UNIT 8 DATUMS, MAP PROJECTIONS AND COORDINATE SYSTEMS

Structure

8.1 INTRODUCTION

You have studied about maps in your school text books and you know that mapping involves determining the geographic locations of features on the

Earth. This is followed by transforming them into positions on a flat map using map projection. The idea of depicting the Earth on a flat map is easy to imagine and even simple to understand but the actual process of a map projection is really a complex task. The Earth is a complex three-dimensional object with physical dimensions, including height, width, depth, mass, and density. An important physical characteristic is its non-straight, curvilinear surface. This curvilinear property adds complexity to a map projection's transformation of the Earth model or globe to a flat surface. Due to the Earth's curvilinear surface, map projections always hold a degree of distortion. In this unit, we will discuss about the concepts involved in transforming three dimensional Earth into two dimensional maps on paper. First, we will discuss about the concepts of geoid, ellipsoids, datum and then about the types of map projections and their characteristics. At the end of the unit, we shall briefly discuss about how to choose an appropriate projection system for your area of interest.

Objectives

After reading this unit, you should be able to:

- discuss about the basics of geodesy, Earth model and datum;
- define the coordinate systems;
- describe the map projection and its classification;
- list different types of distortion and measures to minimise them; and
- choose a suitable map projection.

8.2 GEODESY

Geodesy, also known as *geodetics,* is the scientific discipline that deals with the measurement and representation of the Earth in geometrical and gravimetrical sense. ***Geodesy*** is defined as the science concerned with study of shape and size of the Earth. In geometric sense, it deals with the study of shape and size of Earth's surface, the measurement of the position and motion of points on the surface. In gravimetric sense, it deals with the study of equipotential surfaces of the gravity potential. Every geodetic measurement is a function of gravity field which may vary from point to point. Unlike plane survey, geodetic survey involves complicated three-dimensional spherical geometry.

Geometry deals with the shape, size, relative position of figures and the properties of shape.

Gravimetry is the measurement of the strength of a gravitational field of the Earth.

There are two types of surveys i.e. plane survey and geodetic survey, that may be used in compiling large scale maps. Plane (or cadastral) survey, is commonly performed for a limited area. Thus, Earth's curvature may not be taken into account because it is relatively insignificant over a small area. The lines of a plane survey are determined from ground observations, and are mapped as observed without being referred to as a spheroid. On the other hand, topographic maps are based on a framework developed by geodetic survey, and the ground observations are referred to as a spheroid as it involves a large area. Consequently, the two kinds of survey usually do not match.

Here, equipotential/ isopotential surface refers to points having uniform gravity potential on Earth.

The study of geodesy began with mere curiosity and the never-ending human will to explain the Earth's unknowns through logic. There exists the first remnant of geodetic analysis starting as far back as the early Greeks. Homer, Pythagoras, Plato, and others all had put efforts to describe the shape of the Earth. Homer (8th century B.C.), for instance, held the idea that Earth was a large flat disc, while Pythagoras (6th century B.C.), a mathematician, viewed Earth as a perfect sphere. Aristotle's (4th century B.C.) arguments for a spherical Earth was more compelling as he noted that sailing ships always disappear into the horizon rather than becoming smaller dots. You will be interested to know that ancient Indians contributed significantly to geodesy. Indian mathematician Aryabhatta (476 to 550 A.D.) was the first person to determine the circumference of the Earth with a remarkable accuracy. The discovery that Earth rotates on its own axis from west to east is described in his famous book *'Aryabhatiya'*.

Until about seventeenth century, the Earth was thought to be a perfect sphere. The change came around 1670, as a consequence of Newton's theory of gravity. He proposed that earth's shape slightly bulged out at the equator due to the greater centrifugal force generated by rotation. The equator bulging produces slight flattening at the poles.

8.3 EARTH MODELS

Earth's rugged, irregular surface is difficult to be explained by mathematics for calculating the positions of Earth features. To effectively represent the shape and size of the Earth in scientific and real-life applications, a calculable, formula-driven model of the Earth is required. The closer a model comes to the actual surface of the Earth, the better it is for geographic positioning.

Geoid (meaning earthlike) is an approximated figure of the Earth. In more technical terms, it is a sea level equipotential surface - the surface on which gravitational potential remains equivalent to its strength at mean sea level. However, it is not a smooth surface. The gravitational pull is not uniform throughout the Earth surface. This is mainly because of density variation inside the Earth. For example, gravitational pull of the Earth is stronger in areas rich with iron and other dense materials. This results in undulating geoid surface which is not smooth and regular.

It is clear that Earth is not a perfect sphere, but rather slightly ellipsoidal in shape. For small scale mapping, shape of the Earth may be considered as spherical instead of ellipsoid. The spherical model of Earth is called ***Authalic Sphere***, which may be defined as a sphere of the same surface area as that of ellipsoid. Cartographers use authalic sphere for mapping on a small scale, i.e., 1:1 million or smaller.

A large scale map shows a small land area in great detail whereas small scale map shows less detail but a larger land area.

Today's precision GIS is built upon the advances in modern geodesy and the refreshed knowledge of the Earth's true shape. With satellite imagery and modern physical science, it is evidently established that the Earth is a nearly spherical object that bulges in the middle and flattened at the poles. In this regard, ellipsoid can best represent an Earth as it poses a near resemblance to its shape without the inherent complexities.

8.4 ELLIPSOID MODELS

You have read that Earth could be represented as ellipsoid. The Earth's ellipsoid is an ellipse rotated upon its minor axis, which is functionally called the *axis of rotation* or *axis of revolution*. The purpose of a mathematical model is to simplify calculations. Since an ellipse is a two-dimensional shape and an ellipsoid is a more complex three-dimensional object, we utilise the figure of an ellipse to achieve this mathematical simplicity (Fig.8.1).

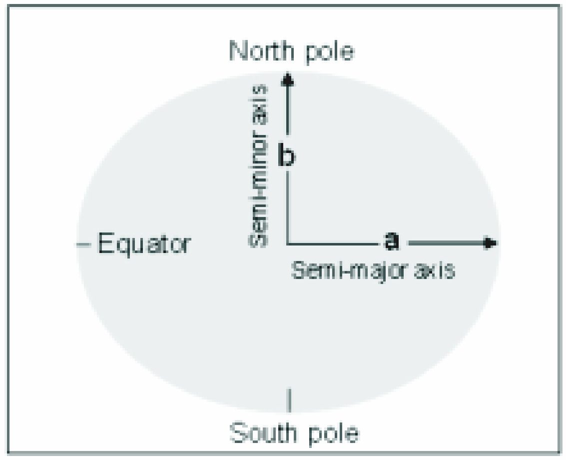


Fig. 8.1: Shape of an oblate ellipsoid

Rotating this ellipse about the polar axis (Fig. 8.2) would outline a three-dimensional model of the Earth called an ***oblate ellipsoid*** (also called *oblate spheroid*).

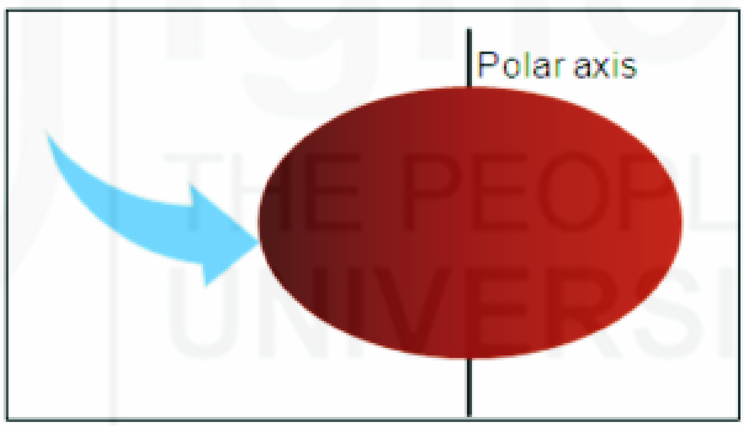


Fig. 8.2: Rotating ellipse about the polar axis in three dimensions

The ellipsoid's flattening causes two unequal axes:

i) a longer axis, and

ii) a shorter axis.

The north-to-south axis through the Earth's core is the shorter axis and, as such, is called the ***minor axis*** or ***polar axis***. The east-to-west axis through the Earth's core is longer and is called the ***major axis*** or ***equatorial axis***. In brief, the flattening of the ellipse is directly related to the differences in both the semi-major axis (half of the major axis) 'a? and semi-minor axis (half of the minor axis) 'b?. It is represented by the formula

Flattening (f) = $(a – b) / a$

Earth's flattening is sometimes displayed as a reciprocal called the *inverse flattening*. For instance, a flattening of 0.003389831 (or 1/295) can be portrayed as an inverse flattening of 295. The formula for inverse flattening is represented as:

Inverse Flattening (f^{-1}) = $1 / f$

$$= a / (a - b)$$

Fig. 8.3 illustrates different flattening scenarios. However, in the case of the Earth, the polar axis measurement is only slightly shorter than the east-to-west equatorial axis measurement. Newton in the seventeenth century had predicted the flattening to be about 1/300th of the equatorial axis. And present day measurements show, it as 1/298th of the equatorial axis.

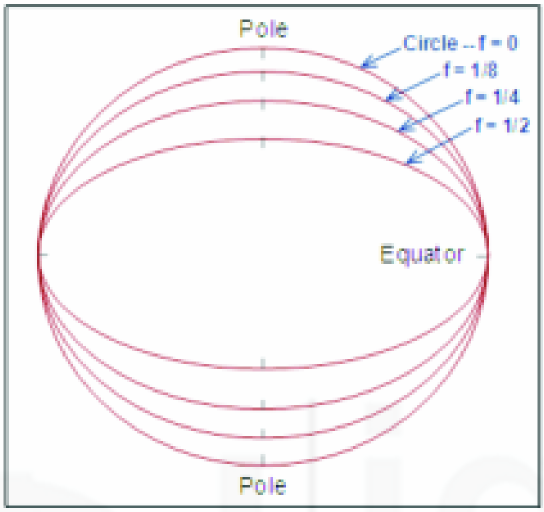


Fig. 8.3: Various levels of flattening (f)

8.5 USE OF GEOID, SPHERE AND ELLIPSOID

After reading the previous sections of this unit, you are now aware of the three approximations i.e. *Geoid*, *Sphere* and *Ellipsoid* are used to define Earth's true shape in different ways. The geoid is considered as a reference from which elevations or heights can be measured. It is the reference surface for ground survey. The horizontal and vertical positions are mapped with reference to the geoid surface. Horizontal positions are later adjusted to the ellipsoid surface, because the irregularities on the geoid surface would make projection and other mathematical computations extremely complex. On the other hand, elevations are determined with reference to the geoid surface.

The authalic sphere is the reference surface for small-scale maps of countries, continents and larger areas. This is because the difference between sphere and ellipsoid is negligible while mapping large areas on page size maps. There is also a significant increase in complexity of map projection equations for the ellipsoid than a sphere. In addition, the spherical and ellipsoidal equations for a particular map projection provide the same results for the small scale (1:1 million) maps. A sphere proves adequate along the equator (equatorial plane) but fails at locations closer to the poles. With large scale (1:10,000) maps, however, the difference between spherical and ellipsoidal approximations can be significant, and hence we need to take the Earth's oblateness into account. The ellipsoid, also referred to as a spheroid, is a much better approximation for the shape of the Earth than the sphere; the poles are slightly flattened and the equator bulges. Unlike the sphere, the ellipsoid can handle dissimilar dimensions.

Spend 5 mins

Check Your Progress I

1) Earth can be represented as ....

2) .... is the equipotential surface which would coincide exactly with the mean ocean surface of the Earth, if the oceans were in equilibrium.

3) .... is the hypothetical non-spherical shape of the Earth resembling flattening the Earth at Poles.

8.6 WHICH IS THE BEST REFERENCE ELLIPSOID FOR YOU?

Basically, every GIS and geographic information product is based upon a reference ellipsoid, which is defined as a ***standard ellipsoid*** with measured parameters. Over the years, numerous reference ellipsoids have been developed by various geodesists with measurements taken at different source points on Earth. From 1800 to till date, at least 20 determinations of Earth's radii and flattening have been attempted from measurements taken widely at different locations. The foremost purpose of these various determinations is to develop a best fit model for specific part of the Earth compromising accuracy over other parts. Values for 10 different ellipsoids used as basis for mapping in various parts of the world are listed in Table 8.1.

Table 8.1: Ellipsoids (source: Snyder, 1995)

Name	Year	Semi-major axis 'a' (meters)	Semi-minor axis 'b' (meters)	Polar Flattening
Airy	1830	6,377,563.40	6,356,256.91	1/299.32
Everest	1830	6,377,299.37	6,356,098.36	1/300.8
Bessel	1841	6,377,276.30	6,356,075.40	1/299.15
Clarke	1866	6,378,206.40	656,583.80	1/294.98
Clarke	1880	6,378,249.10	656,514.90	1/293.46
International	1924	6,378,388.00	6,356,911.90	1/297
Krasovsky	1940	6,378,245.00	6,356,863.00	1/298.3
Australian	1965	6,378,160.00	6,356,774.70	1/298.25
WGS 72	1972	63,781,335.00	656,750.50	1/298.26
WGS 84	1984	6,378,137.00	6,356,752.30	1/298.257

These reference ellipsoids are only slightly dissimilar. Because of the differences in source data locations, certain reference ellipsoids work better than others for certain applications and for certain regions. For example, you should choose reference ellipsoid Everest 1830 for Asian GIS products because this ellipsoid is more suited for Asian region, whereas for other regions some other reference ellipsoid may be more suitable (Fig. 8.4).

Unlike Everest 1830 and other major area specific ellipsoids, WGS84 is more universal. As a result, it is one of the most widely used reference ellipsoids. It is determined from satellite orbital data and considered more accurate than the earlier ground measurement determinations. Although, it gives the best fit ellipsoid for the entire Earth, it may not offer best fit for a particular part of the Earth.

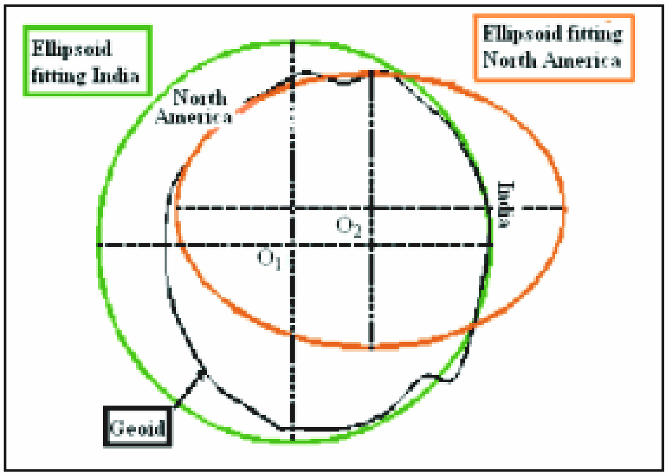


Fig. 8.4: Reference ellipsoid's mismatch

8.7 CONCEPT OF DATUM

Let us now discuss about the concept of datum. Datum describes the position, orientation and scale relationship of a reference ellipsoid to Earth. In surveying and geodesy, a datum is a reference on the Earth's surface against which positions are measured. Datum also defines the origin of coordinate system from where the measurements are made. A datum can be associated with one or more Earth model(s) for determining the positions.

There are hundreds of locally developed reference datums around the world, usually referenced to some convenient local reference point. A specific point on the Earth can have substantially different coordinates depending on the datum used to make the measurement. Referencing geodetic coordinates to the wrong datum can result in position errors of hundreds of meters. Different nations and agencies use different datums as the basis for coordinate systems used to describe positions. Present-day datums, based on increasingly accurate measurements of the Earth, are intended to cover larger areas.

Geodetic datums are generally classified as either a ***geocentric datum*** or a ***local geodetic datum***. A geocentric datum is globally centered and adequately approximates the Earth's size and shape as a whole. The center of the reference ellipsoid coincides with the Earth's center of mass as you can see in Fig. 8.5.

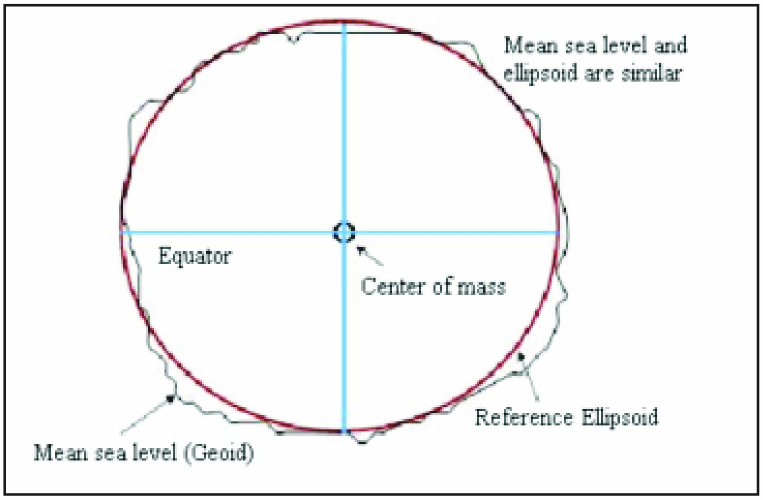


Fig. 8.5: The geocentric datum

Mean Sea Level refers to the average height of ocean surface between the mean high tide and mean low tide levels.

A local geodetic datum is used to characterise a particular region where the reference ellipsoid and the Earth's shape coincide. As you can visualise in Fig.8.6, the center of the ellipsoid is often located away from the Earth's center of mass, thereby providing a poor representation of the Earth as a whole.

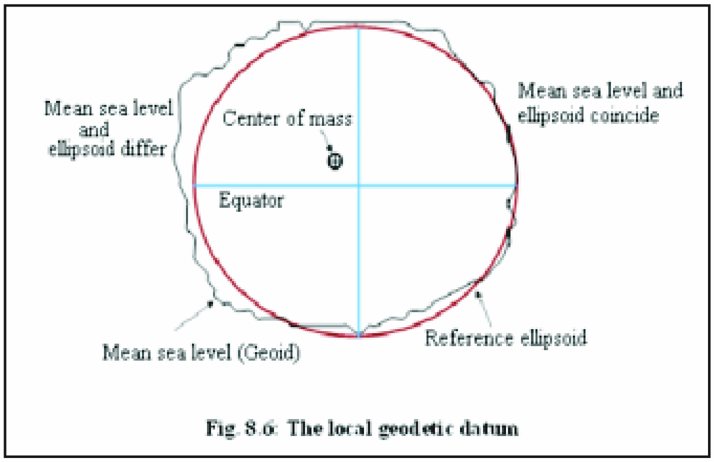


Fig. 8.6: The local geodetic datum

The classification of geodetic datums offers a degree of user control as it relates to which datum suits the application or location best. This form of user control surrounds the act of selecting a location specific datum based on individual characteristics, rather than using a more general datum. There are following two types of datums:

- Horizontal Datum and
- Vertical Datum.

Now, we shall briefly discuss about the two types of datums.

8.7.1 Horizontal Datum

Horizontal datums empower the users with a geographic coordinate system for determining accurate point positions. As we already understand, it can take on both geocentric and local geodetic forms. Standard horizontal datums cover large geographic areas. They are used throughout the world for simplicity and local datum reference (e.g. NAD27, ED50, Modified Everest, etc.). The World Geodetic System 1984 (WGS84) is considered a global datum that defines a fixed global reference frame for the Earth. WGS84 is based upon the ellipsoidal parameters of the Geodetic Reference System of 1980. WGS84 is used for GPS satellites, and is defined by a gravity model of the Earth.

8.7.2 Vertical Datum

Bathymetry is the study of underwater depth of lake or ocean floors as opposed to elevation which describe the height of the Earth surface from a reference level i.e. Mean Sea Level.

Just as a horizontal control datum requires a reference grid on which all coordinates abide, a vertical datum demands a contiguous, consistent reference from which topographical elevation and bathymetric depth are measured. The vertical datum is used to define the positional height as it relates to a determined reference. How can one accurately measure heights on the Earth's surface? What could constitute a valid reference for precise vertical positioning?

In many day-to-day uses, elevations are referred to as height above or below sea level. Height measurements referenced to the ellipsoid are less accurate than when referenced to the sea level. Additionally, since tides are constantly changing, rising and falling, the term *sea level* is considered the average of the tide levels (Fig. 8.7). In terms of datums, we call this average sea level as the *mean sea level* (MSL).

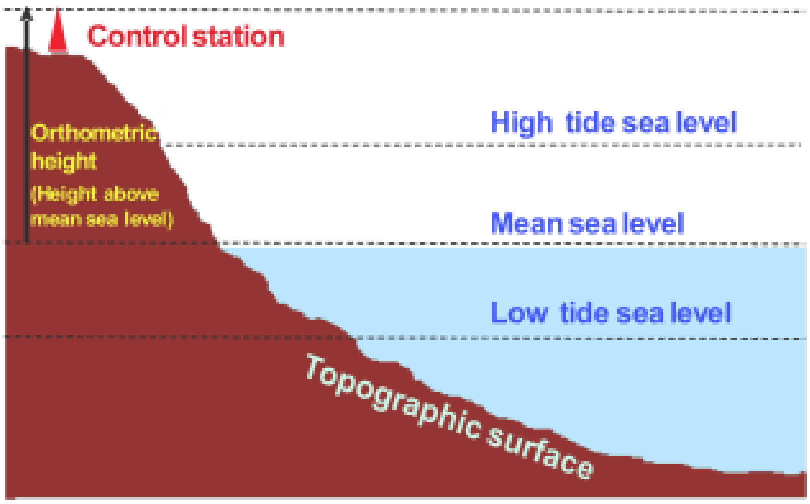


Fig. 8.7: The relationship between sea level and height measurements

The height measurement is the distance from the control station to the geoid (MSL). This is generally called as the *mean sea level elevation* or known formally as the *orthometric height*. To understand the relationship between these reference surfaces, we must examine the separations between the accurate but complex geoid (MSL) and the general but simple ellipsoid. These surface differences are called ***geoid-ellipsoid separations*** or ***undulations.*** The geoid-ellipsoid separation is an important factor in defining the vertical position as it relates to the vertical datum. GPS and certain datums reference the ellipsoid, while other vertical datums reference the geoid. Hence, a firm understanding of their relationship is needed to link datums and determine precise vertical positioning. You can better understand the relationship between the geoidal, ellipsoidal, and topographical surfaces from Fig. 8.8.

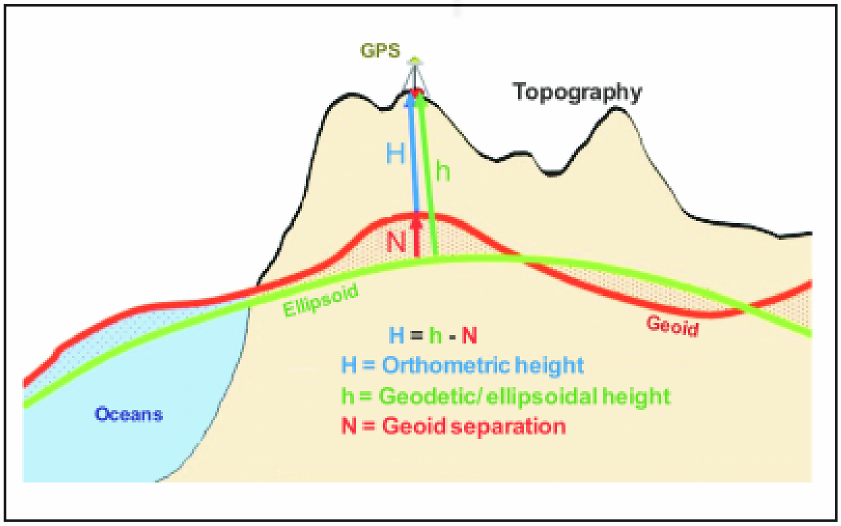


Fig. 8.8: The geoid-ellipsoid-topography relationship (modified from http://priabroy.wordpress.com)

8.8 COORDINATE SYSTEMS

You will require a coordinate system in order to locate points precisely as well as measure distance and direction correctly. A coordinate is a number set that denotes a specific location within a reference system. A coordinate system is often structured in a two-dimensional or three-dimensional plane that consists of a set of reference points and rules that define the spatial position of points.

An ellipsoid, datum, a projection and units make up a coordinate system.

In general, there are following two types of coordinate systems:

- Geographic coordinate system, and
- Rectangular or planar coordinate system

We shall now discuss about them in detail and while reading the following sections you should refer to Figs. 8.9, 8.10 and Fig. 8.11.

8.8.1 Geographic Coordinate System

Geographic coordinate system is a three-dimensional coordinate system that enables every location on the Earth to be specified by a set of numbers. Locations on the Earth surface are defined by three numbers, *latitude* (also known as parallel), *longitude* (also known as meridian), and *ellipsoidal height*. It is the primary system and is used for all basic mapping, such as navigation and fundamental surveying. This system is useful for locating positions on the curved surface of the Earth. You can see Earth's latitude and longitude lines in Fig. 8.9.

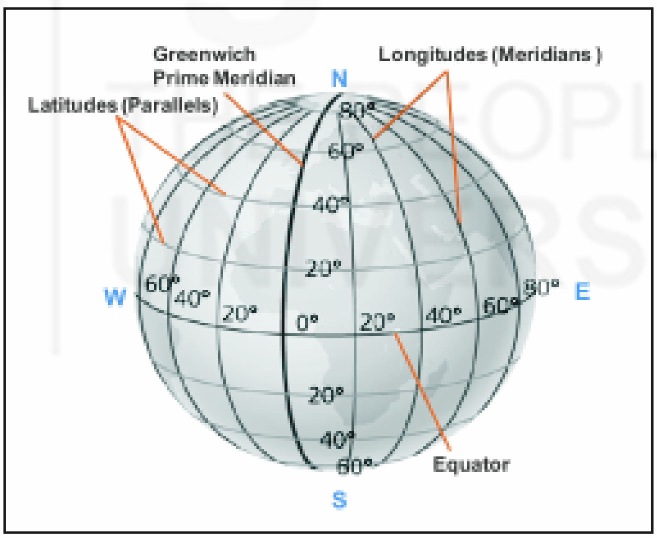


Fig. 8.9: Earth's parallels, meridians and origin of geographic coordinate system

The geographic coordinate system is defined on Earth's three-dimensional ellipsoid or on the authalic sphere through a series of horizontal and vertical reference lines. As you know, the vertical lines are called *longitude*. Since longitudes meet at the poles, they are often called *meridians*. The horizontal lines are called *latitudes*. Latitude lines are parallel to each other, and are therefore often called *parallels*.

Earth's parallels intersect every meridian at right angles. The central parallel, located exactly between the poles, is called the *equator*. This central, standard parallel holds a latitudinal value of 0° (Fig. 8.10). The remaining parallels are assigned angular values relative to their direction and distance from the equator. The lines of latitude progressing toward the North Pole are assigned parallels ranging from 0° to 90° North. Likewise, the lines of latitude progressing from the equator toward the South Pole are assigned parallels ranging from 0° to 90° South.

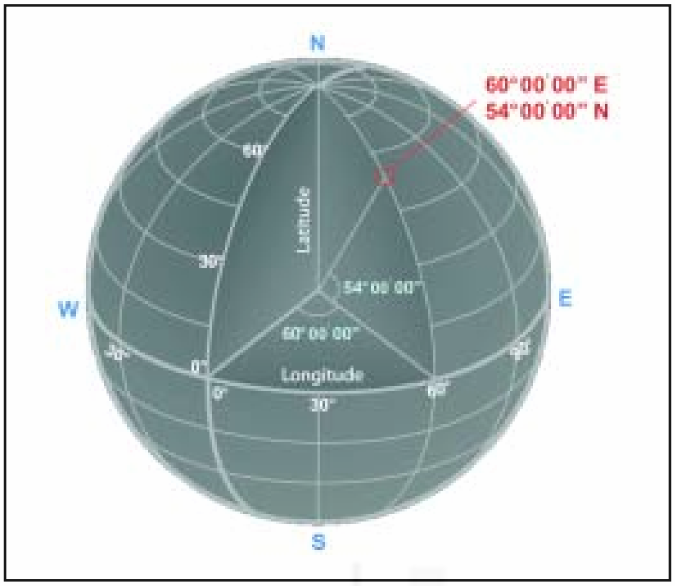


Fig. 8.10: Angular units of latitude and longitude

Likewise, the *Greenwich meridian* is called the *prime meridian* and holds a longitudinal value of 0°. The lines of longitudes progressing eastward from the prime meridian are assigned longitudes from 0° to 180° E (otherwise 0° to +180°) and the lines of longitude progressing westward are assigned longitude from 0° to 180° W (otherwise 0° to -180°).

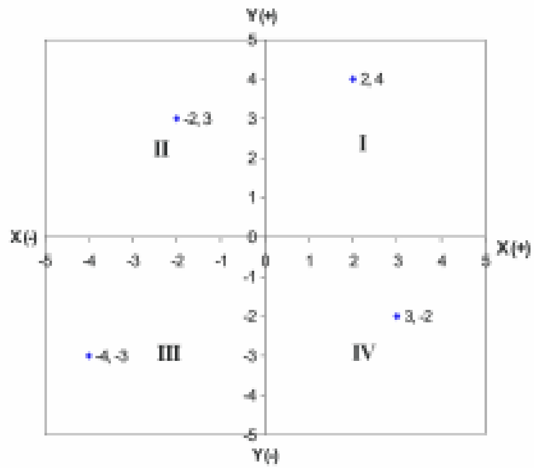
8.8.2 Planar Coordinate System

Planar coordinate system is used to locate positions on a flat map representing Earth's curved surface. It is evolved from the Cartesian coordinates. *Cartesian coordinate* system is the most popularly used reference system in mathematics, science, and GIS. The Cartesian system is a reference structure in which point positions are measured along intersecting planes in two and three dimensions. The intersection point for all planes is called the *origin*. All measurement, distance increments are consistent on every intersecting plane and throughout the system. These planes are detailed along the X, Y, and/or Z axes.

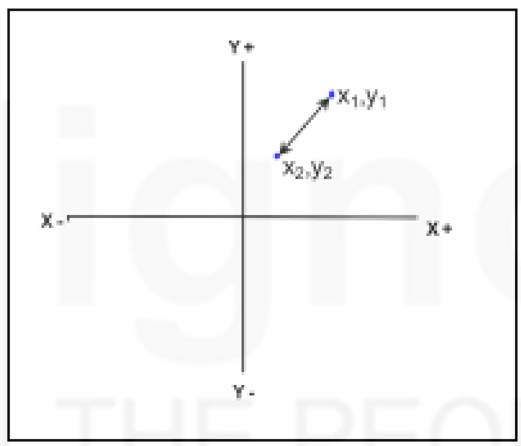
The Cartesian two-dimensional coordinate system involves two axes: the horizontal X axis and the vertical Y axis. These axes are subdivided into units of the specified distance from the origin (0, 0). This planar, two-dimensional model is divided into four quadrants by the perpendicular, intersecting axes. These are labelled as quadrants I through IV, starting in the upper right-hand quadrant and progressing counter clockwise as you can see in Fig. 8.11(a). These quadrants take on the following structures:

- Quadrant I: x axis is positive; y axis is positive (+x, +y);
- Quadrant II: x axis is negative; y axis is positive (–x, +y);
- Quadrant III: x axis is negative; y axis is negative (–x, –y); and

- Quadrant IV: x axis is positive; y axis is negative (+x, –y).



(a)



(b)

Fig. 8.11: (a) Two-dimensional Cartesian coordinate system; (b) Distance between two points in a Cartesian coordinate system

Cartesian coordinate system is used on maps to take advantage of the simplified calculations. Calculating distances and direction on plane is very basic, which you have studied in school. However, calculating distances and directions on a sphere or on Earth's surface are not that simple because of the complex spherical trigonometry functions. Therefore, working with a planar surface simplifies calculations and promotes the use of maps. This can be used for applications at different scales. As you can see in Fig. 8.11(b), the distance d_{12} between two points in a plane is given by

$$d_{12} = \{(x_1 - x_2)^2 + (y_1 - y_2)^2\}^{1/2}$$

where (x_1, y_1) and (x_2, y_2) are the coordinates of two points in a Cartesian plane.

You should keep in mind that the value of x and y should be taken with proper arithmetical signs. For example, points falling in quadrant I will have both x and y values positive, whereas, points falling in quadrant II will have x value negative and y value positive, and so on. If arithmetical signs are not properly used, the distance calculated using the above formula will not be correct.

Check Your Progress II

Spend 5 mins

1) List the elements of geographic coordinate system.

....

....

....

....

....

2) Geodetic datums are classified as ....

8.9 MAP PROJECTIONS

You have already read about geoid, ellipsoid and coordinate system, now we shall discuss about map projection. Map projection is a systematic transformation that allows the orderly representation of Earth's spherical graticules on a flat map. In other sense, it is a transformation of geographic coordinates (latitude, longitude) into the Cartesian (x, y) coordinate plane of the map sheets. It portrays a three-dimensional object, such as the Earth's globe, in a two-dimensional format. Map projections, by default, are not true or accurate portrayals of the globe because a two-dimensional plane cannot accurately represent large portions of the rounded, curvilinear surface of the Earth. Fig. 8.12 illustrates the geographic model and projection, in which you can notice the shapes of different continents and how different they are from each other. You should note that no projection maintains all the features (i.e. shape, distance, area and direction) in their original form after transformation. Either one or more or all of the above features are compromised based on the type of projection used for transformation.

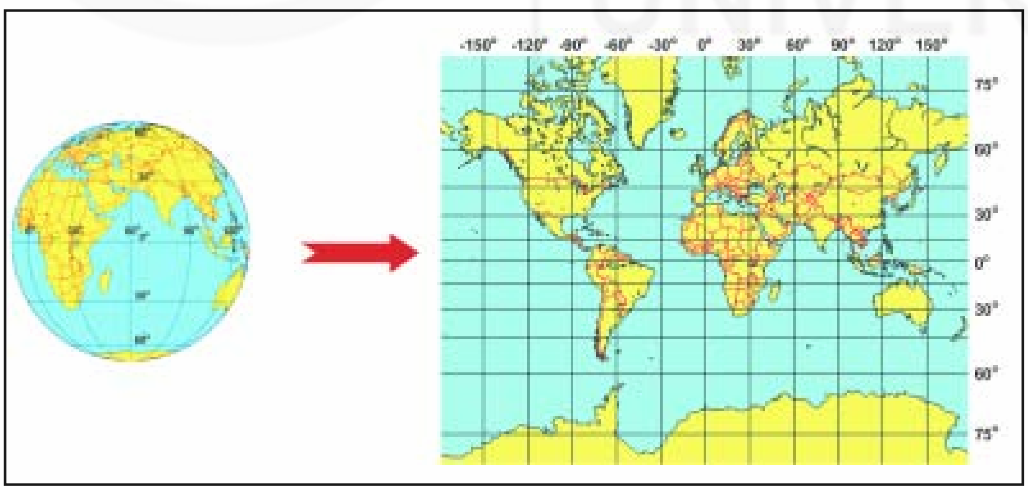


Fig. 8.12: Geographic model and projection to flat map (modified from www.math.ubc.ca/~israel/m103/mercator/mercator.html)

To show regions of the Earth on any appreciable area with accuracy, geographic data must be drawn to compromise the distortions of shapes, distances, area and directions introduced by the spheroid.

8.9.1 Construction of Map Projection

Let us now see how a map projection is created. Creation of a map projection involves two steps:

1) Selection of a model for the shape of the Earth or planetary body (usually a sphere or ellipsoid). You should remember that because the Earth's actual shape is irregular, some degree of information is lost in this step.

2) Transformation of geographic coordinates (longitude and latitude) to Cartesian (x, y) or polar plane coordinates.

Some of the simplest map projections may be easily understood when you place a light source at some definite point relative to the globe and project its features onto a specified surface as shown in Fig. 8.13. However, most of the projections are defined only with the help of certain mathematical formulae and have no direct physical projection.

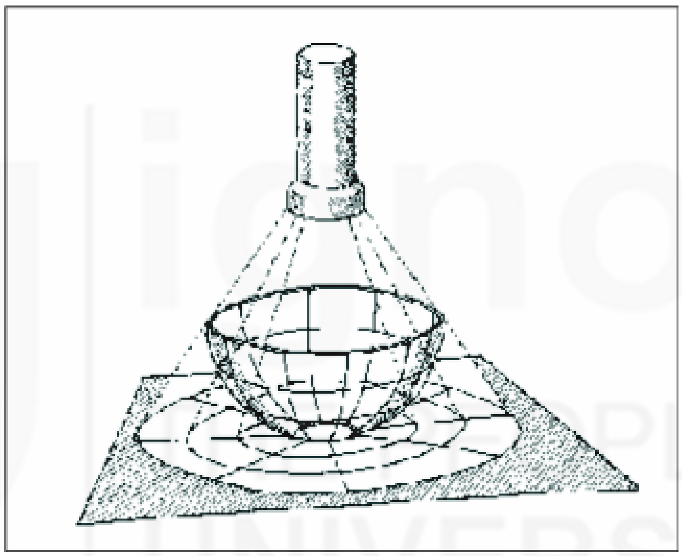


Fig. 8.13: Projection of Earth graticule onto a plane through a light source (source: www.gma.org/surfing/imaging/mapproj.html)

8.9.2 Scale Factor and Transformation

Assume that the Earth has been mapped on a globe reduced to the size (scale) chosen for the flat map. Let us call this as reference globe. The reference globe will have a representative fraction called the *principal scale* (PS) (ratio of radii of globe to that of Earth).

$$\text{Principal Scale (PS)} = \frac{\text{Radius of Globe}}{\text{Radius of Earth}}$$

On the reference globe, the actual scale anywhere will be equal to that of the principal scale. The scale factor (SF) is equal to the actual scale divided by the principal scale.

$$\text{Scale Factor (SF)} = \frac{\text{Actual Scale}}{\text{Principal Scale}}$$

Thus, the scale factor should remain 1.0 throughout the globe.

Now, suppose that the globe's surface is mathematically transformed onto a flat surface, where the actual scale at various places on the map will be either larger or smaller than the principal scale. Thus, SF will always vary from place to place on a flat map. In other sense, scale factor is equal to the unit distance on the projection surface divided by unit distance on the reference ellipsoid.

$$\text{Scale Factor (SF)} = \frac{\text{Unit distance on the projection surface}}{\text{Unit distance on the reference ellipsoid}}$$

On large scale maps, SF at various places varies slightly. It describes the distortions as a result of map projection. Nevertheless, by suitably varying the SF, we can retain some angular relationships or retain relative size of features on a map.

8.10 CLASSIFICATION OF MAP PROJECTIONS

There are several ways to classify or categorise map projections. Classification of map projections with respect to the properties of the projection surface, relative to the globe surface provides four distinct classification types, such as

- Nature of the projection surface or otherwise developable surface
- Coincidence or contact of the projection surface with the globe
- Position or alignment of the projection surface in relation to the globe, and
- Properties of map projection.

Let us now discuss about each of the four types.

8.10.1 Nature of Projection Surface

The simplest scheme is to classify according to the type of developable surface onto which the network of meridians and parallels are projected. A developable surface is a surface that can be laid out flat without any distortion. The sphere and ellipsoid are not developable surfaces. There are three distinct types of developable surfaces and so the type of projections namely, cylindrical, conical, and planar. Each projection type implements a unique pattern of graticule (basically a grid of meridians and parallels) as a base projection surface. Let us further discuss about each of the major projection type.

i) **Cylindrical Projection:** In case of cylindrical projection, a cylinder is assumed to circumscribe a transparent globe (marked with meridians and parallels) so that the cylinder touches the equator through its circumference as you can see in Fig. 8.14. Assuming as if a light bulb is placed at the centre of the globe, the graticule of globe is projected onto the cylinder. By cutting open the cylinder along a meridian and unfolding it, a rectangle-shaped cylindrical projection can be visualised. The globe's longitudinal meridians and latitudes are represented by equidistant, parallel straight lines that intersect one another at right angles. The cylindrical projection is a clear grid representation of the curvilinear surface that is true at the equator and more distorted towards the poles.

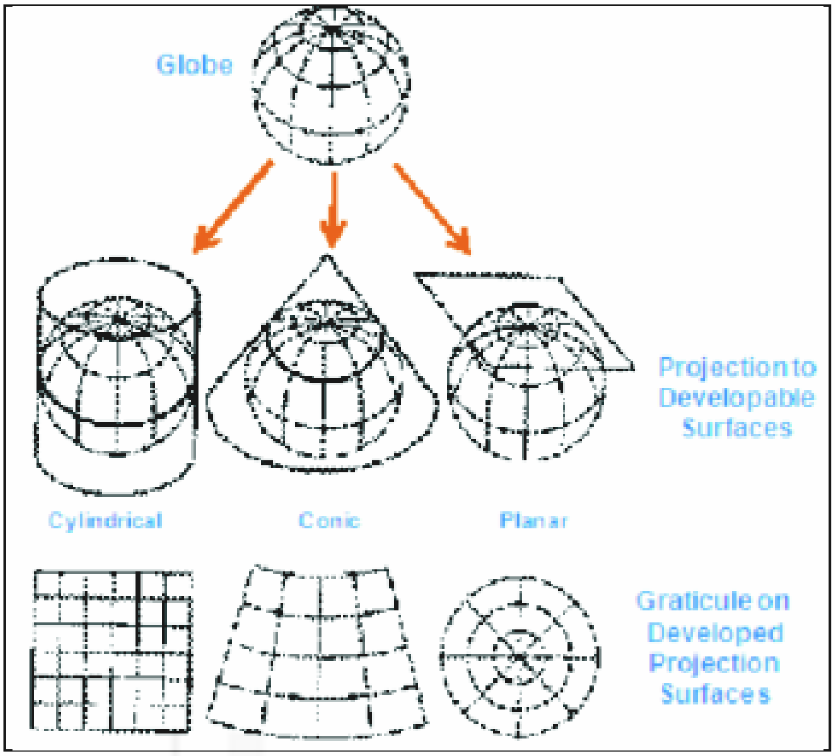


Fig. 8.14: Projections with different developable surfaces (modified from http://hosting.soonet.ca/eliris/gpsgis/Lec2Geodesy.html)

ii) **Conical Projection:** In a conical projection, assume that a cone is placed on the globe in such a way that the apex of the cone is exactly over the polar axis as in Fig. 8.14. The cone must touch the globe along a parallel of the latitude, known as the standard parallel, selected by the user. Along this standard parallel, scale is correct and there is least distortion. When the cone is cut open along a meridian and laid flat, it appears fan shaped. The meridians appear as straight lines radiating from the vertex at equal angles, while the parallels appear as arcs of concentric circles.

iii) **Planar Projection:** In a planar projection, a plane surface is placed so that it touches the globe at the North or South Pole as you see in Fig. 8.14. The projection resulting is known as the ***Polar Azimuthal Projection***. It is circular in shape with meridians projected as straight lines radiating from center of the circle, the pole.

8.10.2 Coincidence of Projection Surface

There are three types of coincidence between the projection surface and the globe, namely, *tangent*, *secant* and *polysuperficial*. In the previous section, the three developable surfaces are assumed to touch the surface of the globe to form a tangent cylinder, a tangent cone, and a tangent plane as seen in Fig. 8.15. Mathematically, it is possible to make the developable surface cut through the globe as a secant cylinder, a secant cone, and a secant plane as shown in Fig. 8.15. The secant case is introduced to increase the contact between the globe and the developable surface and thus the area of minimum distortion. In secant cylinder or cone, two standard parallels are produced, where the scale will be in better control than in other parts of the map.

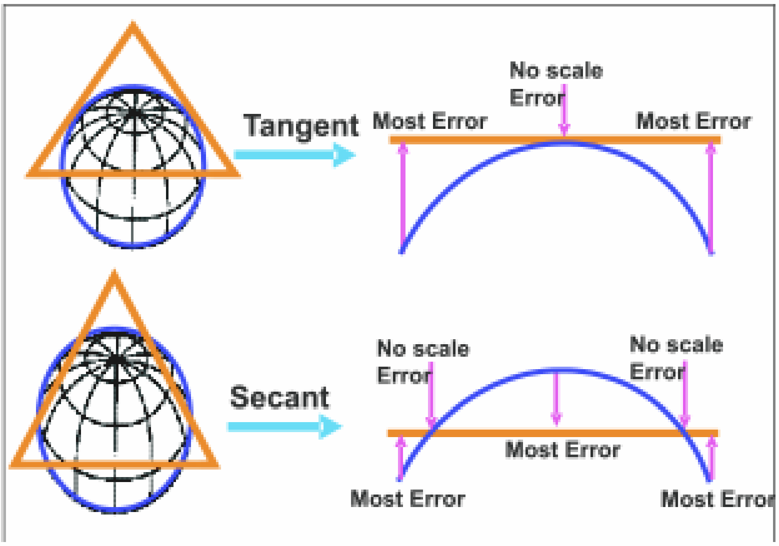


Fig. 8.15: Distribution of distortion with coincidence of projection surface to globe

A still further increase of contact is achieved by employment of superficiality, or in other words, a series of successive projection surfaces. A series of successive planes will produce a polyhedric (multi plane) projection. The best example is polyconic projection as shown in Fig. 8.16, which has been used to generate Survey of India topographic maps. It is created from a series of cones.

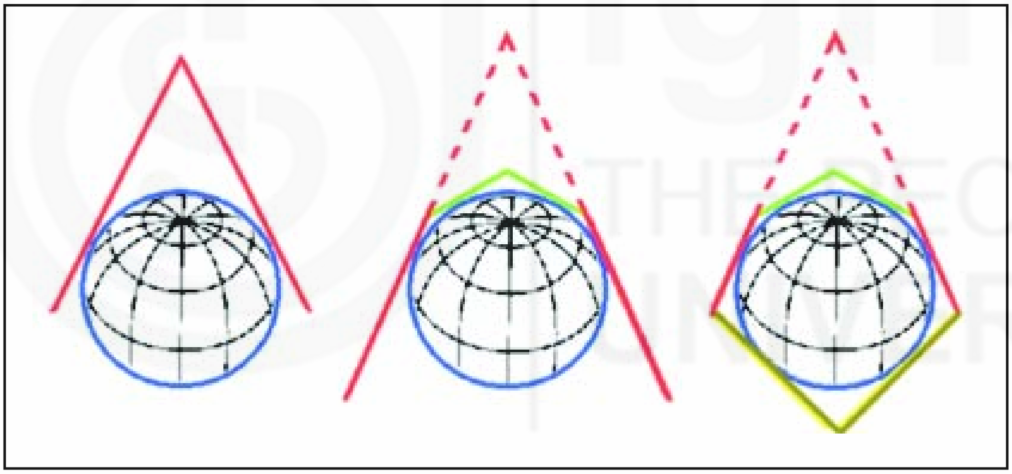


Fig. 8.16: Superficiality in coincidence of projection surface to globe

8.10.3 Position of the Projection Surface

This is also called aspect of the projection. The developable surface may be placed in three different ways relative to the globe: normal, transverse or oblique (Fig. 8.17).

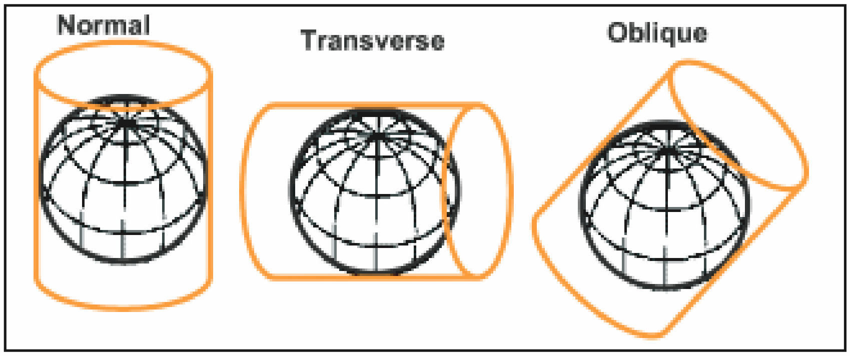


Fig. 8.17: Aspects of map projection

Different aspects of map projections are selected to preserve certain desired properties for particular applications.

8.10.4 Properties of Projection

There are four such cartographic properties to consider:

i) **Conformal or Orthomorphic:** Preserves local shape by retaining correct angles between points. The scale factor (SF) remains constant at a point in all directions on a projection surface. In this condition, the parallels and meridians will intersect at 90^o.

ii) **Equal Area:** Preserves the area displayed so that all regions on the map will be shown in correct relative size. This property is obtained by arranging the SF in the principal direction so that the product of the SFs equals 1.0 at every point. SF varies in every other direction and projection is not conformal.

iii) **Equidistant:** Preserves distances between certain points by maintaining the consistency of scale along the standard lines or meridians.

iv) **Azimuthal:** Preserves direction of all points on the map correctly with respect to the center.

For spherical Earth, all these four properties are correct. However, while transforming the Earth features onto a plane, only some of the properties can be retained. Different map projections are designed to achieve one or two of these properties for specific applications. It is clear that scale requirements for both conformality and equivalence are contradictory and cannot be obtained.

8.11 FALSE ORIGIN OF A PROJECTION

The actual origin of a projection lies at the intersection of the defined central meridian and centrally occurring latitude. As discussed in the previous sections, Quadrants II, III and IV in a Cartesian coordinate system display negative coordinates. This makes the mathematics complex for calculating parameters, like distance, area, etc. To make the entire coordinate values positive, origin of the Cartesian coordinates is shifted in such a way that the total map area falls in Quadrant I. This is achieved by adding certain values to the x and y coordinates. The value added to x coordinate is called *False Easting* and the value added to y coordinate is called *False Northing*. Now, all the points are measured with reference to the shifted origin called as *False Origin* (Fig. 8.18).

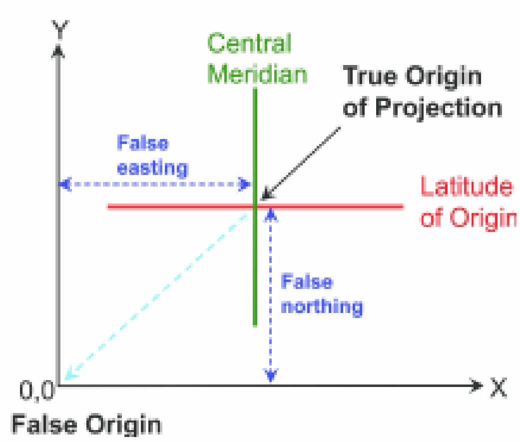


Fig. 8.18: True origin and false origin of a common projection

Check Your Progress III

Spend 5 mins

1) List the steps in defining a map projection.

....

....

....

....

8.12 COMMONLY USED MAP PROJECTIONS

A map projection converts geographic data from the 3D Earth onto a flat plane. There are several projections developed for projecting different regions, for different applications and for managing distortions in specific patterns. Each map projection is designed for a specific purpose and distorts the geography differently. It is worth for the geoinformatics community to know about them for their efficient use. Let us now discuss some of the commonly used projections listed below:

- Mercator Projection
- Transverse Mercator Projection
- Universal Transverse Mercator Projection
- Lambert Conformal Conic Projection
- Polyconic Projection

8.12.1 Mercator Projection

It is the most famous map projection ever devised. This was introduced by Gerardus Mercator in 1569, specifically for nautical navigation for the early navigators (Fig. 8.19). The equator is selected as the standard line along which the scale is true. The rate of change in SF is relatively small near equator, however, enlarges features at a rapidly increasing rate as we go away from the equator. There is no angular distortion.

The graticule consists of equally spaced meridians but unequally spaced parallels. As the parallels get closer to the poles, the spacing between each becomes wider. Therefore, the distortion increases toward the poles. This projection type also safeguards the integrity of features at the equator and the regions in proximity. The poles cannot be represented. It is useful in determining direction in standard sea navigation charts. It is used for other directional uses, like, air travel, wind direction, ocean currents.

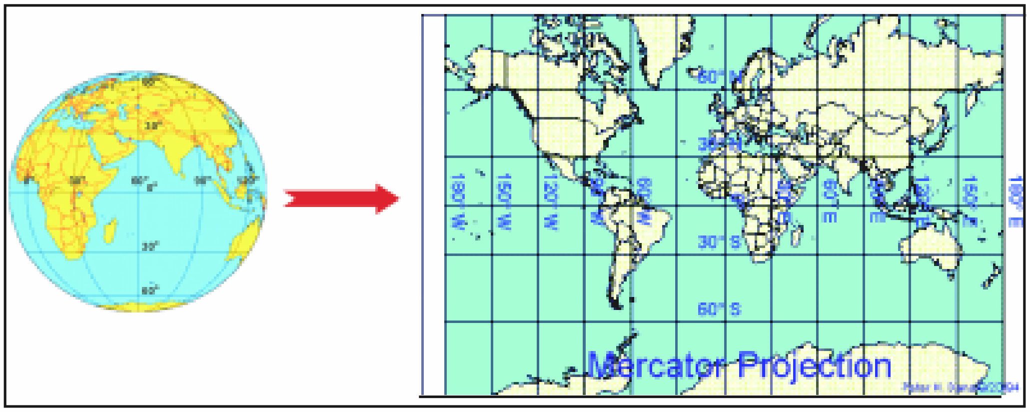


Fig. 8.19: Mercator projection (modified from www.colorado.edu)

8.12.2 Transverse Mercator (TM) Projection

Transverse Mercator Projection is similar to the Mercator projection except that the developable cylinder is longitudinal along a meridian instead of the equator (Fig. 8.20). Usually, there are two parallel lines of contact, equidistant from the central meridian. Hence, it is a cylindrical, conformal and secant projection. The distortion increases with the increase in distance from the standard parallels. TM tiles are always designed smaller than $6^{o}\times6^{o}$ size. There is a popular TM projection called as *Universal Transverse Mercator* (UTM) projection.

8.12.3 Universal Transverse Mercator (UTM) Projection

The Universal Transverse Mercator (UTM) is a global map projection. The UTM coordinates extend around the world from 84° North of the equator to 80° South. It is extended to an extra 4° in the North to cover the northernmost land on Earth. This bounded coverage is because the UTM projection distorts exceedingly near the north and south poles.

The UTM coordinate system is set upon a zoned grid, which divides the Earth into 60 equal zones that are all 6° wide in longitude (east-west). The UTM zones are numbered 1 through 60, starting at the international date line (zone 1 at 180° west longitude), progressing east past the prime meridian (zone 30), and back to the international date line (zone 60 at 180° east longitude). Fig. 8.21 illustrates the complete UTM zones and rows.

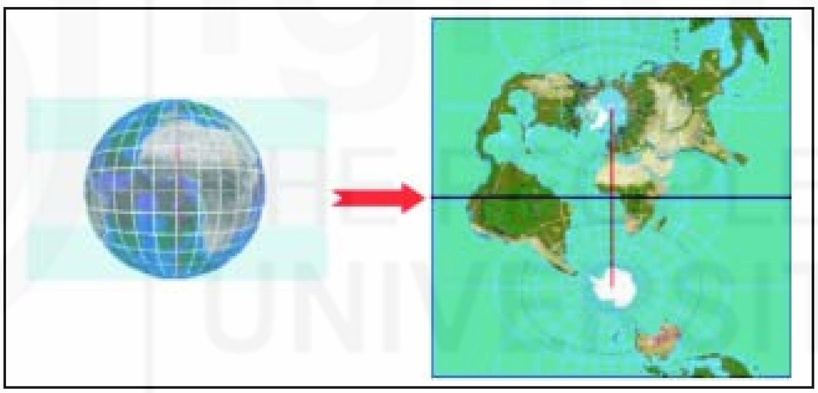


Fig. 8.20: Transverse Mercator projection (modified from http://en.wikipedia.org/ wiki/ Transverse_Mercator_projection)

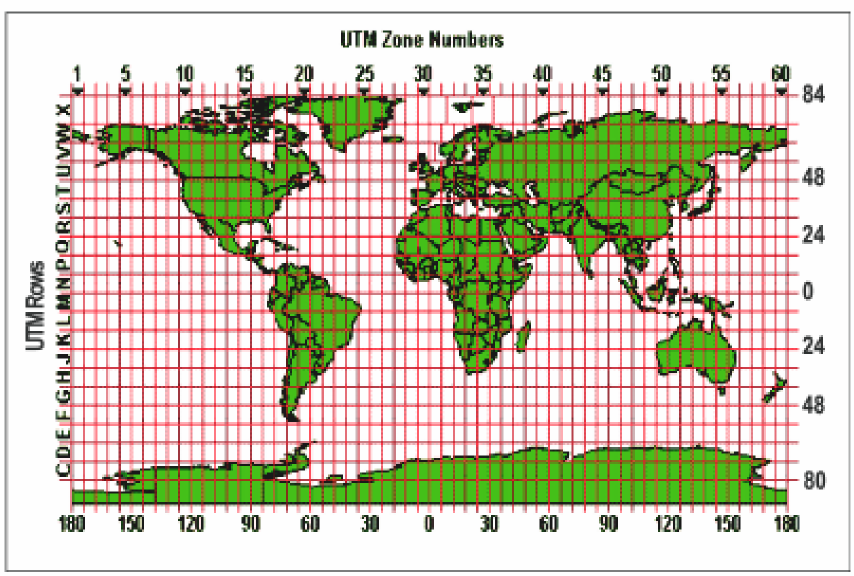


Fig. 8.21: Universal Transverse Mercator Coordinate system (modified from www.utas.edu.au/spatial/locations/spautm.html)

As depicted in Fig. 8.21, each zone extends in both the northern and southern hemispheres. Each UTM zone refers to its own central meridian located at each zone's middle (3°). The origin of each zone is the point on the equator where it is intersected by the central meridian of the zone. To eliminate negative values of coordinates, 500,000 m false easting (X coordinate) and a zero northing (Y coordinate) for the northern hemisphere and 500,000 m false easting (X coordinate) and false northing of 10,000,000 m for the southern hemisphere are used. Thus, the equator separates each zone into two sections: a north section and a south section. Each section then employs its own rectangular grid system using the equator and the zone's central meridian. In total, the UTM comprises 120 separate predefined projections.

The distortion increases for regions that cover more than one UTM zone. The scale is constant on the central meridian with a scale factor of 0.9996 to minimize the lateral distortions within each zone. Most appropriate for regions with north-south expanse. Many countries use the local UTM zones based on the official geographic coordinate system. Recently, Survey of India (SOI) has also released the Open Series Maps (OSM) in UTM coordinate system.

8.12.4 Lambert Conformal Conic (LCC) Projection

It is a conformal conic projection presented by Johann Heinrich Lambert in 1772. The scale is true usually along the chosen standard parallels. In general, for equal distribution of scale error, the standard parallels can be placed at one-sixth and five-sixths of the range of latitudes (Deetz and Adams, 1934) but there are more refined means of selection. It is free of distortion only along the standard parallels and the distortion is constant along any given parallel. It is conformal everywhere except at the poles. The only limitation of this projection is that it can normally be used only for a single hemisphere (Fig. 8.22). The latitude of origin and longitude of origin are specified with reference to the mid-extent of the study area.

Johann Heinrich Lambert was a Swiss mathematician, physicist, philosopher and astronomer. He was the first mathematician who addressed the general properties of map projection. He discussed about seven new map projections under the tittle *Anmerkungen und Zusätze zur Entwerfung der Land- und Himmelscharten*. This paper was later translated into *Notes and Comments on the Composition of Terrestrial and Celestial Maps* by Waldo R. Tobler in 1972 which was published by University of Michigan Press.

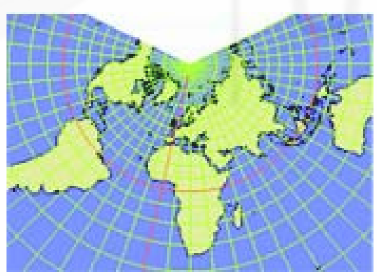


Fig. 8.22: Lambert Conformal Conic projection (source: Google Earth user guide)

8.12.5 Polyconic (PC) Projection

The name of the projection itself explains the use of multiple cones. The projection is created mathematically by lining up an infinite number of cones along the central meridian tangent to an infinite number of parallels. Parallels of latitude (except for Equator) are arcs of circles but not concentric. It is neither conformal nor equal area but a compromise projection. Central meridian and Equator are straight lines. Scale is true along each parallel and along central meridian but no parallel is standard. Directions are true only along central meridian. Distances are true along each parallel and the central meridian. It is more suitable for large scale maps (such as topographic quadrangles).

Spend 5 mins

Check Your Progress IV

1) List the names of commonly used map projections.

....

....

....

....

....

2) How can the distortions caused by map projection be minimised? List out the major points.

....

....

....

....

....

8.13 CHOOSING A SUITABLE MAP PROJECTION

When we try to atten a spherical Earth onto a piece of paper, it is bound to introduce distortions irrespective of the projection selected. Maps will have distortions of one or more of the following properties:

- distance
- direction
- area, and
- angle

All the map projections have distortions but map makers try to minimise them so that the projected map represents features similar to that on the globe. Secondly, types of distortions are not same for all projections. They are different for different types of projection. Selection is made to suite the purpose for which maps are created.

There exists a traditional rule described by Maling (1992) that a country in the tropics asks for a cylindrical projection, whereas a country in the temperate zone asks for a conical projection. Selection of a map projection for an application is established on the basis of assessment of the distortions caused by the projection, and the extreme distortions should be tolerably smaller than would occur in any other projection. The final projection choice would seem to be a fairly straight forward function of minimised distortion. In the end though, there are several other factors that will influence the selection, the main factors are:

- **Location of the study area of interest on the Earth surface** (You should choose a projection depending upon where your area of study is located on earth because different kinds of map projections have been designed for different areas of the Earth. For mapping polar region Azimuthal projection is often used. For mid-latitude regions, you should prefer to use Conic projection and similarly, Cylindrical projection for equatorial region.)
- **Size or extent of the study area** (It is important to know the area and scale of your study area and the outputs, as this would guide you to choose a specific map projection. Distortion becomes a significant problem for large areas (sub-continental) however, selection of projection is less critical for smaller areas.)
- **Shape and orientation of the study area** (For study areas, extending mainly in the east-west direction, Albers Eaual-area Conic, Lambert Conformal Conic projections should be preferred, whereas for study areas extending in north-south directions, Transverse Mercator projection is preferred. For mapping the areas having circular shape, Azimuthal projection is more suitable.)
- **Type of application** (You should choose a projection depending upon which important spatial properties (i.e. area, distance, direction, and angle) you would like to preserve to meet your study objectives. Navigational maps are generally prepared using Mercator projection as it requires to show true direction and/or equidistant. Maps prepared for presentation purpose are usually in conformal projection whereas equal area cylindrical projections are used for showing global distribution of a variety of geographic phenomena.)
- The accuracy requirement/acceptable tolerance limit of the application.
- All data used for same study should have similar projection. Otherwise, there are chances of positional shift from one map to another.

8.14 ACTIVITY

Download and install any open source GIS software (OpenJUMP, Quantum GIS, GRASS GIS, etc.) in your system. Download and open a vector file of state boundaries or a raster image of India. Try to define a projection system starting with geographic coordinate system and then with a projected coordinate system. Observe the shape and orientation of the vector or raster image after defining different projections, like Mercator projection, Lambert Conformal Conic projection by calculating the projection parameters, and also Universal Transverse Mercator projection with different zones (central meridians). Now try to explain the differences from the properties of the projection.

8.15 SUMMARY

Let us now summarise what you have studied in this unit:

- Different applications of geography, geology, GIS, etc. deal with information having spatial reference to Earth's surface that requires spatial coordinates.

- The coordinates can be defined on a model that represents the true shape of Earth. Since Earth's physical shape is bulging out at the equator and flattened at the poles, an ellipsoid can be selected as a true earth model.
- A spheroid (ellipsoid) is a mathematical model that describes the shape of the Earth. It is defined by equatorial radius and the relationship between equatorial and polar radii.
- Geoid is a representation of the Earth as an equigravitational surface. Due to variations in gravity, the geoid does not follow the ellipsoid exactly. The difference between the geoid and ellipsoid is known as geoidal height.
- Ellipsoidal model is generally used to prepare topographic maps and for other maps that needs to accurately portray the earth surface features.
- There are many reference ellipsoids, which are in use by different countries and agencies.
- Datum is a set of control points whose geometric relationships are known either through measurement or calculation. In other words, datum is a known and constant surface which can be used to describe the location of unknown points on Earth. It defines the position of the ellipsoid relative to the centre of the Earth.
- Datum helps to define the position and orientation of the ellipsoid in relation to the earth surface, and also the origin of the coordinate system.
- Coordinate system uniformly measures the Earth's surface, and it has an origin. The geographic coordinate system is defined on an ellipsoid or a sphere, and is then projected (with the help of a map projection) onto a flat map with Cartesian coordinate system for different applications.
- Map projection is a systematic transformation of the 3D spherical surface onto a 2D plane surface of a flat map. In other words, it is a process of fitting a spherical object to a flat surface.
- A map projection uses mathematics to relate the spherical coordinates of the globe to planar coordinates of a map. However, this process causes distortion of one or more different properties such as area, distance, direction, and angle. Learning how to manage distortion is an important step for using projections in GIS.
- The projections are classified based on the developable surface used; aspect and coincidence of the developable surface with respect to globe or datum surface; and also properties.

Spend 30 mins

8.16 UNIT END QUESTIONS

1) Elaborate geoid and ellipsoid.
2) What do you understand by datum and what is the importance of horizontal as well as vertical datum?
3) Briefly elaborate the geographic coordinate system.
4) List out the advantages of planar coordinate system.
5) Discuss properties of map projection.
6) List important criteria for selecting a suitable map projection.

8.17 REFERENCES

- Deetz, C.H. and Adams, O.S. (1934), *Elements of Map Projection with Applications to Map and Chart Construction, 4th ed.*, Washington, DC: U.S. Coast and Geodetic Survey Special Pub., 68p.
- Maling, D.H. (1992), *Coordinate Systems and Map Projections* (2nd Ed.), Pergamon Press, Oxford, 476p.
- Snyder, J.P. (1995), *Map projections: A reference manual*. Taylor and Francis, London, 328p.

8.18 FURTHER/SUGGESTED READING

- Maling, D. H. (1973), *Coordinate Systems and Map Projections*, George Philip and Son Ltd., London, 255p.
- Robinson, A. H., Morrison, J.L., Muehrcke, P.C., Kimerling, A.J. and Guptill, S.C. (1995), *Elements of Cartography* (6th Ed.), John Wiley & Sons Inc., New York, 674p.

8.19 ANSWERS

Check Your Progress I

1) Ellipsoid.
2) Geoid.
3) Ellipsoid.

Check Your Progress II

1) Longitude, latitude and ellipsoidal height.
2) Geocentric datum/local geodetic datum.

Check Your Progress III

Your answer must include the following points

1) Selection of an ellipsoid or sphere for defining the geographic coordinate system
2) Transformation of geographic coordinates to Cartesian coordinates (x,y) defined on a grid with eastings and northings.

Check Your Progress IV

1) Mercator Projection
   Transverse Mercator Projection
   Universal Transverse Mercator Projection
   Lambert Conformal Conic Projection
   Polyconic Projection
2) By selecting a suitable map projection.

   By manipulating the scale factors by redistributing the standard lines.

   Changing the aspect of projection.

Unit End Questions:

1) Your answer must include the following points:
   - Geoid, a surface of equal potential gravity and equivalent to mean sea level
   - Ellipsoid, a three dimensional model used for earth's surface for defining the geographic coordinate system.
2) Your answer must include the following points:
   - A model that defines the position and orientation of the ellipsoid relative to Earth's surface and origin of the coordinate system and provides a frame of reference for measuring locations on the surface of the Earth.
   - Horizontal datum is the base reference in a coordinate system for accurate measurement of positions on earth's surface.
   - Vertical datum is the base reference in a coordinate system for measurement of topographical elevations and bathymetric depth.
3) Your answer must include the following points:
   - Defined on a three dimensional curved surface: an ellipsoid or a sphere.
   - Positions are defined through longitude (x) and latitude (y) pair
   - Positions are defined by their angular distance from the origin
4) Your answer must include the following:
   - Easy for calculating distance, direction and area on the surface.
   - Promotes use of maps, as it is easy to carry, print, and reproduce.
   - Can be used for applications at different scales.
5) Your answer must include the following points:
   - Conformality
   - Equal area
   - Equal distance
   - True direction
6) Your answer must include the following points:
   - Analysing the location of the study area
   - Analysing extents and shape of the study area.
   - Analysing the accuracy requirements of the study
   - Analysing type of application.