Unit Four

Coordinate System and map projection

Unit objectives

At the end of this unit, you will be able to:

- ✓ Understand the Earth's reference surfaces and datums
- ✓ Explain types of coordinate systems
- ✓ Identify map projections classes (types) and distortion properties
- ✓ Explain importance of map projections
- ✓ Identify commonly used map projection in GIS
- ✓ List factors to choose suitable map projections
- ✓ Understand how to georeference entities in a GIS

4.1. Introduction

The unique characteristics of GIS as compared to other information systems is the spatial data. These spatial data include coordinates that define location, shape, and extent of geographic objects. The primary requirements of coordinate is to give a unique reference of location. However, the earth has an irregular shape. This also affects how we best map the surface of the Earth. This unit explains coordinate systems and map projection for positioning of spatial data.

4.2. Reference surfaces

The physical surface of the earth contains variety of irregular landforms, such as plains, valley, water, mountains, and so on. Due to these variations, there are two main reference surfaces used to approximate the shape of the Earth (or Earth figures). They are *the ellipsoid,* and *the Geoid.*

4.2.1. The Geoid

The geoid is defined as the equipotential surface of the earth's (mass) gravity field. It is approximately the same as mean sea level (MSL), used to measure heights. Unlike the ellipsoid, it is difficult to mathematically model the Geoid due to its irregular shape caused by the changing earth's gravitational field. Hence, the Geoid deviates from a mean ellipsoidal shape, called *geoid separation* (N) or *geoid undulation*. Several MSL measurements on ocean coastal areas using tide-gauge stations helped to approximate the geoid. The height determined with respect to MSL called *orthometric height* (height H above the Geoid). Heights above the ellipsoid are ellipsoidal height (h), adjusted using geoid undulations (N); i.e., $H = h - N$.

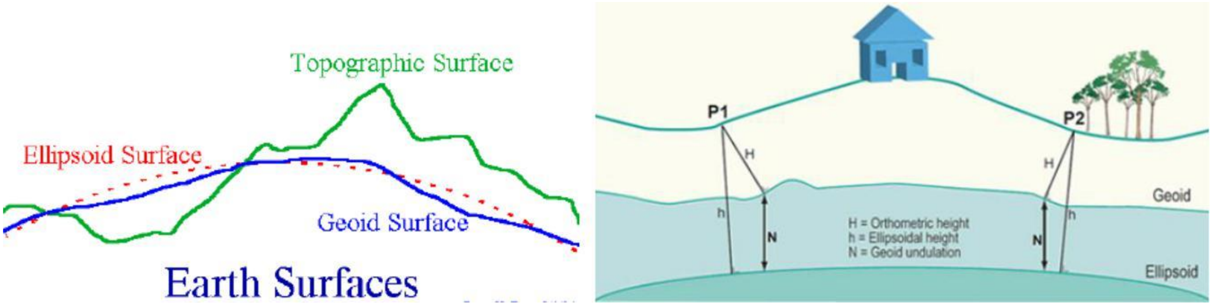


Figure 4-1: Ellipsoidal, orthometric, and geoidal height reference surfaces

4.2.2. Reference ellipsoid

The ellipsoid provides a relatively simple mathematical figure to approximate the Geoid, though for small scale mapping a *sphere* may be used. A sphere is based on a circle, while a spheroid (ellipsoid) is based on an ellipse. The most convenient geometric reference (Earth's shape) is an *oblate ellipsoid* (Figure 4-2, middle). The semi-major axis *(a)*, and the semi-minor axis *(b)* are the two radius of the oblate ellipsoid. An ellipsoid is a horizontal reference frame used to measure locations of points in terms of latitude and longitude. A reference ellipsoid is generated by rotating an ellipse about its *minor axis (b)*. Note that ellipsoid and spheroid used here interchangeably.

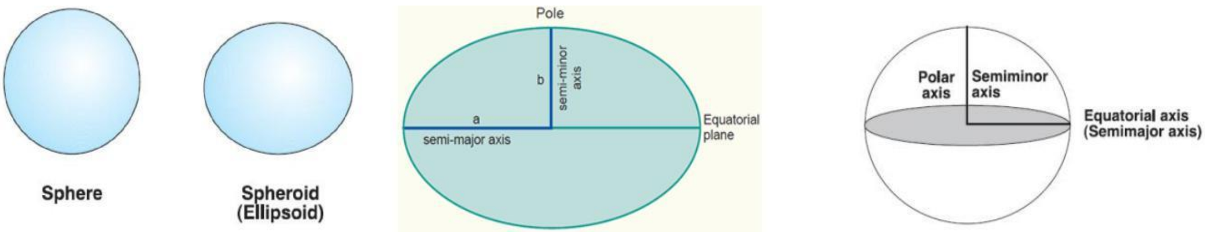


Figure 4-2: Sphere and ellipse of earth (left); major and minor axes of an ellipse (middle); semi-major axis and semi-minor axis of a spheroid (right)

In geodetic practice, the shape of an ellipsoid is defined by its semi-major axis 'a', and flattening 'f'. Flattening indicates how much the ellipsoid departs from spherical shape. It is the ratio of the semi-major axis 'a', minus the semi-minor axis 'b', to the semi-major axis 'a'. Since flattening is a small value, often expressed as a fraction (1/f). The value ranges between '0' and '1', where 'f' ='0' means the two axes are equal, resulting a sphere. Typical values of the parameters for an ellipsoid are a = 6378135:00m; b = 6356750:52m; f≈ 0.003353 = 1/298

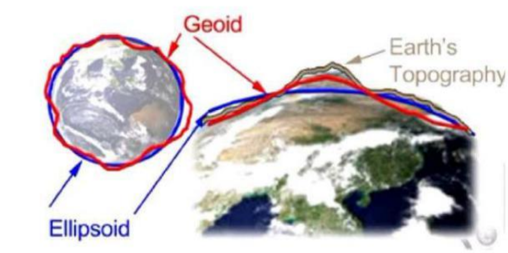


Figure 4-3: The relationship between the Geoid surface and a reference ellipsoid.

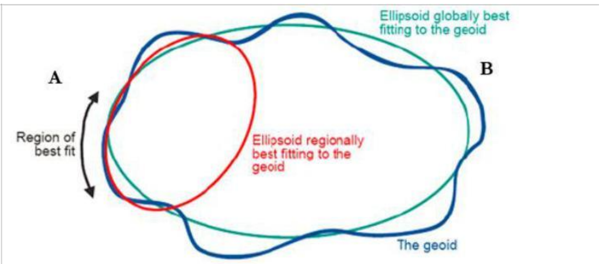


Figure 4-4: An ellipsoid that fits well in one portion of the Earth may fit poorly in another

Due to Geoidal variation in the Earth's shape, there are different ellipsoids developed in the world to approximate the geoid. They may have different origins and orientations to best fit the surface area of interest. For example, in Figure 4-4, ellipsoid **'A'** best fits over one region of the geoid, and ellipsoid '**B'** fits globally to the geoid, but both provide a poor fit in many other areas. In general, for global measurements, it is ideal to use a global reference ellipsoid (e.g., WGS84) that best fits the entire globe. For local applications, locally fitted ellipsoids are preferable. Table 4-1 shows commonly used ellipsoids and their parameters.

Table 4-1: Examples of reference systems and associated ellipsoids

Reference system	*Ellipsoid*	*Semi-Major axis 'a' (m)*	*Semi-Minor axis 'b' (m)*	*Flattening (1/f)*
WGS84	WGS84	6378137.0	6,356,752.30	1/298.26
NAD83	GRS80	6378137.0	6,356,752.30	1/298.26
NAD27	Clarke1866	6378206.4	6,356,583.80	1/294.98

4.3. Datum

While a spheroid approximates the shape of the earth, a datum defines the position of the spheroid relative to the center of the earth. A datum is a mathematical model of the earth that is used in mapping. Datum consists of a series of numbers that define the shape and size of the ellipsoid, as well as the origin and orientation of latitude and longitude lines in space. This provides a reference frame for measuring locations on the surface of the earth. A datum is chosen to give the best possible fit to the true shape of the Earth. Whenever you change the datum, or the geographic coordinate system, the coordinate values of your data will change. This is because the datums and spheroids express the underlying shape of the earth differently. There are many datums in use (Table 4-2), but divided in to two types: local and global datums.

4.3.1. Global or geocentric datum

A global datum is a geocentric datum that uses the earth's center of mass as the origin. Satellite data used to define the best earth fitting of the spheroid, which relates its coordinates to the earth's center of mass. The most recently developed and widely used global datum is WGS1984. It serves as the framework for locational measurement worldwide. WGS84 ellipsoid datum is assumed to be identical with GRS80 datum (Table 4-1).

4.3.2. Local datum

The local datum aligns its ellipsoid to fit closely the Earth's surface in a particular area. The origin of the coordinate systems is selected for the best-matched point on the surface of the ellipsoid to a particular position on the surface of the earth. Its coordinate origin point offsets from the center of the earth. There are hundreds of local reference datums around the world. NAD 1927 and the European Datum of 1950 (ED 1950) are examples of local datums. Ethiopia uses also a local datum known as Adindan, in which its reference ellipsoid is Clark 1880. Its origin shift is Δx = 165, Δy =11 and ΔZ= - 206. This datum is located in Southern Egypt and used by six African countries. Table 4-2 lists some local datums and their area of use.

Table 4-2: Local datums and their principle areas of use

Datum	Ellipsoid	Origin	Area
WGS1984	WGS84	Earth center of mass	Global
NAD 1983	GRS80	Earth center of mass	North America, Caribbean
NAD 1927	Clarke 1866	Meades Ranch	North America
European 1950	International	Potsdam	Europe, Middle East, North Africa

4.4. Coordinate systems

Coordinate systems measure positions on the ellipsoid in a two-dimensional (2D) or three-dimensional (3D) space. They used to establish a common reference framework for display and analysis of spatial data. There are two common types of coordinate systems: **geographic (global)** coordinate systems and **projected (planar)** coordinate systems.

4.4.1. Geographic or global coordinate system (GCS)

GCS defines locations on the earth using a 3D spherical surface. GCS is often incorrectly called a datum, but a datum is only one part of a GCS. GCS includes an angular unit of measure, a prime meridian, and a datum (based on a spheroid). A feature is referenced by its latitude (ϕ) and longitude (λ) values measured from the earth's center to a point on the surface of the Earth.

- Latitude (ϕ) is the angle measured relative to the equator towards the Poles. It is zero on the equator, and increases towards the two Poles. Maximum values of ϕ=+90° at the North Pole and ϕ=-90° at the South Pole. Lines of equal latitude called parallels and form circles.
- *Longitude (λ)* is the angle measured relative to the prime meridian either in Eastwards through λ=+180°, or Westwards through λ=-180°. Lines of equal longitude called *meridians,* and form meridian ellipses. The prime meridian, called the *International Date Line* that passes through Greenwich, England.

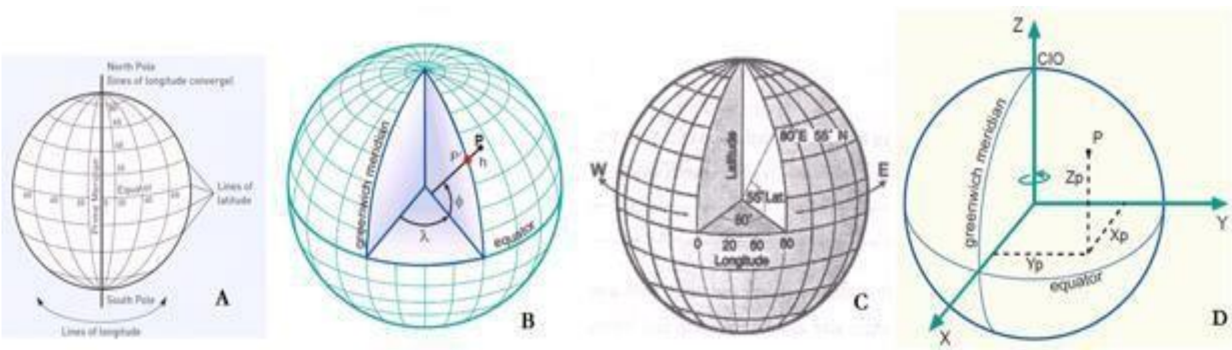
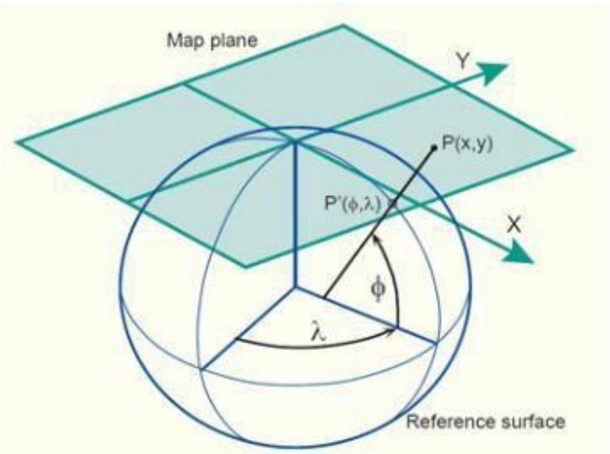


Figure 4-5: 3D GCS representations

Intersection of equator and prime meridian *defines the origin (0, 0)* of the globe; and then *divided into four geographical quadrants* based on compass bearings from the origin. Above and below the equator are north and south, and to the left and right of the prime meridian are west and east. GCS values often recorded in degrees-minutes-seconds, e.g., N42°35'20" of latitude; or in decimal degrees; e.g., longitude 80°E and latitude 55°N as in Figure 4-5-C. A 3D coordinate is obtained by introducing an ellipsoidal height '*h'* to the system (e.g., point '**P'** in Figure 4-5-B). Another method of defining a 3D position on the surface of the Earth is geocentric coordinates of (X, Y, Z); also known as a 3D Cartesian coordinate system (Figure 4-5-D). The system has its origin at the mass-center of the Earth, where the three axes are mutually orthogonal. The X- and Y-axes in the plane of the equator, while the Z-axis coincides with the Earth's axis of rotation.

4.4.2. Projected or planar coordinate systems

A projected coordinate system is defined on a flat, 2D surface, like a flat piece of paper. Map projections transform coordinate locations of latitude (ϕ) and longitude (λ) of into (grid) Cartesian (x, y) coordinate locations. Unlike GCS, a projected coordinate system has the advantage that lengths, angles, and areas are constant across the two dimensions.



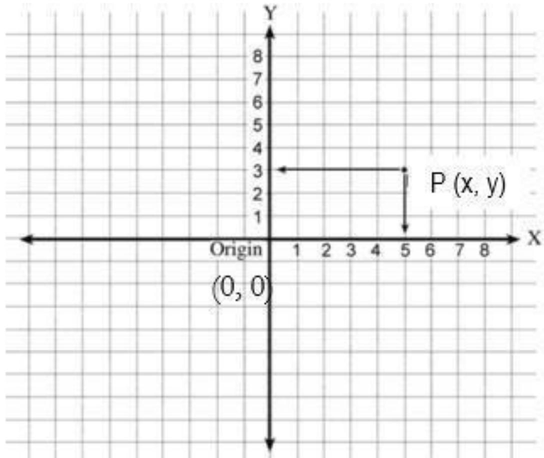


Figure 4-6: Geographic coordinates (ϕ, λ) projected onto a 2D plane of (x, y) Cartesian coordinates

A projected coordinate system is always based on GCS. Normally, the center of the grid is given values of (0, 0) as an origin of (x, y) coordinates. However, sometimes, the origin of the grid is located inside or far away from the area of interest. In this case, large values are added to the origin coordinate values to avoid negative values of (x, y) coordinates or to reduce the range of (x, y) coordinate values. The adjusted origin x = 0 and y = 0 coordinate is called *false origin.*

4.5. Types of projection

There is no universal criterion to classify projections types. However, map projections described in terms of their *developable surfaces* (cylindrical, conical or azimuthal); in terms of their distortion properties (equal-area, equidistant, and conformal); and in terms of their orientation of projection plane relative to the globe, called **aspects** (normal, transverse and oblique), all discussed as follows.

4.5.1. Types of projection based on their distortion properties

Map projection types, such as cylindrical, conical or planar determine the type of distortion properties when compared to the original curved reference surface. Common distortion properties of the projected map are equal-area, equidistant, and conformal. They are described according to what is not distorted in area, shape, direction, angle, distance (bearing), and, scale

4.5.1.1. Conformal (Orthomorphic) projections

Conformal map projections correctly represent angles and shapes (of small areas). A straight line drawn on the map has a constant bearing angle. However, as the region becomes larger, they show area distortions. The scale of a map at any point is the same in any direction on the globe. An example is a **Mercator projection**. Area distortions are significant towards the Polar Regions. For example, Greenland appears to be larger, but is only one-eighth the size of South America (Figure 4-7, left). Maps measure angles (e.g. topographic maps) use *conformal projections.*

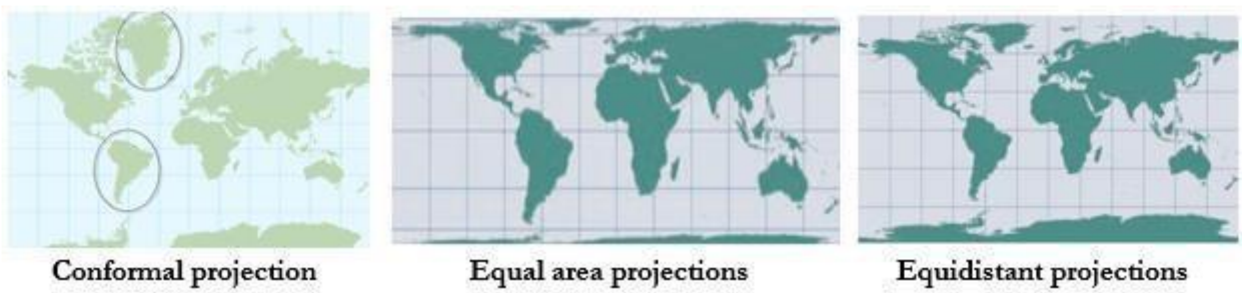


Figure 4-7: Three types distortion properties of cylinderical projection

4.5.1.2. Equal area projections

Equal-area map projections correctly represent area sizes. However, as the region becomes larger, it shows considerable distortions of angles and shapes (Figure 4-7, middle). For example, *Africa shown on this projection suitably as* the equator cuts it into two halves. The *equal area* property ensures that areas measured on the map are always in the same proportion to areas measured on the Earth's surface. Maps for measuring areas (e.g. distribution maps) often use an equal area map projection. No map projection can be both conformal and equal-area.

4.5.1.3. Equidistant projections

Equidistant map projections correctly represent distances (true to scale). A projection can only be equidistant at certain places or in certain directions. A map could be true to scale (with no distortion) along the meridians (i.e. North-South direction), e.g. the equidistant cylindrical projection in figure 4-7, right; or along all parallels (i.e. in East-West direction). Maps that require correct distances measurements (e.g. air-route maps) or maps that require reasonable area and angle distortions (several thematic maps) often use an equidistant map projection.

4.5. 2.. Types of projection based on developable surface

A developable surface is the one that can be flattened and receive projected lines from a globe. Cylindrical, conical and azimuthal map projections are classes of developable surface projections.

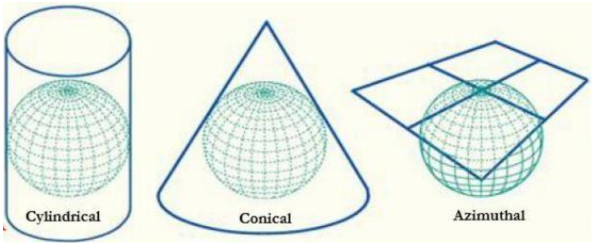


Figure 4-8: Three classes of developable surfaces map projections

4.5.2.1. Cylindrical projection

Cylindrical projection is analogues to wrapping a sheet of paper around the Earth in a cylinder; and then projecting of features onto it and unrolled into a flat map. The equator is tangent or secant to the cylinder's inside. There are three types of such projections: *normal, transverse,* and *oblique.*

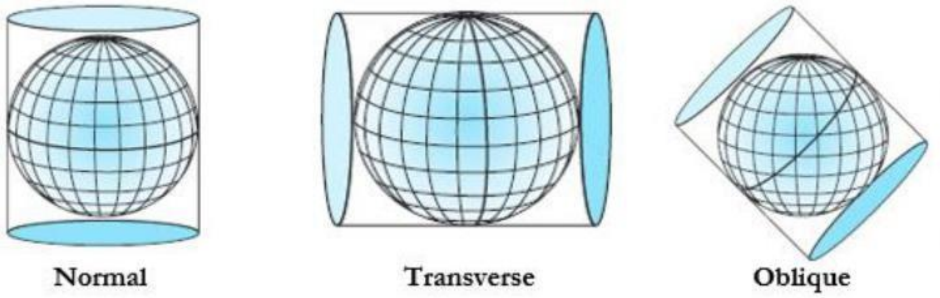


Figure 4-9: Normal, transverse, and an oblique cylindrical projection

In a *normal cylindrical projection*, the main orientation of the projection surface is parallel to the Earth's axis, where the equator is the standard line. This projection has no distortion on the equator, low distortion nearby, and are suitable for tropical countries. In a transverse cylindrical projection, the main orientation of the projection surface is perpendicular to the Earth's axis. Such map projections have no distortion on the central meridian and low distortion nearby, and they are suitable for countries at any latitude. An oblique *cylindrical projection* is non-parallel and non-perpendicular. *Transverse* and *oblique projections* used for most parts of the world. Cylindrical projections may introduce distortions in *area, angle,* *distance* or *direction.* Cylindrical projections have continuous picture of the Earth, Polar Regions distorted, and area of most parts preserved.

4.5.2.2. Conic projections

A conic projection is analogues to wrapping a sheet of paper around the Earth in a cone, and then transferring of points from a globe grid. In a normal aspect, the axis of the cone coincides with the axis of the sphere, and line of latitude are tangent to the globe, called the standard parallels. Points and lines projected radially onto the cone without distortion. In the secant aspect, however, the cone intersects the sphere along the two parallels (Figure 4-8-B). An example is Lambert conformal conic projection. Scale is preserved for the most parts. Area and distance distortions increase away from the standard parallels. Thus, cutting off the top of the cone produces accurate projection. However, for small areas, distortion is minimal. Conic projections are unsuitable for Polar Regions, but they best map areas of great east-west extent than north south (mid-latitude) zones.

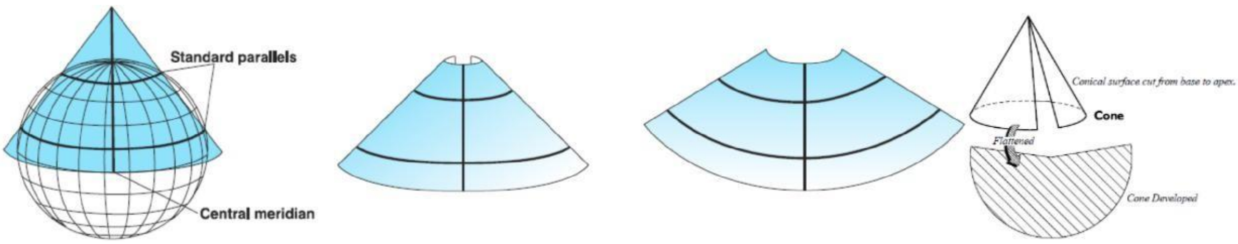


Figure 4-10: Conic projection

4.5.2.3. Azimuthal or planar projections

Azimuthal (planar) projections project map data (parallels, meridians, and points) onto a flat surface touching the globe. This type of projection is usually tangent to the globe at one point but may be secant. The point of contact may be the two Poles, a point on the equator, or any point in between (Figure 4-11). Tangency at the poles is **a normal aspect**, at the equator, it is **an equatorial aspect,** and at mid-latitude, it is **an oblique aspect**.

Figure 4-11: Azimuthal or planar projections

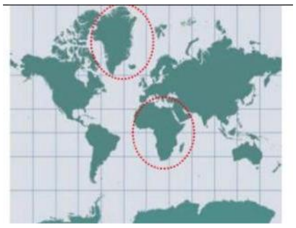
All meridians radiate from the pole at their correct angular distance in straight lines. Parallels are concentric circles with the pole as their centers. In this projection, only one part of the globe is visible; and distortions occur at all directions. Distance preserved for most parts; yet distortion increases with distances from the standard lines. They maintain true directions from the map's center.

4.6. Commonly used map projections

Probably one of the best-known projection is Mercator's cylindrical projection. The Transverse Mercator and the Universal Transverse Mercator (UTM) projections are the best known.

4.6.1. Mercator projection

Mercator projection is a normal cylindrical projection with a conformal property. Parallels and meridians are straight lines intersecting at right angles, and preserved around all locations. Meridians are equally spaced. The parallel spacing increases with distance from the equator.



It was originally designed to display accurate compass bearings for sea travel; and recommended for mapping of regions near the equator However, it is inappropriate for global mapping as it exaggerates areas far from the equator. For example, Greenland appears to be larger than Africa, when in reality Africa's area is 14 times greater.

4.6.2. Transverse Mercator projection

Transverse Mercator (TM) projection, also called Gauss conformal projection constructed from a Transverse cylinder. It maintains scale, shape, area and bearing angles of small areas. This explains why it has become a popular projection for conformal topographic mapping of small areas with north-south extent at scales from 1: 20,000 to 1: 250,000. The right figure shows part of the world mapped on the TM projection. Ghana uses TM projection with central meridian located at 1°W of Greenwich.

4.6.3. Universal Transverse Mercator (UTM) projection

The Universal Transverse Mercator (UTM) projection uses a transverse cylinder, but secant to the reference surface (Figure 4-12). It divides the world into 60 narrow *longitudinal zones* of 6^o. The latitudinal interval is 8^o with the extent from 84^oN to 80^oS. Regions above 84°N and below 80°S are excluded from the system due to distortion. The narrow zones of 6^o (and the secant map surface) make distortions so small that they can be ignored for maps at a scale of 1:10,000 or smaller. Each zone has a false easting central meridian assigned a value of 500km to avoid negative values. For the northern hemisphere, the equator is assigned a value of 0. The UTM grid system is a metric system and widely adopted for topographic maps.

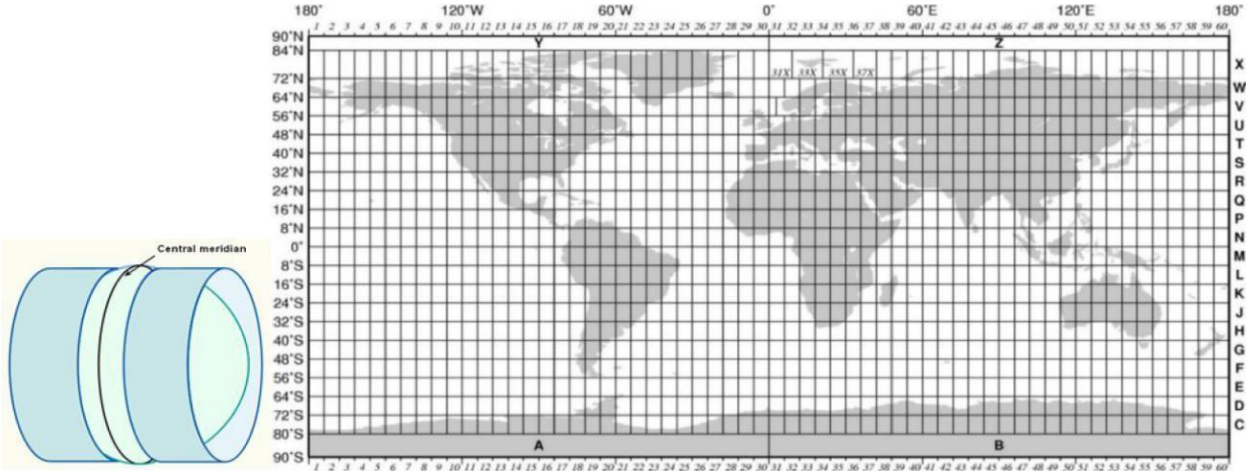


Figure 4-12: UTM projection plane, a cylinder in a transverse position (left), and its zone layouts (right)

4.7. Choosing a map projection

One map projection might be used for large-scale of a limited area, while another used for a small-scale map of the world. A well-chosen map projection considers **scale distortions** remain within limits and **map properties** match the purpose of the map. In general:

- Normal cylindrical projections used to map the entire world (areas near the equator).
- Conical projections often used to map different continents (mid-latitudes regions).
- Polar azimuthal projections used to map the polar areas.
- Transverse and oblique aspects of many projections used for most parts of the world.
- Conformal property for maps that require measuring, angles (e.g. topographic maps).
- Equal-area property for maps that require measuring areas (e.g. land use thematic maps).
- Equidistant property for maps that require reasonable area and angle distortions (thematic).
- The position (orientation) of the projection plane is optimal when the projection center coincides with center of the area.

4.8. Projections within the GIS Environment

In ArcGIS, coordinate systems integrate different datasets with common coordinate framework for display and analysis. If a dataset has GCS of latitude and longitude, ArcMap employs an on-the-fly projections and transformations, and draws the data by simply treating latitude and longitude coordinates as planar X, Y coordinates. However, this tool only changes the display, and it does not alter the original dataset. If our datasets are in GCS of latitude and longitude, they need to be in a projected coordinate system to make measurements and spatial analysis. Therefore, it is important to understand the difference between GIS tools that merely define projections and the tools that actually perform a projection. However, if the datasets do not have coordinate systems as in the case of scanned images, it should be georeferenced. Georeferencing is the act of assigning or establishing of coordinates to locations of entities (ESRI, 1999; Goodchild et al., 2005). ArcGIS contains a *"Georeferencing" toolbar* allows users to shift, pan, resize, rotate, and add *ground control points* (GCPs) to assist georeferencing of spatial datasets.

After georeferencing, the image aligned (rectified) exactly within the map coordinate system. New pixel coordinate values computed using the least squares adjustment. It generates *Root Mean Squares Error* (RMSE) *that* measures the accuracy of the adjustment. Scanned maps and historical data usually do not contain spatial reference information. In these cases you will need to use accurate location data to align or georeference your raster data to a map coordinate system. A map coordinate system is defined using a map projection-a method by which the curved surface of the earth is portrayed on a flat surface. When you georeference your raster data, you define its location using map coordinates and assign the coordinate system of the map frame. In ArcMap, the **georeferencing tools** on the georeference tab allows you to georeference any raster dataset. In general, there are four steps to georeference your data:

1. Add the raster dataset that you want to align with your projected data.
2. Use the georeference tab to create control points, to connect your raster to known positions in the map
3. Review the control points and the errors
4. Save the georeferencing result, when you are satisfied with the alignment.