The SatNav Users Guide to Navigation and Mapping Using GPS

by

John Maris

This preview serves as a test for Adobe Reader on your computer, so good, good, if you are reading this you have it. Furthermore, to give you an idea of the content there are snips from the beginning of each section and from within sections. The latter is to give you a feel for the diagrams and maps that are used to explain and demonstrate the concepts involved.

A school of dolphins

A flock of birds

A pride of lions

A lost of navigators

Preface

This book is intended for those who are satnav users, whether car satnav, mobile phone or leisure users of hand held GPS, who have a desire to learn more about what is behind the scenes of GPS, navigation and mapping. The intent is to provide an understanding of how positioning and navigation, with emphasis on GPS, is related to the real world, either by maps or other applications.

Much of the content was in the past confined to a small group of specialists but the advent of GPS and electronic mapping systems has opened the field of navigation to a wider community of users. Previously, positioning and navigation was a local concern, to the surveyor with local based electronic equipment and to the leisure user with a local based map and a compass. Now there is global positioning and a wider application of navigation there is a need to understand the relation between local and global. To fit GPS to your area corrections are likely to be needed, not always, some are lucky. These corrections may be inbuilt into the application but in other cases they need to be manually applied.

There will be some technical terms and names, which we cannot get away from. They will be used but backed up by easy explanation. Understanding the principles does not always require memory of the terms. There are also some horrible sums involved. These formulae will not be included in the content but, for the brave, will be in appendices. This is just to demonstrate further what is involved. Do not think us guys are clever. Yes, many of us in this field learned these sums in study, but I bet for most they are forgotten, other than awareness. Professionally we do use them. We input some parameters and the computer does the rest. Trust me, I’m a navigator. Any sums included in the text will be just plus, minus, times and divide.

The information covered starts with the basics of what GPS consists of, how it works, techniques to increase GPS accuracy and an explanation of how accuracy is defined. This will be followed by how we transfer the sphere onto the flat, projections, and once done how we relate our local area to the globe, coordinate systems and datums. For those of a Latin background, in this field the plural is datums not data. So far this covers the hand held GPS leisure user’s requirements. We will then proceed to applications of GPS positioning to electronic mapping, the realm of the car satnav and mobile phone.

Finally, this book is only a behind the scenes info. It will not directly help you to find the best system for you, or sort the problems you have with your system. It will not directly help you to not get lost. Trust me, I’m a Navigator.

Having said that, you may work out your own solutions based on this information, with respect to which system you are looking at or using. In field though, if you get lost, that’s misuse of equipment or it’s broken. I am a professional but trust me I am a navigator and I can be as lost as you. But, I can find my way back on track, eventually.

Contents

The Preview is navigable forwards and backwards.

Introduction

One of mankind's oldest requirements and importance to many activities is knowing where you are and where you want to go. Frequently this has been awkward to achieve, probably in the past without understanding why. Over time several systems have been developed to simplify navigation but all have their disadvantages.

The simplest is the use of landmarks, once commonly used by the motorist. "Excuse me, can you tell me where…?" Did this ever work well for you? In a way it is a good system, if the directions are simple and accurate and you can remember them. It tends to break down if the landmark has been removed or destroyed. "Where's that bar gone?"

Dead Reckoning (DR), calculated guesswork, estimates position based on distance travelled in a direction from a previous position. The better the estimate of speed is and also wind or current vectors, if you are in a marine or air environment, the better your estimated position. Errors accumulate as new positions are calculated from successive previous positions. You may think that DR is now outdated, but no. It lives on in a more sophisticated way in the form of Inertial Navigation Systems (INS) often used along with GPS to provide continuing navigation when GPS drops out.

Celestial Navigation or Astronavigation is relatively complicated requiring the use of almanacs for the stars position and Sight Reduction tables. This gives you latitude. For the commonest method an accurate timepiece is required for longitude due to the rotation of the Earth at 15° an hour. Have you seen the TV drama 'Longitude'? Let's not forget, a clear sky is critical.

Several electronic navigation systems have been developed, many being radio navigation systems. Radio waves emitted from beacons of known position are received and the time difference converted to distance. Special charts were often required for plotting. Few, if any, provided full global coverage as they were useable only within range of the beacons. Under the known laws of physics, for a certain power, greater range is given by lower frequency. Lower frequency means lower accuracy, further degraded by atmospheric effects that are enhanced by following the curvature of the Earth.

The requirement for global positioning has resulted in Global Navigation Satellite Systems (GNSS), the generic term for satnav systems with global coverage. The first development was TRANSIT or NAVSAT, designed for the US Navy. It used Doppler shift of its radio signals. The high pitched oncoming siren and the low pitched receding is a Doppler effect. Satellite passes were infrequent, therefore there was no high speed real time positioning.

The first real GNSS is the NAVSTAR GPS, (NAVigation by Satellite Timing And Ranging) again developed by the US military. Other GNSS are the Russian GLONASS, the European Space Agency's Galileo and the Chinese Compass (BeiDou-2). In 2016 GPS and GLONASS are fully operational. Galileo and Compass are in development. Some higher end receiver systems integrate GPS and GLONASS. Future systems will probably incorporate the others when operational. Most car satnavs and hand helds only use GPS.

Originally GPS had high accuracy for military use but accuracy was downgraded for civil use. This Selective Availability (SA), giving 30 - 200m accuracy depending when and where in the world, was good for general navigation. As with other systems it would get you close enough to use the best navigation system in the world. The Mark I Eyeball. This has its disadvantage also. If it's foggy, it's of little use. For survey use this accuracy level was not high enough so techniques, referred to as augmentation systems, were developed to enhance accuracy. Whether a final consequence of this, commercial pressure, competitive pressure or other reasons, SA is now switched off. Accuracy is now in the order of 2 – 15m Stand Alone, no augmentation, depending on the type of receiver. This is in moving or dynamic mode. This accuracy level has enabled commercial developments such as the car satnav. Remember though, although unlikely, Selective Availability can be switched on again.

A last point to clear up. Despite the term ‘satnav’ and my common use of it, GPS is not a navigation system, as such. It is what it says on the label, Global Positioning System. The reference as ‘GPS navigation system’ is, however, common, accepted and understandable. The navigation system is some piece of software that GPS position is input to, so GPS gives the ability to navigate. Very frequently this software resides in the same box that holds the GPS receiver. This receiver, particularly for car satnav, hand held GPS or mobile phone, contains the antenna, various electronics and software for processing the signal and software that will calculate position from this processing. Also within the unit is software that will receive the position and use it to navigate, usually giving you direction and speed or even put it on a map. Input of a destination position will enable calculation of direction to go from the last input GPS position. All this can happen in updates of a second or even 1/10th of a second depending on the sophistication of the device.

1 What GPS Is

GPS is a radio based system providing Position, Navigation and Time services (PNT). It consists of 3 segments:

Space Segment: Satellites or Space Vehicles (SV).

Control Segment: Monitoring ground stations.

User Segment: You and your receivers.

PNT sounds like a contradiction to the Introduction. It does not do the navigation for you but gives you the capability (within the software) to navigate from the position given.

1.1 Space Segment

The Space Segment consists of 24 SVs in use (21 plus 3 active spares). There are usually more in orbit as new ones are launched ready to replace older ones. The satellites orbit at 10900 nm (1 nautical mile = 2000 yards or 1852m) in Medium Earth Orbit, every 12 hours with a 55° inclination to the equator, Fig 1. Note that the length of a nautical mile is deliberate. It equates to 1 minute (1’) of arc, 1/60 of a degree, along a meridian such as a line of longitude, the equator or any circle cutting the Earth in half. That’s navigation.

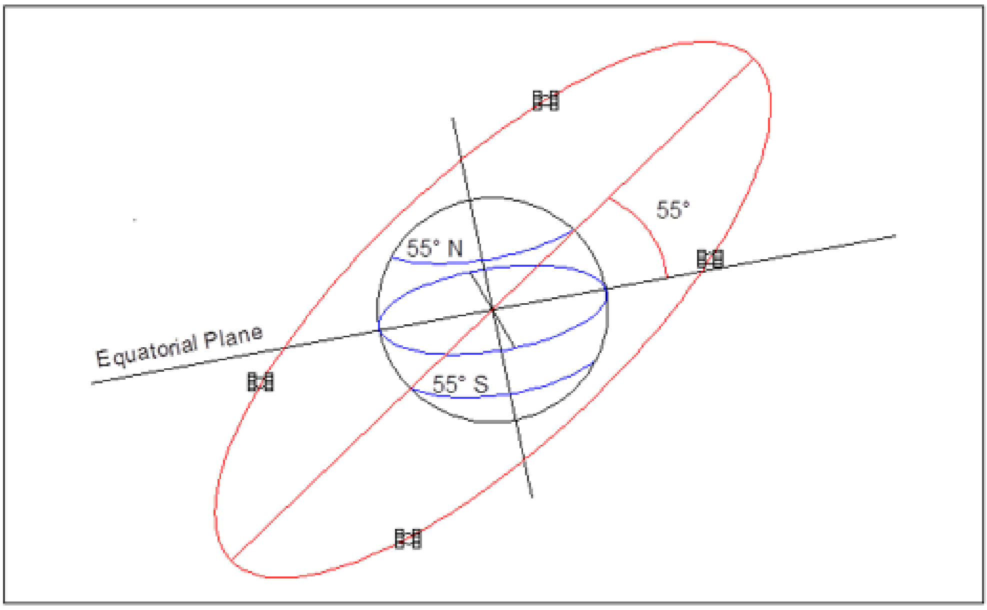

Fig 1. Inclination of Orbital Plane to Equatorial Plane.

There are 6 orbits 60° apart in the equatorial plane each with 4 satellites. This enables 5 – 8 satellites visible at any time for most areas. There may be less visible from north and south of 55° as this angle is as high as the satellites go. Between these latitudes satellites can be overhead (zenith) and all around. This is not so for higher than 55°. The system accuracy can be somewhat reduced as satellites are lower in the sky, the geometry is degraded and atmospheric effects increased due to longer travel path of radio waves through the atmosphere.

The orbital geometry takes some 3D thinking. It may have crossed your mind that 6 orbits at 60° doesn't fit, with orbits overlying. The orbits are 60° from the previous going round the circle and as one orbit is inclined 'up' its opposite at 180° is inclined 'down'. Figs 2a, b and c explain.

2 How GPS Works

A GPS receiver determines its position (really the antenna, which may or may not be attached) using a ranging method based on travel time of radio signals transmitted from the SVs. There are various factors involved in calculating these ranges, some inherent to the system's method of function and some external. This section covers how the range is used, how it is determined and how the signal works to produce it.

2.1 Positioning Method

Each satellite transmits coded messages to determine range at the receiver. Additional navigation messages contain information necessary for position calculations and error corrections that are done by software in the receiver. Sorry, but more and more we have to try and think 3D.

2.1.1 Basic Principle

The receiver position is derived from a process that is called triangulation. I say 'called' because that is what it's commonly referred as. It is not correct as the process is trilateration. Maybe it's because triangulation was, before GPS, the more common form of surveying control points, or, perhaps it's just more catchy. The difference is that triangulation involves, guess what, measuring angles. Trilateration involves measuring ranges, Fig 4a, b and c.

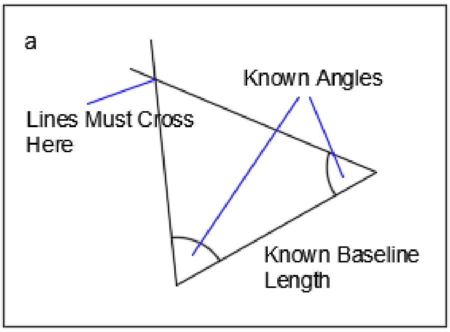

Fig 4a. Triangulation
If the baseline endpoints are known there is orientation, and the other apex can be found. Otherwise the triangle can point anywhere.

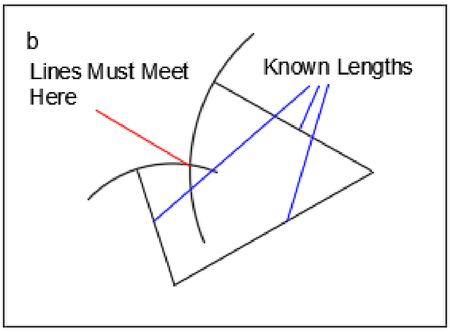

Fig 4b. Trilateration
Finding a third point but with ranges. Again, if the points measured from are known there is orientation.

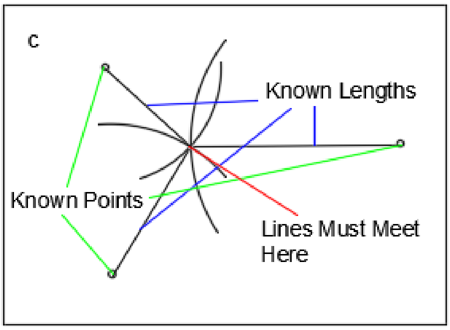

Fig 4c. Trilateration
Ranges from 3 known points defines a fourth

In both cases angles and lengths can be calculated afterwards. The measurements decide the process.

2.1.2 GPS Geometry

The GPS system resembles Fig 4c in 3D. Knowing the range from one satellite the receiver must be anywhere on a sphere of radius equal to range. A range from a second satellite puts the receiver on its sphere. Should, by chance, the spheres just touch, that point of contact is your position. Usually the spheres intersect and the surfaces in contact define a circle anywhere on which is the receiver, Fig 5.

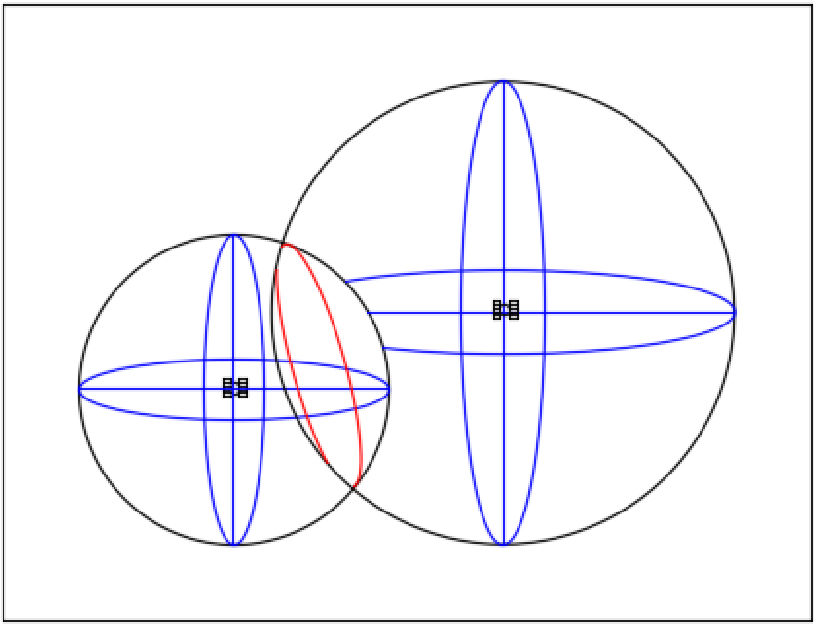

Fig 5. Two Spheres Intersecting
The red ellipse, the result of the angled view, defines the circle of intersection of the two sphere surfaces.

A third sphere, Fig 6, narrows down to two points, one of which may obviously be wrong, such as a point in space, and we know we are on the Earth. A fourth will give the correct point, the closest to the fourth sphere. In any case the receiver knows roughly where it is from the last computed position even if that was the last before turning off. What if you have moved half way around the world? Many modern receivers from switch on will not take too long to catch up with a reasonable output. Cheaper ones may take some time. Usually a rough position can be manually entered as a latitude and longitude or selected from an electronic map in the receiver menu. A position within 300km is usually sufficient.

4 What Does Accuracy Mean?

How accurate is your GPS and what does it mean? Equipment adverts and sometimes the specs do not always give a clear idea of the accuracy. What does accurate to 10m mean? Is the worst case 10m from what it should be or is the spread around the true position, from one side to the other, 10m. One would be ±10 the other ±5.

Alright, GPS accuracy is a fairly complicated topic and a detail would be out of place in ads and specs. However, I do feel that sometimes the specification is too loose.

Alright, again. Before we go on, I myself am guilty, for throughout this missive I have also used the term ‘accuracy’ loosely. There is more than just accuracy.

4.1 [Accuracy, Precision and Resolution]

To assess the quality of measurements there is more than one factor involved but in normal talk they all come under ‘accuracy’ and can even have cross definition. Once you have read this bit, to clear up the differences, you can go back to being loose with accuracy, just as I have been.

5 [From Round to Flat: Projections and Datums]

Now that we are looking at a lat and long on our GPS we may need to see how it fits with the world, or our part of it, in relation to a map. This leads to some complications.

5.1 [Overview and a Bit of History]

Once upon a time life was easy, for the world was flat. Then we had to go and invent a round world creating difficulties. As we delved further into these problems we discovered that the Greeks or even civilisation before had beaten us to the round world concept. It had just been forgotten.

The Greeks theorised a round earth from observations of stars, how their altitudes differed from different locations, from the curved earth shadow on the moon and the simple act of watching ships approach port. At first the topsails become visible then more sail then the whole mast and then the hull. Skippers and crews swore that they had not come out of the sea. “Nah, mate” said Capt Pugawasades “The submarine won’t be invented for another 2 ½ thousand years”.

5.2 [Projections]

What is a projection? It’s any method of flattening onto a 2D plane a surface having curvature in all 3 spatial dimensions. This cannot be done without distortion. Also, forget the geography lesson explanation of a light at the centre of a globe projecting the surface onto a sheet of paper. This perspective is generally too simplistic and of minimal use. The projections we are talking about here result from mathematical functions transforming coordinates on a sphere to the plane. This section will describe methods of doing this and we will use a sphere as it is a good enough approximation even though the reality requires best fit spheroids and the appropriate spheroidal sums.

The use comes into it. Projecting from a spherical model will not have noticeable difference on small scale maps such as a world atlas. Here the presentation is likely to be only visual with no great need for computations using the map. For larger scale maps such as national grids a spheroidal model is more appropriate.

Note that large and small scale used here may seem the wrong way round as small scale covers a bigger area. A world atlas may be 1:50 000 000 and a local map 1:1000. It’s the scale expressed as a fraction that counts, 1/1000 being a bigger fraction.

5.2.1 [Some Concepts to Consider]

Almost an infinite number of projections are possible each needing its own definition parameters. Different projections have different uses. Which you use depends on what you want from your map. As we can’t escape the distortion we need to decide features that need preserving for our use at the expense of distortion in other features. Here is time to make note of a difference. There are projection methods, examples you may have heard of, Mercator or Lamberts Conformal. There are projected coordinate systems which comprise of the method and the defining parameters. National grids would be an example covering a defined area.

Scale Factor

A measure of distortion is given by the scale factor, SF, a ratio of distance on the projection to that on the sphere. A single value of this ratio is usually only valid for a small area as it can vary over the whole projection. It is a result of projecting and no real relation to the map scale as such. The map scale will be constant even if inaccurate due to scale factor. SF will typically be larger at the edge of the projection so SF of 1.004 means 1km on the sphere is represented by 1004m on the projection. On the map you will measure this distance as 1004m, at the scale of the map, but the real distance, on the ground is only 1000m. Commonly a form of rescaling or overall scaling is applied to keep SF closer to 1 throughout the projection.

Preserving Features

The rules used to determine how features are taken from the sphere go hand in hand with how scale factor is applied. The features are generally those of shape, area and size with the aim to preserve one at the expense of the others, or to compromise between some features. These are projection methods.

5.2.2 Cylindrical Projections

Probably the commonest known cylindrical is the normal equatorial projection. The cylinder is upright, its centre line coinciding with the rotation axis, and its circumference touching the equator. This is a tangent surface. Secant surface projections also exist where the cylinder slices through the sphere at two latitudes known as standard parallels (SPs). This is how the overall scaling is achieved.

On the normal cylinder meridians are equally spaced parallel vertical lines. Parallels of latitude are straight lines at 90° to meridians. These are stretched east – west. Distance is stretched. The shortening latitude circumferences approaching the poles are stretched to the length of the equator, Fig 24a and b. Scale factors increase along parallels towards poles as a multiple of the equator scale (1) expressed as secant (sec) of latitude, 1/cosine of latitude. Common symbols for latitude and longitude are ϕ,λ phi,lambda. You may see phi as φ.

The Mercator

As a conformal or orthomorphic projection conserving shape and angles the Mercator is one of the most important and widely used for navigation. Meridian and parallel scale factors must be equal all over the map and at any point the same in all directions. Keeping to cylindrical rules, meridian scale factors increase towards the poles to keep in line with parallel increases. Although shape is preserved sizes increase greatly towards poles, Fig 28. Does anyone believe that Greenland is the same size as South America? I believe it’s about an 1/8th.

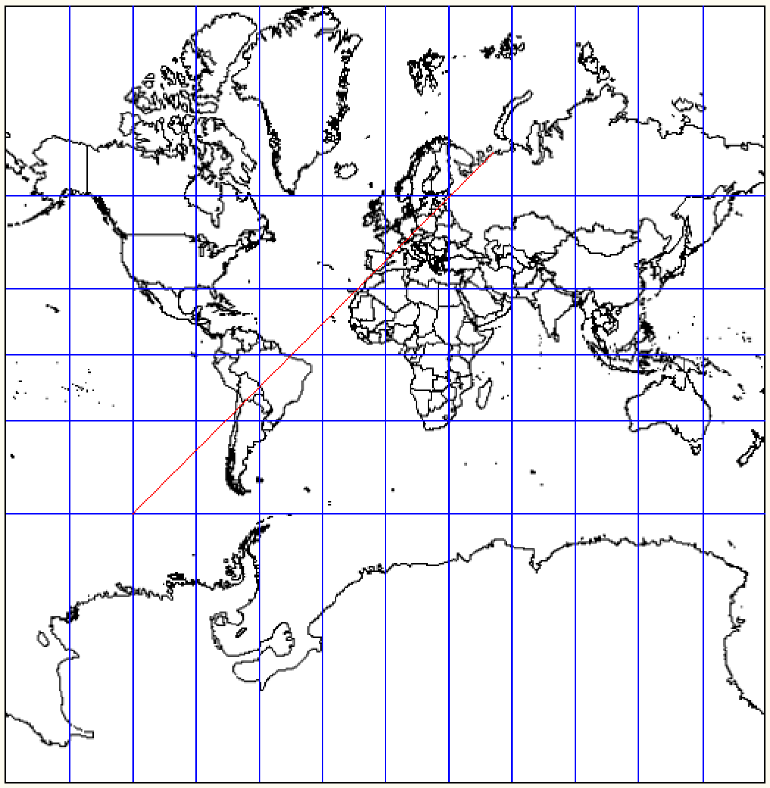

Fig 28. Mercator
The red line is a bearing of 045°

The bearing line cuts meridians at 45°. Don't expect them to pass through intersections, it's not a square graticule although meridian and parallel SF are equal. The angle is constant. This is known as a rhumb line or loxodrome. This is not the shortest route, that would be a great circle which would be a curve on the Mercator.

5.2.3 Conic Projections

Conics have similar issues as regards to preserving distance, area or shape and also scale factor behaviour. Cylinders and planes can be considered as extreme cases of the cone. Let's take the common normal aspect or polar projection where the apex is above the pole and the tangent case where the cone touches the sphere along a single standard parallel. The projection will look triangular with meridians converging towards a point. If the standard parallel is the equator meridians cannot converge. The cone cannot slope to an apex. You have a cylinder. If the standard parallel is a pole, now a point, the cone is flattened so it is a plane at this extreme.

On the more common normal conics (there are pseudo, polyconics and bipolar obliques) meridians are radiating straight lines that converge towards the pole. Whether they meet at the pole depends on the type, but even so, this point is not the pole but represents it. The cone apex is some distance above the pole. The full circle is compressed into a sector depending on the standard parallel and map extent. This means that angles between meridians are smaller than the actual longitude difference and depend on this difference and a factor called the cone constant. This itself has varying relations to latitude depending on the type of conic and the number of standard parallels.

Albers Equal Area Conic

The commonest equal area was introduced by Albers in 1805. Again, meridians are equally spaced radii of the parallel concentric arcs centred at the apex. Meridians cut parallels at 90°. Poles are arcs. Parallels are not equally spaced getting closer to the north and south map extents, Fig 38. To maintain area meridian SF must reduce as parallel SF increases.

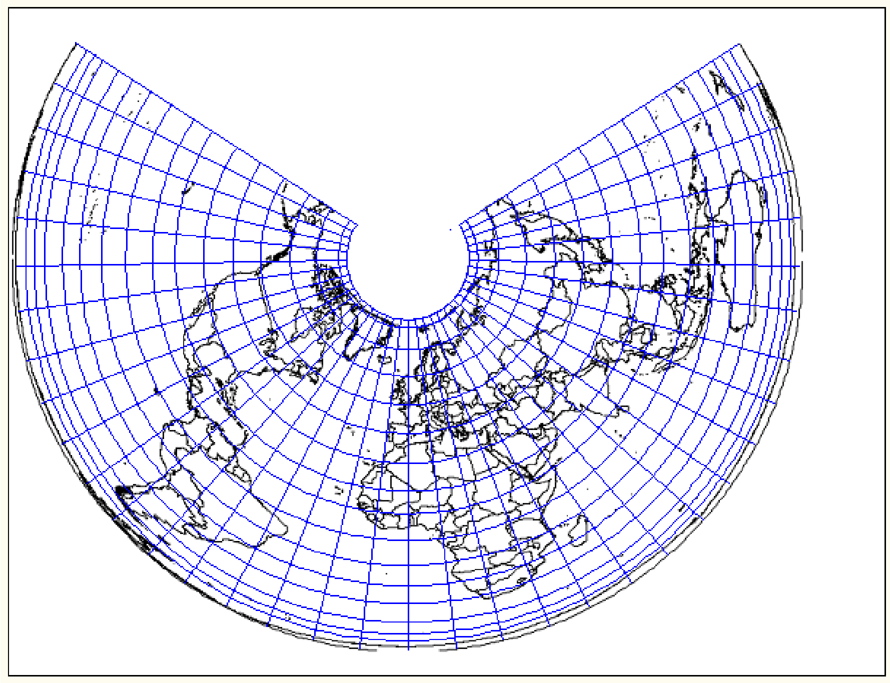

Fig 38. Albers Equal Area Conic with 30°N and 60°N Standard Parallels and 0° CM
The graticule is 10° latitude and longitude.

5.3 Datums

Now, we are in our local area with our GPS lat and long looking at a map that has been projected from a spheroid, more appropriate for large scale mapping than from the sphere. Depending on where you are in the world will depend on which spheroid the map is projected from and which datum is used, of which there are many.

GPS works on a datum based on an internationally agreed world reference system. This datum is the World Geodetic System 1984 (WGS84). Your map datum may be completely different so in plotting your WGS84 position you could be hundreds of metres out. You need to convert. A point can have different lat and long. It depends on what datum.

Most GPS will have a selection of commonly used datums in the menu which, if selected, will give the position on that datum as lat/long or grid coordinates. I don't believe that any car satnav gives this option but by some method the map for your region will be, or should be, correct. It has already been mentioned that projections onto a map are nowadays mostly by mathematical processes but we will see later other methods of making your satnav map.

5.3.1 What is a Datum?

It's a reference system for expressing coordinates of spatial data in 3D, 2D or 1D.

3D is Cartesian XYZ or lat/long and height.

2D is lat/long, rectangular XY(east,north) or polar(ρ,θ) distance and angle (rho, theta), Fig 58.

1D is height eg. Tide datum.

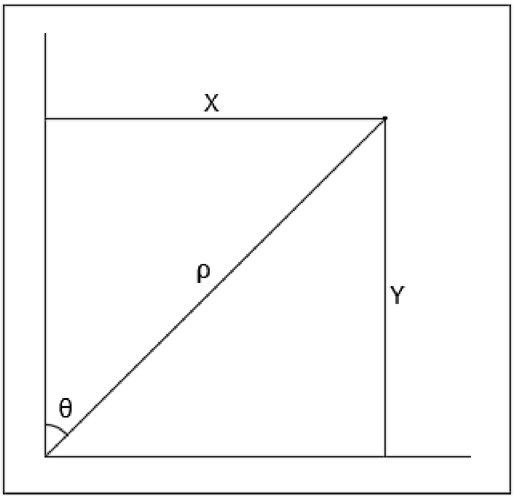

Fig 58. Rectangular and Polar Coordinates

The datum establishes an origin and defines a system with a conventional set of parameters. It is based on a spheroid of particular size, shape, location and possibly orientation with respect to the earth. The spheroid is not a datum. Different countries may use the same spheroid but have different origins for their mapping system. Datums on a spheroid are geodetic datums also known as Terrestrial Reference Systems(TRS).

6 Making Digital Maps

Up to now we have our position on a paper map referenced to a particular datum. We now need to integrate that position into the electronic map on our satnav, mobile or other device to give us visual meaning and the computer mathematical meaning to where we are and determine direction instructions to a destination. We will now look at methods to achieve this bringing us into the realms of digital mapping, Geographical Information Systems (GIS) and databases.

6.1 Why SatNav?

Someone believed that people would like car navigation systems and they could make money providing them. We do not have these devices because of GPS, as such. Attempts at developing automotive navigation systems based on dead reckoning using INS preceded GPS by maybe a decade. They are more a result of developments in computer technology, enabling the application of colour to digital mapping, and increased data storage capability such as the CDROM. The technology was already there for GPS, waiting for the day. That day probably came in 2000 when selective availability was turned off giving GPS its higher accuracy. Then such developments continued from car to other systems and devices. GPS has given the term satnav but not the original principle.

Business has formed and is increasing to further improve the technology. Therefore, note that information on companies and methodologies is believed valid in 2010. Business arrangements change as do technical developments.

A major issue is things change. Maps become outdated. As the number of users increase world wide better ways are needed to map and disseminate updates. It looks achievable through increases in higher resolution satellite mapping of the world, leading to availability of cheaper data, automated on road surveys with cameras and other devices, coupled with web based user error reporting and faster processing, leading to faster updates via, nowadays, on the fly wifi.

6.5 Map Structures and Databases

There are two types of digital map, raster and vector. Both have their own uses for various methods of display but vector is the commonest for real time use, it is more practical and versatile. Raster maintains the visual beauty and art of the original paper map as it is a picture. For demonstration I have used my GIS package. Mappers will use similar principles in a more sophisticated way.

6.6 Navigating a Route

In most other environments the best navigation solution is a straight line from A to B. We are restricted to roads, which means the system has to follow a set of lines and nothing else.

Firstly, we have to be on a road. Well we probably are, however, inaccuracy in GPS may position us off the road. A snap to road feature solves this forcing the GPS position onto the vector that defines that section of road. This may give a false sense of security on GPS accuracy but usually we will not care as it improves the total system accuracy. Sometimes though, when the signal is particularly bad, you may get snapped to the wrong road, a wrong A for direction instruction to B.

> I took my old cheap Navman into the middle of a field. On screen I was on a road 200m away. The actual given lat/long was within 5m of the national grid and hand held Garmin.

To get to B following road line segments a search algorithm is used.

End of preview. Please click to return to main page.