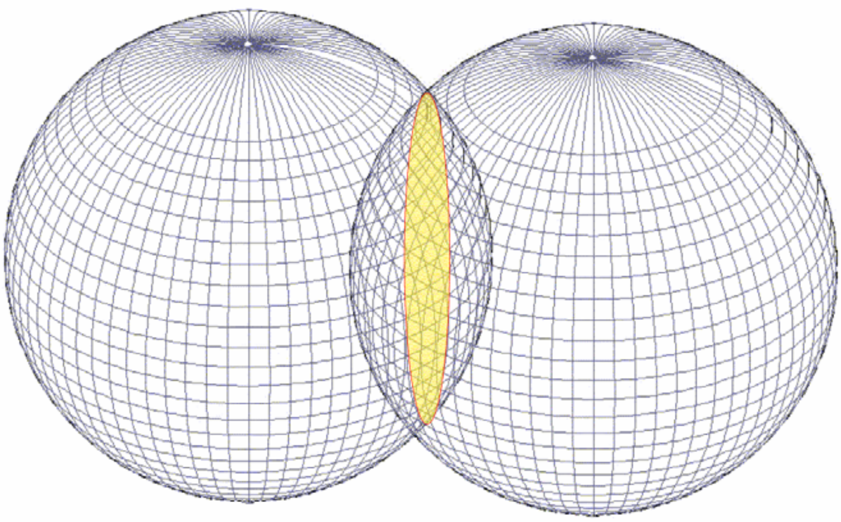


Figure 6-8 Basic Trigonometry - Two Satellites.

A third measurement reduces this to the intersection of a plane with the circle. This reduces the possible location to two points. Only one of these can be on the Earth's surface.

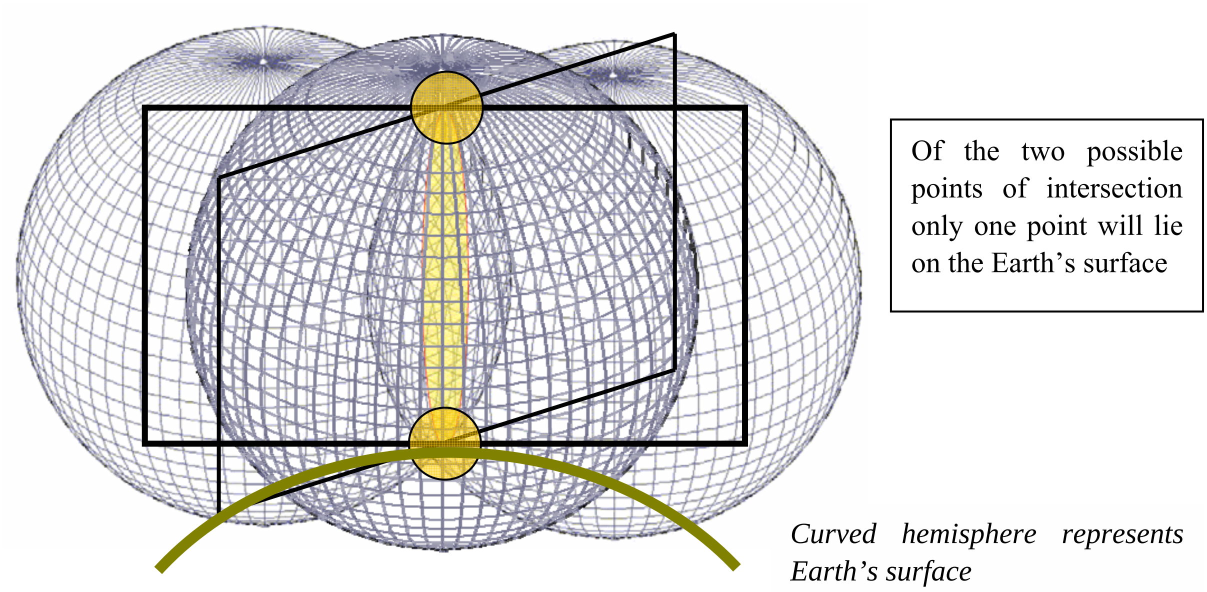


Figure 6-9 Basic Trigonometry - Three Satellites.

Unfortunately, the above description is an oversimplification. This method of triangulation requires the receiver to know the precise time that the signal was transmitted and received. Even though time at the satellite (t_{to}) is known precisely because it is time stamped by the atomic clock on board the satellite, time at the receiver (t_r) is not known because this is generated by the internal receiver clock. To determine positional fixes to metre accuracy requires the GPS receiver to measure time accurately to 10^{-10} of a second. To keep the cost of GPS receivers below several thousand dollars per unit, atomic clocks are not used in the handsets. Due to these inaccuracies in timing the margins of error in calculated positions are very large. The way GPS receivers circumvent this problem is by using an additional measurement. The internal clock of the receiver will measure t_r incorrectly for all satellites. Therefore, because the offset is the same for all satellites, the receiver can use an additional satellite to bring all the points to one location.

The number of satellites a GPS receiver can talk to at one time affects the accuracy and the speed at which the system can function. 12 channel are the most commonly used receivers today, and are both quicker and more accurate than older models.

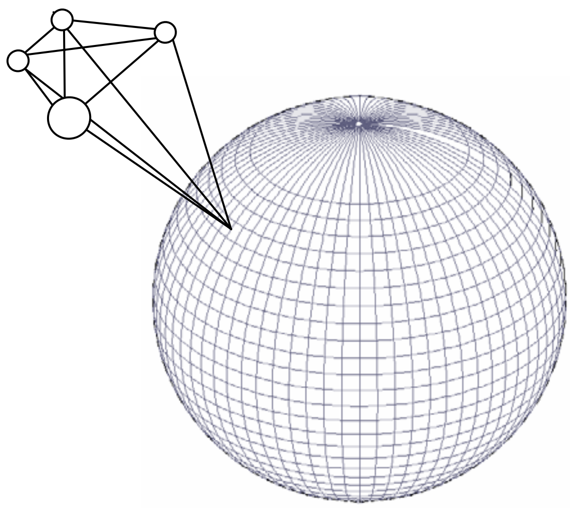


Figure 6-10 Basic Trigonometry - Four Satellites.

6.5 GPS accuracy

The signal transmitted by the satellites has a potential accuracy of <1 m but several factors influence this and reduce the actual resolution. The US military designed the end user of the SPS to be able to resolve a position 95.4% of the time (two standard deviations) to an accuracy of 100 m in X and Y (longitude and latitude) and 156 m in Z. Using the PPS service the end user should be able to resolve 22 m in X and Y and 27 m in Z. These are very conservative estimations and actual accuracy will lie between the theoretical resolution and these design schematics.

6.5.1 Factors affecting GPS accuracy

The reason why the actual locational position is significantly less accurate than the data transmitted by the satellite is due to various influences on the signal. These can be collectively termed *local* and *atmospheric* effects. Local effects are detrimental conditions on the ground near the receiver or in the receiver's software while atmospheric effects are problems with the medium through which the signal passes.

Table 6-1 Common factors that affect GPS accuracy.

Local Effects	Atmospheric Effects
Receiver Clock Error	Ionospheric Effects
Percentage Sky Visible	Tropospheric Effects
Satellite Geometry	
Multipath Error	
Ellipsoid	

6.5.2 Local effects

Receiver Clock Error: This is the error in the offset of the GPS measurement of the pseudo random code and the time recorded by the satellite for the data. The receiver attempts to compensate for this with additional measurements but it remains the single largest error that affects positional accuracy.

Potential Error: 0-10,000 m   Common Error: 3-10 m

Percentage Sky Visible: This is of concern when getting an initial fix and generally causes the second largest error in calculated positions. It is linked to *satellite geometry* (below) and is a measure of how obscured the sky is. In areas where large parts of the sky are out of sight to the receiver, such as beneath a cliff or when surrounded by buildings, the error in the calculated position will be very large. This is also an issue in areas where the receiver antenna is beneath a thick forest canopy when the signal can be lost altogether.

Potential Error: 0-100 m   Common Error: 5 m

Satellite Geometry: GPS receivers are only accurate when the quality of the data they receive is of a high standard. When the satellites being used for determining position are clustered together or all within one hemisphere the quality of the data will be poor. For accurate positions GPS receivers require satellite coverage from across the sky.

Potential Error: 0-20 m   Common Error: 5 m

Multipath Error: When the receiver calculates the length of time the signal has taken to travel from the satellite to use in determining distance to the satellite it assumes the signal has taken the shortest path i.e., a geometric straight line. In actuality the signal may have bounced off a surface before reaching the receiver and the travel time could be slightly longer because of this. In these occasions the receiver will overestimate the distance to the satellite.

Potential Error: 0-10 m Common Error: <5 m

Ellipsoid: As discussed below, GPS receivers are designed to function in the WGS84 ellipsoid. Any other datum displayed by the receiver is a product of applying the Molodensky formula to the data. This gives a good approximation to the resultant datum but is not perfect.

Potential Error: ~5-10 m Common Error: 5 m

6.5.3 Atmospheric effects

Ionospheric Effects: All GPS signals travel through the charged plasma of the ionosphere. This can cause the signal to be attenuated (slowed down). Any changes in the signal involve changes in the travel time and thus affect calculated positions similar to multipath errors.

Potential Error: 2-30 m Common Error: 5-10 m

Tropospheric Effects: The water particles in the upper atmosphere cause very slight changes to the signal. These are very small but can affect minor changes.

Potential Error: 0-5 m Common Error: <2 m

Many of these errors can be quite easily compensated for and the section below will deal with the correct use of GPS receivers. Best practise with receivers involves using them in areas where their view of the sky is unobstructed, buildings or other corner reflectors are not present and that data is only recorded when the satellite geometry is of an acceptably high standard.

Table 6-2 Magnitude of errors in calculated GPS position.

Error Type	Compensation	Typical error (m)	Max. error (m)
Atmospheric / Ionospheric	WAAS or Differential	5	30
Receiver clock error	None (Differential)	5	10,000
Percentage full sky visible	Averaging	5*	100
Satellite geometry	Averaging	5*	20
Multipath	None (Covariance)	<5	10
Ellipsoid	Manual / use WGS84	~5	10
(Selective Availability)	*None Applicable*	*20-60 (23 RMS)*	*100*

** Percent sky visible will affect geometry and these two errors are not necessarily cumulative.*

The final error in Table 6-2, Selective Availability, has not been discussed and is a historical feature of only limited significance. When the GPS project was announced, as well as encrypting the PPS signal the US DoD applied a signal scrambling code to the SPS signal. This was to eliminate foreign powers using the signal to plan and orchestrate military attacks and to safeguard the US from precision attacks. Selective Availability was designed to ensure civilian grade GPS receivers were never more than accurate to 100 m. In May 2000, US President Bill Clinton signed a decree ending SA. However, the US DoD reserves the right to reactivate the system in times of war. This can be done over specific regions leaving other areas unaffected.

6.5.4 Real world accuracy

Most manufacturers quote receiver accuracy as <15 m. The total effect of the typical errors shown in Table 6-2 is closer to 30 m but repeated tests show that under good conditions the accuracy of a standard civilian set using the SPS signal on L1 should be considerably better than this.

The distance of any given point from the actual location is called the dilution of precision. Sometimes data is quoted in circular error probability (CEP). The CEP describes a circle of a radius containing 50% of the data. A typical GPS might have a CEP of 3 metres. More commonly the 2σ dilution of precision is quoted. The symbol σ is equal to the standard deviation of the data set. This is equal to the square root of the sum of the values of a data set minus the average value of the data squared divided by the number of points in the series. The 2σ dilution of a 12 channel GPS receiver is often assumed to be a circle with a radius of ± 7.5 m (a 15 m diameter circle around a point's true location).

To see the significance of the standard deviation of a dataset, imagine plotting all values in a dataset against the frequency with which each value occurs. Figure 6-11 gives an example, using the error found in many GPS readings taken at a point. In this case, the mean value is 0, and the other values are distributed to either side in a characteristic bell-shaped curve. This shape indicates a 'normal' data distribution (also called Gaussian); many statistical measures and tests assume that data are normally distributed in this way. In the case of the standard deviation, if data are distributed normally, then we can say that:

- 68.2% of all values lie within $\pm 1\ \sigma$ of the mean
- 95.4% of all values lie within $\pm 2\ \sigma$ of the mean
- 99.7% of all values lie within $\pm 3\ \sigma$ of the mean

So, for example, when GPS are described as having 7.5 m accuracy to 2σ this means that 95.4% of readings are within error margin.

In many cases data will not be distributed normally. For example, if the data in the distribution is skewed to one side, has more than one peak, or if the number of values is relatively small then the curve may not be Gaussian. In such cases, the standard deviation must be used more conservatively, e.g. 2σ contain 75% of the data, rather than 95.4%, or may not be applicable at all. The nature of the distribution can be assessed graphically, as in Figure 6-11, although statistics text books give more rigorous tests.

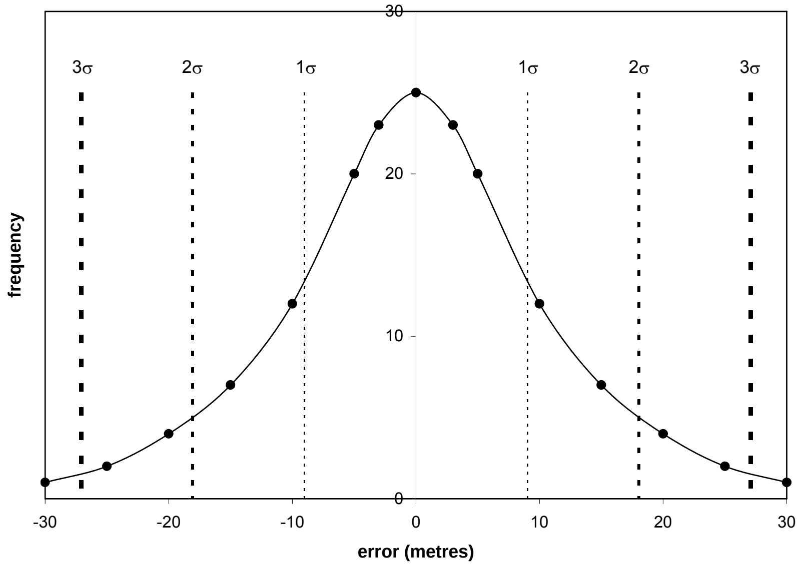
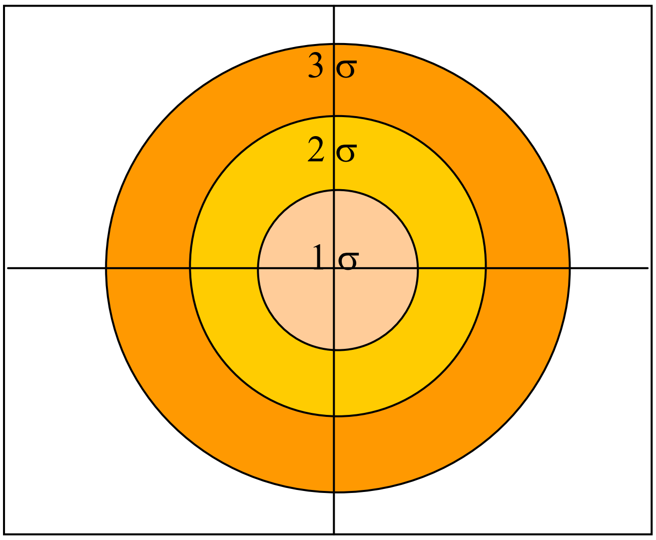


Figure 6-11 Distribution of errors in a set of GPS readings taken at one point. The line drawn through the points has a bell-shape, indicating a Gaussian or 'normal' distribution of the data. The vertical bars show the number of GPS readings lying within different standard deviations of the mean: 1σ (68.2% of all values), 2σ (95.4%) and 3σ (99.7%).

The scatter distribution of a GPS receiver is shown below in Figure 6-12.



The circles in the diagram are concentrically centred about the intersection of two lines. If this intersection is the actual location of a point, then a Gaussian distribution says that:

68.2% of values lie within the inner circle

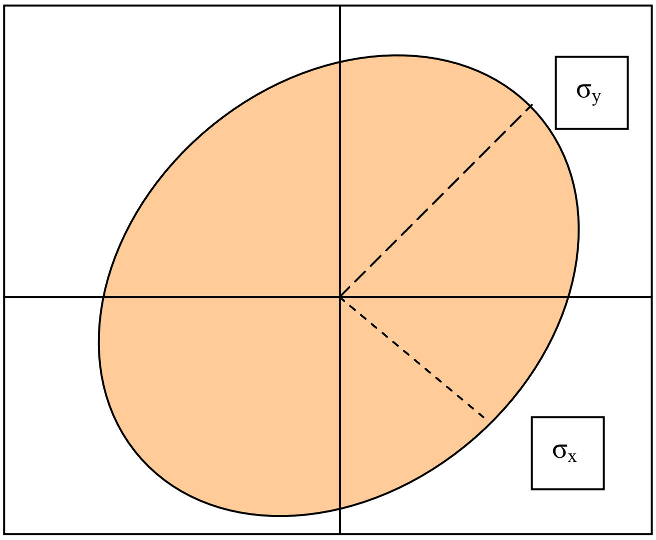
95.4% of values lie within the middle circle

99.7% of values lie within the outer circle

The size of these circles allows a determination of GPS accuracy. The design of the Standard Positioning System was for the 2σ circle to have a diameter of 100m but in reality the system is significantly better.

Figure 6-12 Error circles for SPS signals.

A 2σ circle means 95.4% of the readings are in the circle as shown above. Extensive tests on GPS receivers show that the data actually scatters with a slightly elongated shape. This 'error ellipse' is normally orientated with the semi-major-axis directed northeast - southwest with the following dimensions:



Error ellipses are different for the different types of GPS receivers. The statistics below are for generic receivers for standard deviations 1-3 (2σ - 6σ)

Twelve Channel Sets (%, 2σx, 2σy)

68.2% 3 m / 4 m

95.4% 6 m / 8 m

99.7% 9 metres /12 metres

Figure 6-13 Error ellipse for 12 channel GPS.

Standard deviation can be shown in a clearer format using either of the following equations:

$$\sigma = \sqrt{\frac{\sum(\mathrm{x} - \bar{\mathrm{x}})^2}{n}}$$

However,

Standard Deviation is more commonly expressed in the form :

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

For GPS analysis where the data has a positional nature, standard distance is commonly used instead of standard deviation. The equation looks the same but is a description of the spread of points around the mean centre. In standard distance, *x* is the *x* co-ordinate (*longitude*) of any individual point and *x*-bar is the mean centre of the distribution. An identical equation for *y* (latitude) is also required.

$$\text{Standard Distance} = \sqrt{\frac{\sum(x - \bar{x})^2}{\mathrm{n}}} \text{ where } x \text{ is x co-ordinate or longitude.}$$

Elevation is a more complicated variable because it requires more satellites and because the ellipsoids are more difficult to calculate. As such elevation data will be considered separately later on but is normally quoted with a dilution of precision 1.5 to 2 times that of the *x, y* value.

6.6 Correct GPS handling

As discussed above GPS receivers are only accurate when used correctly. Improper use or failure to consider environmental factors such as canopy cover or urbanisation will result in severely degraded data. This section discusses how the GPS should be used in the field and important considerations when using them on expeditions.

To calculate its position, a receiver needs to know which satellites it is talking to and where they are. The time taken to do this can vary considerably and is an important consideration for fieldwork. The time taken from switching the receiver on to the time of the first fix is referred to as start-up time. This start-up time is controlled by the capability of the receiver (how many satellites it can 'talk' to, commonly 8 or 12) and the accuracy of the GPS almanac. The GPS almanac is a digital record kept by the receiver of where satellites are in the sky and what satellites it should be receiving data from. It is essential the receiver knows the satellites it is using in order for it to accurately calculate its position. Determining this from fresh takes a significant length of time (up to 12 minutes for a complete download from a satellite), so the GPS keeps a record of the almanac data it has previously collected to speed up the start-up time. Almanac data is not very precise; it is accurate for geographic regions up to ~300 km diameter and for up to two months. If the receiver moves further than a hundred kilometres from the last time it was operational or if it is left inactive for a period of longer than two months then the almanac will not be valid. In these cases the receiver will have to determine this data from fresh. This is known as a 'cold start' and can take up to 12 minutes. If the Almanac is accurate the GPS receiver can initialise a 'warm start'.

The GPS receiver collects data for its almanac from information transmitted by the satellites. Each satellite transmits its own locational data. This data is referred to as ephemeris data and is very precise. The ephemeris information is broadcast every 30 seconds but it can take up to 24 transmissions to completely describe the orbit in detail for the GPS. The difference between warm and cold starts is therefore quite significant. Before a GPS can use a satellite, it must have a complete 'packet' of ephemeris information. Any glitches while acquiring this information will cause the GPS receiver to start over again for that satellite. This means a satellite that is in a difficult position or affected by multipath or similar errors may take a very long time to send an uninterrupted stream of information for the receiver to use. This is why some satellites remain inaccessible even though the GPS can see that they are in the sky. The time taken for modern sets to acquire a lock is usually considerably less than for older models. These time differences are summarised in Table 6-3.

Table 6-3 Realistic acquisition times.

Receiver	Almanac is correct ('warm start')	Almanac is incorrect ('cold start')
8 channel receiver	6 minutes	12 minutes
12 channel receiver	1 minute	4 minutes

After acquisition, the data from the receiver will fluctuate for a period of time. This fluctuation will generally continue for 5 minutes after the GPS is first activated and will then settle to give a better fix more inline with the data quoted in Section 6.5.2. The satellites being tracked, and the quality of information coming from them is commonly shown as a sky view. This view is important in determining both the quality of data and whether the receiver has acquired a satellite lock.

When a GPS receiver has a satellite fix it can be used in one of two ways. Location points can be recorded at a user's discretion by clicking a button on the receiver (usually Mark or similar) or automatically at given time or distance intervals. These two methods are referred to as Waypoints and Trackpoints respectively.

Trackpoints: The behaviour of a receiver when recording a trackpoint is different for each model. Typically the receiver will store a co-ordinate value and location ID either at time intervals or when it detects a significant change in user direction. This option is often configurable on the more expensive receivers and this is a useful option because it avoids filling the memory with unnecessary points. Less commonly the receiver will record altitude with the trackpoints. The receiver will often not quote how many points it can store but with increasing memory capabilities over 5000 points is not uncommon. Once this limit has been reached the receiver will begin to write over its initial track (often with no warning). To avoid this, the receiver can store the track in a more stable form as a tracklog. Most receivers can store around 10 tracks as tracklogs but these are simplified descriptions of the original data used to save memory. They are not as accurate as the original data and may only use 30 points from the original ~1000 points. Modern Magellan models such as the Sportrack and Meridian ranges overcome this and have a detailed tracklog option. This can be used to create a fully detailed tracklog referred to as a 'backtrack'. The major disadvantage with a trackpoint is that the receiver will take a reading regardless of the satellite constellation at the time. If the receiver is set up to take a reading every minute and the receiver is moved under dense canopy for a period of time, then the readings it records may be very inaccurate (>50 m). Trackpoints can be useful to a team because they allow large amounts of data to be collected very quickly. If the team is visiting an area where the road network is poorly mapped or not known, then by driving the roads for a day the GPS can accurately map the road network. Even with the problems associated with trackpoints such as taking readings in non-ideal conditions, the map would still be c. 1:5,000 to 1:10,000 scale.

Waypoints: Waypoints are recordings of a location's co-ordinates, commonly with user descriptions and elevation data. Waypoints can usually use an associated symbol for displaying on map views. This is sometimes downloaded as label that can be re-associated with a graphic at a later time. Waypoints can be given a text string to accompany them of between 8 and 12 characters that can be used to annotate the waypoint.

Waypoints are only recorded with user interaction and are not automatic. This is a severe disadvantage, as user interaction with the unit may be difficult due to local conditions or difficult terrain. This often means relying on trackpoints is safer and easier. User interaction does, however, mean that only accurate positions are recorded. The major disadvantage of trackpoints is that they record data whether or not the GPS constellation is good. Waypoints will always be of a higher standard. Waypoints have many other advantages and disadvantages. Receivers can commonly store 500 waypoints or more, with newer sets with greater memory capabilities becoming available. However, the storage capacity of waypoints will only be 1/10 that of the trackpoint storage. The collected data can be used individually or can be collated into a route. Routes are selected waypoints that define a path to follow. Around 50 waypoints can be selectively chosen to add to a route and the GPS can instruct the user on how to travel between each one with direction and distance supplied.

It is important to be able to associate a waypoint location with any field description. This should involve noting the waypoint in the field log, however, the GPS receiver can also be used to take notes. Though the waypoint description in a GPS is often limited to 6 (ETREX) 8 (Magellan 310) or 12 characters, new Magellan models such as the Sportrack Pro offer the ability to add 'messages' to the waypoint locations. The Sportrack can store 500 waypoints at an 8 character limit but 204 of these can have an associated 30 character message.

Trackpoints and waypoints have distinct advantages and disadvantages. The most significant difference is that trackpoints are automated whereas waypoints require user interaction with the set. User interaction can be difficult in certain terrains where handling a set can be hazardous. However, user interaction means that only data of a high standard is acquired and the overall data set is of a high quality. Trackpoints record points automatically and give no consideration to the quality of the data. This means that points can be recorded at times of poor satellite geometry. When studying these points later, they may have a very large error but because this error is unknown interpretations may be made based on data points that are not correct.

Table 6-4 Relative advantages and disadvantages of waypoints and trackpoints.

	TrackPoints	WayPoints
Advantages	No interaction with the set is necessary Large numbers can be stored to create accurate routes.	Waypoints can be recorded when accuracy is high. Additional data such as a name and comment can be put with the point.
Disadvantages	Although you can set a time or distance interval for the tracklog it automatically records position regardless of satellite geometry and visibility.	The set must be handled repeatedly. Only a small number of points are likely to be recorded making routes less detailed.

As with any part of expedition kit GPS receivers need to be treated with respect and due care. GPS should never be the only form of navigation tool taken into the field but if treated correctly they can be very reliable complimentary tools. GPS receivers require battery power and this is the most important factor that needs to be considered in remote areas. In addition, most GPS receivers have LCD screens that are susceptible to cold conditions. Temperatures below 0°C may irreparably damage a receiver's screen. How to combat these problems is discussed in the Field Technologies chapter. A GPS should never be left exposed at night during cold conditions and should always be packed into a rucksack or similar, to protect its screen.

6.7 Assessing data quality

GPS is a valuable tool for expeditions but the ease with which buttons can be clicked and data collected can lead to poor scientific practise. The team should always be conscious of what the GPS is recording and whether the data is of a high or low standard.

When using waypoints, the user has the opportunity to select readings of a high quality. The quality of data can be very important to an expedition. If a position is to be recorded in the field for future visits then poor quality data may make the location ambiguous. The difference between a good fix (±5 m) and a poor fix (±20 m) may place the waypoint on the wrong side of a river or crevasse making relocating the point or making interpretations very difficult. The GPS receiver will show the quality of data within the sky view as an estimated positional error (EPE).

The EPE is based on the satellite geometry and should not be considered an accurate gauge of the actual error. The EPE is a good indication of data quality and care should be taken to record waypoints only when it is low and data is of a high standard. Ideally, the reported position and its associated estimated error should also be recorded in a notebook or handheld PDA computer as a backup precaution. Figure 6-14 below shows how the geometry of the satellites affects the EPE. As skilled users of GPS the team members should be able to approximate the EPE just from the satellite geometry.

The receiver display to the left has very poor satellite constellation geometry and correspondingly, a very high estimated positional error.

Even though the receiver on the right has fewer satellites, they are positioned better and therefore the error would be correspondingly lower.

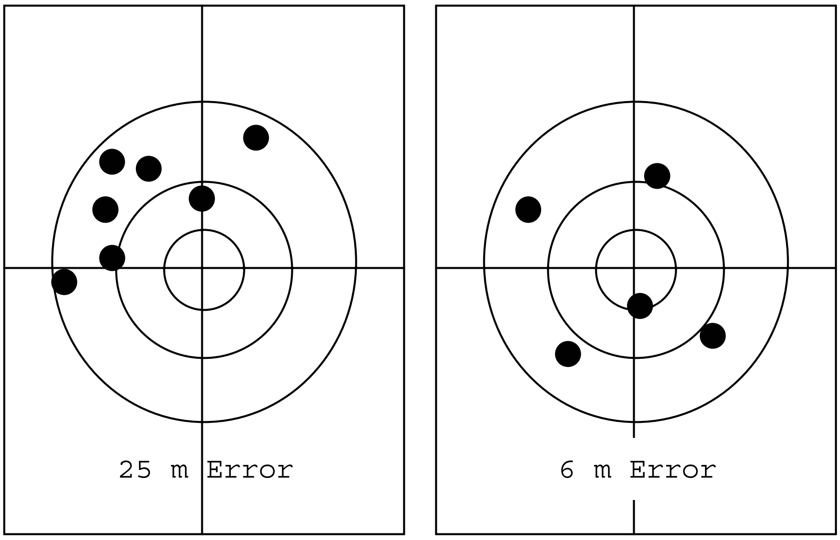


Figure 6-14 The difference between good and bad constellations.

When using GPS it is important to check that the receiver is giving the data that is required. The receiver has to be set to the correct projection and datum and it is always important to be conscious of the data it is producing. The nature of the GPS signal means that it is possible for any one reading to be wrong. As can be seen in Plate 4, some readings can be a long way out due to the nature of the errors associated with GPS. As a user in the field it is important to be aware of the previous data to ensure that a reading does not diverge too much from the expected value. Obviously, detecting small fluctuations is very difficult but it can be a good warning that something has gone wrong with your settings or the GPS system as a whole.

On 1 January 2004 one of the older satellites PRN 23 experienced a system failure error. Most satellites operate with triple redundancy to mitigate the problem of a system failure but some of the older satellites are now operating without redundancy. This lack of redundancy means that errors can occur in the system. As the GPS IIR satellites replace them this will not be an issue because they are both newer and they are configured to use their back up systems as a real time check on the satellite status. This should mean problem satellites are taken off line much more quickly but it is always worth checking readings to make sure they are consistent and that there is no sudden change in the calculated position. In the case of PRN 23 the error was around 200-300 m in Dover Straits and up to 40 km in Scotland. The problem was mitigated within three hours by setting the satellite to 'unhealthy' but any results collected at such a time would be useless for scientific work. The satellite was brought back on line by 20 January 2004 running on its backup caesium clock. The SPS and PPS performance standards require ≤ 3 failures per year. This incident was the first major failure since 2001, so for most purposes the GPS signals are incredibly reliable. More detailed reports can be found at the US National Coast Guard Navigation Centre *www.navcen.uscg.gov*.

The receiver should always be used in accordance with the instructions here and in the receiver manual to ensure the best quality of data. The receiver should be held away from the body with the antenna (either in the head of the unit or in the aerial) held towards the sky. This is shown below in Figure 6-15.

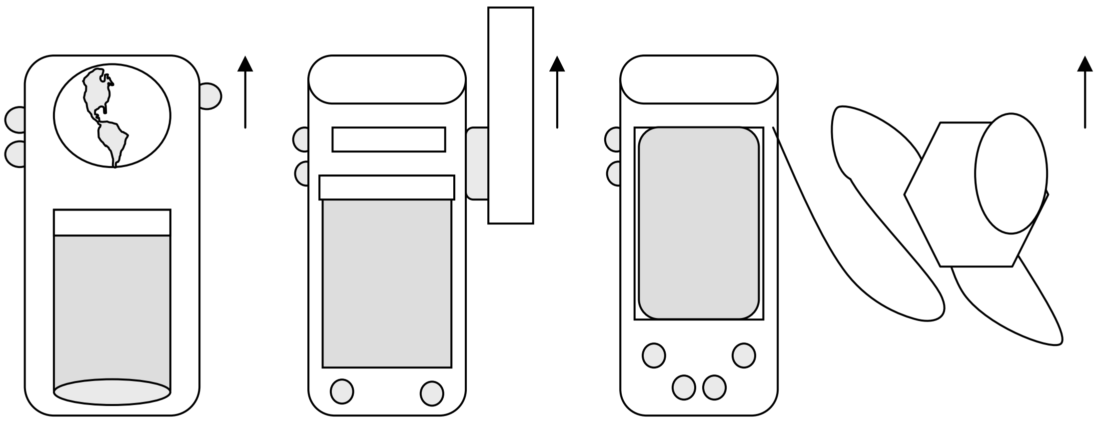


Figure 6-15 Location of GPS antennas. All GPS antennas should be directed towards the sky even when housed internally. Far left GPS is typical of modern units that house the antenna internally in the device head. Middle and right devices have different forms of external antenna.

While discussing antennas and their positioning, it is worthwhile mentioning the specific type of antenna in the unit. There are essentially two different types of antenna used in common GPS models. These are the patch antenna and the quad helix antenna. Though there are a number of reports and publications stressing the advantages of quad helix over patch in areas of weak signal strength, the actual differences are negligible. Many newer receivers such as the Magellan Meridian and Garmin 76 use quad helix as opposed to the older Garmins and Magellans that use a patch antenna. There is a slight performance gain with these newer units under canopy cover but not enough in itself to warrant purchasing new units. Attaching an external antenna is a better solution than using either forms of internal antenna.

The most sensitive receivers available use the SIRF Star III hardware. This is a combination of sensitive patch antenna combined with additional signal processing firmware located inside the GPS. The SIRF Star III can lock onto 20 satellites at one time and can reprocess signals usually discarded by receivers as being too noisy or too weak. SIRF Star III can reprocess these and can work indoors. The receiver has only just come onto the market (mid 2005) and the practical 'real-world' results with the set are still not fully known. The improved sensitivity is believed to offer improved accuracy. Whereas normal 12 channel sets usually have an error of around 15 m, 20 channel Star III sets claim an accuracy of between 5 – 25 cm CEP.

Most importantly when assessing data quality is to make sure the GPS is being used in an appropriate manner. Even the most sensitive GPS will record poor data when used incorrectly. The constellation of satellites should also be checked as described above to make sure the data is being collected at the appropriate time. Using a GPS blindly without any appreciation for the values recorded is poor practice and should be avoided in all but extreme cases.

6.8 How GPS calculates and stores positional data

When assessing data quality it is important to understand how the GPS is arriving at its results. The NAVSTAR system was designed to work with a mathematical model of the Earth. This model was the shape of an ellipsoid (a three dimensional ellipse flattened at the poles and elongated at the equator) and the GPS calculates positions relative to this model. The GRS80 (Geodetic Reference System 1980) ellipsoid is the model that was taken for use with NAVSTAR. This is a good approximation of the shape of the Earth. It was modified slightly to be the World Geodetic Survey 1984 (WGS84) and this is the common reference system that all receivers use. This best fit of the whole Earth is not accurate for all areas and most countries use their own ellipsoid. In Britain the ellipsoid used is the Airy 1830 ellipsoid. When displaying data the GPS will by default display co-ordinate information according to the WGS84 ellipsoid. This is acceptable for latitude and longitude but would often give unexpected and erroneous height data. The height of topography above or below a hypothetical ellipsoid is often of limited use and a more conventional description is often required. Height is most commonly expressed as elevation above mean sea level. Mean sea level can be approximated by a geoid. As described in Chapter 2 a geoid is a model of the earth defined as a surface where the lines of gravitational force are perpendicular. Mean sea level itself is not constant across the globe; it can alter by as much as 2 m, depending on where it is measured. Heights in Britain are measured relative to the tide gauge at Newlyn, Cornwall. Even the best geoid available will still not tally with every country's maps and heights, because there is often a discrepancy in the zero altitude used. The geoid and mean sea level are commonly very close to one another, so heights against the geoid are an acceptable measure of heights against mean sea level. These two measurements do not diverge by large amounts (up to a maximum of 2 m but most commonly sub-metre) though they can both diverge by up to +85 m or –102 m against the WGS84 ellipsoid.

The GPS receiver always measures heights relative to the WGS84 ellipsoid. It can apply a transformation to get to a better fitting ellipsoid for a given area, but always works internally to WGS84. When working outside of WGS84 the GPS needs to know the conversion factors defined by the variables δX, δY and δZ. These will be built in for most typical datums so all the user needs to do is select the correct setting and the map and GPS will correlate almost perfectly. Section 11.4 describes how this can be done when mapping in an area with maps that are not compatible with the GPS. These can be used for calculating better approximations of the height at a given area. Some models also allow correct height to be input at start up to aid in calculating elevation.

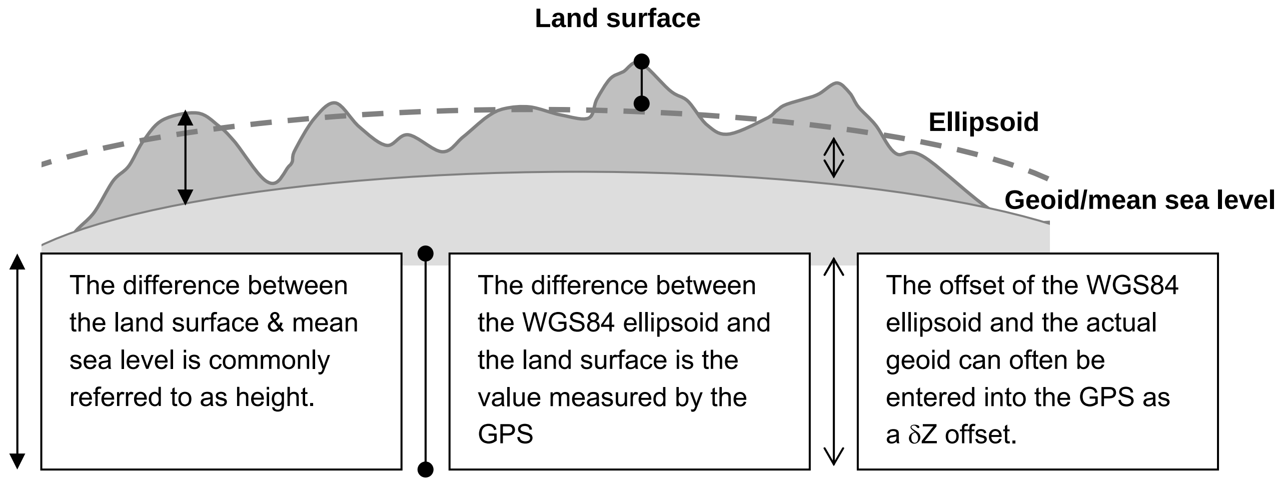


Figure 6-16 Offsets between geoid, ellipsoidal and land surface.

The height above mean sea level is called orthometric height. To calculate orthometric height the GPS requires an elevation above the WGS84 ellipsoid and a knowledge of how much the WGS84 ellipsoid and the geoid differ at that location. Unlike the WGS84 ellipsoid, which is a comparatively simple shape, the WGS84 geoid is very complicated requiring many megabytes of data to store it. This is beyond most GPS receivers so a crude approximation of the WGS84 geoid is included with spot heights at a number of locations in a look-up-table. This method is how all modern GPS receivers' measure height. Though height is described as metres above sea level, it is in fact metres above the low-resolution geoid look-up-table that approximates sea level. A true height above sea-level is therefore very difficult for the GPS receiver to calculate. To gain accurate height the receiver needs the standard constellation of satellites (usually 4) plus an additional satellite to fix elevation. Commonly, this means a good view of five satellites. Because of this, the error in elevation is generally greater than the error in X or Y by at least 1.5-2 times. Realistically manufacturers quote an error of at least 25 m in Z. The conversion from the WGS84 ellipsoid to the WGS84 geoid downgrades the accuracy of the height, but this is generally of limited concern because the inherent errors in generating height are so large.

6.9 NMEA sentences and stored information

NMEA is a standard format that can be used to download information from a GPS unit in real-time (NMEA = National Marine Electronics Association, which defined the standard). Almost all receivers generate NMEA information called sentences. These are a middle layer of information that is less processed than the onscreen data but more refined than the internal GPS calculations. This hierarchy of layers is shown schematically in Figure 6-17.

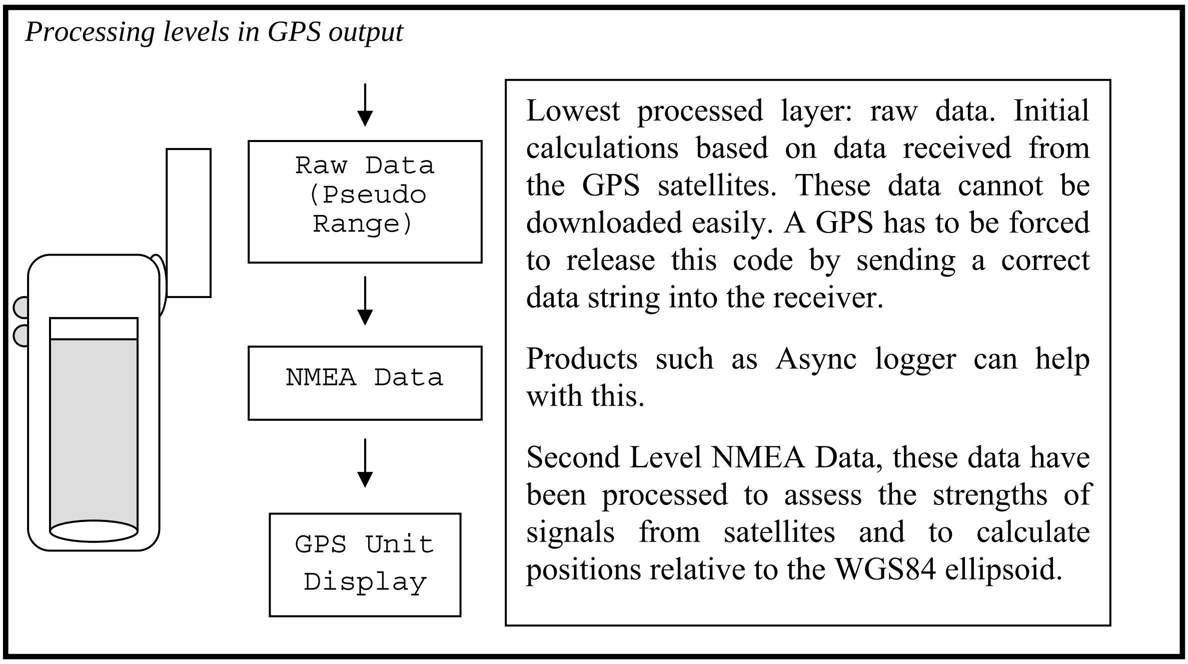


Figure 6-17 Processing tiers in the GPS positioning architecture.

If a GPS uses NMEA, a sentence is available that shows the current ellipsoid and how it differs from the WGS84 ellipsoid. This can be of vital importance if elevation is required as seen in Figure 6-16. Downloading NMEA data requires a computer to be connected to the receiver. The data streams as ASCII information at a standard bit rate as discussed below. Each NMEA sentence is prefixed with a $ symbol and a 4-5 letter code denoting the information contained within the sentence and a comma separates each value. If a value is unavailable it is left blank and a comma prefixes the next variable. NMEA data is very valuable because it can be used as a step back from the standard GPS interface as shown in the figure over the page. In standard WGS-84 mode, the corrections and transformations the unit would normally do to display the data have not been done. The data is therefore provided in a 'raw' or 'native' state.

Downloading NMEA data is comparatively easy; it requires the GPS to be connected to a computer with its interface set to NMEA data. The following example shows how to do this using a typical Garmin ETREX. On the ETREX models go to the Menu page ⇒ Setup ⇒ Interface set the input output (I/O) type to NMEA and the baud rate to 4800 (4800 is the default NMEA standard). A software interface with the RS232 connection must then be established on the computer. The computer interface must be set to recognise 8 bit data at the correct baud rate (4800), with no parity and 1 byte checksum stop values. This can be achieved comparatively easily on PCs using the MSComm control. More recently MSComm.ocx, an ActiveX component, has become available and can be easily distributed amongst PCs for this purpose.

Most GPS manufacturers add additional sentences to the NMEA output. These are prefixed differently. Garmin use a 'P' for Prefix. These sentences give additional information applicable to the manufacturer's model. One of the most important NMEA sentences is the GGA sentence. This sentence is examined in Figure 6-18.

One of the most important sentences is GGA (sometimes) GPGGA. This sentence describes the X,Y and Z co-ordinates of the calculated fix. It is also very useful because it shows the difference in the height values. An example sentence is shown below.

Sentence Structure
Time
$GPGGA,**172704**,5126.7759,N,00020.6673,W,1,05,1.3,13.8,M,47.4,M,,*67
Northing
$GPGGA,172704,**5126.7759,N**,00020.6673,W,1,05,1.3,13.8,M,47.4,M,,*67
Easting
$GPGGA,172704,5126.7759,N,**00020.6673,W**,1,05,1.3,13.8,M,47.4,M,,*67
Fix Type
$GPGGA,172704,5126.7759,N,00020.6673,W,**1**,05,1.3,13.8,M,47.4,M,,*67
Number of SVs tracked
$GPGGA,172704,5126.7759,N,00020.6673,W,1,**05**,1.3,13.8,M,47.4,M,,*67
Approximate dilution of precision (horizontal)
$GPGGA,172704,5126.7759,N,00020.6673,W,1,05,**1.3**,13.8,M,47.4,M,,*67
Height above sea level (geoid)
$GPGGA,172704,5126.7759,N,00020.6673,W,1,05,1.3,13.8,M,**47.4**,M,,*67
Orthometric correction
$GPGGA,172704,5126.7759,N,00020.6673,W,1,05,1.3,13.8,M,47.4,M,,***67**
End of sentence checksum.

Figure 6-18 Analysis of an NMEA sentence.

The data in the sentence in Figure 6-18 shows the information output from the GPS in NMEA mode during a test using OSGB, British Grid. Changing the projection and co-ordinate system has no effect on the output data, as can be seen in the example below, using UTM degrees, minutes and seconds:

British Grid: $GPGGA,172704,5126.7759,N,00020.6673,W,1,05,1.3,13.8,M,47.4,M,,*67

OSGB: $GPGGA,172706,5126.7759,N,00020.6673,W,1,05,1.3,13.9,M,47.4,M,,*64

The GPS makes all calculations relative to the WGS84 ellipsoid, but stores and outputs height relative to the mean sea level, as calculated by the receiver. NMEA data is vital for conducting user-defined transformations and for ascertaining accurate height information from the GPS. NMEA data records both the orthometric and ellipsoid height information. By taking the raw ellipsoidal height and correcting it with a known MSL-Ellipsoid offset (this can be found on any good map) yields better height data. Though NMEA data is very useful, it cannot be used alone to give cheap differential corrections. Post processing NMEA is not possible because the data from each SV is not stored in the sentence. For individual SV data, a pseudo range output is required. Pseudo range is the name given to the pre-processed data and is discussed more fully in Section 6.10.2.

NMEA data can be very useful for making some manual corrections to a GPS. Though GPS receivers are capable of applying an equation to warp their results to any system co-ordinate system, they do this by using the Molodensky formula. This technique is a simplified three co-ordinate shift that has at best an accuracy of 5 m. The current advice from the Hydrographic Office and the Admiralty (*Admiralty List of Radio Signals Vol. 8*) is to keep GPS receivers in WGS84 mode and make any shifts manually. This is significant for height data, as the only way an accurate height measurement can be obtained is by stripping out NMEA data to a PC and applying manual calculations. Accurate height calculations require a measurement against the WGS84 ellipsoid and then a calculation of offset between the ellipsoid mean sea level. It must be noted that the output NMEA data are relative to the geoid model selected in the receiver. A GPS receiver will always store data in WGS84 but the NMEA output will have a co-ordinate transformation. As such, the receiver should always be set back to WGS84 before NMEA data logging. If a simple and convenient method of communicating with a GPS is required without the need for complex data filtering, then the communications programme HyperTerminal, supplied as part of the Microsoft Windows operating system, should suffice (see Appendix 2).

6.10 Understanding precision and improving accuracy

GPS accuracy is better than 15 m (around 6 m by 8 m). GPS receivers are often called upon to report data more precisely than this. The UTM system is a measurement of metres from a datum in 1 m intervals. The common degrees minutes and seconds system is also often more accurate than the resolution of the GPS receiver. The Earth is approximately 6,378,200 m in radius at the equator giving an equatorial circumference of 40,075,413 m. If this is divided into degrees minutes and seconds each degree is 111,320.6 m, each minute is 1,855.34 m and each second is 30.92 m. When divided decimally each fraction of a second is 3.1 m. This means that the data is recorded to 3 m intervals, which is still more accurate than the GPS is actually capable of. This level of precision (how many decimal points quoted) is often incorrect and caution should be taken when using seemingly accurate information.

6.10.1 Improving accuracy (standard methods)

There are a number of techniques for reducing dilution and improving data quality. The signals sent from the NAVSTAR satellites are accurate to the centimetre scale. This accuracy is downgraded by the various factors discussed in Section 6.5.1. In October 2001 the US Military released their first indication of the post-SA quality of the data. Their findings indicate that though the signal is now accurate to at least 13 m, local errors will often reduce this. If a GPS receiver can be used in an area of unobstructed sky the position should be as accurate as the diagrams shown in Section 6.5.4. If a reading has to be taken in an area of obscured sky, then a GPS can be combined with more traditional surveying techniques. If a point of clear sky can be found, then the expedition can mark this point using the GPS as normal. The co-ordinates of the desired point can then be calculated by surveying a line back from the known point to the required location. To do this the expedition would have to use a compass to measure the angle from the known point to the desired location then accurately measure the distance between them. The exact method of transforming this information to the GIS will depend on the co-ordinate system used. This process is shown in Plate 5.

Combining basic surveying and GPS work can be an effective method for surveying points and entering them back in the GIS later. It is, however, a very time consuming method and the accuracy reported on the GPS should be considered to see if the additional work is required.

As well as using the GPS in areas of unobstructed sky, there are additional methods for improving the SPS accuracy without the need for the PPS information. The most common methods for improving data are listed below and discussed in turn in the following section.

Methods for Improving GPS (Standard Positioning Service – SPS)	
Averaging	A simple method using standard receivers
Differential Corrections	A more involved process using specialised equipment
WAAS	A cheaper more accessible version of differential
Carrier Wave	A more complex version of differential signals

Averaging: When a GPS records a location it will lie a certain distance from the true location. This 'dilution of precision' is random but multiple readings will plot within concentric ellipses forming Gaussian distributions along each axis as shown above in Section 6.5.4. The data plotted in Plate 4 shows real world examples of this. In the left hand plot there are outliers of information shaded dark red with inner lighter areas. The dark areas show low concentrations of data and the light areas show high concentrations. Because this data forms a Gaussian distribution it is probable that the actual location lies closer to one of these light areas. This is shown on the plot on the right. The error in the GPS data is plotted along the abscissas and the frequency of this measurement is shown on the ordinate. The vertical line shows zero error. It is clear that the most frequently reported co-ordinate is very close to the line of zero error.

The most commonly occurring value in a distribution is referred to as the modal. By taking the modal value of the distribution, a point's most likely true location should be significantly more accurate. This requires a substantial amount of data to be collected at each point for a significant statistical analysis to take place. Table 6-5 compares an averaged GPS reading, computed from taking readings every 30 seconds for 10 minutes after successfully acquiring 4 satellites, to a standard single recorded waypoint value.

Table 6-5 Effects of averaging on GPS accuracy.

GPS Type	Single Plot Error in X	Single Plot Error in Y
8 Channel	**20 m**	**40 m**
12 Channel	**6 m**	**8 m**
GPS Type	**Averaged Error in X**	**Averaged Error in Y**
8 Channel	3 m	14 m
12 Channel	3 m	4 m

Averaging improves the accuracy considerably but requires more time to be spent at a site to record data than standard single point data collections.

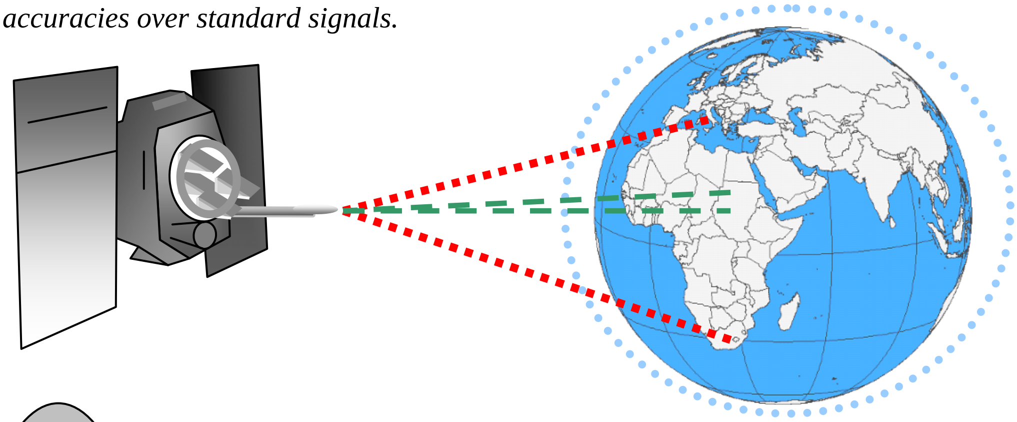
Differential GPS (DGPS): The signal from a satellite passes through a section of the ionosphere that attenuates it and introduces an error of between 2 and 30 m. The actual effect of the atmosphere is to increase the travel time of the signal. For receivers located in geographically small areas, the signal to each receiver will have passed through essentially the same part of the ionosphere and be attenuated in the same way. If a point whose location is known very accurately can be used for a GPS base-station, then the travel time to any satellite should be known. Any deviations from this will be because of errors generated by the ionosphere. These errors can then be transmitted to properly equipped GPS receivers in the local area and these errors can be removed.

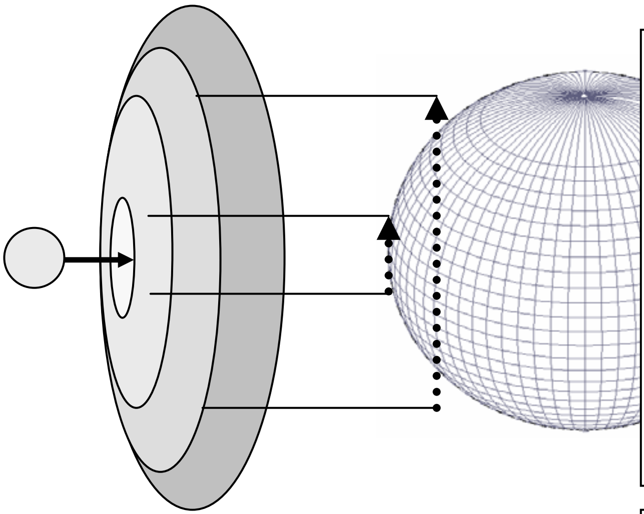
The station records all the travel delays for all the satellites and re-transmits them. The receiver in the field can then make corrections for the delay. This reduces the GPS error to between <1-5 m, depending on the sophistication of the hardware used. Differential corrections can be conducted in real-time, i.e. the correction signals from the station are used by the receiver in determining its location. Alternatively the receiver data can be post processed. This means logging all the GPS data then downloading the timing errors from a station and applying corrections later.

The final method sometimes cited is inverted differential where the receiver takes a standard reading and transmits it to the station that corrects the data without sending information back to the receiver. Many standard GPS units are differential ready but adding the additional hardware is more expensive than using the GPS in its standard mode. Differential equipment ranges in price from £500 - £40,000 and is dependent on the accurately positioned 'base stations'. The differential kit can also be cumbersome in the field and politically sensitive, as some countries may object to the use of such high accuracy surveying equipment being used in the field. However, the accuracy and reliability of DGPS is substantially better than any other form of GPS technique. The main principles of differential GPS are shown in Figure 6-19.

Wide Area Augmentation System: WAAS is the American name for a system similar to differential GPS with corrections based on deviations of GPS estimations compared to known locations. Geo-stationary communication satellites are used to transmit this correctional information to GPS receivers. The European system is known as EGNOS and the Japanese system currently under development is called MSAS. An EGNOS prototype system (ESTN) is currently broadcasting signals as part of its testing phase. Sometimes the satellite will stamp them with a '*Do not use*' signal and GPS receivers will ignore these signals when making calculations. Many modern 12 channel sets include the capability to receive WAAS corrections and they dedicate one of their channels to receiving these. WAAS is of particular interest to the US Federal Aviation Administration (FAA) who developed the system to aid in landing planes. The system is designed to offer a guaranteed position of better than 7 m and down to 3 m near airports (see *http://gps.faa.gov/*).

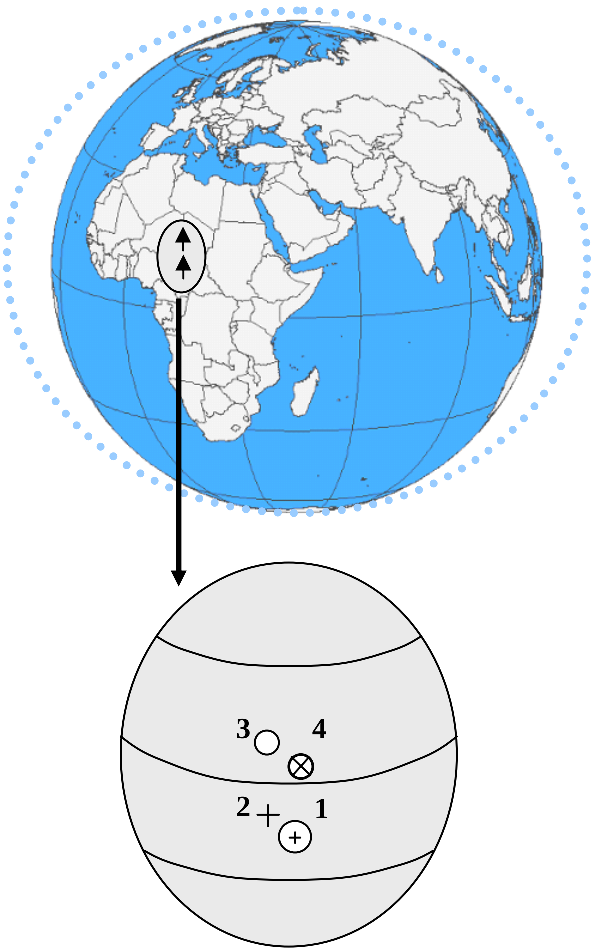
Figure 6-19 Fundamentals of differential GPS. Differential GPS offer significantly improved accuracies over standard signals.





GPS signals are distorted as they pass through the ionosphere. Though the ionosphere is different in density at every point, areas of the ionosphere close to one another will normally have similar electron densities.

The two locations on the ground receiving the dashed lines are receiving data passing through similar points on the ionosphere. The two locations receiving the dotted data are receiving data passing through distant points on the ionosphere. Close points will usually have similar ionospheric errors.



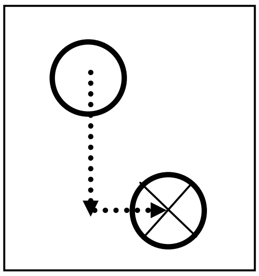
It follows that close areas on the ground (<100-200 km) will have similar dilutions. A differential signal is calculated by using the current error at a point of known location. By applying the inverse of the known error to the recorded position in the field allows a much-improved estimate of the location. The quality of this correction is dependant on the type of equipment and the proximity of a base station. An error of 1-5 ppm (part per million) of the distance of the receiver to the base station should be added to the stated accuracy of the device. I.e. if the nearest base station is 50 km away then the DGPS error = 50/1,000,000 = 0.00005 = 5 cm. This gives an error of approximately 5-25 cm depending on the equipment.

(1) Point of Known Location

(2) Offset of known location

(3) Reported position in the field

(4) Corrected position using the inverse of the offset of the point of known position. (*See right*)



Carrier Wave Post Processing: Surveying grade differential GPS stations use this feature. It is an augmentation that analyses the actual radio signal sent from the satellite and not the information it carries. Because the pseudo random code only emits a 1 or 0 every microsecond it is very difficult for a receiver to transform this into a precise time. Within a microsecond light travels 300 m, which is a very large error. High quality receivers can measure to within 1 or 2% of this but that error is still very large (3-6 m). The carrier wave itself has a frequency of 1,570 MHz, meaning it transmits about a billion times a second. In this length of time light travels less than 30 cm. The carrier wave alone is useless for timing because each wave looks essentially identical but carrier wave processing combines the pseudo random code and the carrier wave to determine a more precise location. Carrier wave processing can get positional resolutions down to 1-3 % of the frequency of the signal (1-2% of 30 cm = 3-6 mm) but this requires significant post processing. The realistic maximum resolution is the 20-30 cm wavelength of the signal, but is commonly around 1 m. Carrier wave post processing requires a laptop or PDA to be connected to the GPS at the time of data collection and requires a lot of processing to determine a location.

For inexpensive use carrier wave processing can sometimes be achieved by using a standard set of output files from the receiver (known as RINEX) but this requires access to unprocessed, poorly-accessible pseudo-range data held within the GPS. This method is explored in more detail in Section 6.10.2. The various techniques and their accuracy levels are summarised in Table 6-6.

Table 6-6 Accuracy improvements for GPS.

Type	Error	Cost
Standard quoted GPS Accuracy	15 m	Free
Typical GPS Accuracy (post SA)	6-8 m	Free
Averaging	~3-4 m	Free
Differential	<1-5 m	£500-£40,000
WAAS	~3-7 m	Free on compatible GPS receivers
Carrier Wave	<1 m	Free or Professionally done (>£500)

6.10.2 Improving accuracy (manual post processing)

As this chapter has stressed on numerous occasions high accuracy GPS data is often no real benefit to an expedition. However, if this accuracy is needed and the party cannot borrow equipment or purchase a several thousand pound DGPS system then there are alternatives.

Many methods for improving GPS accuracy rely on accurately sited base-stations in close proximity to the expedition and or a degree of expense. In remote expeditions on a tight budget these two facets make many of the techniques inaccessible. As we have seen from Section 6.10 a differential GPS can be corrected by using signals from a point of known location. The section on averaging also suggests that if a GPS left for a length of time any point can be surveyed to a high degree of accuracy. Based on these two facts it might seem logical to assume that a GPS could be setup at a base-camp and used to calibrate a roving receiver in the field. If all the readings of the base-camp GPS were recorded for a day then at the end of the day the roving GPS could have all of its readings post-processed to remove the errors at all the waypoints. Unfortunately, this is not possible. Because not even NMEA data (the complete output of all the calculated data performed by a GPS as shown in Section 6.9) contains the details about the SV used when making the calculation you can not know if the calculated error from the base-camp GPS has any relevance to the waypoints recorded by the GPS. Proper differential GPS record the pseudo-range data. This pseudo-range contains the information about the individual satellites and can be used to correct the signals appropriately. If a user tries this without the complete data then the processed information will be meaningless and less accurate than a standard waypoint. If the pseudo-range data can be obtained then there is the possibility of post-processing the data.

Until recently the most common inexpensive software for pulling out pseudo-range data was Async Logger. Async Logger outputs a RINEX (Receiver Independent Exchange) file; this is a data standard used in many applications. RINEX dumps generate large amounts of data and care should be taken if only small amounts of storage space are available. A 5 min dump will create somewhere in the region of 250 Kb of data. Where several GPS are being used in the field these dumps will be creating several Mb files per day. This compares against the few Kb generated by normal Waypoints and track points. The entire Bogda Shan GPS file is about 988 Kb. This was over a period of several weeks using up to five GPS receivers. It is easy to see how RINEX dumps can get out of hand if storage space is a premium.

Async Logger is a difficult product to use that has not been updated for a while. A much easier to use modern equivalent is the Delorme GPSPostPro 2.0 software. This software comes bundled with a roving GPS and a base-station GPS. The base-station can be used to calibrate the roving unit. The only disadvantage of this system is that the base station unit is a self contained box that has no user configuration features on the outside and no I/O functions except Bluetooth. This means that to use this you will need either a PDA with Bluetooth or an expedition laptop similarly configured.

The Delorme system costs around $300 which is very cheap for a differential solution. This system claims sub-meter accuracy and is both very cost effective and easy to use. Both the base-station and the roving receiver are complemented by WAAS and can search several FTP sites for Continuously Operating Reference Stations (CORS) or Orbit and Permanent Array Center (SOPAC) signals in the vicinity of the expedition base-camp. CORS and SOPAC are differential signal stations that post their data onto FTP sites. These can be accessed from the Internet and post-process expedition data. An idea for how this hardware can be used in the field is shown below in Figure 6-20. The advantage of using the equipment shown is that it can all be purchased very cheaply. The laptop can be acquired for £199.74 through SterlingXS. A yellow Garmin EXTREX retails for £84.84 through GPS Warehouse (*www.gpsw.co.uk*). The Delorme differential GPS set including the roving antenna and the calibration set costs $349.95, the external antenna costs $32.95 and additional PDA software costs $39.95. Shipping from the US for the Delorme set is $40.00, import duty comes to £11.00 and VAT comes to £36.74. Using an exchange rate of 1.91 (current at time of writing) gives a total cost of £279.07 for the GPS equipment. The total amount for the equipment is £563.65 for sub metre accuracy and a field hardened laptop.

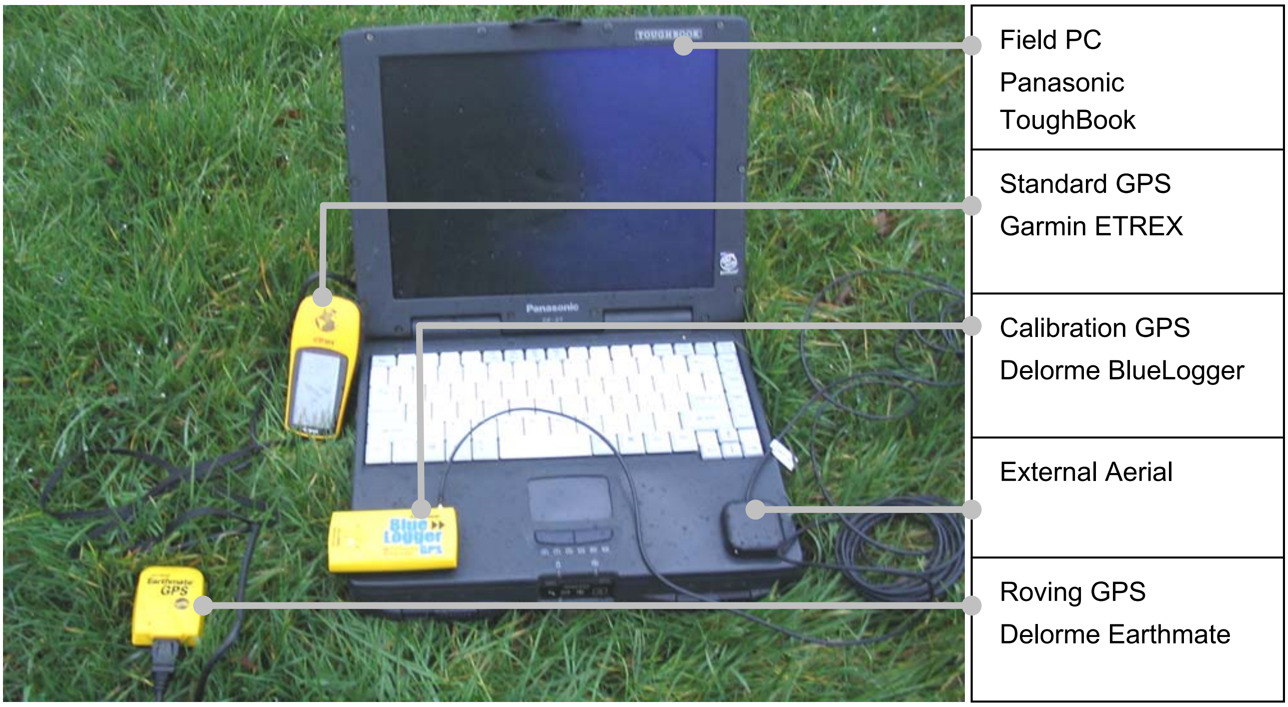


Figure 6-20 Delorme Earthmate hardware with GPS PostPro on a field laptop.

Five hundred pounds is a lot of money for an expedition but it is the only realistic way of obtaining a full hardware solution capable of delivering sub-metre accuracy GPS data.

6.11 Future developments of the Global Positioning System

The NAVSTAR system is continually evolving and it is worthwhile to be aware of the new systems that are coming online over the next few years so that any expedition is best placed to take advantage of them.

The biggest current development in GPS accuracy has been the satellite based WAAS. Though these systems were due to come online in 2003-2004 their current (2005) coverage remains poor (*www.esa.int/EGNOS*). The US is further ahead with their plans while the European system is still undergoing testing. There have been issues with the European WAAS satellites (EGNOS) picking up US based corrections. As can be seen from Section 6.5.3, a differential correction for America will have passed through a significantly different part of the ionosphere and received different levels of attenuation. A firmware update for Garmin units was issued in 2003 to stop the unit confusing US and European corrections. There are plans to launch WAAS satellites for Japan, China, Australia, India and Brazil. This will significantly improve the NAVSTAR system throughout the world. The timings of these launches are not yet officially known and the testing phases of the system are often longer than stated. Any expeditions to these territories should check in advance on the status of the corrections.

In 2005 the satellites should begin broadcasting extra signals that will allow a form of basic differential correction to be conducted using a single receiver. Two military signals will be added to the L1 band and a civilian signal will be added to the L2 band. The specifics of this system are not yet available but in future even standard non-WAAS GPS signals will be more accurate and better able to cope with an obscured sky, such as under dense tree cover.

Another significant development is a third GPS constellation due to be launched within five years; this is a European GNS called Galileo aimed at commercial use. To encourage take up, Galileo will not be susceptible to SA. Should the US rescramble any signals the Galileo system will guarantee global coverage. Though the system is no more accurate than NAVSTAR, having access to more satellites will speed up acquisition time and improve accuracy.

Eventually, by 2008, a new generation of NAVSTAR satellites will begin broadcasting a third wavelength in the L5 band. The principle behind the inclusion of the L5 band is that the difference in retardation between L1 and L5 will tell the receiver the electron density in the ionosphere. If the receiver knows the status of the ionosphere it can eliminate this error. Eventually the only errors in the GPS system will be from local effects. This will offer real-time inexpensive centimetre accuracy.

The expedition can keep up to date with recent advances by checking the websites of the various service providers.

Table 6-7 GPS Service Providers.

System	Internet address
GPS	*www.gps.losangeles.af.mil*
GLONASS	*www.glonass-center.ru*
Galileo	*www.europa.eu.int/comm/dgs/energy_transport/galileo/*
EGNOS	*www.esa.int/EGNOS*