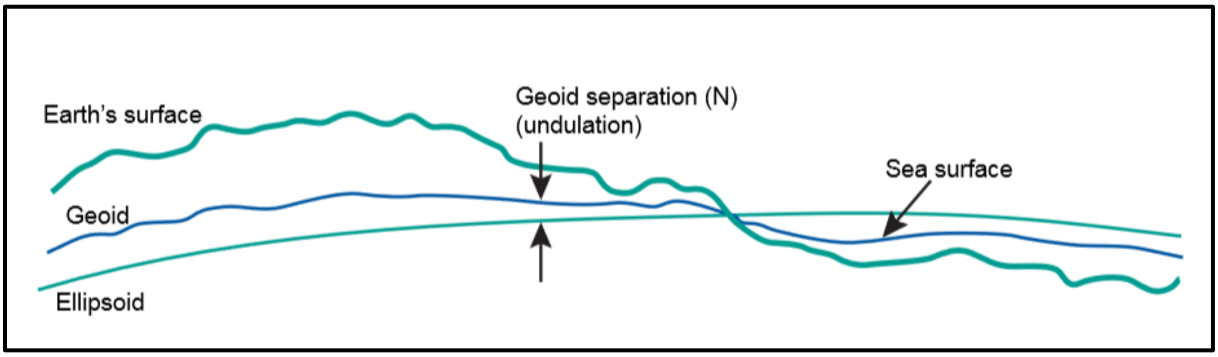
Chapter 4 :- Spatial Referencing and Positioning

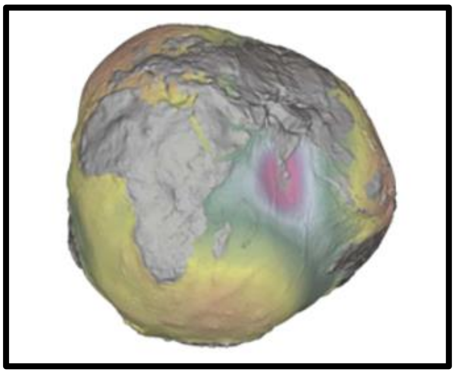
4.1 Reference Surface for mapping

The surface of the Earth is anything but uniform. The oceans can be treated as reasonably uniform, but the surface or topography of the land masses ex- hibits large vertical variations between mountains and valleys. These variations make it impossible to approximate the shape of the Earth with any reasonably simple mathematical model.

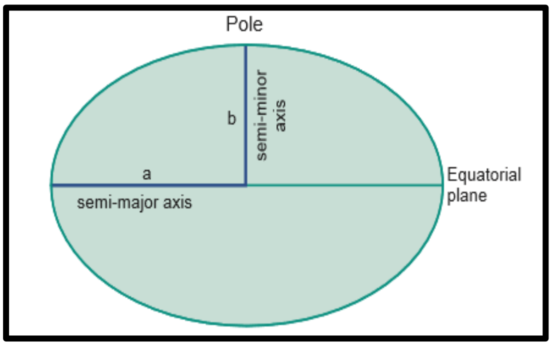


Consequently, two main reference surfaces have been established to approximate the shape of the Earth.

GEOID :- a hypothetical solid figure whose surface corresponds to mean sea level and its imagined extension under (or over) land areas.



ELLIPSOID :- surface that may be obtained from a sphere by deforming it by means of directional scalings, or more generally, of an affine transformation.



Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

The Geoid and the vertical datum

- Assumption:-
  - *Entire earth surface is largely covered by water*
  - *Water surface is affected by gravity (ignoring tidal and current effects)*
- Therefore, direction of gravity (plumb line) is dependent on the mass distribution inside the earth
- The plumb line through any surface point is always perpendicular to it.
- Geoid is used to describe heights
- For this, the ocean’s water level is registered using tide gauges (mareographs).
- The registrations are averaged and are called mean sea level.
- For eg:- For Netherlands and Germany, the local mean sea level is realized through the Amsterdam tide gauge (zero height).
- The height of any point is determined through a technique called as geodetic levelling,
- The height determined with respect to tide gauge is known as orthometric height.
- The local vertical datum is implemented through a levelling network.
- A levelling network consists of benchmarks whose height above mean sea level has been determined through geodetic levelling.
- So, the surveyors need not start from the beginning everytime they need to determine the height of a new point

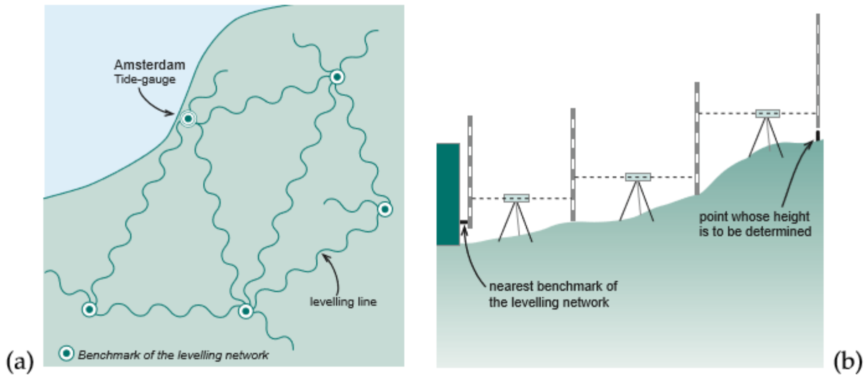
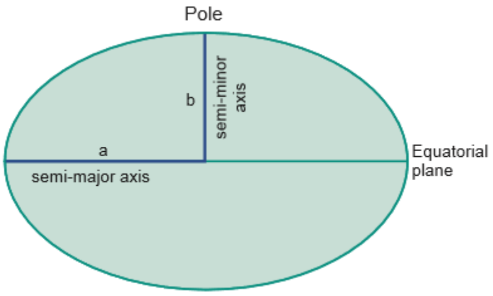


Figure 4.3: A levelling network implements a local vertical datum: (a) network of levelling lines starting from the Amsterdam tide-gauge, showing some of the benchmarks; (b) how the orthometric height (H) is determined for some point, working from the nearest benchmark.

The ellipsoid and the horizontal datum

- Reference surface for heights – geoid
- Reference surface for horizontal co-ordinates – oblate ellipsoid
- An ellipsoid is formed when an ellipse is rotated about its minor axis

Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By



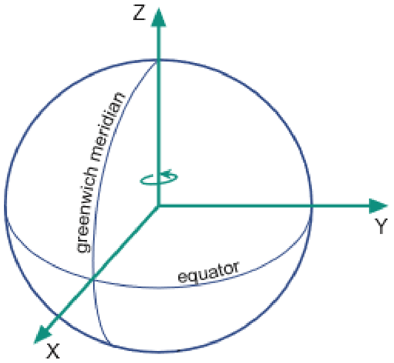
- Local Horizontal Datum

Ellipsoids have varying position and orientations. An ellipsoid is positioned and oriented with respect to the local mean sea level by adopting a latitude (φ) and longitude (λ) and ellipsoidal height (h) of a so-called fundamental point and an azimuth to an additional point. We say that this defines a local horizontal datum. Several hundred local horizontal datums exist in the world. The reason is obvious: Different local ellipsoids with varying position and orientation had to be adopted to best fit the local mean sea level in different countries or regions.

A local horizontal datum is realized through a triangulation network. Such a network consists of monumented points forming a network of triangular mesh elements (Figure 4.6). The angles in each triangle are measured in addition to at least one side of a triangle; the fundamental point is also a point in the triangulation network. The angle measurements and the adopted coordinates of the fundamental point are then used to derive geographic coordinates (φ, λ) for all monumented points of the triangulation network.

- Global Horizontal Datum – (X,Y,Z)

The most important global (geocentric) spatial reference system for the GIS community is the International Terrestrial Reference System (ITRS) . It is a three-dimensional coordinate system with a well-defined origin (the centre of mass of the Earth) and three orthogonal coordinate axes (X, Y, Z). The Z-axis points towards a mean Earth north pole. The X-axis is oriented towards a mean Green- wich meridian and is orthogonal to the Z-axis. The Y -axis completes the right- handed reference coordinate system



4.2 Co-ordinate Systems

- Spatial :- are used to locate data either on the Earth's surface in a 3D space or Earth's reference surface on 2D space
- Planar:- are used to locate data on a flat surface of a map in a 2D space.

2D Geographic coordinates (ϕ, λ)

The most widely used global coordinate system consists of lines of geographic latitude (phi or φ or ϕ) and longitude (lambda or λ). Lines of equal latitude are called parallels. They form circles on the surface of the ellipsoid4. Lines of equal longitude are called meridians and they form ellipses (meridian ellipses) on the ellipsoid.

The latitude of a point P is the angle between ellipsoidal through the point P and the equatorial plane; Latitude is 0 on the equator; max value is +90 degrees at North Pole and -90 degrees at South Pole.

The longitude of a point P is the angle between the meridian ellipse which passes through Greenwich and the meridian ellipse containing the point

It is measured in the equatorial plane from the meridian of Greenwich (it is zero), eastwards +180 degrees and westwards -180 degrees.

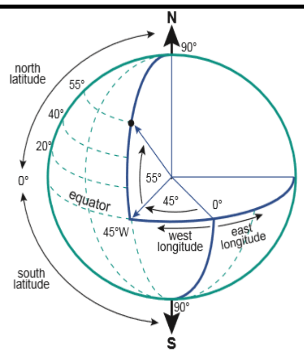


Figure 4.9: The latitude (ϕ) and longitude (λ) angles represent the 2D geographic coordinate system.

2D Cartesian coordinates (X, Y)

A flat map has only two dimensions: width (left to right) and length (bottom to top). Transforming the three dimensional Earth into a two-dimensional map is subject of map projections and coordinate transformations (Section 4.1.3 and Section 4.1.4). Here, like in several other cartographic applications, two-dimensio- nal Cartesian coordinates (x, y), also known as planar rectangular coordinates, are used to describe the location of any point unambiguously.

The 2D Cartesian coordinate system is a system of intersecting perpendicular lines, which contains two principal axes, called the X- and Y -axis. The horizontal axis is usually referred to as the X-axis and the vertical the Y –axis. The intersection of the X and Y -axis forms the origin. The plane is marked at intervals by equally spaced coordinate lines, called the map grid.

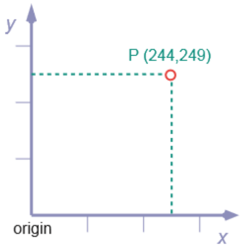


Figure 4.12: An illustration of the 2D Cartesian coordinate system (see text for further explanation).

2D Polar coordinates (α, d)

Another possibility of defining a point in a plane is by polar coordinates. This is the distance d from the origin to the point concerned and the angle α between a fixed (or zero) direction and the direction to the point. The angle α is called azimuth or bearing and is measured in a clockwise direction. It is given in angular units while the distance d is expressed in length units.

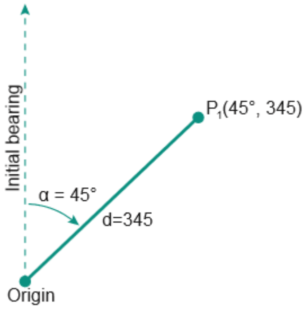


Figure 4.14: An illustration of the 2D Polar coordinate system (see text for further explanation).

3D Geographic coordinates (ϕ, λ, h)

3D geographic coordinates (φ, λ, h) are obtained by introducing the ellipsoidal height h to the system. The ellipsoidal height (h) of a point is the vertical distance of the point in question above the ellipsoid. It is measured in distance units along the ellipsoidal normal from the point to the ellipsoid surface. 3D geographic coordinates can be used to define a position on the surface of the Earth (point P in Figure below).

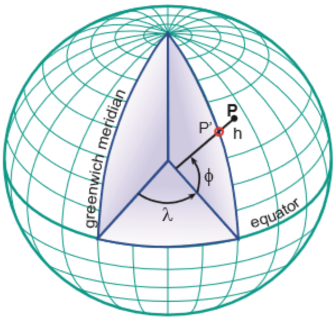


Figure 4.10: The latitude (ϕ) and longitude (λ) angles and the ellipsoidal height (h) represent the 3D geographic coordinate system.

3D Geocentric coordinates (X, Y, Z)

- Also called 3D Cartesian Coordinates
- Origin is the center of the earth; X-Axis passes through the meridian of Greenwich;Y-Axis in the plane of the equator; Z-Axis coincides with the Earth's axis of rotation

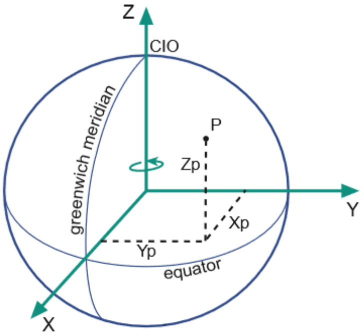


Figure 4.11: An illustration of the 3D geocentric coordinate system (see text for further explanation).

4.3 Map Projections

Maps are one of the world's oldest types of document. For quite some time it was thought that our planet was flat, and during those days, a map simply was a miniature representation of a part of the world. Now that we know that the Earth's surface is curved in a specific way, we know that a map is in fact a flattened representation of some part of the planet. The field of map projections concerns itself with the ways of translating the curved surface of the Earth into a flat map.

"A map projection is a mathematically described technique of how to represent the Earth's curved surface on a flat map."

To represent parts of the surface of the Earth on a flat paper map or on a com-puter screen, the curved horizontal reference surface must be mapped onto the2D mapping plane. The reference surface for large-scale mapping is usually anoblate ellipsoid, and for small-scale mapping, a sphere. Mapping onto a 2D mapping plane means transforming each point on the reference surface with geographic coordinates (φ, λ) to a set of Cartesian coordinates (x, y) representing positions on the map plane

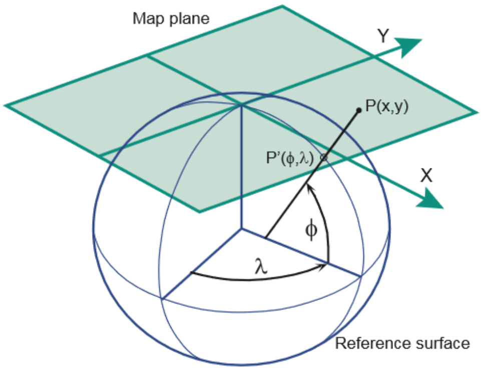


Figure 4.15: Example of a map projection where the reference surface with geographic coordinates (ϕ, λ) is projected onto the 2D mapping plane with 2D Cartesian coordinates (x, y).

Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

The actual mapping cannot usually be visualized as a true geometric projection, directly onto the mapping plane. This is achieved through mapping equations.

A forward mapping equation transforms the geographic coordinates (φ, λ) of a point on the curved reference surface to a set of planar Cartesian coordinates (x, y), representing the position of the same point on the map plane: (x, y) = f (φ, λ)

The corresponding inverse mapping equation transforms mathematically the pla- nar Cartesian coordinates (x, y) of a point on the map plane to a set of geographic coordinates (φ, λ) on the curved reference surface: (φ, λ) = f (x, y)

4.4 Satellite Based Positioning

Satellite-based positioning was developed and implemented to address military needs, somewhat analogously to the early development of the internet. The technology is now widely available for civilians use.

The requirements for the development of the positioning system were:

- Suitability for all kinds of military use: ground troops and vehicles, aircraft and missiles, ships;
- Requiring only low-cost equipment with low energy consumption at the receiver end;
- Provision of results in real time for an unlimited number of users concur- rently;
- Support for different levels of accuracy (military versus civilian);
- Around-the-clock and weather-proof availability;
- Use of a single geodetic datum;
- Protection against intentional and unintentional disturbance, for instance, through a design allowing for redundancy.

A satellite-based positioning system set-up involves implementation of three hardware segments:

1. The space segment, i.e. the satellites that orbit the Earth, and the radio sig- nals that they emit,

2. The control segment, i.e. the ground stations that monitor and maintain the space segment components, and

3. The user segment, i.e. the users with their hard- and software to conduct positioning.

In satellite positioning, the central problem is to determine values (X, Y, Z) of a receiver that receives satellite signals, i.e. to determine the position of the receiver with a stated accuracy and precision.

Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

4.4.1 Absolute Positioning

The working principles of absolute, satellite-based positioning are fairly simple:

1. A satellite, equipped with a clock, at a specific moment sends a radio mes- sage that includes

   (a) the satellite identifier,

   (b) its position in orbit, and

   (c) its clock reading.

2. A receiver on or above the planet, also equipped with a clock, receives the message slightly later, and reads its own clock.

3. From the time delay observed between the two clock readings, and knowing the speed of radio transmission through the medium between (satellite) sender and receiver, the receiver can compute the distance to the sender, also known as the satellite's pseudorange.

"The pseudorange of a satellite with respect to a receiver, is its apparent distance to the receiver, computed from the time delay with which its radio signal is received."

Such a computation determines the position of the receiver to be on a sphere of radius equal to the computed pseudorange (refer to Figure 4.24(a)). If the receiver instantaneously would do the same with a message of another satellite that is positioned elsewhere, the position of the receiver is restricted to another sphere. The intersection of the two spheres, which have different centres, determines a circle as the set of possible positions of the receiver (refer to Figure 4.24(b)). If a third satellite message is taken into consideration, the intersection of three spheres determines at most two positions, one of which is the actual position of the receiver. In most, if not all, practical situations where two positions result, one of them is a highly unlikely position for a signal receiver. The overall procedure is known as **trilateration**: the determination of a position based on three distances.

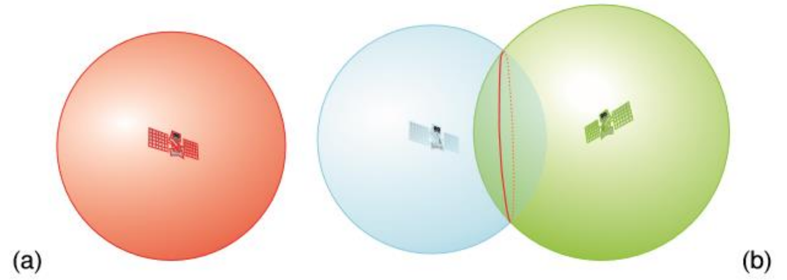


Figure 4.24: Pseudorange positioning: (a) With just one satellite the position is determined by a sphere, (b) With two satellites, it is determined by the intersection of two spheres, a circle. Not shown: with three satellites, it is the intersection of three spheres.

It would appear therefore that the signals of three satellites would suffice to determine a *positional fix* for our receiver. In theory this is true, but in practice it is not. The reason is that we have made the assumption that all satellite clocks aswell as our receiver clock are fully synchronized, where in fact they are not. The satellite clocks are costly, high-precision, atomic clocks that we can consider synchronized for the time being, but the receiver typically has a far cheaper, quartz clock that is not synchronized with the satellite clocks. This brings into play an additional unknown parameter, namely the synchronization bias of the receiver clock, i.e. the difference in time reading between it and the satellite clocks.

Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

Our set of unknown variables has now become (X, Y, Z, Δt) representing a 3D position and a clock bias. By including the information obtained from a fourth satellite message, we can solve the problem (see Figure 4.25). This will result 3D positioning in the determination of the receiver's actual position (X, Y, Z), as well as its receiver clock bias Δt, and if we correct the receiver clock for this bias we effectively turn it into a high-precision, atomic clock as well!

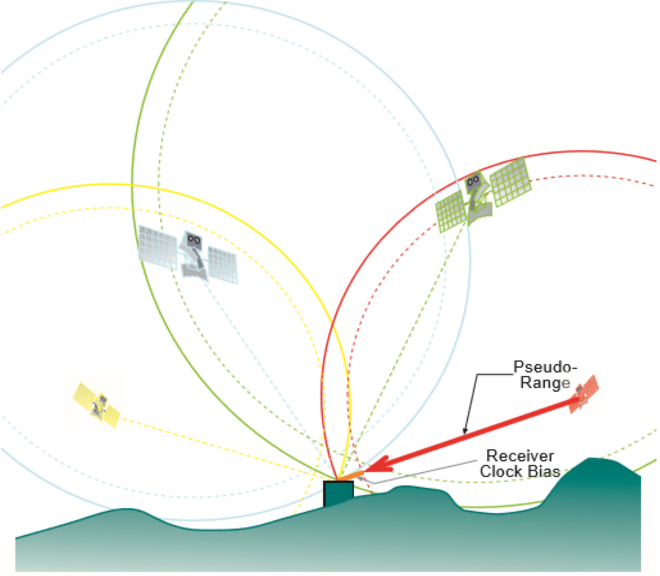


Figure 4.25: Four satellites are needed to obtain a 3D position fix. Pseudoranges are indicated for each satellite as dashed circles representing a sphere, as well as the actual range as a normal circle, being the pseudorange plus the range error caused by receiver clock bias.

Time,Clocks and World Time

During most of human history, the determination of time and position have gone hand in hand.While latitude was determined with a sextant from the position of the Sun in the sky, they carried clocks with them to determine the longitude of their position. Before any notion of standard time existed, villages and cities simply kept trackof their local time determined from position of the Sun in the sky. When trains became an important means of transportation, these local time systems became problematic as the schedules required a single time system. Such a time system needed the definition of time zones: typically as 24 geographic strips between certain longitudes that are multiples of 15◦. This all gave rise to Greenwich Mean Time (GMT). GMT was the world time standard of choice. It was a system based on the mean solar time at the meridian of Greenwich, United Kingdom, which is the conventional 0-meridian in geography.

GMT was later replaced by Universal Time (UT), a system still based on meridian crossings of stars, but now of far away quasars as this provides more accuracy than that of the Sun. UT actually has various versions, amongst which are UT0, UT1 and UTC. UT0 is the Earth rotational time observed in some location.

Coordinated Universal Time (UTC) is used in satellite positioning, and is maintained with atomic clocks. By convention, it is always within a margin of 0.9 sec of UT1, and twice annually it may be given a shift to stay within that margin.

Errors in Absolute Positioning

1. Errors related to space segment

-- Selective availability - meaning that the military forces allied with the control segment *will* still have access to undisturbed signals—may cause er- ror that is an order of magnitude larger than all other error sources combined.

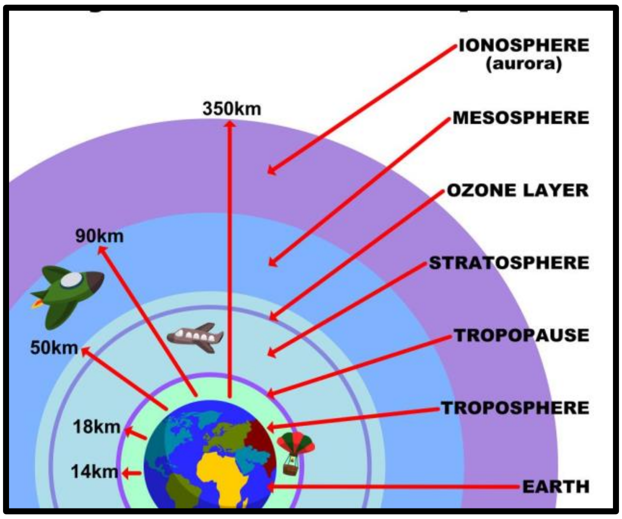
-- Incorrect information

- Incorrect clock reading:- A clock reading that is off by 0.000001 sec causes an error in the satellite's pseudorange of approximately 300m
- Incorrect orbit position:- The orbit is easy to describe if both bodies are considered as point masses. In reality, they are not stationary and there are disturbances due to solar and lunar gravitation.

Control segment in the ground is responsible to control the errors. If unacceptable errors found, a satellite is labelled unhealthy.

2. Errors related to medium

- Stratosphere and mesosphere are harmless
- Troposphere:- approx. 14 km high just above the surface of the earth. It holds atmospheric oxygen and delays radio waves.
- Ionosphere:- outward part of the atmosphere that starts at an altitude of 90 km. It holds many electrically charged atoms, forms protection against various forms of radiation from space.



Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

3. Errors related to receiver's environment

The error occurring when a radio signal is received via two or more paths between sender and receiver, some of which typically via a bounce off of some nearby surface, like a building or rock face. The term applied to this phenomenon is multi-path; when it occurs the multiple receptions of the same signal may interfere with each other (see Figure 4.26). Multi-path is a difficult to avoid error source.

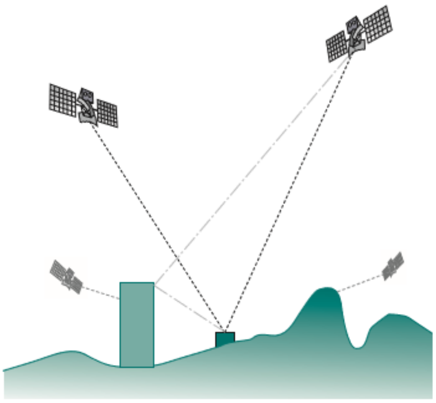


Figure 4.26: At any point in time, a number of satellites will be above the receiver's horizon. But not all of them will be 'in view' (like the left and right satellites), and for others multipath signal reception may occur.

4. Errors related to relative geometry of satellites and receiver – GDOP error

There is one more source of error that is unrelated to individual radio signal characteristics, but that rather depends on the combination of the satellite signals used for positioning. Of importance is their constellation in the sky from the receiver perspective. Referring to Figure 4.27, one will understand that the sphere intersection technique of positioning will provide more precise results when the four satellites are nicely spread over the sky, and thus that the satellite constellation of Figure 4.27(b) is preferred over the one of 4.27(a). This error source is know as geometric dilution of precision (GDOP). GDOP is lower when Geometric dilution of satellites are just above the horizon in mutually opposed compass directions

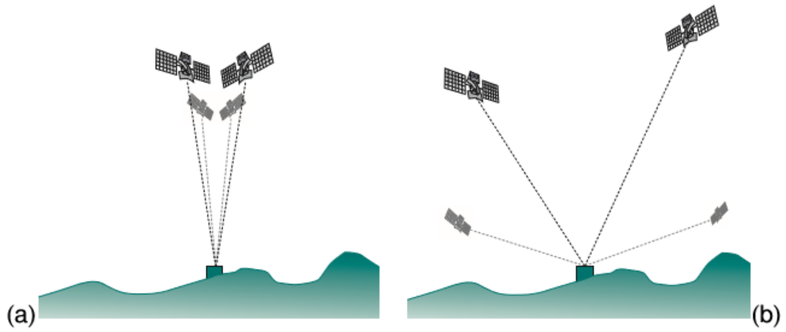


Figure 4.27: Geometric dilution of precision. The four satellites used for positioning can be in a bad constellation (a) or in a better constellation (b).

4.4.2 Relative Positioning

- Solution for errors in absolute positioning:- Perform many position computations and determine the average over solutions
- This will address random errors like signal noise, selective availability,multipath to some extent; but not address systematic errors like incorrect satellite data, atmospheric delays and GDOP effects.
- Reference receiver and target receiver; 70-200 km apart
- Reference receiver will determine its pseudorange error.
- Target receiver will apply the corrections for each of the four satellite signals that it uses for positioning.
- Inverted relative positioning:- Target receiver does not correct the error, but uses its data link to upload its positioning/timing information to a central repository where corrections are applied. Not useful where many target receivers are needed

4.4.3 Network Positioning

- An integrated, systematic network of reference receivers covering a large area like the whole globe.
- Network of reference stations – each monitoring the signals and their errors.
- One/more control centers receive the reference station data,verify the correctness and relay this information to a geostationary satellite.
- This satellite will retransmit the correctional data to the area it covers.
- Target receivers in that area correct the errors and obtain accurate position fixes.

4.5 Positioning Technologies

GPS

- NAVSTAR Global Positioning System – 1994 - American
- Providing PPS (Precise Positioning System) to military and US govt agencies and SPS (Standard Positioning System) to civilians throughout the world.
- Space segment - 24 satellites – each orbit our planet at an altitude of 20200 km.
- These satellites are organized in six orbital planes, with an angle of inclination of 55-63 degrees with the equatorial plane, normally having 4 satellites each.
- This means the receiver on earth will have between 5-8 satellites in view at a point of time.
- Control segment – master control in Colorado,US and monitor stations in a belt around the equator.
- The NAVSTAR satellites transmit 2 radio signals:- L1 frequency at 1575.42 MHz and L2 frequency at 1227.60 MHz

Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

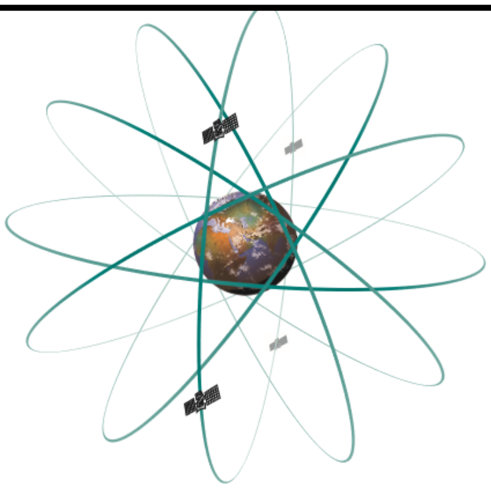


Figure 4.28: Constellation of satellites, four shown in only one orbit plane, in the GPS system.

GLONASS

- Global Orbiting Navigational Satellite System
- Russian Space Force
- GPS better than GLONASS due to its application for civilians
- Space segment – 24 satellites organized in 3 orbital planes with an inclination of 64.8 degrees with the equator orbiting altitude – 19130km, period of revolution of 11 hrs 16min.
- Radio frequencies – L1 frequency – 1605 MHz ; L2 frequency – 1248 MHz

Galileo

- European Union
- 27 satellites, orbiting in 1 of 3 equally spaced circular orbits at an elevation of 23222 km inclined at 56 degrees with the equator.

Navigation with Indian Constellation (NavIC)

- ISRO has established a regional navigation satellite system called Navigation with Indian Constellation (NavIC).
- NavIC was earlier known as Indian Regional Navigation Satellite System (IRNSS).
- NavIC is designed with a constellation of 7 satellites and a network of ground stations operating 24 x 7.
- Three satellites of the constellation are placed in geostationary orbit, at 32.5°E, 83°E and 129.5°E respectively,
- four satellites are placed in inclined geosynchronous orbit with equatorial crossing of 55°E and 111.75°E respectively, with inclination of 29° (two satellites in each plane).
- The ground network consists of control centre, precise timing facility, range and integrity monitoring stations, two-way ranging stations, etc.

Source:- Principles of Geographic Information Systems,Otto Huisman and Rolf A.de By

- NavIC offers two services: Standard Position Service (SPS) for civilian users and Restricted Service (RS) for strategic users.
- NavIC SPS signals are interoperable with the other global navigation satellite system (GNSS) signals namely GPS, Glonass, Galileo

****