Chapter 5.
Object Space Coordinate Systems

5-1. Introduction

- In this chapter, three-dimensional *object space coordinate systems* are presented and described
- Object space coordinate systems have always been important for specifying the relative positions of points in surveying, photogrammetry, and mapping
- Object space in photogrammetry refers to the three-dimensional region that encompasses the physical features imaged in the photographs
- When mapping the earth's terrain and natural and cultural features, it is important that all mapped objects be accurately located with respect to an accepted geographic frame of reference
- If any of the spatial data sets are not accurately defined in an accepted frame of reference, then gaps, overlaps, and mismatches will occur

Several accepted **reference systems** exist:

geodetic	geocentric	local vertical	map projection

5-2. Concepts of Geodesy

- The field of ***geodesy*** involves the study of the size, shape, gravity, rotation, and crustal movement of the earth ⇨ It is a highly refined science which provides the basis for earth-related reference coordinate systems
- To understand its basis, three "reference surfaces" must be defined
  ⇨ The ***physical earth***, the ***geoid***, and the ***ellipsoid***
- The ***geoid*** is an equipotential gravity surface, which is generally considered to be *mean sea level*
- The *geoid* is a gently undulating surface which is everywhere perpendicular to the direction of gravity
  ⇨ These gentle undulations are due to gravity variations caused by the nonhomogeneous mass distribution of the earth
- The amplitude of geoid undulations depicted in Fig.5-1 are greatly exaggerated
- The shape of the geoid, in fact, results from the ***net attraction***, comprised of gravity and the effect of the earth's rotation

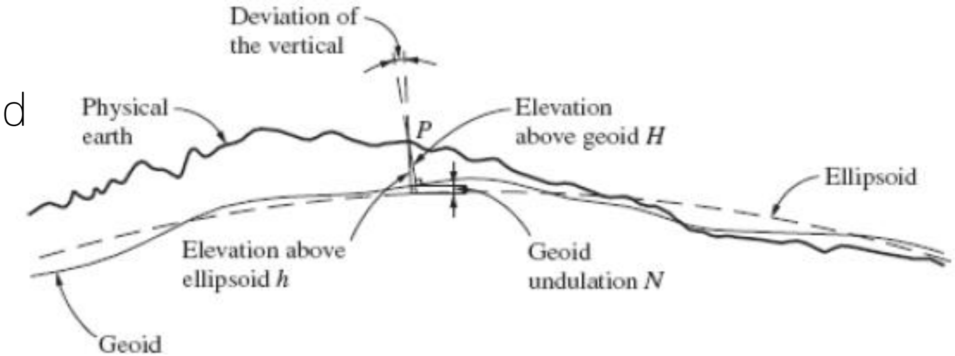

Figure 5-1. The three fundamental geodetic reference surfaces: physical earth, geoid, and ellipsoid.

5-2. Concepts of Geodesy

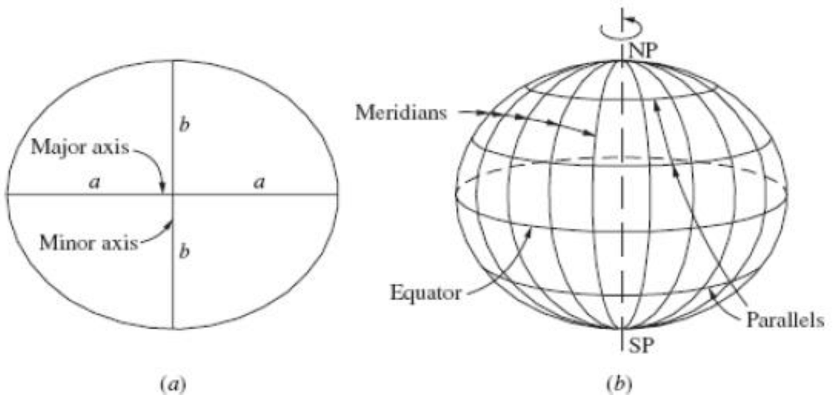

Figure 5-2. Definition of a reference ellipsoid. (a) Two-dimensional ellipse showing major and minor axes. (b) Three-dimensional ellipsoid formed by rotation of ellipse about the minor axis.

- ❖ A reference *ellipsoid* is a mathematically defined surface which approximates the geoid either globally or in a large local area such as a continent
- ❖ This surface is formed by rotating a two-dimensional ellipse (shown in Fig. 5-2a) about its minor axis
- ❖ Rotation of an ellipse in this manner generates a three-dimensional ellipsoid, as shown in Fig. 5-2b
- ❖ This figure shows the (curved) lines which pass through the north and south poles (NP and SP, respectively), known as *meridians*, and (curved) lines which are parallel to the equator, called *parallels*
- ❖ A meridian is formed by the intersection of the ellipsoid with a plane containing the pole
- ❖ A parallel, however, is formed by the intersection of the ellipsoid with a plane that is perpendicular to the pole (i.e., parallel to the equator)

5-2. Concepts of Geodesy

- To define the size and shape of the reference ellipsoid, at least two constants(from actual measurements) are required
  ⇨ Generally, the semimajor axis of the ellipsoid a and the flattening f are specified

- The following equations give relationships between the ellipsoid constants of the semimajor axis- $\boldsymbol{a}$, the semiminor axis- $\boldsymbol{b}$, the flattening- $\boldsymbol{f}$, the first eccentricity- $\boldsymbol{e}$, the second eccentricity- $\boldsymbol{e'}$

$$f = 1 - \frac{b}{a} \quad (5\text{-}1) \qquad e = \frac{\sqrt{a^2 - b^2}}{a} \quad (5\text{-}3) \qquad e' = \frac{\sqrt{a^2 - b^2}}{b} \quad (5\text{-}5)$$

$$b = a(1 - f) \quad (5\text{-}2) \qquad e^2 = f(2 - f) = \frac{a^2 - b^2}{a^2} \quad (5\text{-}4) \qquad e'^2 = \frac{e^2}{(1 - f)^2} = \frac{a^2 - b^2}{b^2} \quad (5\text{-}6)$$

- The flattening f is a parameter for an ellipsoid (or ellipse) which quantifies how much it departs from a true sphere (or circle)
  ⇨ The value of f for an ellipse can range from 0, which corresponds to a circle, to 1, which corresponds to a completely flattened ellipse

- The accepted value of f for the earth is roughly 0.0033, which implies that the earth is very nearly spherical

- The first and second eccentricities $\boldsymbol{e}$ and $\boldsymbol{e'}$ are also parameters which quantify how much an ellipse departs from a true circle, with values near 0 denoting near circularity

5-2. Concepts of Geodesy

Table 5–1. Parameters for Select Reference Ellipsoids

Reference Ellipsoid	Semimajor Axis *a*	Flattening *f*
Clarke 1866	6,378,206.4 m	1/294.9786982
GRS80	6,378,137 m	1/298.25722210088
WGS84	6,378,137 m	1/298.257223563

- Table 5-1 gives semimajor axis a and flattening f values for three commonly used reference ellipsoids
- The values for a and f are selected on the basis of geodetic measurements made in different locations of the world
- The Clarke 1866 ellipsoid is an example of this type of local surface, which was a best-fit to the geoid in North America.
- Recently, reference ellipsoids have been derived which give a best-fit to the geoid in a worldwide sense
- The GRS80 and WGS84 ellipsoids are examples of worldwide reference surfaces

5-3. Geodetic Coordinate System

- Geodetic coordinates for specifying point locations relative to the earth's surface are latitude ϕ, longitude λ, and height h
  ⇨ These coordinates all depend upon a reference ellipsoid for their basis

- Latitude and longitude are horizontal components, while the vertical component is height

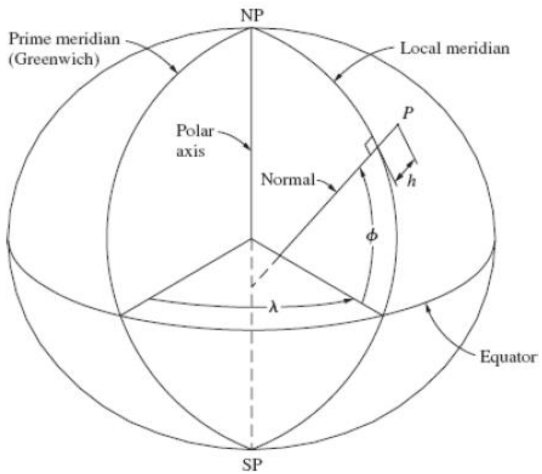

Figure 5-3. Geodetic coordinates of latitude ϕ, longitude λ, and height h.

- These three coordinates are illustrated in Fig. 5-3

- This figure shows a point P with a line passing through it, perpendicular to the ellipsoid and extending to the polar axis ⇨ This line is called the *normal*

- The longitude l of a point is given by the angle in the plane of the equator from the ***prime meridian*** (usually the meridian through Greenwich, England) to the ***local meridian*** (meridian passing through the normal line)

- Values of longitude range from -180° to +180° with those west of the prime meridian being negative and those to the east being positive

5-3. Geodetic Coordinate System

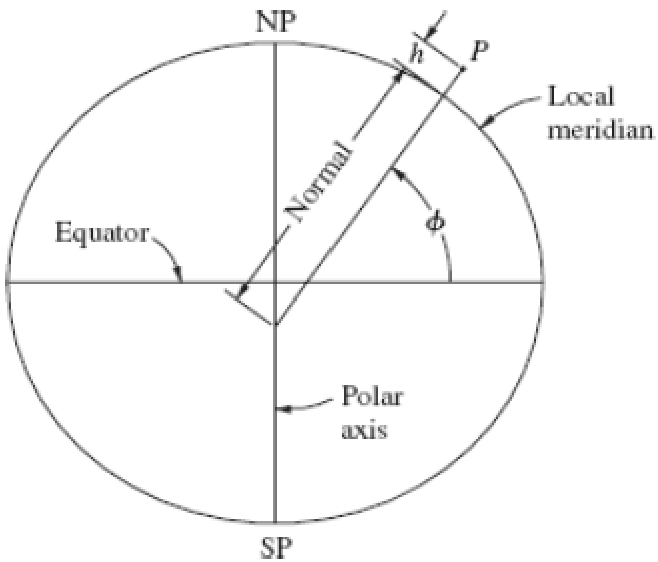

Figure 5-4. Illustration of latitude ϕ, normal, and height h in the plane of the local meridian.

- ❖ The latitude (ϕ) of a point is the angle from the equatorial plane to the normal line
- ❖ Values of latitude range from -90° to +90° with those north of the equator being positive and those to the south being negative
- ❖ The latitude is more clearly illustrated in Fig. 5-4, which shows a plane section through the ellipsoid containing the local meridian and the normal

- ❖ As also illustrated in Fig. 5-4, height h is the distance from the surface of the ellipsoid to the point P, in the same direction as the normal
- ❖ This value specifies the elevation of a point above the ellipsoid, also known as the *ellipsoid height*
- ❖ The elevation of a point above the geoid H, also known as *orthometric height*, is commonly considered to be the *mean sea level* elevation

5-3. Geodetic Coordinate System

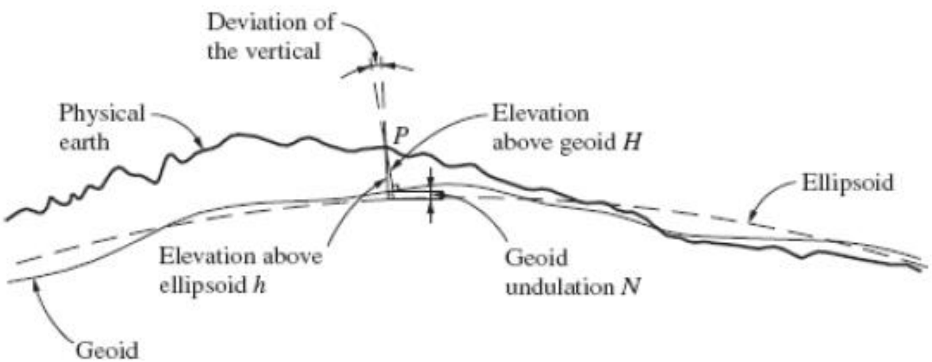

Figure 5-1. The three fundamental geodetic reference surfaces: physical earth, geoid, and ellipsoid.

❖ Figure 5-1 illustrates the relationship between these two height definitions

❖ The difference between the two heights is referred to as the ***geoid undulation*** or ***geoid height*** and is indicated in Fig. 5-1 by the value N

❖ The relationship between ellipsoid height h, orthometric height H, and geoid undulation N is specified in Eq. (5-7)

$$h = H + N \qquad (5\text{-}7)$$

❖ Figure 5-1 shows two lines through point P perpendicular to the ellipsoid and geoid
⇨ These lines intersect at an angle known as the ***deviation*** of the *vertical*

❖ The deviation of the vertical can be determined by precise surveying techniques, and its value never exceeds 2 arc minutes anywhere on the earth

5-4. Geocentric Coordinates

- While geodetic coordinates $\phi\lambda h$ provide an earth-based definition for a point's threedimensional position, they are related to a curved surface (reference ellipsoid)

- The geodetic coordinates are nonorthogonal and as such are unsuitable for analytical photogrammetry, which assumes a rectangular or cartesian coordinate system

- The geocentric coordinate system, on the other hand, is a three-dimensional XYZ cartesian system which provides an earth-centered definition of position, independent of any reference surface
  - ⇨ This system has its XY plane in the plane of the equator with the Z axis extending through the north pole
  - ⇨ The X axis is oriented such that its positive end passes through the prime meridian

5-4. Geocentric Coordinates

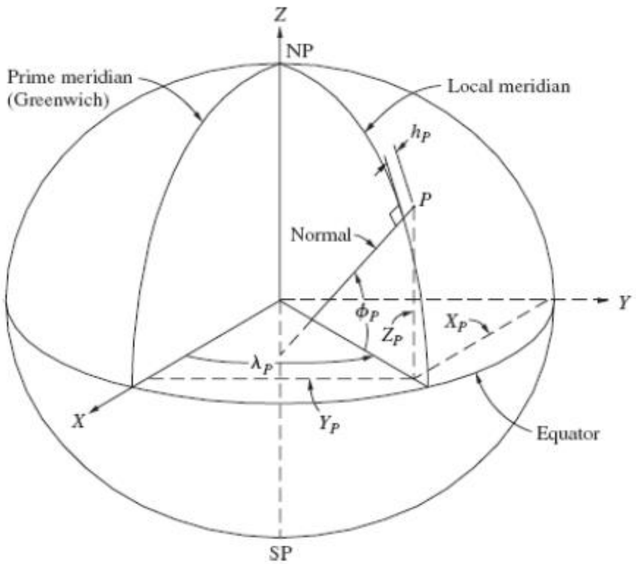

Figure 5-5. Relationship between geocentric and geodetic coordinates.

- Figure 5-5 illustrates **the geocentric coordinate** system and its relationship to geodetic coordinates
- The geocentric coordinate system is a convenient system for many worldwide geodetic applications such as satellite geodesy

- Values of the coordinates are very large and have no obvious relationship to the cardinal directions in a local area
- The direction of the camera axis would be quantified relative to the earth's pole instead of the local vertical

⇨ For these reasons, a special coordinate system, the local vertical system, with its origin in the local project area is generally used

5-5. Local Vertical Coordinates

- A local vertical coordinate system is a three-dimensional cartesian XYZ reference system which has its origin placed at a specific point within the project area
- At this local origin, the Z axis extends straight up from the ellipsoid in the same direction as the normal at the origin
- The positive X and Y axes are tangent to the ellipsoid and point to the east and north, respectively

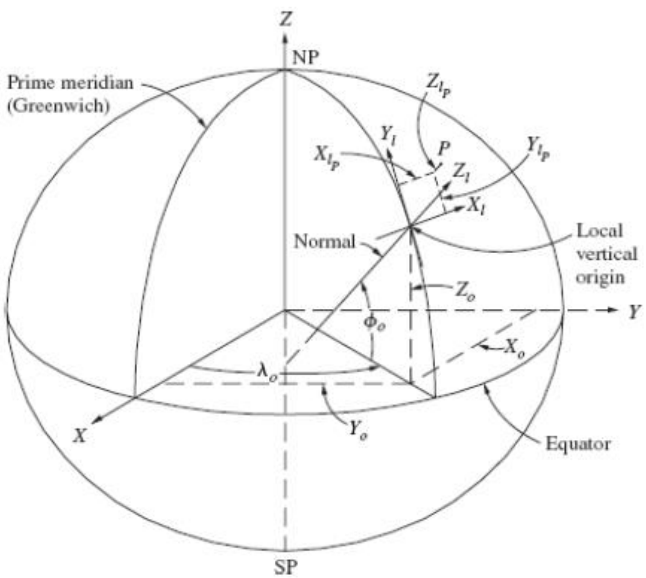

Figure 5–6. Local vertical coordinate system relative to geocentric and geodetic systems.

- Figure 5-6 shows the local vertical system and its relationship to geocentric and geodetic coordinates
- In this figure, the position of the local origin is specified in terms of geodetic coordinates ϕ_0, λ_0, and h_0, with the last equal to zero
- As shown in Fig. 5-6, the local origin has geocentric coordinates X_0, Y_0 and Z_0, and point P in the project area has local vertical coordinates X_{lP}, Y_{lP}, and Z_{lP}

5-5. Local Vertical Coordinates

- A local vertical coordinate system is a three-dimensional cartesian XYZ reference system which has its origin placed at a specific point within the project area
- At this local origin, the Z axis extends straight up from the ellipsoid in the same direction as the normal at the origin
- The positive X and Y axes are tangent to the ellipsoid and point to the east and north, respectively

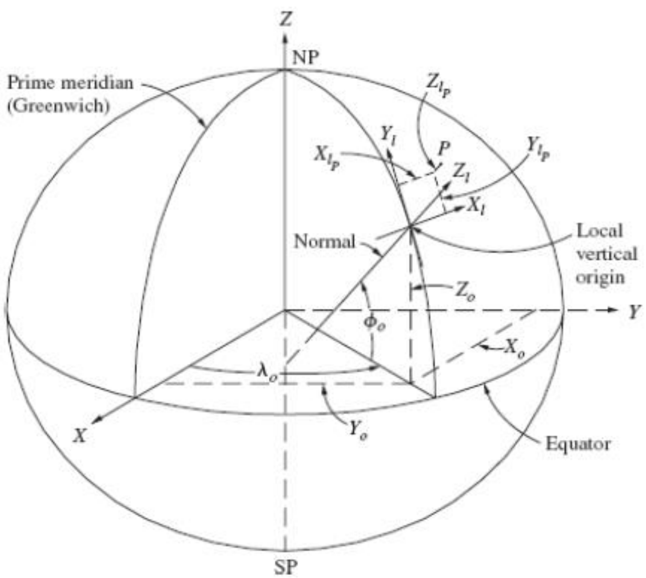

Figure 5-6. Local vertical coordinate system relative to geocentric and geodetic systems.

- Local vertical coordinates have the characteristic that $X, Y,$ and Z values will correspond roughly to eastings, northings, and heights above the ellipsoid, respectively
- This will no detrimental effect on coordinates computed through analytical photogrammetry, since the local vertical coordinates can be accurately converted to useful geodetic coordinates or map projection eastings and northings

5-6. Map Projections

- A map, in general, consists of points, lines, arcs, symbols, or images which are placed on a flat, two-dimensional surface such as a sheet of paper or computer display
- It is preferable that the map present the viewpoint from directly overhead throughout the area
- While there are several types of map projections to choose from, those most often used in photogrammetric mapping applications are ***conformal***, meaning that true shape is maintained
- Two particular conformal map projections which will be discussed here are the ***Lambert conformal conic*** and the ***transverse Mercator***
- These map projections both employ the concept of the ***developable surface***

  ⇨ A **developable surface** is a surface that may be three-dimensional in its natural form, but can be "unrolled" and laid flat
- The developable surface is created which nearly coincides with the ellipsoid in the region being mapped
- Points are then projected from the ellipsoid onto the developable surface or vice versa

5-6. Map Projections

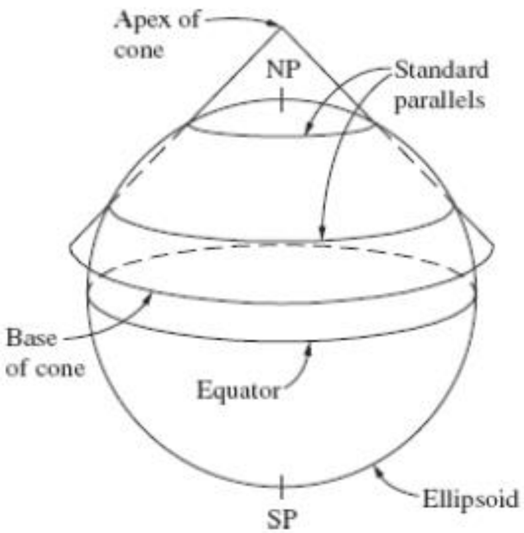

Figure 5-7. Cone used in the Lambert conformal conic projection.

❖ The Lambert conformal conic projection uses a cone as its developable surface

❖ The axis of the cone is made to coincide with the minor axis of the ellipsoid and will pass through the ellipsoid along two parallels of latitude, called the *standard parallels*

❖ Figure 5-7 shows a cone superimposed on the reference ellipsoid of the earth

❖ In the region between the standard parallels, the conic surface is below the ellipsoid
⇨ Therefore, lines that are projected from the ellipsoid to the cone will be made shorter, and those outside the standard parallels will be made longer
⇨ This change in dimension can be quantified by a *scale factor* which is less than 1 between the standard parallels, greater than 1 outside, and exactly equal to 1 at the standard parallels

❖ Since this scale factor varies in the north-south direction but remains the same in the east-west direction, the Lambert conformal conic projection is appropriate for areas of limited extent north-south, but wide extent east-west

5-6. Map Projections

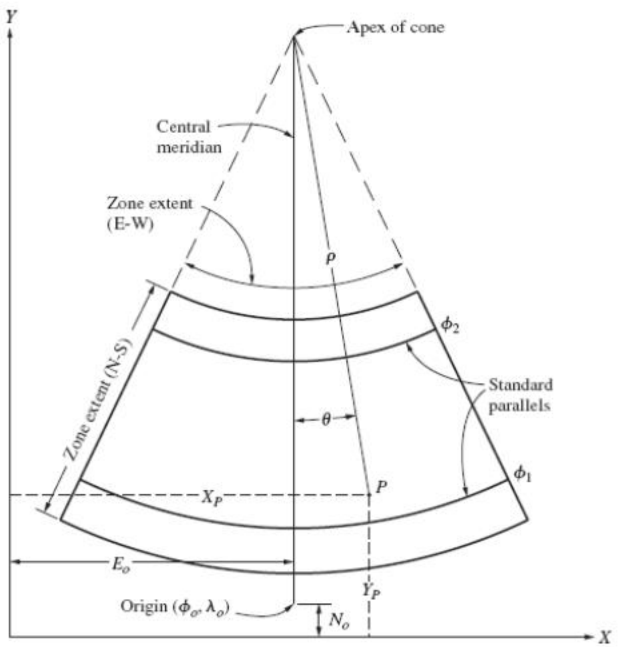

Figure 5-8. The Lambert cone unrolled and laid flat.

- The *XY* (easting, northing) coordinates of points within a Lambert conformal conic projection are based on the cone after it has been unrolled and laid flat
- Figure 5-8 shows how the flattened cone is placed relative to the *XY* axes
- In this figure, parallels of latitude form concentric circular arcs centered on the apex of the cone, and meridians appear as lines which are radial from the apex
- Different Lambert conformal conic projections can be set up for specific local areas or "zones"

- When a Lambert conformal conic map projection is developed for a specific area, a *central meridian* is selected whose longitude is equal to that of the approximate center of the zone
- The *origin* for the map projection is also selected
- It lies on the central meridian at a location below the coverage of the zone

5-6. Map Projections

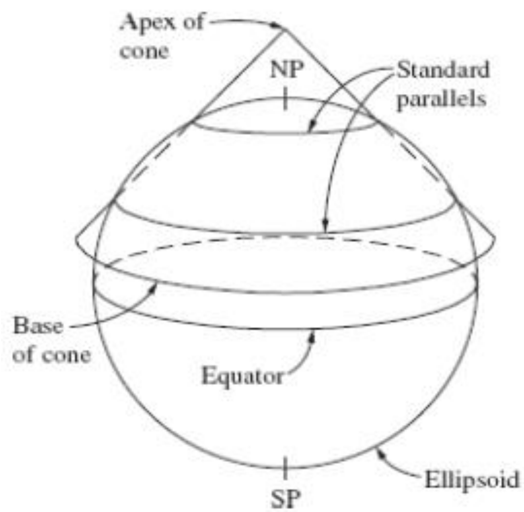

Figure 5-7. Cone used in the Lambert conformal conic projection.

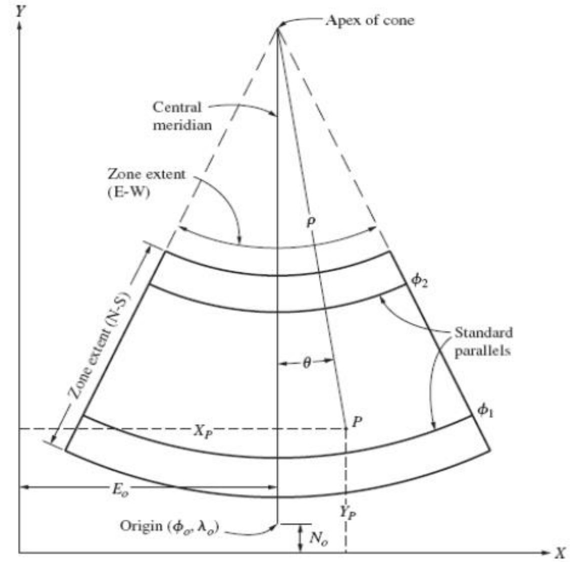

Figure 5-8. The Lambert cone unrolled and laid flat.

- ❖ **Six parameters** define a Lambert conformal conic map projection for a specific zone
  1. ① The latitudes ϕ_1 and ϕ_2 of the two standard parallels depicted in Figs. 5-7 and 5-8
  2. ② The latitude and longitude ϕ_0 and λ_0 of the grid origin
  3. ③ The false easting E_0 and false northing N_0 of the origin

- ❖ These latter parameters, as shown in Fig. 5-8, are included so that all coordinates within the zone will be positive

- ❖ A point P having a particular latitude and longitude (ϕ_P, λ_P) will have corresponding map projection coordinates (X_P, Y_P)

5-6. Map Projections

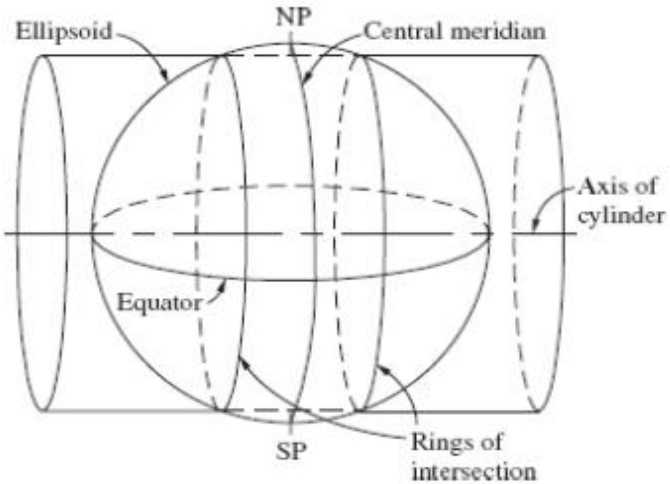

Figure 5-9. Cylinder used in the transverse Mercator projection.

- The axis of the cylinder is defined so as to lie in the plane of the equator, transverse to the minor axis of the ellipsoid
- Computation of the forward conversion for the Lambert conformal conic projection involves complex mathematics
- Conversion from XY to $\phi\lambda$ is referred to as the inverse conversion

- Another commonly used map projection is the transverse Mercator, which employs a cylinder as its developable surface
- Using an ellipsoidal representation of the earth requires the use of a cylinder that has been flattened slightly to conform to the shape of the ellipsoid
- Conversion of latitude and longitude ($\phi\lambda$) of a point into map projection coordinates (XY) is known as the *forward* or *direct conversion*

5-6. Map Projections

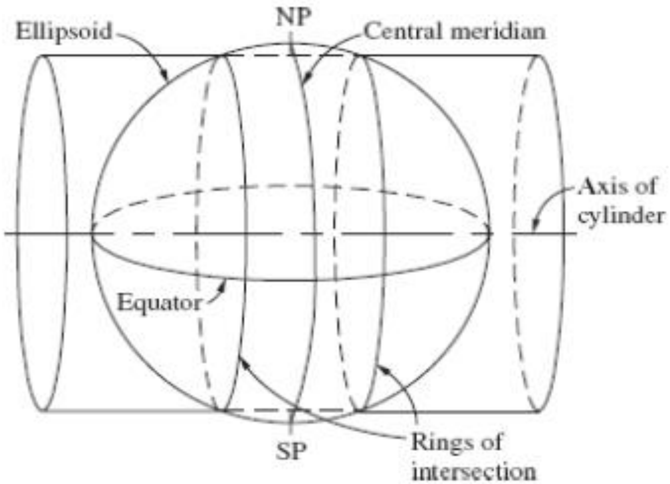

Figure 5-9. Cylinder used in the transverse Mercator projection.

❖ Figure 5-9 shows a cylinder superimposed on the reference ellipsoid of the earth

❖ The cylinder intersects the reference ellipsoid along two rings which are nominally oriented in the north-south direction

❖ In the region between the rings of intersection, the cylinder is below the ellipsoid and therefore lines that are projected from the ellipsoid to the cylinder will be made shorter, and those outside the rings will be made longer

❖ This change in dimension can also be quantified by a scale factor which is less than 1 between the rings of intersection, greater than 1 outside, and exactly equal to 1 at the intersecting rings

❖ Since this scale factor varies in the east-west direction but remains approximately the same in the north-south direction, the transverse Mercator projection is appropriate for areas of limited extent east-west, but with long extent north-south

5-6. Map Projections

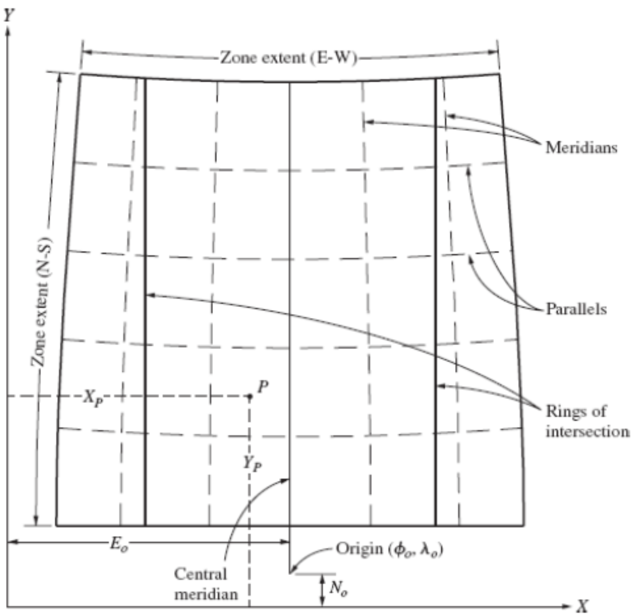

Figure 5-10. The transverse Mercator cylinder unrolled and laid flat.

- The XY (easting, northing) coordinates are based on the cylinder after it has been unrolled and laid flat
- Figure 5-10 shows how the flattened cylinder is placed relative to the X and Y axes
- In this figure, parallels of latitude and meridians of longitude appear as lines which take the shapes of complex curves, symmetric about the central meridian
- Different transverse Mercator projections can be set up in local areas or zones, also defined in terms of their ranges of latitude and longitude
- To develop a transverse Mercator map projection for a specific area, a central meridian is selected in the approximate center of the zone
- An origin is also defined which lies on the central meridian at a location below the coverage of the zone (see Fig. 5-10)

5-6. Map Projections

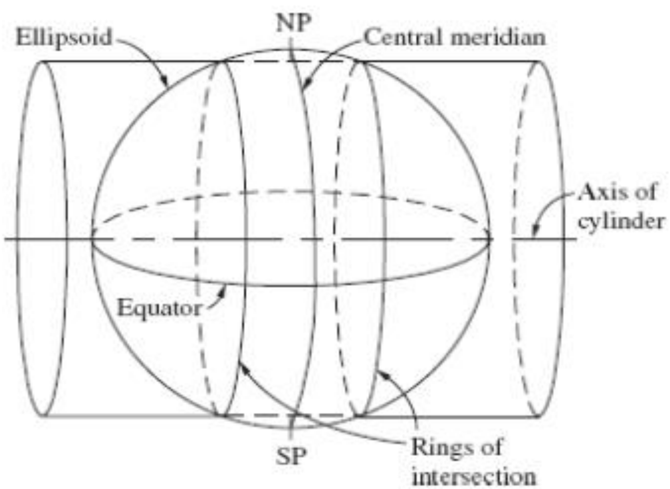

Figure 5-9. Cylinder used in the transverse Mercator projection.

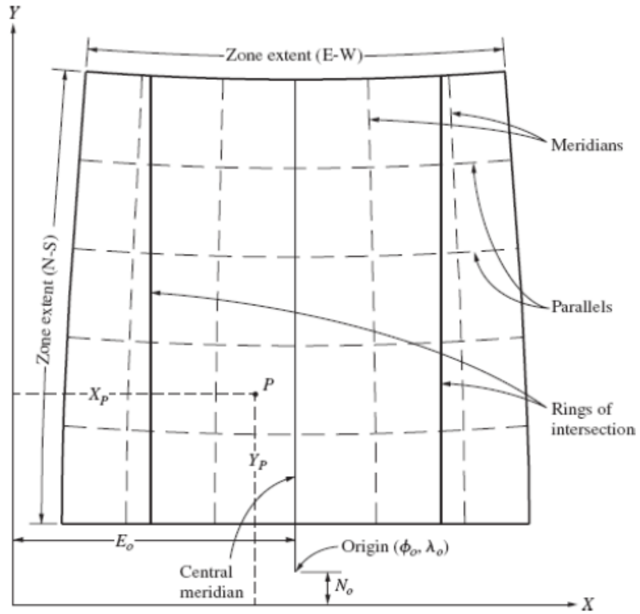

Figure 5-10. The transverse Mercator cylinder unrolled and laid flat.

❖ **Five parameters** uniquely define a transverse Mercator map projection for a specific zone

① the latitude and longitude, ϕ_0 and λ_0 of the grid origin

② k_0, is the scale factor along the central meridian

③ a false easting E_0 and false northing N_0 of the origin are included to keep the coordinates positive

❖ A point P having a particular latitude and longitude (ϕ_P, λ_P) will have corresponding map projection coordinates (X_P, Y_P)

5-6. Map Projections

- Both the Lambert conformal conic and transverse Mercator projections are used in *state plane coordinate* (SPC) systems in the United States
  ⇨ These SPC systems were established to provide convenient local coordinate systems for surveying and mapping

- In the SPC system, each state is divided into one or more zones chosen so that the maximum scale distortion is no more than 1 part in 10,000

- To achieve this distortion limit, the north-south dimension of each Lambert conformal conic zone and the east-west dimension of each transverse Mercator zone are limited to approximately 254 kilometers (km)

- Each zone has its own unique set of defining parameters, with some states, such as Rhode Island, having a single zone, and other states, such as Alaska, having as many as 10 zones (Note that one of the Alaska zones uses the oblique Mercator projection)

5-6. Map Projections

- Another common map projection system is the *universal transverse Mercator*(UTM) system
- This system was established to provide worldwide coverage between 80° S and 80° N latitude by defining 60 zones, each having a 6° longitude range
- UTM zone 1 extends from 180° west longitude to 174° west longitude, with a central meridian of 177° west
- Zone numbers increase to the east, at an equal spacing of 6° longitude
- The value of the scale factor along the central meridian k_0 is equal to 0.9996 for every zone, resulting in a maximum scale distortion of 1 part in 2500
- Each zone has its origin (ϕ_0, λ_0) at the intersection of the equator with the central meridian
- The false easting for each zone is 500,000m
- The false northing is 0m for latitudes north of the equator, and 10,000,000m for latitudes south of the equator
- Other projections which may be routinely encountered are the *polyconic*, *polar stereographic*, and *space oblique Mercator*

5-7. Horizontal and Vertical Datums

- A *datum* is a system of reference for specifying the spatial positions of points
  - ⇨ These spatial positions are generally expressed in terms of two separate components, *horizontal* and *vertical*
  - ⇨ Thus datums have also traditionally been of two kinds, *horizontal* and *vertical*

- In a physical sense, a datum provide the basis for specifying the relative positions of points both for surveying operations and for mapping purposes

- Common horizontal datums encountered in the United States include the *North American Datum of 1927* (NAD27), the *North American Datum of 1983* (NAD83), the *World Geodetic System of 1984* (WGS84), various statewide *high-accuracy reference networks* (HARNs), and the *International Terrestrial Reference Framework* (ITRF)
  - ⇨ They provide a means of relating horizontal coordinates derived through surveying and mapping processes to established coordinate reference systems

- Although the theoretical development of horizontal datums is rather complicated, conceptually they can be considered to be based on three primary components as minimum constraints ⇨ a reference ellipsoid, an origin, and an angular alignment

5-7. Horizontal and Vertical Datums

- The North American Datum of 1927, for example, uses the Clarke 1866 ellipsoid as its reference surface which was a best-fit to the geoid in North America
- In a physical sense a horizontal datum such as NAD27 consists of the monument points along with their published coordinates. Since these coordinates were computed from a large number of measurements, each of which contains a certain amount of error, NAD27 likewise contains distortions due to these errors
- With advances in instrumentation, particularly accurate electronic distance-measuring devices, the distortions inherent in NAD27 began to cause difficulties in constraining the newer, more accurate measurements to the distorted system
  ⇨ To address this problem, the U.S. National Geodetic Survey created a new datum known as NAD83
- The World Geodetic System of 1984 was established by the U.S. Department of Defense during the same time period that NAD83 was being developed
- WGS84 is also the datum to which the broadcast ephemeris of the Global Positioning System (GPS) is referenced

5-7. Horizontal and Vertical Datums

- At the same time that NAD83 was being completed, GPS was beginning to be widely used for geodetic surveys
- Due to the exceptionally high accuracy of GPS, discrepancies were being revealed in the newly created NAD83
- As use of GPS expanded, these discrepancies became a significant nuisance for geodesists, and newer, more accurate datums were sought
- As a response, the National Geodetic Survey, in cooperation with individual states, began to establish high-accuracy reference networks (HARNs)
- The HARNs were individually connected to a highaccuracy worldwide network
  ⇨ therefore even though these networks have a high degree of internal consistency, there are discontinuities along the borders between states

5-7. Horizontal and Vertical Datums

- The horizontal datums are static systems
  ⇨ i.e., the coordinates of the monument points are based on a specific moment in time
- It has been well established that the surface of the earth is dynamic
- In addition, the rotational axis of the earth is continually on the move at a slow but detectable rate
- The International Earth Rotation Service established the *International Terrestrial Reference Frame*, or ITRF.
- The ITRF is essentially a three dimensional reference datum of high accuracy which is commonly used as a basis for precise GPS orbit determination
- Periodically, WGS84 has been refined so as to closely coincide with the ITRF

5-7. Horizontal and Vertical Datums

- In order to eliminate the need for separate HARNs for individual states, the National Geodetic Survey has readjusted the NAD83 datum using GPS and other measurements connecting the HARN points along with those for the existing NAD83 points
  ⇨ The result is the unified *National Spatial Reference System* (NSRS)

- Because of most recent readjustment, the new datum is commonly designated as NAD83(2011)
  ⇨ Compared to the location of the earth's mass center in the ITRF, there is a displacement of approximately 2 m to the NAD83(2011) origin

- Likewise, since WGS84 is periodically updated to the latest ITRF epoch, there is a displacement of approximately 2 m between WGS84 and NAD83(2011) origins

- A single common reference system should be used in order to produce consistent spatial data products

5-7. Horizontal and Vertical Datums

❖ A vertical datum is a reference system for giving elevations of points relative to the geoid (i.e., orthometric heights)

❖ Two primary vertical datums are currently in use in the United States
⇨ the *National Geodetic Vertical Datum of 1929* (NGVD29)
the *North American Vertical Datum of 1988* (NAVD88)

❖ The NGVD29 can be considered to be a mean sea level datum

❖ The NGVD29 evolved in much the same way as the NAD27 in that many additional vertical surveys were connected to the network in local areas
⇨ This fact, in addition to distortions in the datum due to measurement errors and constraint to the tide gauging stations, led to a vertical datum which was not sufficiently accurate for modern applications
⇨ Also, the increasing use of GPS dictated the use of a vertical datum that more nearly corresponded to the geoid
⇨ The NAVD88 was established based on a worldwide gravity model which is the geoid, and the NAVD88 is more compatible with worldwide horizontal datums

5-7. Horizontal and Vertical Datums

- It is often necessary to convert (transform) points that have been referenced in one datum to another
- These **transformations** have become especially commonplace with the increasing use of geographic information systems
- These systems often utilize information from different dates and different sources, and frequently the information is based on different reference coordinate systems
- But the information must all be coordinated in a common reference system before being integrated for analysis and use in a GIS
- A number of different **mathematical procedures** have been used for making these conversions
- Unless the transformation procedure appropriately accounts for the distortions in the datums, however, errors on the order of several meters can result in the converted positions

5-7. Horizontal and Vertical Datums

Several examples of conversion programs:

NADCON

- To aid in making accurate horizontal datum conversions, the NGS has developed a program called NADCON
- It can convert horizontal datum coordinates between NAD27 and NAD83 to an accuracy of approximately 15cm with occasional errors as high as 50cm

VERTCON

- A related program called VERTCON, performs vertical datum conversions between NGVD29 and NAVD88 to an accuracy of approximately 2cm

GEOID12A

- Another useful program is GEOID12A which can be used to compute geoid undulation values N within the area encompassed by the NAVD88
- These programs are available from the NGS on their website