Basic Geodesy

Unit: II

Semester: I

Paper Code: GIS 05
Name of Paper: Earth Positioning System

PG Diploma in RS & GIS

Dr. SHYAMA PRASAD MUKHERJEE UNIVERSITY, RANCHI

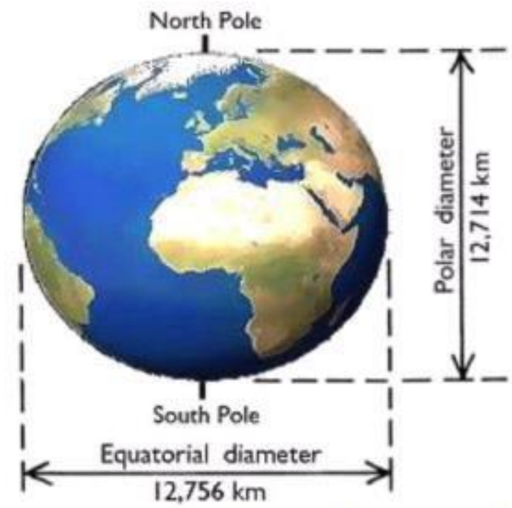
BASIC GEODESY

1.1. The shape of the earth

The first measurements on the size and shape of the earth came from Eratosthenes (276-195 B.C). He is regarded as the first geodesist for his work in determining a radius for the spherical earth. He assumed that that the light rays coming from sun to earth are parallel and are perpendicular to the earth. Then he measured the arc length from Syane to Alexandria and the angle subtended by the arc at the centre of the earth using astronomical observations.

$$\frac{\theta}{360^0} = \frac{D}{2\pi R}$$

Where; θ = angle subtended by the arc at he centre of the earth
D = Arc length from Syane to Alexandria
R = Radius of the earth



Shape of the earth as Oblate spheroid

Assuming the shape of the earth as a sphere; he obtained an earth‘s radius of 6267 km using above trigonometric equation. This value departs from the radius of a mean spherical Earth (6371 km) by -2% only. Later Sir Isaac Newton suggested that the earth‘s shape was ellipsoidal. Newton‘s idea was that the planet bulged around the equator and flattened at the poles. Precision of the measurements taken on the Earth has been increased tremendously. A series of gravity measurements were carried out from 1734-41 to confirm the shape of the earth and found that those measurements are not exactly tallying to the spherical model for the earth. They were closer to an oblate spheroid which is an ellipsoid of revolution. That is; it is generated by rotation of an ellipse around its shorter axis and then flattened at its poles.

1.2. Reference surfaces

The physical surface of the earth contains variety of land forms like plains, valley, mountains, water features etc. It has excursions of +8,000 m (Mount Everest) and −11,000 m (Mariana Trench). Although this is the surface on which actual earth measurements are made, it is irregular and complex showing large vertical variations. Topographers and hydrographers are interested in this topographic surface as it provides topographic information for them. Due to irregular topographic nature of the Earth, there are two reference surfaces used to approximate the shape of the Earth: the Geoid and the Ellipsoid.

- ✓ The geoid or mean sea level - the level surface to determine heights (vertical reference).
- ✓ The ellipsoid - the reference frame to determine locations (horizontal reference).

1.2.1. Geoid:

The true shape of the Earth varies slightly from the mathematically smooth surface of an ellipsoid. Differences in the density of the Earth cause variation in the strength of the gravitational pull, in turn causing regions to bulge above or below a reference ellipsoid. This undulating shape is called a geoid as a representation of the earth's gravity field. Figure 4-1 shows the undulations, greatly exaggerated, in the Earth‘s gravity, and hence the geoid. The Geoid is the equipotential surface of the earth's gravity field which best fits, in a least squares sense, global mean sea level (MSL) used for measuring heights. Essentially this is a representation of the surface of the earth in terms of sea level for every position on earth, in a more complex manner than an ellipsoid. The starting point for measuring these heights is MSL points established at coastal places represent the Geoid.

Note:

A GPS receiver on a ship may, during the course of a long voyage, indicate height variations, even though the ship will always be at sea level (tides not considered). This is because GPS satellites, measure heights relative to a geocentric reference ellipsoid (WGS84). To obtain receiver‘s geoidal height, a raw GPS reading must be corrected. Modern GPS receivers facilitate to implement geoid models inside (e.g. Earth gravitational model 96, EGM-96). These geoid models include the height differences of the geoid with respect to WGS84 ellipsoid. Then the user is able to correct the heights above WGS ellipsoid to the heights above geoid. In that case when the height is not zero on a ship it is because of the tides.

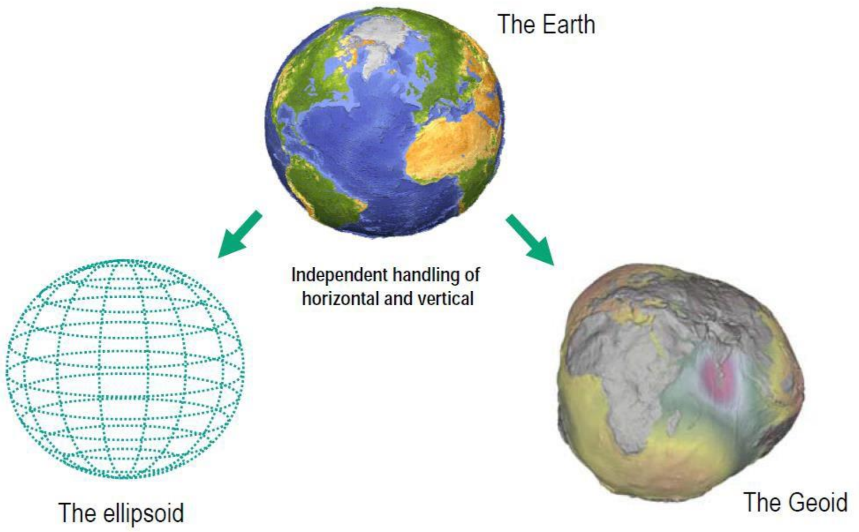


Figure 1.1: The Geoid, exaggerated to illustrate the complexity of its surface and its relationship with the ellipsoid

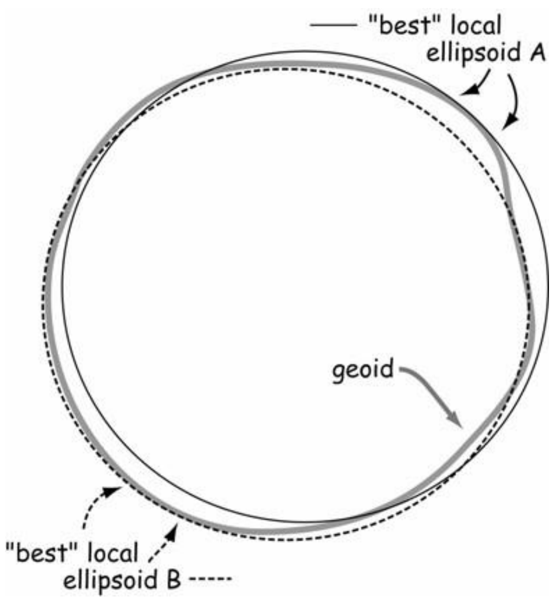


Figure 1.2: An ellipsoid that fits well in one portion of the Earth may fit poorly in another

1.2.2. Reference ellipsoid:

Since both topographic surface and geoid are irregular in shape, and complex in mathematically, they are unsuitable for exact mathematical computations. If the calculations take into account these irregularities, the formulas would be unnecessarily complicated. Therefore, the necessity for the reference model that allows such topographic irregularities to be recorded was in demand. Then the mathematical representation or horizontal reference of this physical earth was generated by rotating an ellipse about its shorter axis (minor axis) and is called the reference ellipsoid.

The reference ellipsoid can be defined by using two parameters; the semimajor axis and flattening. The size is represented by the radius at the equator; the semi-major axis of the cross-sectional ellipse and designated by the letter a. The shape of the ellipsoid is given by the flattening, f, which indicates how much the ellipsoid departs from spherical shape. Flattening ‘f ‘indicates how much the ellipsoid departs from spherical shape.

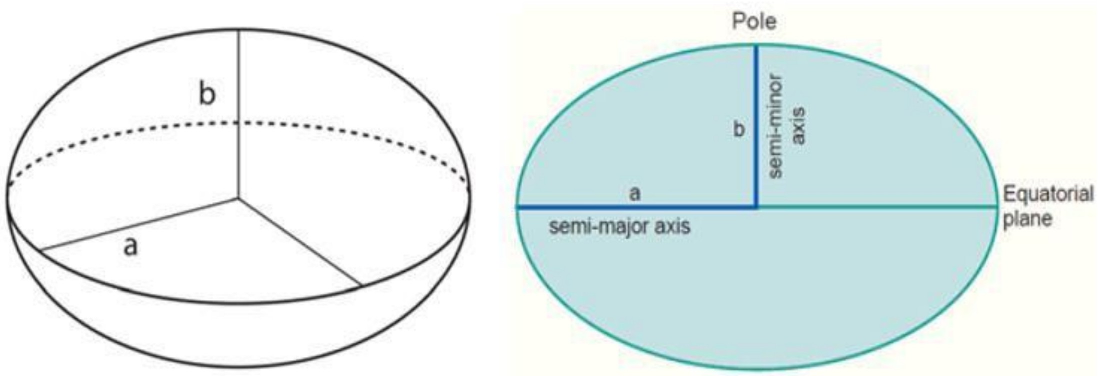


Figure 1.3: A cross section of an ellipsoid, used to represent the Earth surface, defined by its semi-major axis ‘a ‘and semi-minor axis ‘b ‘.

The flattening is the ratio of the semimajor axis a, minus the semi minor axis b, to the semimajor axis; and often expressed as a fraction - written as:

$$f = \frac{a-b}{a}; OR \quad f = 1 - \left(\frac{a}{b}\right)$$

The Earth is slightly flattened, such that the distance between the poles is about 1 part in 300 less than the diameter at the Equator. The ellipsoid also defined by its semi-major axis a, and eccentricity e. Eccentricity measures how much an ellipse deviates from a true circle.

$$e^2 = (1 - (\frac{b^2}{a^2})) = \frac{(a^2 - b^2)}{a^2} = 2f - f^2$$

Eccentricity is measured as the square root of the quantity 1.0 minus the square of the ratio of the semi minor axis to the semimajor axis. It measures how much an ellipse deviates from a true circle. Typical parameter values of an ellipsoid are:

$a = 6378137.0m$ $b = 6356752.31m$
$f = 1/298.26$ $e = 0.0818187$

Flattening: $f = (a-b)/a$ Eccentricity: $e^2 = (a^2 - b^2)/a^2$

1.2.3. The ellipsoid as horizontal reference surface:

The ellipsoid is a horizontal reference frame; known as horizontal datum or geodetic datum used to measure locations of points of interest in terms of latitude (□) and longitude (λ). These locations on the ellipsoid are then projected onto a mapping plane. Different ellipsoids were adopted in
various parts of the world, primarily because there were different sets of measurements used in each region or continent. There are locally fitted ellipsoids than using one reference ellipsoid for the entire globe. There are many different ellipsoids defined in the world.

For the global measurements, it is ideal to use global reference ellipsoid which fits as good as possible to entire globe. It approximates the earth as a mean ellipsoid (or sometimes alternatively uses the world spheroid). The International Union for Geodesy and Geophysics (IUGG) plays a central role in establishing these reference figures. Rather than using one reference ellipsoid for entire globe, there are some locally fitted ellipsoids, that is; they best fit over some specific areas only. They have been established to fit the geoid (mean sea level) well over an area of local interest. Therefore, when selecting the ellipsoid for any reference, ellipsoid is chosen such a way that it best fits the surface of the area of the interest. Table 4-1showns some well-known reference systems and their associated ellipsoids.

Table 1.1: Examples of reference systems and associated ellipsoids

Reference system	Ellipsoid	Semi-Major axis a	Semi-Minor axis b (m)	Flattening (1/f)
WGS 84	WGS 84	6378137.0	6,356,752.30	1/298.26
NAD 83	GRS 80	6378137.0	6,356,752.30	1/298.26
NAD 27	Clarke 1866	6378206.4	6,356,583.80	1/294.98

1.2.4. The geoid as vertical reference surface

The equipotential surface, the geoid is used to describe heights. In order to establish the geoid, the ocean's water level is monitored at coastal areas over several years using tide gauges and the resultant water level is the approximation to the geoid or mean sea level.

Since this height refers the geoid or mean sea level, it is called orthometric height. Heights above the ellipsoid are often referred to as ellipsoidal height (h). The difference between the ellipsoidal height (h) and geoidal height (N) at any location is called geoidal separation/undulation. Ellipsoidal heights have to be adjusted before they can be compared to orthometric (mean sea level) heights using Geoid undulations (N).

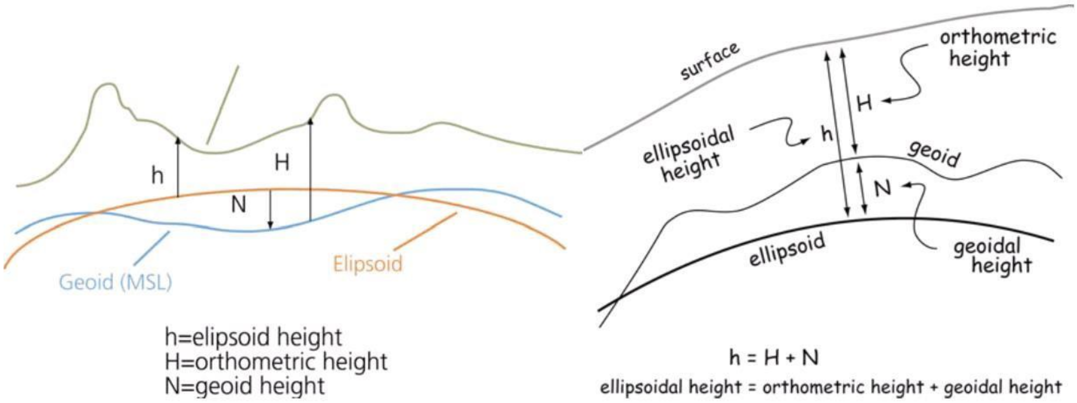


Figure 1.4: Ellipsoidal, orthometric, and geoidal height are interrelated

Nowadays, gravity satellite missions help to derive global mean sea level surfaces with the accuracy of few centimeters. One of the most common models is called ―Earth gravitational model 96 (EGM96). Data has been collected from ESR-1 and GeoSAT missions to derive this model. This can be used to calculate the undulations better than one-meter accuracy.

1.3. Datum

The ellipsoidal figures of the earth or the reference ellipsoids have different origins and orientations with respect to their area of interest. A datum defines the position of the ellipsoid relative to the centre of the earth including the ellipsoid; its origin and orientation; and reference frame (defines axes).

Table 4-2: Datums and their principle areas of use

Datum	**Area**	**Origin**	**Ellipsoid**
WGS 1984	Global	Earth center of mass	WGS 84
NAD 1983	North America, Caribbean	Earth center of mass	GRS 80
NAD 1927	North America	Meades Ranch	Clarke 1866
European 1950	Europe, Middle East, North Africa	Potsdam	International

In general, a datum provides a frame of reference for measuring locations on the surface of the earth and it defines the origin and orientation of latitude and longitude lines. A geodetic datum is a set of constants specifying the coordinate system used for geodetic control, i.e. for calculating coordinates of points on the earth‖. These constants include parameters to specify the location of the origin of the coordinate system, the orientation of the coordinate systems and the reference ellipsoid. There are hundreds of locally-developed reference datums around the world. The most widely used datum is WGS 1984. It serves as the framework for worldwide positional measurements. There are two types of datums: local and global datums.

1.3.1. Local datum

With respect to the local reference ellipsoids, local datums have been developed. The origin of the coordinate systems is selected in such a way that the best matched point on the surface of the ellipsoid to closely fit a particular position on the surface of the earth. This point is the origin point of the datum. The coordinates of the origin point are fixed, and all other points are calculated from it. For example, Ethiopia uses local datum known as Adindan, in which the reference ellipsoid or spheroid is Clark 1880. It is non-geocentric datum with shift of the origin $\Delta x = 165$, $\Delta y = 11$ and $\Delta Z = -206$. This datum is located in Southern Egypt and used by six African countries.

1.3.2. Global datum

A global datum is the reference ellipsoid and its center coincides with the center of the earth. Satellite data has provided geodesists with new measurements to define the best earth-fitting spheroid, which relates coordinates to the earth's center of mass. The global datums use the earth's center of mass as its origin. The axes are oriented in such a way that the Z-axis directed to the mean rotation axis of the earth, the X-axis towards the intersection of Greenwich meridian and equator from the origin and Y-axis obeys the right hand rule with respective Z- and X-axis (Figure 4-4). There are two types of Global datums.

1.3.2.1. International Terrestrial Reference Frame (ITRF)

ITRF is the most precise Earth-Centered-Earth-Fixed (ECEF) datum maintained by the International Earth Rotation and Reference Systems Service (IERS). An extensive global network of accurate coordinate points derived from geodetic observations using Integrated GPS Satellites (Figure 4-4 (b)) has realized it. ITRF is based on the GRS80 (a geocentric ellipsoid), designed to approximate the geoid on a global scale.

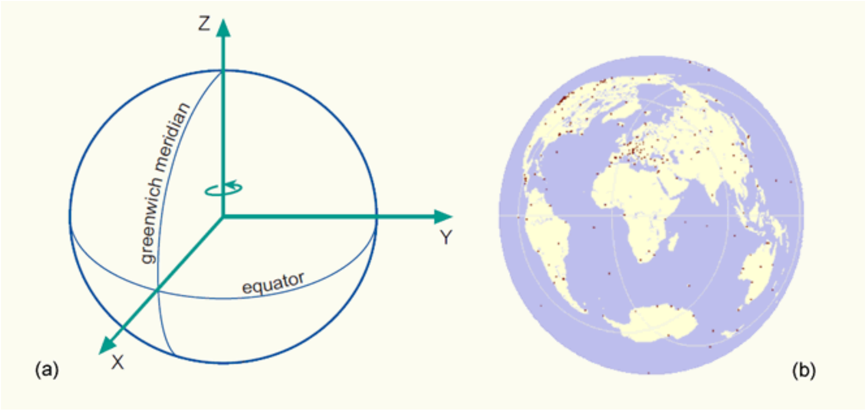


Figure 1.5: ITRF visualized as a distributed set of ground control stations represented by points (right)

1.3.2.2. World Geodetic System 1984 (WGS84):

WGS84 ellipsoid datum is assumed to be identical with the GRS80 (Table 1.1). It is the most recently developed and widely used datum. This datum has been refined several times to coincide with the current ITRF2000 within a few centimeters at the global level for all mapping and charting purposes (2002). GIS uses WGS84 as a reference coordinate system. Satellite-based positioning equipment such as GPS helps to determine heights with an accuracy of a few centimeters w.r.t a reference ellipsoid (e.g. WGS84). It serves as the framework for positional measurements worldwide.

1.4. Coordinates systems

Coordinate systems are based on mathematical rules used to measure distances and/or angles in order to identify the location of points by means of unique sets of numerical values defined in

3D or 2D reference ellipsoid spaces. On the ellipsoid, positions are either expressed in Cartesian coordinates (X, Y, Z) or in curvilinear coordinates (ϕ, λ, h), i.e. geodetic latitude, longitude and ellipsoidal height (Figure 1.5B). The purpose is to enable geographic datasets to use common locations of reference system. Each coordinate system is defined by:

- ✓ Its origin, the orientation of the axes, and units of measurement (typically feet or meters for projected coordinate systems; or by decimals or degrees of latitudes–longitudes for GCS);
- ✓ The datum and projection parameters such as one or more standard parallels, a central meridian, and the shifts in the X- and Y-directions.

There are two types of coordinate systems in use: geographic or global coordinate systems and projected or planar coordinate systems.

1.4.1. Geographic or global coordinate system (GCS)

Geographic coordinate system is used to define locations either in a 3D space or in 2D on the Earth‘s reference surfaces (ellipsoid or sphere). A GCS is often incorrectly called a datum, but a datum is only one part of a GCS. In fact, a GCS includes an angular unit of measure, a prime meridian, and a datum. The concept of geographic coordinates is applied to a sphere as the reference surface. Geographic coordinate systems consist of latitude (phi or ϕ) which varies from north to south and longitude (lambda or λ) which varies from east to west.

- ✓ The latitude (ϕ) is the angle between the ellipsoidal normal and the equatorial plane. Lines of equal latitude are called parallels. They form circles on the surface of the ellipsoid. Latitude is zero on the equator (ϕ = 0°), and increases towards the two poles to maximum values of ϕ = +90 (90°N) at the North Pole and ϕ = - 90° (90°S) at the South Pole.
- ✓ The longitude (λ) is the angle between the equatorial plane from the meridian of Greenwich λ = 0°) either eastwards through λ = + 180° (180°E) or westwards through λ = -180° (180°W). Lines of equal longitude are called meridians and they form ellipses (meridian ellipses) on the ellipsoid. The prime meridian of longitudes is called the International Date Line, which passes through Greenwich, England. Both lines form the graticule when projected onto a plane.

The necessary parameters to determine locations using either 2D or 3D are:

- ✓ Geographical latitude (ϕ), longitude (λ), ellipsoidal height (h), and
- ✓ Geocentric or Cartesian X, Y and Z coordinates.

3D geographic coordinates (ϕ, λ, *h*) are obtained by introducing the ellipsoidal height ***h*** *to the system; and can be used to define a position on the surface of the Earth (point ‗P' of Figure 1.5 (B))*

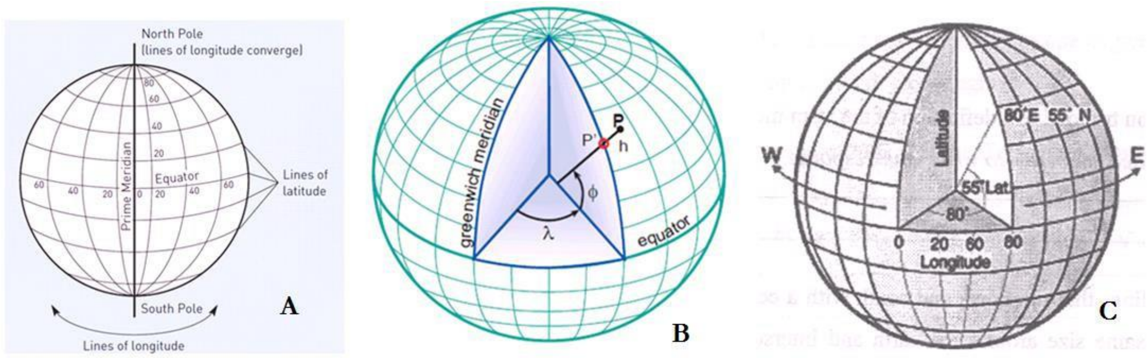


Figure 1.6: The latitude (ϕ) and longitude (λ) angles and the ellipsoidal height (h) represent the 3D GCS

An alternative method of defining a 3D position on the surface of the Earth is by means of geocentric coordinates (X, Y, Z), also known as 3D Cartesian coordinates. The system has its origin at the mass-center of the Earth with the X- and Y-axes in the plane of the equator. The X-axis passes through the meridian of Greenwich, and the Z-axis coincides with the Earth's axis of rotation. The three axes are mutually orthogonal and form a right-handed system. Geocentric coordinates can be used to define a position on the surface of the Earth (point P Figure 1.6).

Spherical coordinates are most often recorded in a degrees-minutes-seconds (DMS), for example N42º 35‘ 20”’, signifying 42 degrees, 35 minutes, and 20 seconds of latitude. Minutes and seconds range from 0 to 60. Alternatively, spherical coordinates may be expressed as decimal degrees (DD). For example, Figure 1.5 (C) shows a geographic coordinate system, where a location is represented by the coordinate longitude 80° E and latitude 55° N. A point with longitude can reference any location on earth and latitude coordinates. The values can have measurements units of decimal degrees (DD); degrees-minutes-seconds (DMS); linear unit (meter), and so on.

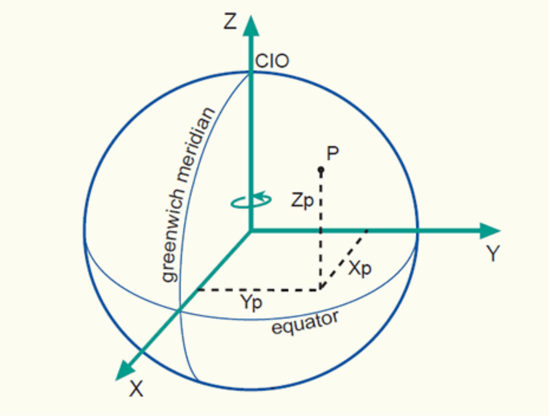


Figure 4-7: An illustration of the geocentric coordinate system