NCSR
EDUCATION FOR A SUSTAINABLE FUTURE
www.ncsr.org

GEOGRAPHIC INFORMATION SYSTEMS

NORTHWEST CENTER FOR SUSTAINABLE RESOURCES (NCSR)
CHEMEKETA COMMUNITY COLLEGE, SALEM, OREGON
DUE # 9813445

FUNDING PROVIDED BY THE NATIONAL SCIENCE FOUNDATION
OPINIONS EXPRESSED ARE THOSE OF THE AUTHORS AND NOT
NECESSARILY THOSE OF THE FOUNDATION

GIS

The Northwest Center for Sustainable Resources is an Advanced Technological Education project funded by the National Science Foundation.

Geographic Information Systems was developed at Central Oregon Community College, Bend, Oregon. Materials were prepared by Art Benefiel, Lead Program Developer for NCSR. Benefiel holds a M.Sci. degree in Forest Resources and a B.Sci. degree in Forest Management from University of Washington, and an A.S. degree in Forestry Technology from Mount Hood Community College.

Technology education programs in which this course is incorporated are described fully in the Center's report entitled, "Visions for Natural Resource Education and Ecosystem Science for the 21st Century." Copies are available free of charge.

Course materials will also be posted on our website:

www.ncsr.org

Please feel free to comment or provide input.

Wynn Cudmore, Principal Investigator

Susie Kelly, Project Director
Northwest Center for Sustainable Resources
kels@chemeketa.edu/503-315-4583

Pre-Introduction for Instructors

This course is designed to provide instructors with quality lecture materials and overheads for a beginning course in GIS (with a forestry/natural resources emphasis). For information on technical support for GIS and training opportunities, and other educational materials and GIS-related sites, we suggest you research the following:

- www.esri.com

ESRI (Environmental Systems Research Institute, Inc.) is a company who prioritizes GIS in education. They offer numerous training workshops all over the country on ArcView® GIS.

> *We have a host of materials on our K12 website expressly for educators, and a page from which they can request a CD containing ArcVoyager Special Edition, a free GIS software for Windows and Mac. In addition, our "Community Atlas" project is one where classes can earn free GIS software for the school by submitting a complete project. If you're seeking training events, you might check out www.esri.com/ncge. Also, you might want to look at the Virtual Campus and see the courses there—students can download ArcView if they take the Introduction to ArcView class (120 day copy). There are other courses and applications on line plus the Library and other resources for use by students (even if they don't take a VC class). And look at the Searchable Database of programs on the Higher Ed page. That will give you some ideas about what is going on at other colleges. Join the Geography Generation!*

> www.esri.com/industries/university/university.html = Higher education
> www.esri.com/k-12 = info on GIS in schools
> www.esri.com/communityatlas = special projects
> www.esri.com/arclessons = GIS Lesson repository
> www.esri.com/gisedconf = ESRI Education Conference
> www.gisday.com = GIS Day

Charlie Fitzpatrick, ESRI Schools & Libraries, 1305 Corporate Center Drive, Suite 250, St. Paul, MN 55121-1204; USA; phone: 651/994-0823 x.8349; fax: 651/454-0705; cfitzpatrick@esri.com

Ann B. Johnson, Community College Manager, ESRI; 380 New York Street; Redlands, CA 92373-8100; (909) 793-2853, ext. 1-1793; FAX (909) 307-3039; ajohnson@esri.com

- pacificmeridian.com

> *Pacific Meridian Resources' activities in GIS and remote sensing include using satellite imagery and advanced image processing techniques to create land use and land cover maps for a number of markets. Pacific Meridian Resources has created information-rich databases and classifications as well as geospatial software for quantitative and statistical analysis, spatial modeling, growth simulation, change detection and various GIS applications. Pacific Meridian Resources also develops customized Web interfaces for vertical markets. With its corporate headquarters in Emeryville, Calif., Pacific Meridian Resources has 70 employees located in eight different offices across the United States. The company was founded in 1988 to pioneer the use of satellite imagery, geographic information systems (GIS) integration, and remote sensing application development. The company's services include online mapping services, software development for land use assessment and management, GIS and remote sensing services and market consulting.*

Table of Contents

Geographic Information Systems

Geographic Information Systems (GIS)
3 Credits—2 hours lecture, 3 hours lab

Text: Course Packet available in bookstore, Central Oregon Community College, Bend, Oregon.

Prerequisites: F190 *Introduction to Computer Applications in Forestry* or BA131 *Business Data Processing*. Students are expected to be comfortable using DOS and Windows.

COURSE DESCRIPTION:

Geographic Information Systems will introduce students to the principles and practice of GIS, while providing experience using ArcView^{R} and Idrisi^{R}. This course will develop both a theoretical understanding of GIS and experience in accessing GIS datasets. Students will be exposed to raster and vector GIS.

Geographic Information Systems combine spatial data (maps) with tabular data (databases) for the purpose of analyzing the environment. That environment can be as localized as an individual building, where one might analyze, for example, room use—or a forest where one might analyze the effect of beetle kill on fire spread; or nationally or internationally where you might analyze literacy rates and their effects on economic development.

At Central Oregon Community College, GIS is designed primarily for students in the GIS and Forestry programs. This class will also be beneficial to others interested in GIS, including geography, business, and health students.

TOPICS:

I. INTRODUCTION TO GEOGRAPHIC INFORMATION SYSTEMS

A. What is GIS?
B. Basic Geography
C. ...Continued
D. Hardware and Software components of a GIS

II. MODELS OF REALITY

A. Spatial Databases as Models of Reality
B. Raster and Vector Data Models
C. Networks and Fuzzy Logic

III. THINKING SPATIALLY

A. Sampling and Measuring
C. Maps and Map Analysis

IV. SPATIAL DATA

A. Environmental and Natural Resource Data

V. IMPLEMENTATION OF A GIS

A. Needs Assessment, Functional Requirements Studies
B. Requests for Proposals, Benchmarking, Pilot Studies

What is GIS?

OBJECTIVES (Figure 1-1)

- Define Geographic Information System.
- Outline origins of the field.
- Discuss GIS terminology and concepts.
- Compare the differences between GIS, CADD, and Automated Mapping.

I. WHAT IS GIS?

GIS = GEOGRAPHIC INFORMATION SYSTEM (FIGURE 1-2)

GIS can be defined as "any sequence of interrelated functions that achieves the input, storage, processing, and subsequent generation of spatial data."

1. A key element is "subsequent generation" via identifying spatial relationships between map features and producing new features previously not identified.
2. Most GIS are computer based; but it doesn't need to be.
3. ***Geographic*** because *locations of features in space* is important.
4. ***Informational*** because *attributes, or characteristics of the space,* is important.
5. ***System*** because there must be *a tie from the information to the geography* in a seamless operation.

II. WHY IS THERE SUCH A CURRENT INTEREST IN GIS?

There is current interest in GIS because of high levels of interest in new developments in computing. GIS gives a "high tech feel" to geographic information, maps are fascinating—especially in computers, and there is increasing interest in geography and geographic education. And GIS is a uniquely important tool in understanding and managing the environment.

III. WHAT IS THE LINEAGE OF GIS? It is the result of the marriage of many “old” and “new” disciplines. GIS is sort of a “technical offspring” that has had an influence from all of the specialties listed below. The key to most of these is the need to associate attributes with spatial data.

Geography—the descriptive science dealing with the surface of the earth. The physical features (especially the surface features) of a region, area, or place.

Cartography—the art of working with maps or charts.

Remote sensing—the science of deriving information about the earth’s land and water areas from images acquired at a distance. It usually relies upon measurement of electromagnetic energy reflected or emitted from the features of interest.

Photogrammetry—the process of surveying (measuring), or mapping with the help of photographs.

Surveying—the determination of the location, form, or boundaries of a tract of land by measuring the lines and angles in accordance with the principles of geometry and trigonometry.

Geodesy—the study of the shape of the earth.

Statistics—the science of assembling, classifying, and tabulating facts or data of a numerical kind, so as to present significant information about a given subject.

Operations research—the study of the processes or actions that are a part of a series of work.

Computer science—the study of devices and techniques used for computing; specifically electronic machines which, by means of stored instructions and information, perform rapid, often complex calculations, or compiles, correlates, and selects data.

Mathematics—the group of sciences dealing with quantities, magnitudes, and forms, and their relationships, attributes, etc., by the use of numbers and symbols.

Civil engineering—the branch of engineering dealing with the design and construction of highways, bridges, and waterworks, harbors, etc.

Automated Mapping/Facilities Management (AM/FM)—business management applications to track facilities [data control (FM), and mapping (AM)] that are spatially located.

Computer Aided Design/Drafting (CADD)—the use of computers to create digital representations of two-dimensional drawings and three-dimensional models.

Land Information Systems (LIS)—a term you might hear dealing with engineering applications of survey information (corner locations, etc.).

IV. GIS TERMINOLOGY AND CONCEPTS

A. Terminology

Features—landscape characteristics such as a road, forest stand, spotted owl nest, etc. On a map we depict features with graphical abstractions composed of points, lines, and areas. We further delineate features within these "feature classes" using shading and symbols.

Types of features (Figure 1-3)

Points—single coordinate pairs represented by *(x,y)* coordinates (e.g., spotted owl nest)
Lines—connected sets of points (e.g., a road system)
Areas—a tract identified by the coordinates surrounding the border (e.g., a forest stand)

Vector format—the above data model using points, lines, and areas to identify landscape features. Location data is only stored where there exists a feature of concern.

Raster format—an imaginary grid of cells are used to represent the landscape features. All space is stored, but much of the space may have no attribute value. The cells are delineated by a column and a row entry.

Types of features (from above) could be represented in the Raster Data Model as follows:

Points—individual *column, row* location cell.
Lines—a set of connected cells.
Area—all the cells within the interior of each feature.

Note: the cell is the smallest discernable unit in this format; any spatial detail smaller than this size is lost.

B. Electronic linkage (Figure 1-4)

Identification number—ties map features to the attribute table.

Attribute tables are databases that store information in rows (records) and columns (items or fields). *Note:* this is the vector method.

Raster method has a similar link. Each cell has an *implied ID number (#)* based on its column, row position in the grid. One normally talks in terms of row#, column# when locating a cell. Cell ID# 1 is the upper-left corner (row 1, column 1), then moving right along the row is cell ID# 2 (row 1, column 2). **Standard raster order** would therefore move from left to right and from top to bottom. A common item is an attribute in one database that serves as an identification number to access another database **(Figure 1-5).**

C. Data Structures

When you view maps, you make interpretations such as flow of water, boundaries that are alongside each other, etc. A GIS cannot do that without structure. This structure is termed **topology.**

Topology is the organization of data such that location, direction, adjacency, and connectivity of features can be determined by the computer.

Elements of topology in the vector world (Figure 1-6):

- coordinate systems
- tics—map registration points (i.e., control points)
- discrete points—locations you want to document (wells, bench marks)
- nodes—beginning and end points of a line
- vertices—directional change locations in a polyline
- arcs—line features
- polygons—areas enclosed by arcs
- annotation—text to clarify
- coverage—all the files that incorporate the above items to digitally identify a landscape feature

Elements of topology in the raster world:

- cell—the basic unit of the structure
- whole cell—a single characteristic throughout the interior of the cell
- partial cell—contains a mixture of characteristics
- lines—connected series of whole and/or partial cells
- areas—sets of whole and/or partial cells
- surfaces—describes the continuous distribution of gradient data (e.g., elevation)
- Coordinate System

In raster format we call this a Digital Elevation Model (DEM) for elevational data; it also can be called a Digital Terrain Model or Surface Image.

V. COMPUTER-AIDED DRAFTING/DESIGN (CADD)

Digital storage of maps/The Overlay Concept (Figure 1-7)

Digital means "in computer form." CADD Systems allow for easy editing but are limited in ability to have attributes tied to space. CADD is applicable primarily for drawing features.

VI. AUTOMATED MAPPING (AM)

Ties geography to databases

- displays data on map
- doesn't allow analysis; hence no "generation of subsequent sets of data"
- doesn't allow adding features

VII. COMPUTER-AIDED MAPPING—NEW WAYS OF CONSIDERING MAPS *"WHAT IT IS AND WHAT IT ISN'T"* (FIGURE 1-8)

A. GRAPHIC ELEMENTS VERSUS DRAWN LINES

Dynamic entity types

- *point*—includes text & blocks
- *polygon/area*
- *line/arc*
- *attributes*

Computer generated maps are created from different entities that have unique values and can be given unique identifiers.

B. DATA LINKS—INTEGRATING GRAPHIC AND ATTRIBUTE DATA

AutoCAD entities can be linked to databases using handles, extended entity data, block attributes, the new SQL extension (ASE) available in R12, and AutoCAD Data Extension.
Links to data still lack the topology needed for GIS.

C. FLEXIBILITY

Computer mapping allows for layers and provides different types of information. Different purposes can be contained in one map (refer back to **Figure 1-7**).

D. Mapping in real size/real world coordinates—the key to Computer-Aided Mapping

Real size/real object makes *Topological Relationships* possible.

- Adjacency/Connectivity
- Direction/Distance
- Area

Real World coordinates

- Different projections are available so work can be done in the projection system needed
- Can switch between systems without changing data.

E. Computer (digital) maps can serve more purposes than traditional mapmaking, thus have higher value (Figure 1-9)

The advantages of digital maps are:

- lower costs
- faster production
- accuracy—plotters, digitizers, equipment very accurate
- greater flexibility in output—easy scale or projections change; easy selection of features allow for greater flexibility

The disadvantages of digital maps are:

- not as automated as once thought, thus not as cost-effective as predicted
- computer methods do not ensure high quality

GIS and Computer Cartography

CAM was primarily designed to produce maps. CAM was not designed as an analytical tool because spatial and attribute relationships are not always stored with maps.

VIII. GIS COMPARED TO ANALOG (PAPER) MAPS—*KEY DIFFERENCE IS GIS'S ABILITY TO CREATE TOPOLOGY* (FIGURE 1-10)

A. Data Stores

Spatial data stored in digital form allows for fast retrieval. The nature of maps creates difficulties when they are used as sources for digital data. Most GIS takes no account of differences between datasets. Map generalization can become locked into GIS and into data derived from the GIS.

However, maps can be designed to be easy to convert to digital form. Maps can be produced from GIS.

B. Data Indexes

The GIS's ability to provide multiple and efficient cross-references and searches is ideal for indexing data.

C. Data Analysis Tools

The GIS can quickly and efficiently measure area, create overlays, and perform other analysis functions that are too cumbersome by hand. New techniques in spatial analysis are becoming available.

D. Data Display Tools—electronic display has many advantages

Data Display Tools allow:

- the ability to browse areas without constraints of map sheet boundaries
- the ability to zoom and change scale
- potential for animation of time-dependent data
- display in three dimensions for perspective views
- ability to control and change color for different purposes and makes it easy to produce one-of-a-kind products

References

Cartensen, Laurence W. and Henry, Norah F. 1995. *Digital Mapping and Geographic Analysis: An Introduction to Geographic Information systems.* Unpublished manuscript.

Cowen, D.J. 1989. *GIS versus CAD versus DBMS: What are the differences?* Photogrammetric Engineering and Remote Sensing 54:1551-5.

Environmental Systems Research Institute. 1995. *Understanding GIS: The ARC/INFO Method.* John Wiley and Sons, Inc. New York, New York.

Goodchild, Michael F. and Kemp, Karen K. 1990. *Introduction to GIS: NCGIA Core Curriculum: Units 1, 2 University of California Santa Barbara.* Santa Barbara, California.

Star, Jeffrey and Estes, John. 1990. *Geographic Information Systems: An Introduction.* Prentice Hall. Englewood Cliffs, New Jersey.

Basic Geography

OBJECTIVES (Figure 2-1)
Understand the basic concepts of geography pertaining to:
- Geodesy
- Map projections
- Map-based grid systems

I. THE EARTH

A. GEODESY IS THE STUDY OF THE SHAPE OF THE EARTH.

The original model of the earth was flat; then it was spherical. Isaac Newton thought the earth was ellipsoidal (flattened at the poles due to centrifugal force). There was much argument from 1670 to the 1730s over whether the earth's shape was spherical or ellipsoidal. Two French expeditions measured the earth at widely separate latitudes and found it was indeed ellipsoidal.

Ellipsoid (Figure 2-2). The term comes from *ellipse*, which means the path of a point that moves so that the sum of its distances from two fixed points (called foci) is constant; or the closed curve produced when a cone is cut by a plane inclined obliquely to the axis and not touching the base. The radii of an ellipse are not equal **(Figures 2-3 & 2-4)** because of flattening due to centrifugal force.
- semi minor axis through the poles
- semi major axis through the equator
- our flattening factor is 299:300 for minor:major

Clarke 1866 or 1880 was a very popular ellipsoid for the U.S. until recently:
- good fit for north America
- semi minor axis radius is 6,376,583.8 meters
- semi major axis radius is 6,378,206.4 meters

Geodetic Reference System of 1980 (GRS80) is now widely used in US, Canada, and Mexico:

- semi minor axis radius is 6,356,752.3
- semi major axis radius is 6,378,137.3

Geoid (Figure 2-5) is the hypothetical figure of the earth with the entire surface represented as taken at mean sea level.

- gets rid of terrain, removes weather, removes moon effect on tides, covers the surface with water
- surface is not smooth; it still varies due to gravitational pull of the earth, density of earth’s materials, and hydrostatic forces.

We use different ellipsoid and geoid measurements depending on our location on the earth **(Figure 2-6)**. Elevation can be either *Height Above Ellipsoid* (HAE) or *Height Above Geoid* (HAG). In Bend, Oregon, the ellipsoid is about 64 feet <u>below</u> the geoid.

Graticule (Figure 2-7). Carstensen(1995) defines the graticule as “the basic grid that defines position on Earth.” It is made up of two angular measures called *latitude* and *longitude*. The graticule is the set of lines drawn on most globes.

- **Latitude** measures the angular distance north and south from the equator to 90º N and S. These grid lines are also called **parallels**. Latititude lines run *parallel* to each other at almost equal spacing around the earth.

- **Longitude** measures the angular distance east and west from the prime meridian (the 0º starting point running through Greenwich, England) to 180º E and 0 to -180º W. These grid lines are also called **meridians**. Starting at either pole as a single point, the meridians diverge to the equator (at which point they are furthest from each other), then begin converging to meet at a single point at the opposite pole. This results in a 7 ½ minute quadrangle getting narrower as we go north.

Both latitude and longitude are measured in **degrees (Figure 2-8)**. Each degree (written as xx^0) may be broken down into 60 minutes (written as 60’). Thus, 45’ = three quarters of a degree. Each minute can be broken down into 60 seconds, written as xx. Thus, 15” is one quarter of a minute.

A. MAP PROJECTIONS (FIGURE 2-9)

The earth is three-dimensional; a map is two-dimensional. With a map projection, we must somehow get the 3-D “real world” into a 2-D representation.

- globes are hard to carry around
- globe scales are very small
- you can’t see all the earth at one time on a globe

A projection is a transformation of the features on the curved surface of the earth into a two-dimensional surface. It is defined by type of projection, coordinate units, and datum (and other parameters depending on type of projection).

Options would include FIPS (Federal Information Processing System) zone, spheroid, x-shift, y-shift, UTM zone, or semi-major-minor axis.

Desirable properties of projections are (Figure 2-10):

- **equivalence** (equal area; areas of features is accurate)
- **conformality** (true shape, lack of angular deformation)—shapes should be accurate
- **azimuthality** [true direction; the straight line path between two points should be the shortest actual route (i.e., the "great circle route")]
- **equidistance** (true scale; distances measured on the map, when adjusted for scale, represent correct distances between features)

Types of projections:

Planar (Figure 2-11)

- *azimuthal map* is only accurate from tangent point out to other points; it's good for small areas close to point of tangency

  Place a plane tangent to the earth—shine a light from either: a) center of the earth, b) antipode (opposite point of tangency), or c) infinite distance from the earth

Developable surfaces—project onto a geometric form that may be cut and flattened.

- **Cylinders (Figure 2-12).** Think about wrapping a cylinder oriented north/south to touch all the way around the earth at the equator; the equator would be known as the standard line (also called standard parallel). Shine a light from the center of the earth onto the cylinder. Cut the cylinder and spread it out **(Figure 2-13)**. The error in the properties of the map would be the least at the standard parallel (i.e., east and west) and increase as one moves away.

  Alternative ways to orient the cylinder would be: 1) to make the diameter of the cylinder smaller than the equator. This gives two standard parallels. The features are *compressed* between the parallels, and *expanded* outside the parallels; and 2) transverse align with a meridian of longitude **(Figure 2-14)**; the error in map properties would be least along the line from north to south, increasing as one moves east and west.

- **Cones (Figure 2-15).** Place a cone over the north or south pole (depending on area to map); tangent at a parallel line. The parallel that touches the cone is called the *standard line*. Project onto the cone. Then slice and unfold the cone **(Figure 2-16)**. The error would be least east and west along the standard line and increase north and south. Alternative ways to orient the cone could consist of: 1) use taller cones to have standard lines nearer the equator, 2) use squatter cones to have standard lines closer to the poles, or 3) have cones cut through the earth, producing two standard lines.

- **Projection Distortion (Figure 2-17).** *Zone of compression* (on the map), where the secant is within the earth (features on the map are smaller than they actually are) [*A|B to a|b on the figure*]. *Zone of expansion* (on the map),where the secant is outside the earth. (features on the map are larger than they actually are) [*C|D to c|d on the figure*].

For more information on types of projections and what properties are preserved in the various types, contact: US Geological Survey, 507 National Center, Reston, VA 22092,
703-860-6045 or **1-800-USA-MAPS.**

C. MAP-BASED GRID SYSTEMS

Grid systems allow various users to have a common reference on the ground.

REFERENCING MAPS

How are they referenced? From "known" points that collectively are called a **datum** or Geodetic Control Network (NGS). "Datum" are cross-country networks established using trilateration surveys, EDMs, Bilby towers, and lots of ground work. The datum serve as a starting point.

What do the networks look like? **(Figure 2-18)**
- North American Datum of 1927 (called NAD27)
- Specific known monuments
- Clarke ellipsoid 1866 used
- Meades Ranch, Kansas—taken as initial point, the 25,000 known monuments were adjusted to this initial point
- Until NAD83 came out, all stations established were adjusted to be consistent with NAD27

North American Datum of 1983 (called NAD83) includes stations, plus nearly 2,000,000 geodetic surveying measurements on record. It was initiated in 1974 for completion in 1983 (hence the name, NAD83), but it wasn't finished until 1986. The *GRS80 ellipsoid* was used (Geodetic Reference System). It: 1) fits the earth globally better than Clarke, 2) is consistent with GPS (Global Positioning Systems), and 3) matches better with established worldwide lats./longs. There were shifts from NAD27 to NAD83) due to the different ellipsoid used and additional measurements. *Shifts in latitude:* about 2 arc seconds in the north to about 1 arc second in the south **(Figure 2-21)**.

Shifts in longitude: 4-5 arc seconds (about 400') in the west to 0 in the central U.S. to 1 arc second (about 100') in the east **(Figure 2-22)**.

The advantages of NAD83 are 1) control is now consistent and uniform in accuracy, and 2) it is practically free from distortion.

What do the NGS control points look like? **(Figure 2-23)**

The Federal Geodetic Control Committee (FGCC) has published a detailed set of survey accuracy specification. *All the NGRS control points are First Order* (read further for more information on NGRS).

Horizontal Control Accuracy Standards **(Figure 2-24)**

First Order:		
		1 in 100,000
Second Order:		
	Class I	1 in 50,000
	Class II	1 in 20,000
Third Order:		
	Class I	1 in 10,000
	Class II	1 in 5,000

Vertical control accuracy (tied to Benchmarks)

First Order		
	Class I	$0.5\text{mm} \times \sqrt{K}$
	Class II	$0.7\text{mm} \times \sqrt{K}$
Second Order		
	Class I	$1.0\text{mm} \times \sqrt{K}$
	Class II	$1.3\text{mm} \times \sqrt{K}$
Third Order		$2.0\text{mm} \times \sqrt{K}$

K is the distance between benchmarks, in kilometers.

The National Geodetic Reference System (NGRS) now has more than 270,000 horizontal control monuments, and approximately 600,000 vertical control monuments (called bench marks).

High-Accuracy Reference Networks (HARNS); also called High Precision Geodetic Networks (HPGNs): All states are moving to adopt this as a result of GPS technology. Accuracies of NAD83 are less than GPS.

How does one tie GPS data to a less accurate system?

- Establish "Order A" sites (about 5 per state); tie to the NAD83 stations known to be the most highly accurate.
- Add additional "Order B" stations relative to the "Order As."

The HARNs create an updated datum. Wisconsin calls theirs NAD83(1991). It allows for better distribution of stations than triangulation. The future will densify the stations.

•Horizontal control accuracy

Order A	1 in 10,000,000
Order B	1 in 1,000,000

Two Cartesian coordinate systems are used for the majority of maps: *State Plane Coordinate System and Universal Transverse Mercator*.

I. The State Plane Coordinate System (SPCS), maintained by the U.S. Coast and Geodetic Survey has an allowable error of 1 in 10,000 (0.01%) between the map and ground location. A state uses as many zones as necessary to meet error requirement (e.g. Oregon has two zones—north and south; California has seven zones, I-VII; Indiana has two zones—east and west) **(Figures 2-25 & 26)**. Each state designates a grid; a false origin is established outside each zone to insure positive values for all locations. Locations are measured in feet from origin (some now in meters). Locations are measured to the right (called *false eastings*) and upward (called *false northings*).

For example, Oregon's origins are:

- north zone 43°40'N latitude, 120°30'W longitude; x(ft) 2,000,000, and y(ft)0.
- south zone 41°40'N latitude, 120°30'W longitude; x(ft)2,000,000, and y(ft)0.

The two types of projections most commonly used are: **(Figure 2-27)**.

Lambert Conformal Conic (cone is secant to the ellipsoid)

- two standard parallels are placed at one-sixth the zone width from the north and south zone limit
- the standard parallels for Oregon are: north zone 44°10' & 46°00', south zone 42°20' & 44°00'
- a central meridian located near the center of the area to be mapped establishes "grid" north
- lines parallel to the central meridian reference "grid" north—not "true" north, since "true" north lines would converge at the top of the cone.
- "conformal," hence, no angular deformation.
- for states oriented east/west (e.g., Washington and Oregon).

Transverse Mercator (transverse-oriented cylinder tangent to the ellipsoid)

- the central meridian is located near the center of the area to be mapped
- establishes "grid" north
- points are measured perpendicular to and parallel with the central meridian
- all parallels of latitude and meridians (except for central meridian) are curved
- conformal
- for states oriented north/south (e.g., Indiana)

II. The Universal Transverse Mercator (UTM)—originally developed by the military **(Figure 2-28)**. There are 60 zones encircling the globe.

Each is 6°; beginning at longitude 180°W; Zone 1 includes 174°W to 180°W and is numbered easterly. The US is covered by zones 10 (west coast) to 20 (east coast).

Accuracy at edges is 1 in 2500 (0.04%). The central meridian is at the center of each zone (i.e., 3° on either side) **(Figure 2-29)**. The Central meridian of zone 1 is 177°W, assigned the value 500,000 m. The equator acts as a division between the north and the south part of each zone; north portion has equator a 0 m, and south portion has equator as 10,000,000 m. This avoids any negative coordinates. Each zone overlaps adjacent ones by 0°30'. The grid north reference to the central meridian includes parallel lines to the central meridian point to grid north, and true north converges at the top of each zone (all measurements are in meters).

References

American Society of Civil Engineers, American Congress on Surveying and Mapping, American Society for Photogrammetry and Remote Sensing.1994. Glossary of the Mapping Sciences. American Society of Civil Engineers, American Congress on Surveying and Mapping, American Society for Photogrammetry and Remote Sensing. Bethesda, Maryland.

Campbell, John. 1993. Map Use and Analysis. Wm. C. Brown, Dubuque, Iowa.

Carstensen, Laurence W. and Henry, Norah F. 1995. Digital Mapping and Geographic Analysis: An Introduction to Geographic Information Systems. Unpublished manuscript.

Environmental Systems Research Institute. 1992. ArcCAD Command Reference. Environmental Systems Research Institute, Redlands, California.

Goodchild, Michael F. and Kemp, Karen K. 1990. Introduction to GIS: NCGIA Core Curriculum: Units 26,27,29 University of California Santa Barbara. Santa Barbara, California.

Wolf, Paul R., Brinker, Russell C.1994. Elementary Surveying: ninth edition. Harper Collins College Publishers, New York, New York.

GIS Hardware and Software

OBJECTIVES (Figure 3-1)
- Become familiar with various hardware associated with GIS.
- Become familiar with various software components of GIS.

I. BASIC GIS HARDWARE (FIGURE 3-2).

***Data Input* includes:**
- **Digitizers** which are a grid of wires imbedded within a board with an electrical charge running through the wires; a puck produces a magnetic pulse that can be pinpointed by the grid
- **Keyboard**
- **Scanners** are an optical device that records images in a digital form (e.g., a photocopy machine is a scanner)
- **CD-ROM**
- **Voice** (more futuristic at this point)

Central Processing Unit (CPU)
Volatile storage—Random Access Memory (RAM) (data is lost when machine is turned off; why "volatile")
Non-volatile storage—hard drives, floppy disks, and tapes
Archival—read/write optical
Backup—tapes used for daily/weekly/monthly backups
Output—monitors, printers/plotters, disks/tape drives, Internet, and servers

II. CATEGORIES OF GIS HARDWARE (FIGURE 3-3).

A. Personal Computers (PCs)

Personal computers are the general term for microcomputers such as IBM or IBM-compatible microcomputers. Desktop types range in price from $3,000-$7,000. Users include learners with small areas and small problems (e.g. counties, district offices). PCs are generally used for front end or limited processing, such as digitizing and usually feed up to a higher level. However, now with faster processors, more ram, and greater storage, pc's are performing at the level of a workstation of just a few years ago.

B. Workstations

Workstations usually consist of a single-user computer station designed for high performance graphic and numerical computations. These computers are designed for the desktop with the power of a larger computer ($10,000-$75,000). Workstations don't appear to look much different from a PC on the outside. Currently the platform of choice for most commercial GIS products is Arc/INFO[R], Intergraph[R] using workstations.

C. Mainframes (e.g., VAX, PRIME)

Mainframes are central computers with many terminals tied to them. They are high level computers designed for the most intensive computational tasks. The most powerful are called "super-computers." Users include researchers, theoreticians, and space applications.

III. GIS SOFTWARE

There is great variability in software products. A way to look at how software might be structured is to consider the following:

A. Functional Elements of a Geographic Information System

Five elements are identified by Star and Estes (1990), based on Knapp (1978) **(Figure 3-4).**

1. **Data Acquisition**—issues are costs (time and money), accuracy and precision (i.e., lineage of the data and age), data structure (does topology exist? is importing possible?), sources of data—collection of new data (surveys, generating maps, aerial photo flights), and existing data (maps, aerial photos surveys, documents, and other remotely sensed imagery).
2. **Preprocessing**—elements are conversion to *digital format* (digitizing and scanning)s, and *sanitation* (undershoots, overshoots, unclosed polygons, sliver polygons, and unjoined lines), and consistency of system for spatial locations (UTM, SPCS), and consistency for recording data (attributes), which ensures similar attributes, and eases error checking and verification.
3. **Data Management**—using the *database theory,* creation of the database (structure and propagating the database); having access to the created data (doing queries—is it "user friendly"?); file management for the database should describe its contents, list, copy, etc.
4. **Manipulation and Analysis** (often called **geoprocessing**) includes *projections* and *transformations* (e.g., from digitizer inches to UTM map coordinates)—this entails *translation* (adding, subtracting a value), *scaling* (multiplying by a value), and *rotation* (aligning the X,Y axis). Deriving "new" information involves geographic queries, interpolation, distance calculations, screen queries, etc. Output to/results back from, and external systems such as spreadsheets and other numerical programs are also part of geoprocessing.
5. **Product Generation—**includes statistical reports, maps, graphics (bar charts, pie charts), hardcopy and softcopy (i.e., files vs. printed material).

This seems to follow Carstensen (1995), who identifies five necessary components of GIS software: 1) data input and verification, 2) data storage and management, 3) data presentation, 4) transformation systems, and 5) User interface.

V. PERSONNEL

Personnel must have an understanding of the data, the hardware, and the models used. Most systems are not used to their full potential due to lack of trained personnel.

References

Carstensen, Laurence W. and Henry, Norah F. 1995. Digital Mapping and Geographic Analysis: An Introduction to Geographic Information Systems. Unpublished manuscript.

Environmental Systems Research Institute. 1995. Understanding GIS: The ARC/INFO Method. John Wiley and Sons, Inc. New York, New York.

Goodchild, Michael F. and Kemp, Karen K. 1990. Introduction to GIS: NCGIA Core Curriculum: Units 1, 2 University of California Santa Barbara. Santa Barbara, California.

Knapp, E. 1978. Landsat and Ancillary Data Inputs to an Automated Geographic Information System. Report No. CSC/tr-78/6019. Silver Springs, Maryland: Computer Science Corp.

Star, Jeffrey and Estes, John. 1990. Geographic Information Systems: An Introduction. Prentice Hall. Englewood Cliffs, New Jersey.

Spatial Databases as Models of Reality

OBJECTIVES (Figure 4-1)

- Become familiar with the concept of *modeling reality* in a GIS.
- Understand various options for data models underlying GIS.
- Understand how data structures contribute to GIS functionality.
- Understand how file structures may ultimately effect GIS software efficiency.

I. INTRODUCTION

A Geographic Information System is attempting to *model reality* (the world), through selection of a *data model* (a way to conceptualize reality in the computer). Then a *data structure* (how to order and display the spatial and attribute data) must be decided on—which requires a *file structure* (how the computer scientist writes the data to files). To understand how a GIS works, it is necessary to have insight into how GIS software might undertake these charges.

II. REALITY

"Reality" is the world, what you want to model, actuality, trueness, existence—what is real or absolute; it can be simplified into different dimensions. We can generalize reality and define it by the following (in order to model reality):

- **Points (Figure 4-2)** are single locations or occurrences; they are dimensionless; having no length, width, or area. They are depicted as only an x,y coordinate (a "z" value can be added).
- **Lines** (also called **"arcs"**) **(Figure 4-3)** have linear features that follow a path between two points. They have length only (no width or area). There are different kinds of lines such as networks, trees, bi-directional, unidirectional, and disconnected.
- **Areas (Figure 4-4)** are enclosed space composed of a set of connected lines (arcs). They are either isolated, adjacent, or nested. There is homogeneity within the area.

Examples of feature classes **(Figures 4-5, 4-6, & 4-7)**.

There are problems with trying to model reality. Any model we use attempts to locate something in space. Often it is not easy to pin down an actual location; e.g., a building—which point do you locate if it is to be located as a point entity? (it is not easy to describe)—what attributes should you store?

Relationships among entities are numerous and complex (adjacency, connectivity, nearness). Which do we use and maintain in a database? The many relationships are difficult to maintain. *Computers only know absolutes, not relatives* (e.g., "far" means two different things to two different people.)

Space (as we model it) is 2-dimensional at a minimum, 3-d in reality, 4-d with time. Computers are linear—they have a tough time with 3-dimensions, much less adding temporal considerations.

Three goals of modeling reality are:

1. **Completeness**—you must decide on how much reality to capture
2. **Robustness**—the model must be flexible so unforeseen uses can be matched, e.g., *are we ready for 3-D?*
3. **Efficiency**—data must be small and easily/readily accessible

III. SPATIAL DATA MODELS

What is the best way to store the data? The conceptual description of *a database* is the method of storing the X,Y data about which one is concerned.

A. Maps (Figure 4-8)

Much GIS and database design is based on maps. A good model of the reality of the earth, is data coded for retrieval; data is simplified, and data is efficiently stored (i.e., in a map file drawer). Spatial relationships are maintained. Problems with data storage is that it is often difficult to update, hard to interpret, and requires skill.

B. Vector (Figure 4-9)

The vector data model will be discussed in much more detail in *"Raster and Vector Data Models"* in this curriculum packet.

Reality is stored by object. Finite locations are based on desired features. Only these locations should be coded using points, lines, and areas. Features can be located at any location.

Types of vector data models:

- **Spaghetti**—just points, lines, areas with no connectivity, adjacency, etc. AutoCAD^R .dxf files, or IDRISI^R .vec. files. These models would only exhibit graphic properties, duplicate coordinates, and entity or blocks exist independently.
- **Topologic**—these modes store, in addition to what was stated in spaghetti, connecting lines and polygons associated with them. Using these models, we no longer use the term "line" but rather "arcs." **Topology** is the concept that expresses the spatial relationships between arc features including connectivity of lines, direction of lines, adjacency (contiguity) of polygons, and island polygons.

C. TESSELLATION'S

In Tessellation's model, reality is stored by space. Using Tessellation's, space is divided up into discrete, mutually exclusive areas. An imaginary grid produces a finite number of locations on which to collect data. Tessellation's uses "cells" to depict points, connected cells to produce a line, and adjacent cells to demonstrate an area. Features may only be located at the discrete points.

Types of tessellation's:

Grid and Regular Tessellation's (called **"raster"**): each subunit of space is of the same type. The raster data model will be discussed in much more detail in *"Raster and Vector Data Models."*

Squares are probably the best known Raster grid. They are compatible with human conceptions of breaking up space and are good for row/column types of hardware such as monitors and printers.

Triangle (Figure 4-10): all neighbors are equidistant; all are not the same orientation. Often time directions are not very good. There are only three same distance neighbors (four in square grids).

Hexagons have six equally spaced neighbors. They have six directions for flow (i.e., slopes, networks). They have the most compact shape, and provide more accurate orienta tion. There are no right angles; no rows/columns. Hexagons are hard to program or display. There are no commercial GIS uses of this tessellation, but some governmental coastal studies have used it.

Hierarchical Tessellation's are recursive models of subdivisions; units are the same shape but not the same size; only squares have been used because they allow subdivisions to maintain orientation.

Hierarchical Tessellations: *Quadtree* (Figure 4-11)—represents the four times increase of members as units are subdivided and the fact that the structure; it is easily pictured as a tree.

How do you make a quadtree? (this example is from the top down)

1. Start with a Raster map of 0's and 1's.
2. Pad Raster to a power of two (i.e., edges must be a power of 2 with row and columns equal).
3. Check entire map (i.e., if all 0's or 1's—do nothing; if not—continue); recall and subdivide, recheck entire map again, recall and keep subdividing until there are no mixed cells.

Characteristics of a quadtree

Variable resolution.
In a regular tessellation there is wasted space with lots of repetition. The large features have coarse resolution and the fine features may have a finer resolution.

Storage of a quadtree.
Each node can be represented by two bits (11 means terminator is "in"; 00 means terminator "out"; 10 means terminator is "in" at the current level, and 01 means terminator is "out" at the current level). Therefore, our example could be saved in 36 bits (6 bytes)—versus the 64 bytes used for standard raster order (thus, the quadtree results in *raster compression*).

Search in a quadtree. **(Figure 4-12)**
It uses a *"Morton" sequence* which is a numbering method. This method identifies a search pattern for the quadtree (e.g., a "Z pattern"). It stores color, level, and location of each cell, (i.e., the *Morton number*).

Advantages of quadtree.
Some processing does not require all the resolution of the original Raster (e.g., analysis between two layers of different spatial resolution). Processing may minimize certain searches through the data-base (e.g., search for a "high" point, only look for the high point of each level). However, it doesn't ensure the highest point of all data points.

Other methods of raster compression include:

- **Chain codes (Figure 4-13)**: the boundary is depicted as a cardinal direction from a starting point (e.g., east = 0, north =1, west = 2, south = 3). Then, using our example—moving clockwise, starting in row 2, column 7; the descriptor would be: 3^5, 2, 3^2, 2, 1, 2, 1, 2, 1^2, 0^3, 1^3, 0.
- **Run-length codes (Figure 4-14)**: all 14 cells would be coded with the following seven numbers: Row 2—7,7; Row 3—7,7; Row 4—7,7; Row 5—4,7; Row 6—4,7; Row 7—5,6; Row 8—6,6.
- **Block codes (Figure 4-15)** are coded by three numbers: the origin (X,Y), and the radius of the square. These codes are called a *medial axis transformation* or *MAT*. For example, a region could be stored in 6, 1-squares plus 2, 4-squares; 18 numbers would be needed to code the data; and 16 numbers would be needed for the coordinates plus two for the square sizes. Our example requires more space, but you can see if you had large, simple areas this would be a good method.

IV. DATA STRUCTURES

Data structures are utilized to order and display the spatial or attribute data. This discussion is limited to two structures—**hierarchical** and **relational**. There are many others (e.g., whole polygon, arc-node(in which relational is a form), and network. See Star and Estes (1990) or Burrough (1986) for places to start if you are interested in further information on data structures.

A. Hierarchical (Figure 4-16)

Hierarchical data are different levels of data such that parent levels are aggregates of the child levels; this allows reclassification of the data across categories (i.e., US Census, state, county, township, census tract, block group, block, street, address, name of person; Public Land Survey System township/range, section, 1/4 section, 1/4-1/4section).

Advantages of hierarchical is that it is easy to understand, updates easily, and is easy to expand.

Problems with hierarchical is the data distance is dependant on the query—only vertical queries are possible (i.e., there could be a long way through the data structure to satisfy a query—edges of a single polygon might not be too bad, but how about the bounding rectangle of an entire map?); and often data is repeated.

B. Relational (Figure 4-17)

Relational data is conceived as 2-dimensional tables. This data links tables together by repeating variables from one table to another. This concept is called a "common item" in some relational databases, or the "primary" and "foreign" keys in others.

Terms:
tuple: one row of data
field: one column of data
record: one value for a given tuple/field location
relation: an entire table
relational join: logical combining of two relations based on a primary key in the first relation and its match (or foreign key) in the second relation.

Considerations:

- no two tuples can be duplicated in one relation
- at least one column must be unique to be the primary key
- no primary key field may be empty or null
- there would be no way to tie to another relation

Advantages of relational data are that it is flexible for queries (mathematical, and boolean—true/false); addition and removal of data is good; different data types are searched, combined, and compared; there is less data distance than hierarchy.

Disadvantages include large files (sequential searches are slow); multiple joins and/or long tuples slows the system due to tracking pointers between the data; sometimes it's just not the best structure—networks are, after all, networks.

V. File Structures

File structures are what, or how, the computer scientist has written data to the disk (e.g., pointers, end of file markers, etc.). File structures are the most basic element of data storage and retrieval; data in a computer is linear and sequential. RAM is ordered from 0—640K in a sequence; disk sectors and tracks are lined up.

How do we load our non-sequential data into memory?

Simple lists (Figure 4-18): there is at least one structure—you begin at the top, and go to the bottom (this is the structure); new data is added at the end. The file gets longer and longer—it can be compared to a shopping list. *What is the structure? What is the source of the structure? Is there a better structure? How is the list used?* (i.e., eggs, buns, milk, cheese, sugar, apple, taco shells). Retrieval is inefficient (e.g., *find cereal in your shopping list of 50 items*); if a polygon is in a simple list and the order of the verticies changes, the polygon changes. The only search, other than to guess, is to start at the start and go to the end.

Advantages of simple lists are that these lists are simple and there is no wasted file space.

Disadvantages include inefficient searches—to find something takes n/2 +1 searches on average (n being the number of items in the list).

Ordered lists (Figure 4-19) are simple lists with a second structure. You are able to sort and itemize (e.g., a phone book's first structure is start to end; its second structure is alphabetized by last name); binary searches are possible—you can split the number of records to search in half after each search. For example, using a phone book, and looking for letter "E," open to the middle (one half of the data), if greater than "E"—"M," for example, split the lower half (one-half of remaining data)—now you get "C," so take the upper portion of remaining data and split, finally getting to "E." To find something takes $\log_2(n)+1$ searches at most (e.g., 1000 items in an ordered list takes $\log_2(1000)+1 = 11$ searches). Each *doubling* of the amount of items in your list only adds *one* more search.

Advantages of ordered lists are that they are fast, efficient, and searching is good for the sorted field.

Disadvantages are that there is lots of overhead (i.e., pointers); data updates must occur in the proper way (i.e., you can't just add to the end of the list); searching is not good for all but the sorted field (e.g., think about finding a first name in the phonebook).

Indexed lists (Figure 4-20) produce a logical sort on the list—they have a second file (called an "inverted file") that reads a simple list and tells the order in which the data should be arranged. It allows us to get around the disadvantage of the ordered list of trying to search on something other than the sorted field (*note:* if you are familiar with dBase this is what *indexing* does—the .ndx file extension). It takes more storage space adding the second file, but it is much smaller than storing the entire original file as many times as needed to have it sorted on all items.

Advantages of indexed lists are faster searches.

Disadvantages are that updates not straight forward—additions/deletions must be made to both the file and the index; often data can only be accessed from the key file contained in the index file; other data is only retrievable with a sequential search. You might get around this by copying the key field from the original file into the index. Pointers could then be calculated mathematically and a binary search performed.

References

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford University Press. New York, New York

Carstensen, Laurence W. and Henry, Norah F. 1995. Digital Mapping and Geographic Analysis: An Introduction to Geographic Information Systems. Unpublished manuscript.

Environmental Systems Research Institute. 1991. Introducing ArcCAD Release 11: Course Manual. Environmental Systems Research Institute, Redlands, California.

Goodchild, Michael F. and Kemp, Karen K. 1990. Introduction to GIS: NCGIA Core Curriculum: Units 10,11,12,13,14,21,30,31,35,36,37,43, and 44. University of California Santa Barbara. Santa Barbara, California.

Star, Jeffrey and Estes, John. 1990. Geographic Information Systems: An Introduction. Prentice Hall. Englewood Cliffs, New Jersey.

Raster And Vector Data Models

OBJECTIVES *(FIGURE 5-1)*

- Be able to discuss the raster data model.
- Be able to discuss the vector data model.
- Understand some of the various vector data models that have been used over time to store polygons.
- Understand the differences between raster and vector data models.
- Understand the situations when one of the models might be better to use.

I. INTRODUCTION

Maps were the original spatial database, But, they were not so good for updating and analyzing (i.e., crunching numbers).

We now use computers because they are: fast (once we have the data in the proper formats), accurate (plotters, digitizers, etc.), and relatively cheap. Computers were expensive in the past, and labor was cheap. Now computers are relatively cheap while labor is expensive. Computers allow more flexibility. Different versions may be produced, and you are able to choose the best one. ***(Figure 5-2)***

We have to tell computers how to perform most of the interpretation of maps that humans do that lead to ways to represent and model reality. The things we do cognitively manually often become difficult to teach a computer to do (e.g., inferences). We must define all terms and features, quantify. At the start of the teaching process is the requirement that we operators decide how reality is going to be modeled. The "Spacial Database" section of this course outline discusses the overall concepts of modeling spatial data, here we will cover **raster** and **vector** data models in more depth. ***(Figure 5-3)***

II. THE VECTOR DATA MODEL

The American Heritage Dictionary defines a vector as a mathematical concept wherein "a quantity is completely specified by a magnitude and direction." In surveying, a vector can be thought of as where we go when following a certain azimuth for a certain distance.

A. Vector Storage *(Figure 5-4)*

The basic unit in vector storage is a point (X,Y coordinate). Connected points make lines. Lines may enclose a polygon. Vector storage is object based (i.e., feature based). It tends to store geographic features as objects; i.e., a line is stored sequentially in the database. Vector storage mimics the manual method. It will digitize (or trace) similar to the way we draw maps. It is important for objects in features to be in order; i.e., points and segments. The order of objects is not important; i.e., one line before another is not important. A discrete data model specifies the distinct location, and has specific edges. Change is abrupt (hence discrete). Emphasis is on the entity. Location is an attribute and location of the entity must be stored. Adjacency is <u>not</u> inherent in a vector file. A typical file tells nothing of what is next to what. We need **topology** for this.

B. Store by object

A finite number of locations are defined by the objects. This method only deals with locations where a feature of concern is.

C. Vector Precision

Vector precision is encoded with any conceivable degree of precision. Precision is limited by the method of internal representation of coordinates [typically 8 (single precision) or 16 (double precision) decimal digits].

Precision is actually measured in significant digits:
AutoCAD^R is double precision (to 13 digits)
ArcCAD^R is single precision (to 7 digits)
Arc/INFO^R is double precision—limits precision to $1/10^8$ or $1/10^{16}$ of the size of the study area, and equivalent raster would need 10^8 by 10^8 or 10^{16} by 10^{16} cells, neither of which is feasible.

✐Note: There may be an "artificial argument" because real vector data accuracy may be much worse that one line width.

Data precision is vector true for certain classes of data. The data is captured from precision surveys (coordinate geometry (COGO). Plat maps are created from land survey coordinates. Few natural phenomena have true edges which can be accurately represented as mathematical lines.

Soil and vegetation types, etc., have fuzzy boundaries. It can be argued that fine lines from the vector system give a false sense of precision and accuracy. Lines on maps are typically 0.5mm wide and are often assumed to represent the uncertainty in the location of the object. In raster-based maps, uncertainty is automatically reflected in the cell size. Data precision allows a concept of accuracy and precision ***(Figure 5-5).***

D. Types of files that have been used over time for storing polygons

The location list ***(Figure 5-6)*** is a response to the original line plotters (SYMAP). Start at one point and proceed around the polygon back to the start. Each polygon is a separate feature (independent entity). There is no tie between them in the database. List a polygon (i.e., "polygon A") then list the points around it (first point = last point). This may be wasteful in that one has to double enter the internal boundary points. Along came pen plotters and showed the problems of the adjacent boundary double entry.

The *points dictionary* ***(Figure 5-7)*** is a response to the pen plotter. It allows us to digitize each point only once and number all the points.

point	X-coordinate	Y-coordinate
1	23560	13254
2	25551	15173
3	15786	16552

We can then have a *polygon definitions* list identifying what points are used for each polygon.

polygon	points
A	1,2,5,7,11,16,1
B	1,9,10,11,12,13,14,4,3,2,1

We do not repeat point coordinates at entry to get a neat edge.

E. Topological Structures

Topological structures are those that can track adjacency, connectivity, etc.

DIME (Dual Independent Map Encoding) ***(Figure 5-8)*** was used by the US Census Bureau. Its first application was used in the 1970's. It was the first structure to really build databases for further analysis; not just to draw a map. DIME uses a table of data rather than a stream of X, Y coordinates, and uses nodes, polygons, and coordinates.

from node	to node	L poly	R poly	X from	Y from	X to	Y to
1	2	0	A				
2	7	C	A				
2	3	0	C				
3	4	0	C				

Any point in this system is called a **node.** In other systems, nodes are only end points. Lines are made up of segments between nodes. Each *segment* requires a record. In the point dictionary method each point has a record.

Advantages of DIME include:

- Connectivity
- Stores lots of info with various retrieval methods and pulls out all "polygon A" references to draw polygon "A"
- Adjacency

There is still repetition in this system; i.e., end points of lines, and big files due to a record for each node.

Chain files (called Arc/Node or String files)**(Figure 5-9).** This file structure utilizes a chain which includes all points between two nodes. A **node** in this system is any point at which the left or right polygon changes; i.e., a point where 3 or more chains are attached.

from node	to node	L poly	R poly	# points	coordinates
1	2	0	102	14	X_0,Y_0 ———— X_{13},Y_{13}
2	1	101	102	6	X_0,Y_0 ———— X_5,Y_5
1	5	103	0	27	X_0,Y_0 ———— X_{26},Y_{26}
1	4	101	103	4	X_0,Y_0 ———— X_3,Y_3

Only <u>nodes</u> (end points of lines) are repeated now. The structure even store coordinates in another file and still maintains adjacency, connectivity.

III. THE RASTER DATA MODEL

A. Raster storage *(Figure 5-10)*

In this model, the basic unit is a cell. A row,/column identifies a cell location. Connected cells form lines. Areas are formed by all the cells within the interior of each feature. A cell size drives the resolution of the grid. Locations can be converted from an X,Y coordinate to the I,J (row/column) of a raster by the following:

$$I \text{ (row)} = \text{integer of } \frac{(Ymax—Y)}{(dY)}$$

$$J \text{ (column)} = \text{integer of } \frac{(X—Xmin)}{(dX)}$$

This storage is space based and stores the attribute of each discrete unit of space defined by a cell. It mimics certain forms of digital data; i.e., remote sensing data, and digital elevation models (DEM's). Objects in the same feature are not stored together. They are stored in *standard raster order* from left to right, top to bottom. A "raster" is actually an entire row of data.

The raster data model is a discrete model. Within a cell you normally assume homogeneity. The new "fuzzy" theory changes this a bit but still assumes homogeneity. There is only one value per cell (need I,J,Z). You actually only need the Z for each cell as I,J distances are the same due to cell size. After the location of the first cell is known the file is only made up of Z values.

Location is emphasized. Entity becomes the attribute (VS attribute is location in vector). You must store the entity of the location (VS store location of the entity in vector).

Adjacency is inherent to the system. Thus, for any location, we know what is next to it, above it, below it, and to the left and right. You are unable to do this in vector without topology.

NW	N	NE
W	?	E
SW	S	SE

B. Data is stored by space

The method takes a continuously varying space (infinite # of points) and converts it to a finite number of points called "cells" (or pixels). All space is stored even if no feature is there.

C. Raster precision

Locational precision is limited by the size of the cells. Acre cell size fits well into the Public Land Survey System, but all features have to be represented as 1/4 mile wide strips.

How about smaller cell sizes? At a scale of 1:24000, the width of a #00 pen line is about 5 meters, and that is not good enough; e.g., it is still not possible to represent objects smaller than 5 meters, e.g. power poles, fire hydrants.

Where is the location of the I,J in the cell (middle?, upper left corner?) it is often unclear where the coordinate is; therefore, locational precision is 1/2 the cell's width and/or height.

IV. VECTOR VS. RASTER *(FIGURE 5-11)*

These two systems are logical duals; they serve different functions. One is not inherently better or worse than the other. They both emphasize different things and ways to store data.

✐**Note: In general, it is probably best to have the data in vector form and convert it into raster as needed for analysis.**

A. Vector

- Vector has a low storage requirement.
- It is only necessary to store the feature. Data is more compact for lines, polys.
- You store by object—location is an attribute
- Potential for better accuracy
- Aesthetic appeal—we are used to the way it looks
- Analysis may be slow and complex—there are many intersections to calculate and many polygons to construct
- Some objects are vector by definition

B. Raster

- High storage requirement—store all the space
- Store by space—entity becomes the attribute
- Neighborhood structure makes certain analysis more functional—friction, viewshed
- Ease of map overlay
- Ease of statistics on general spatial characteristics
- Appealing to remote sensing community—due to "pixel" processing in such systems
- Fixed resolution—may sacrifice detail
- Normally a cheaper system

References

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford University Press. New York, New York

Carstensen, Laurence W. and Henry, Norah F. 1995. Digital Mapping and Geographic Analysis: An Introduction to Geographic Information Systems. Unpublished manuscript.

Goodchild, Michael F. and Kemp, Karen K. 1990. Introduction to GIS: NCGIA Core Curriculum: Units 4, 5,13, 14, 21, 30, 31, 35 38,45. University of California Santa Barbara. Santa Barbara, California.

The American Heritage Dictionary. 1985. Houghton Mifflin Company. Boston, MA

Star, Jeffrey and Estes, John. 1990. Geographic Information Systems: An Introduction. Prentice Hall. Englewood Cliffs, New Jersey.

Networks And Fuzzy Logic

Objectives *(Figure 6-1)*
- Understand the basic components and analysis associated with Networks.
- Become familiar with the concept of "fuzzy" logic and its implication on GIS.

I. INTRODUCTION

Networks are valuable tools when dealing with travel along linear features such as roads, streams, telephone systems, and the Internet. Networks are performed primarily within a vector model.

The types of GIS questions answered by Network analysis include:
1. What is the shortest way to get from X to Y?
2. What is the most accessible point?
3. If a spill is made in a stream, what is its path?
4. How much time should be allocated along a route?
5. How much demand is there for my "good" along this delivery route?

II. COMPONENTS OF A NETWORK *(FIGURE 6-2)*

- *Nodes* are the end points of a link. A *junction point* is where lines come together; an *end point* is the terminus of a line with nothing connected to it.
- Its *links* are the lines along which movement occurs.
- Its *turns* are the allowable movement at nodes.

Streams are fairly simple, with downhill being the most likely movement. Roads allow for more possibilities. At a four-way intersection, there are many possibilities such as left, right, forward, U-turn (times four roads = 16 choices).

Attributes:
- *Impedance*, something which may cause difficulty of movement along a link. It may be just the length and might denote a value for uphill or downhill travel. From-, and to-impedance might be different; e.g., uphill might take twice as long to travel as downhill. The best path is <u>lowest</u> total impedance.

- *Stops* are a node attribute. They are places you pass through on the network similar to stops on a bus route.
- *Centers* are locations that supply a network (where to go to obtain a "good," and how much of a "good" is available.
- *Resource demand* is a link attribute. It identifies where supply would be used up based on demand and supply center; i.e., spraying along a roadside, how far will a tank of material last?
- *Routes* are paths made up of links. Each link that is part of the route would get an attribute relating it to the route. A route is the field; route # is the attribute.
- *Barriers* are where movement along a link is totally prohibited. Impedance is a friction (relative barrier). Barriers are absolute.

III. ANALYSIS IN NETWORKS *(FIGURE 6-3)*

Optimal Routing (also called *optimal paths*) suggests "best." What is best? time?, scenery?, fuel? No algorithm can guarantee the optimal route; so, "optimal" really means good.

Allocation is the delimiting of an area within a certain impedance of a center. Add up impedance moving away from a center until the impedance limit is reached. How far away from a station can a fire engine get and be within an acceptable response time?

Tracing locates all links downstream from a particular node on one directional network; e.g., if cable TV goes out for five people, a trace finds common nodes as source of the problem.

IV. FUZZY LOGIC *(FIGURE 6-4)*

Fuzzy logic is when 100% certainty does not exist. When does it?

The **Traditional Set Theory** has mutually exclusive boundaries between sets.

The **Fuzzy Set Theory** has no absolute boundary definition. It is possible to assign a certain degree of confidence to the set. membership in the set is described by a value (usually between 0 and 1) that is the "possibility" that an entity is actually a member. Various functions are used to identify membership in a set depending on the nature of the phenomenon; i.e., sigmoid, J-shaped, and linear functions are available in the IDRISI package from Clark University.

Error Propagation ***(Figure 6-5)*** is when every overlay becomes more erroneous. To discuss quantification of error in GIS would require a complete class. The discussion here will be limited to a few examples just to give an idea of the importance of awareness of error, and that one will want to pursue more knowledge in the area. Examples claiming independence of categories might be an overlay: forest/open with sandy/silt, or interval/ratio data in overlay process ***(Figure 6-6).*** It may be decided the data is to inaccurate. New data and/or more fieldwork may be needed.

References

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford University Press. New York, New York

Eastman, J. Ronald. 1995.Idrisi for Windows: User's Guide. Clark University, Worchester, Massachusettes.

Kaufman, A. 1975. An Introduction to the theory of fuzzy subsets, Volume 1. Academic Press, New York, New York.

Schmucker, K.J. 1982. Fuzzy Sets, Natural Language Computations and Risk Analysis. Computer Science Press.

Star, Jeffrey and Estes, John. 1990. Geographic Information Systems: An Introduction. Prentice Hall. Englewood Cliffs, New Jersey.

Stoms, D. 1987. Reasoning with Uncertainty in Intelligent Geographic Information Systems. Proceedings, GIS '87, 692-700.

Zadeh, L.A. 1965. Fuzzy Sets. Inf. Control 8 (3), 338-353.

Sampling and Measuring

OBJECTIVES ***(Figure 7-1)***
- Understand how to classify and measure spatial data.
- Understand types of spatial variation
- List and discuss error and accuracy issues.

I. INTRODUCTION

We usually observe conditions, or properties, of features one of three ways: ***(Figure 7-2)***
1. Where something is located is called its **spatial property.**
2. What characteristic varies from one individual to another is called its **thematic property.**
3. How things vary from time to time is called its **temporal property**.

To measure any of these properties, it is necessary to understand the levels of measurement, types of spatial variation, sources of error, and accuracy of the data.

II. SPATIAL VARIATION *(FIGURE 7-3)*

Discrete variation ***(Figure 7-4)*** is the homogeneity within an area, point, or along a line. A value fills up a space, the same throughout the area, then an abrupt change to another value for another area. You are in one ownership and step over a fence and are now in some other ownership.

Continuous variation ***(Figure 7-5)*** is continuous in all directions. The value from any point varies smoothly. High spatial auto-correlation makes interpolation possible. Interpolation includes a variety of techniques to determine probable values for all locations based on values at known locations.

III. CLASSIFICATION OF DATA

A. DIMENSIONS OF DATA *(FIGURE 7-6)*

Points are 0 dimensional, take up no space whatsoever, and there can be an infinite number of points on the head of a pin. They have no length, width, or height.

True points are scale independent regardless of how large, or small, the scale on a map they are always points; e.g., 4 corners (NM, UT, CO, AZ) is a point at any scale.

Cartographic points are scale dependent. They have size but because of scale may be depicted as a point in the database; e.g., a telephone pole is actually 8" in diameter so it takes up space but we treat it as an X,Y location. At a large enough scale it would show up as having space.

Lines are 1 dimensional and provide length only, no width or height. True lines, when on the earth, have no width; e.g., a state boundary. Cartographic lines are scale dependent; e.g., railroad track, trail, creek.

Areas are 2 dimensional having no length and width, and no height. All areas are cartographic areas. If scale changes enough, the area may be reduced to a line or point.

Volumes are 3 dimensional. Volumes are generally not used in GIS; instead, attributes will be added.

B. MEASUREMENT LEVELS *(FIGURE 7-7)*

Nominal measurements would include named classes, equivalence or difference, would have no ranking, and would establish identity, i.e. yes/no, true/false, phone #, SSN. Mathematical possibilities might include =, ≠, mode, number of cases.

Ordinal measurement would be the implied order (rank) by implication (i.e., words). You wouldn't know the difference between classes or how they are different. It is possible to use numbers into the database to represent and ordinal scale (i.e., rank the data) such as 1 = high, 2 = medium, 3 = low. Examples might be 1st place, 2nd place, 3rd place. Mathematical possibilities might include =, ≠, <, >, median, percentiles.

Interval would be the rank with equal steps. The scale is significant in this measurement. Now it is possible to know how much different each individual is because of the equal steps (20° is 10° warmer than 10° but not twice as warm). The temperature scale has arbitrary 0 point. Which is true 0 temperature: Fahrenheit? Centigrade? Kelvin? An example would be temperature, finish time of a race in clock time (not stopwatch time). Mathematical possibilities might be =, ≠, add, subtract, mean, std. deviation, correlation, coefficient of variation.

Ratio measurement is a true numerical rank with an absolute 0 point. A tree's diameter of 20" is 15" larger than 5" and also four times as large. Examples might include tree diameters, trees per acre, elapsed time. A mathematical possibility might be: *all*.

It is important to consider the type of data you have and apply operations properly. You would not take soil type 3 and subtract it from soil type 7 to get soil type 4. They are nominal classes. Be aware that much GIS data comes in as ordinal or nominal. Just because it has a numerical value does not mean it is interval or ratio data. Do not try to perform multiplication, division, or other higher order functions unless all the data you are manipulating is ratio.

IV. ERROR

✐**Note: Error is defined as any discrepancy between a stated value and the real value on the earth.**

This could be in location (X,Y), or attribute (Z). With this definition, most everything is in error. A softer definition would be: an occurrence of information used by a GIS that is not accurate enough to suit the needs of its user. The term *not accurate* could be cost or use.

A. Reasons for error—using the map model *(Figure 7-8)*

Cartographic generalization is a four way process by which a map is produced. *Selection* (compilation) collects features from reality. *Simplification* depends of scale of the map; it requires elimination of certain features or exaggeration of certain features. *Classification* is the grouping of features considered similar and separation of features considered different. Using nominal data, decide how to classify a forest: healthy crop? poor crop?; using ordinal data, what is the level of amenity value (quality of life?) low, medium, high?; or using interval data, how to break into classes of elevation? (what is *low* or *medium* high between 1000m and 5000m?). *Symbolization* is the application of a symbol to represent features and/or attributes.

Several types of errors can occur. Obvious ways include blunders that might occur during data collection; i.e., misreading the instrument, or using poor sample design. Data entry errors can occur from digitizing when tired and typing in wrong attributes.

Fairly obvious errors can occur when settling for "less good" data, such as:

- Age of data—old vs. new
- Definition of the data—may have changed over time; e.g., is the definition of "forest" the same on your source as you define?
- Scale—affects level of generalization, accuracy, ability to match map sheets
- Reliability—Is the source in question?, How about within the map?, Have they interpolated? Extrapolated? **Metadata** is data about the data; i.e., such as source, method of development, lineage, statistics on accuracy.

- Relevance—use of surrogate data, i.e., not our desired object, but another used to estimate ours, such as Interpolation; e.g., estimate an elevation based on other known points, or remotely sensed data; i.e., it's not really data but just a measure of the reflectivity of objects into a sensor.

Natural Variations might occur with position; i.e., ability to measure a position, or ability of a position to be measured. Variation sources might include measurement errors or the difference between actual and measured value. *Accuracy* is the approximation of the true value. It is the opposite of error.

Precision offers the potential for accuracy—how closely can we measure something? ***Precision does not ensure accuracy***. Field errors can be made between crews. Standards and procedures need to be followed. Laboratory errors can occur due to poor lab workup.

Sampling errors can occur due to sampling strategies ***(Figure 7-9).*** Some sampling strategies might include: 1) random—every object has the same chance to get selected, 2) systematic-regular sampling pattern (usually with a random start), and 3) stratified—a prior knowledge of the distribution of the population. Processing errors or computer errors can also occur due to rounding errors—computers only store data at a certain precision; e.g., single or double precision. Translating back and forth from single to double (and back)causes these errors. Digitizing errors are created by an inability to follow a line. *Psychological errors* might include misperception of the line or cursor with respect to the line-parallax through the cursor; not being sure where the line (or center of the line) is. Errors tend to be offset parallel to our desired line. *Physiological errors* might occur when muscular movement of the hand is not steady. Errors tend to make knots along the direction of movement. In line sampling, errors arise because straight line segments always result in a shorter line than a curved line.

Boundary errors ***(Figure 7-10)***—it is necessary to consider not only how well we digitized the boundary, but is the boundary accurately located in reality? Boundaries imply an edge that may be abrupt or transitional. On a map, did the cartographer identify the boundary as an abrupt edge, or when a 50/50 change occurs.

✐**Note: Error is especially important when calculating areas.**

References

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford University Press. New York, New York

Carstensen, Laurence W. and Henry, Norah F. 1995. Digital Mapping and Geographic Analysis: An Introduction to Geographic Information Systems. Unpublished manuscript.

Chrisman, N.R. 1996. Exploring Geographic Information Systems. John Wiley (manuscript in prepartion).

Goodchild, Michael F. and Kemp, Karen K. 1990. Introduction to GIS: NCGIA Core Curriculum: Units 6, 7, 45, 46 University of California Santa Barbara. Santa Barbara, California.

Star, Jeffrey and Estes, John. 1990. Geographic Information Systems: An Introduction. Prentice Hall. Englewood Cliffs, New Jersey.

Zar, Jerrold H. 1984. Biostatistical Analysis: 2nd Edition. Prentice Hall. Englewood Cliffs, NJ.

Maps And Maps Analysis

OBJECTIVES *(Figure 8-1)*
- Be familiar with the various types of maps used to display information.
- Understand the concept of Cartographic Modeling in GIS analysis.
- Be able to name and explain the major analytical functions supported by GIS.
- Understand Cartographic Neighborhoods in Raster GIS.
- Be able to explain various examples of Raster analysis of neighborhoods.

I. TYPES OF MAP

A. Reference maps *(Figure 8-2)*

Reference maps emphasize accurate locations of features; e.g., United States Geological Survey (USGS) Standard Series Maps. Accuracy standards are: 90% of the features are within 1/50" (at the appropriate scale) of their actual location on the surface of the earth; i.e., at 1:24000, 90% of the features found on the map are within 40' of their actual ground locations. At 1:100,000, 90% of the features are within 167' of their actual ground locations.

B. Thematic Maps *(Figure 8-3)*

Thematic maps communicate geographical concepts such as the distribution of population densities, climate, land use, etc. Different thematic maps include:
- *Choropleth Map* ***(Figure 8-4)***—defines data within predefined collection units and uses reporting zones such as counties, or census tracts to show average income, rates of mortality, etc. Think of stair steps, extruded polygons.
- *Isarithmic Map* ***(Figure 8-5)***—interpolated map where data is measured at some points and values at other points are estimated from these. An isopleth map would be this type of map where one would connect points of equal value; e.g., a contour map. Some do not consider these "good" maps in that a discrete map is used to generate a continuous surface map.

- *Proximal map* ***(Figure 8-6)***—a special interpolation that produces (generate) a discrete surface rather than a continuous surface. This type of map creates regions around each data point with all locations within the generaged polygon takes cell out closer to that point than any other data point. Each cell gets the value of the nearest data point and is discrete in that all values within the region gets the value of the region. This map is somewhat similar to a choropleth map except regions are not predefined. You might hear the term **Thiesson polygons** or **Derichlet polygons.**
- *Trend surface map* ***(Figure 8-7)***—statistical process based on a process similar to regression analysis (for a discussion of these analyses consult Davis 1986). This map allows predicting Z (the attribute) value from the spatial location of a data point and can be used to describe the general pattern of the data.
- *Residuals map* ***(Figure 8-8)***—difference between the actual and the trend surface map. The actual minus predicted equals error and it is actually a map of error.

II. CHARACTERISTICS OF MAPS *(FIGURE 8-9)*

- Usually stylized, generalized, or abstracted.
- Usually out of date.
- Show only static situation (one slice of time).
- Easy to use for certain questions—What is at this point? And how far is it as the crow flies from here to there?
- Difficult for other questions—What places can I see from here? What does the thematic map show at this point on the topographic map?

III. CONCEPT OF SCALE

Concept of scale is the ratio between distances on the map and the real world.

Ways to delineate scale ***(Figure 8-10)*** include *representative fraction* (RF) (1:10000), *verbal conversion scale* (1 inch = 1 mile), or *graphic* (scale bar). A "small" scale is a small RF (this is relative). Only large features are shown, e.g. 1:250000. A "large" scale is a large RF (this also is relative). A large scale shows a lot of detail, e.g. 1:2000. Scale controls what is shown and how it is shown. Large scale may show cities as polygons with area. Small scale maps may show cities as points.

Scale also is an indicator of accuracy of data. Smaller scale maps need to generalize and symbolize more than large scale maps.

IV. WHAT ARE MAPS USED FOR?

Maps are used to display data and store data. It's easier to store spatial relationships on a map than a table. The data analysis tool is used to make or test models; e.g., predicting root rot clusters and examining relationships between distributions of different types with overlays.

V. USE OF MAPS FOR INVENTORY AND ANALYSIS

Inventory counts how much of what type existed at the time the map was produced, such as land use and land cover.

Analysis is—what happened over time?, what will happen if?, what is in proximity? and how will it be affected by change? Analysis usually involves comparison (overlay) of different maps. This may be difficult and time-consuming when overlaying and registering maps because usually maps use different map sources and scales. Most GIS and CAM systems can compensate for or change projections or scale. Changing doesn't improve accuracy. Just because we can zoom into a small area doesn't mean we've created a large scale map with corresponding accuracy.

VI. CARTOGRAPHIC MODELING *(FIGURE 8-11)*

Cartographic modeling is using basic spatial manipulation functions in a sequence to solve complex problems.

The three steps in cartographic modeling are:

1. Start with the desired final product. Determine what makes up the final product and information needed. Calculate as much as possible. Collect as little as possible.
2. Determine intermediate steps to get to the final product using major functions supported by the GIS such as reclass, buffering, group, etc.. Step backward from the final product to determine "raw" data needs.
3. Collect needed data.

Advantages of modeling are that they are flexible and deductive. Disadvantages of modeling are that they require more training. Modeling probably involves some inefficiency from a computer standpoint.

VII. MAJOR FUNCTIONS SUPPORTED BY GIS

A. Reclassification

Reclassification is changing from one value to another according to some rationale. Types of reclassification include: *grouping* ***(Figure 8-12)*** (creating fewer categories), *modeling* ***(Figure 8-13)*** (reclass to yes=1, no=0, reclass to what you want=1, don't want=0, and reclass ordinal 2,1,0), *display* ***(Figure 8-14)*** (to set color), and *measurements* ***(Figure 8-15)*** (area calculation with value of area assigned to each pixel).

B. Techniques of reclassification

Techniques of reclassification include *if/then logic*.

Conditional classification—using operators (<, >, =, etc.) to define a reclassification, "if" value = X, "then" assign new value

Complex classification—using connectors (and, or, xor) along with operators to build complex reclass, "if" value<X "or" value>Y, "then" assign new value

Mathematical constant reclass—perform a constant operation on a set of values to create new values (new value = old value + constant)

C. Basis for Reclassification

With **thematic data,** nominal data (qualitative), *Relabelling* (changing the name of the class from one name to another) might be necessary, i.e., change oak to deciduous. *Aggregating* is assigning values that bring individuals with the same value to be the same class; i.e., forest 1, forest 2, forest 3—reclass all to be just "forest."

Ordinal, interval, or ratio data requires relabelling, aggregating, and ranking (assigning new values to order the classes according to their importance). It is helpful to rank for visual display using brightness for more important data. Weighting assigns weights to different components; i.e., for sighting a road the slope is 3x more important than the soil type. Isolating areas with a certain desired characteristic and reclassification to remove non-conforming attributes is helpful (like the opposite of aggregating). Inversing is when the reclassification inverts the selected changing those values that are true to false, those that are false to true. Select to find slope >30% then change the values to false (not desirable).

Locational data uses properties of the features themselves, usually a measurement, where the feature is located or what it looks like on a map. *Size* is used to reclass so an object is displayed by its size. Remember, in raster, there is no concept of a polygon with area. You must find contiguous cells of equal value and add them together. In a forest area, reclass it to have all contiguous a value equal to the number of acres in the area (change from 1 for forest to 125 for 125 acres). *Shape* accounts for the compactness ratio or area to perimeter measure. It compares the area of a polygon to the area of a circle having the same perimeter as that of the polygon being calculated; i.e., C=sqrt(Ap/Ac). The circle is most compact. *Perimeter* measures the perimeter of polygonal areas. Length is also an important measure in Locational data.

VIII. CARTOGRAPHIC NEIGHBORHOODS IN RASTER

A. Definition (Figure 8-16)

Cartographic neighborhoods in raster are the adjacent cells from your cell of concern (NW, N, NE, E,SE,S,SW,W). We analyze the center cell called the **kernel.** Cells can be addressed as row I, column J. "Neighbors" are either one row more or less than the kernel, or one column more or less than the kernel.

I-1, J-1	**I-1, J**	**I-1, J+1**
I, J-1	**I, J Kernel**	**I, J+1**
I+1, J-1	**I+1, J**	**I+1, J+1**

✐**Note: remember—we always know what is next to what in a raster, and proximity and distance is inherent to the structure.**

N,S,E,W, are 1* cell size distant
NW, NE, SW, SE are 1.414 * cell size distant

B. Examples of raster analysis using neighborhoods

The *filter* ***(Figure 8-17)*** is the average of all nine cells. The average value goes into the kernel cell's location in the output.

16	**17**	**14**
15	**12**	**15**
14	**13**	**15**

(1) (16+17+14+15+12+15+14+13+15)/9 = mean
(2) mean = 14 (if integer), 14.55 (if real)

Diversity is the habitat value or the number of different values in the neighborhood. There are 6 different values (12,13,14,15,16,17). Habitat value = 6, using same cell values as above. The value 6 is placed into the kernel cell's location in the output.

The mode filter is the most common class of data. There are 3 occurrences of the value 15, mode value = 15 using same cell values as above. Again, the value 15 goes into the output.

Gradient ***(Figure 8-18)*** is the slope.

SAMPLE PROBLEM:

4355	**4368**	**4312**
4327	**?**	**4315**
4375	**4382**	**4370**

(1) xslope = ([I,J+1]-[I,J-1])/(cellsize * 2)
xslope = (4315-4327)/200 = -0.0600
assumes cell size = 100

(2) yslope = ([I-1,J]-[I+1,J])/(cellsize * 2)
yslope = (4368-4382)/200 = -0.0700

(3) gradient = sqrt($xslope^2$ + $yslope^2$)
gradient = sqrt(-0.06^2 + -0.07^2) = .0922
gradient in percent slope

(4) graddeg = atan(gradient)
graddeg = atan(0.0922) = 5.853°
gradient in degrees

Aspect is the direction of slope
aspect = atan(xslope/yslope)
aspect = atan(-0.06/-0.07) = 45.11°

This is simplified in that the atan function returns a value between 0° and plus or minus 90°. It so happens that this example is actually in relation to 0°, others would need adjustments relative to the quadrant in which they are located. Adjustments for quadrant corrections are based on use of the atan function.

xslope sign	yslope sign	adjustment
+	+	+180°
+	-	+360°
-	+	+180°
-	-	none

C. Analysis through extended neighborhoods

So far we have dealt with local neighborhoods (i.e., a kernel and its adjacent neighbors), we can expand this to neighbors, neighbors (if that makes sense). Watersheds are areas that drain into a river or river system, generated by aspects. Viewsheds are cells that can be seen from a viewpoint cell. Areas may be unconnected spatially. Up to now everything has been spatially connected. While spatially unconnected, they are *functionally* connected.

Grouping reassigns distinct values for all neighbors that are the same value. If you had 4 areas on a map that were classed as forest (value 1), after a GROUP analysis you would have the areas with values from 1 to 4. The map areas are the same but they now have a new value. It is now possible to get perimeter, area, etc., for each.

Accumulated surface values ("costs") continue to accumulate (get added) as one moves away from a cell(s) from which costs should be determined. Costs may be thought of as *friction* (i.e., difficulty moving through a cell). Friction values are considered barriers to movement. Relative barriers are relative to each other. Absolute barriers have no movement possible through the cell. Thinking in 3-D, roads would be the valleys of the surface (value 0) and the ridges are the accumulated surfaces. A higher ridge is a greater cost accumulation.

D. Combinations in Reclassification

Logical (Boolean) operations ***(Figure 8-19)*** look at whether something is true or false.

a) "and"
   (1) the logical intersection
   (2) both must be true

b) "or"
   (1) the logical union
   (2) one and/or the other

c) "not"
   (1) one without the other

d) "xor"
   (1) logical union without logical intersection

e) examples of Boolean Operators ***(Figure 8-20)***

IX. BUFFERING

Buffering is used for creating areas around features such as a point, line, or polygon. Feature classes may be buffered.

Examples for ***vector*** ***(Figure 8-21)***

a) Buffer 1 mile around a spotted owl nest (a point)
b) Buffer 100' visual corridor next to all arterial collector roads (a line)
c) Buffer 200' around a lake (a polygon)
d) Results of a buffer on any feature class in vector data models are polygons

In ***raster,*** there are odd increments of distance when moving to neighboring cells ***(Figure 8-22)***

a) N,S,E,W, are 1* cell size
b) NW, NE, SW, SE are 1.414 * cell size
c) Buffering might not come out quite as expected. You could change cell size to accommodate finer buffering increments

Sources of buffer values include constants and attributes. Constants are when all features receive the same buffer distance; e.g., 50' around all streams. Attributes ***(Figure 8-23)*** may create a variable buffer size; i.e., Buffer class I streams 200', class II streams 100', class III streams 50'.

X. OVERLAY

An *overlay* is a type of reclass where two maps are combined to form one new map. Overlay is one of the most—if not *the most* powerful—function of a GIS. It is used to segregate or delineate areas with certain attributes. The maps must be *spatially registered*. They must occupy the same space, must have the same coordinate systems, and the features must occur at the same locations.

A. Raster Overlay *(Figure 8-24)*

Raster overlay is a simple process compared to vector overlay. Each layer (map) is compared on a cell by cell basis; one can compare what attributes occur at the same place on the two maps. The new map is a combination based on result of the comparison.

Combinations may be performed in a variety of ways. The following are examples from the Idrisi^R (see Eastman 1995) package overlay functions:

- Addition—the corresponding pixels (cells) on each map are simply added together
- Subtraction—values in cells from the second map are subtracted from the first
- Multiplication—cell values are multiplied by one another
- Division—cell values in the first map are divided by cell values in the second map
- Minimizing—take the lesser of the two map values for corresponding pixels
- Maximizing—take the greater of the two map values for corresponding pixels

There are others, see Eastman (1995) for additional Idrisi^R overlay possibilities.

B. Vector Overlay

Vector overlay requires much computational effort because the vector data model allows an infinite number of locations and features, and one map cannot simply be placed over another one with simple mathematics.

Point data is a point on layer number one becomes the point on the output map. You end up with a point on the output map with attributes of the features of the second map included with the point. This is normally limited to a point on a polygon overlay using direct commands in GIS, but point on line, and point on point overlays can be performed indirectly. A point on a polygon overlay results in a point output map with attributes of the polygons which contained the points now associated with them.

Line data has conceivably many intersections to calculate and have topology constructed on the output map. The output map consists of lines with the attributes of the overlay map polygons associated with the lines. This is normally limited to a line on polygon overlay using direct commands in GIS, but line on line can be obtained indirectly.

Area (polygon) data: as with line data, a major difficulty is for the software to calculate the intersections; however, polygonal areas must then be reconstructed. This could conceivably result in thousands of polygons being formed with a polygon on polygon overlay. This is normally limited to polygon on polygon overlay using direct commands in GIS. A problem in vector is that many polygons may be formed that are not actually real because of spatial extents of features not matching up. One might think of this as vector systems being potentially to accurate for their own good. Two people digitize a study area, both select different sample points along lines (it would be impossible to expect operators, or even a single operator, to digitize the exact same line), and when overlain, the maps do not match. The resulting polygons are often small sliver areas known as *sliver polygons*. Much editing is necessary to ensure deletion of errors and retention of true areas.

References

American Society of Civil Engineers, American Congress on Surveying and Mapping, American Society for Photogrammetry and Remote Sensing.1994. Glossary of the Mapping Sciences. American Society of Civil Engineers, American Congress on Surveying and Mapping, American Society for Photogrammetry and Remote Sensing. Bethesda, Maryland.

Berry, Joseph K.1993. Beyond Mapping: Concepts, Algorithms, and Issues in GIS. Montgomery, Glenn E., and Schuch, Harold C. 1993. GIS Data Conversion Handbook. GIS World Books. GIS World, Inc. Fort Collins, CO.

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford University Press. New York, New York

Campbell, John. 1993. Map Use and Analysis. Wm. C. Brown, Dubuque, Iowa.

Carstensen, Laurence W. and Henry, Norah F. 1995. Digital Mapping and Geographic Analysis: An Introduction to Geographic Information Systems. Unpublished manuscript.

Davis, John C. 1986. Statistics and Data Analysis in Geology: Second Edition. John Wiley and Sons. New York, New York.

Eastman, J. Ronald. 1995.Idrisi for Windows: User's Guide. Clark University, Worchester, Massachusettes.

Environmental Systems Research Institute. 1992. ArcCAD Command Reference. Environmental Systems Research Institute, Redlands, California.

Environmental Systems Research Institute. 1995. Understanding GIS: The ARC/INFO Method. John Wiley and Sons, Inc. New York, New York.

Environmental Systems Research Institute. 1991. Introducing ArcCAD Release 11: Course Manual. Environmental Systems Research Institute, Redlands, California.

Goodchild, Michael F. and Kemp, Karen K. 1990. Introduction to GIS: NCGIA Core Curriculum: Units 15, 32, 33, 34, 40, 41, 48. University of California Santa Barbara. Santa Barbara, California.

Star, Jeffrey and Estes, John. 1990. Geographic Information Systems: An Introduction. Prentice Hall. Englewood Cliffs, New Jersey.

The American Heritage Dictionary. 1985. Houghton Mifflin Company. Boston, MA

Sources of Digital Data

OBJECTIVES ***(Figure 9-1)***

- Become familiar with sources and types of attribute data.
- Become familiar with sources and types of spatial data.

I. ATTRIBUTE DATA *(FIGURE 9-2)*

What is it? Attribute data is data with no inherent spatial component. This data must be able to tie to the spatial data.

A. Sources

The US Census Bureau [(301) 763-4100] is one source. This source has been mandated by the US Constitution to ensure equitable representation. The bureau offers an on-line service "CENDATA." Data available includes population and housing, agriculture, and businesses.

Private sources include reworked government data for a specific market; e.g., marketing or sales. Company data might include Strategic Mapping and Planning, or National Planning Data Sources.

Remote Sensed Images or EOSAT are "Earth Observation Satellite system" data. A private firm receives and disseminates data and images of LANDSAT components. Two types include MSS (multi-spectral scanner) which has 4 bands, 80m spatial resolution (LANDSAT #3 and 5 bands), and TM (Thematic Mapper) (7 bands, 30m spatial resolution).

SPOT is Le Systeme Pour l'Observation de la Terre (a French System). This system is multi-spectral, 3 bands, 20m spatial resolution color, 10m black and white.

I. SPATIAL DATA *(FIGURE 9-3)*

Different types of spatial data include:

- DIMECO—initiated in the late 1960's for the 1970 census. This was the first major effort at digital format. This included county boundary files.
- World Data Bank I and II—from the CIA. Version I original database was in MICROCAM. Version II included coastlines, countries, rivers.
- DIME (Dual Independent Map Encoding)—used in 1978 for the 1980 census. It was the precursor to TIGER. This included urban areas (approx. 274 urban areas), streets, addresses, census geography, SPCC and Lat/Long, and had fairly low accuracy.
- TIGER (Topologically Integrated Geographic Encoding and Referencing) encacted in 1988 for the 1990 census. This was to complete the DIME effort and included addressing, census geography in combination with 1:100,000 Digital Line Graph data (based on USGS 1:100,000 quads). Added themes included rivers, railroads, political boundaries, landmarks, statistical boundaries, address ranges. It is a topologic system.

II. PRIVATE SOURCES OF SPATIAL DATA

A. GEOGRAPHIC DATA TECHNOLOGY, LYME, NH

Dynamap 4000 uses enhanced TIGER data and addresses were added. This data updates quarterly and uses digit zip codes.

B. ETAK, MENLO PARK, CA

ETAK was one of the first in digital navigation on street maps. Its emphasis is on planimetric accuracy (claimed to be 40'). It's used for urban areas.

III. US GEODATA *(FIGURE 9-4)*

US Geodata comes from the Earth Science Information Center (ESIC) in Reston VA [(1-800-USA-MAPS), http://edcwww.cr.usgs.gov/doc/edchome/ndcdb/ndcdb.html

A. GEOGRAPHIC NAMES INFORMATION SYSTEM (GNIS)

- [http://www-nmd.usgs.gov/pub/gnis/] for the *Universal Resource Locator (URL).*
- [ftp://www-nmd.usgs.gov/www/gnis/] for retrieving *GNIS digital data.*

This system includes point data which is all known place, feature, and area locations with proper names in the US. It has the largest scale available (i.e., 1:24,000 if available).

The *National Geographic Names Data Base* has over 2,000,000 names of places and features (name, fips code of state and county, lat/long, elevation (if on topo sheet), population (if a census recognized location), up to 4 map codes listing topo sheets on which the place is located.

The *USGS Topographic Map Names Data Base* includes an inventory of names from USGS topo sheets. A *Reference Data Base* catalogs features by type and may include broad categories.

B. Land Use/Land Cover

Most are 1:100,000 and some are 1:250,000. Major classes of land use are urban (or built up), agricultural, range, forest, water, wetlands, barren, tundra, and perennial snow or ice. This data includes polygon data, minimum 4 ha. (10 ac.) urban areas, 16 ha. (40 ac.) rural areas 4,000—5,000 polygons per quad. It is from mid-late 1970's LANDSAT and is available for free on the Internet. This utilizes *Geographic Information Retrieval and Analysis System (GIRAS)* format—vector polygon and raster grid (4 ha. cell size) with *UTM coordinates*.

C. Planimetric Data

Digital Line Graphs (DLG's) utilize vector data which is a digital format of the USGS quadrangles. Its scale is 1:100,000 (corresponding directly to TIGER), and 1:2,000,000. The USGS is working on 1:24,000 DLG's. The 1:24,000 series will be the equivalent to digital 7 1/2 minute quad sheet.

There are also 1:20,000, 1:25,000, 1:48,000, 1:62,500, and 1:63,360 for certain areas of the conterminous US, Alaska, Puerto Rico, and the Virgin Islands. Layers on the base include boundaries (political, administrative), hydrography, Public Land Survey System to the section level, transportation roads, trails, RR, pipelines, and significant landmarks or cultural features.

Other layers that may be available include:

- Hypsography—contour lines. This is where DEM (see next section) is not practical due to flat ground
- Surface cover—vegetation (the green on a topo sheet)
- Non vegetative surface cover—lava flows, sand, gravel
- Survey control—3rd order and higher

D. Elevation Data

Digital Elevation Models (DEM'S) show terrain elevations at regularly space intervals. This is a raster product. Scales available are:

- 7½ minute (1:24,000) with elevations reported every 30 meters (100'). It has a UTM north grid with vertical accuracy within 15 meters.
- 15 minute (1:63,360) scale is used in Alaska with elevations every 2 arc seconds (200') along a profile, 3 arc seconds (300') between parallel profiles. Lat/long grid, NAD27.
- 30 minute (1:100,000) elevations every 2 arc seconds along and between profiles. Lat/long grid, NAD27.
- 1 degree (1:250,000) elevations every 3 arc seconds with lat/long, WGS72. These are available over the Internet.

References

American Society of Civil Engineers, American Congress on Surveying and Mapping, American Society for Photogrammetry and Remote Sensing.1994. Glossary of the Mapping Sciences. American Society of Civil Engineers, American Congress on Surveying and Mapping, American Society for Photogrammetry and Remote Sensing. Bethesda, Maryland.

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford University Press. New York, New York

Campbell, James B. 1987. Introduction to Remote Sensing. The Guilford Press. New York, New York.

Carstensen, Laurence W. and Henry, Norah F. 1995. Digital Mapping and Geographic Analysis: An Introduction to Geographic Information Systems. Unpublished manuscript.

Eastman, J. Ronald. 1995. Idrisi for Windows: User's Guide. Clark University, Worchester, Massachusettes.

Goodchild, Michael F. and Kemp, Karen K. 1990. Introduction to GIS: NCGIA Core Curriculum: Unit 24. University of California Santa Barbara. Santa Barbara, California.

Montgomery, Glenn E., and Schuch, Harold C. 1993. GIS Data Conversion Handbook. GIS World Books. GIS World, Inc. Fort Collins, CO.

Sample V. Alaric. 1994. Remote Sensing and GIS in Ecosystem Management. American Forests: Forest Policy Center. Island Press. Washington D.C.

Star, Jeffrey and Estes, John. 1990. Geographic Information Systems: An Introduction. Prentice Hall. Englewood Cliffs, New Jersey.

Sample Midterm Exam

✐**Note: 100 points possible**

(1)1. If you were to pass an imaginary rod from a hundred miles out in space through the surface of the earth down to the center of the earth, it would pass through three surfaces important to **geodesy**. List and define the three surfaces.

(2)2. What is a geographic Information System?

3. If you measure the distance on the surface of the earth from:
   a. 0° N, 0° W to 10° N, 10° W **or**
   b. 20° N, 20° W to 30° N, 30° W

   (5) Which distance will be the shortest, the first (a.) or the second (b.)?
   (5) Why?

(4)4. How does precision vary between a vector and raster data model?

(5)5. According to the lecture notes, Star and Estes (1990) identified five functional elements of a Geographic Information System. Name the five elements.

(6)6. A Geographic Information System is attempting to **model reality** through the selection of a **data model**. Then a **data structure** must be decided on which requires a **file structure**. A GIS technician must have an understanding of these components to utilize various GIS software. Write a couple of sentences explaining each of the following:

a. reality

b. data model

c. data structure

d. file structure

(7)7. We generalize reality by identifying “feature classes” that simplify characteristics of the landscape into different dimensions. This allows us to model reality. Name the three different feature classes in the vector data model and identify the associated dimensions.

(8)8. Name and explain the desirable properties of a map projection.

(14)9. Vector and Raster data models are logical duals (they serve different functions). One is not inherently better or worse than the other. They both emphasize different approaches to data modeling and data storage. <u>Compare and contrast the vector vs. raster data model.</u>

Sample Final Exam

✐**Note: 150 points possible**

1. U.S. Geodata from the Earth Science Information Center provides various sources of data. The two most commonly used (arguably) are DLG's and DEM's. What do the initials stand for, and, give a few sentences of explanation of each.

2. We use logical (Boolean) operations for many of our reclassification combinations in GIS. Name and explain at least three Boolean operations.

3. We discussed two types of maps used to display geographic information. One emphasizes accurate locations of features, the other communicates geographical concepts. Name the two types of maps.

4. We usually observe conditions, or properties, of features in one of three ways: as **spatial, thematic,** or **temporal** property. To measure any of these properties, it is necessary to understand the levels of measurement and types of spatial variation associated with any feature.

Name and explain the two types of spatial variation discussed in class.

Name and explain the four measurement levels.

5. What is "fuzzy logic"?

6. It is said that the vector data model "**stores by object"** while the raster data model "**stores by space."** What is meant by each of these?

7. In lecture we discussed various types of thematic maps: **choropleth, isarithmic, proximal, trend surface,** or **residuals**. Select four of these five thematic map types and provide a three-sentence explanation of each.

8. What is "Geodesy"?

9. What does **TIGER** stand for? What is the significance of these files?

10. What is an **“overlay”** in GIS analysis? How does the process differ between a raster and vector data model?

11. What is the difference between psychological and physiological errors when digitizing?

12. Cartographic generalization is a four-step process by which a map is produced. Name and explain the four steps.

Objectives

A. Define Geographic Information System.

B. Outline origins of the field.

C. Discuss GIS terminology and concepts.

D. Compare the differences between GIS, CADD, and Automated Mapping.

FIGURE

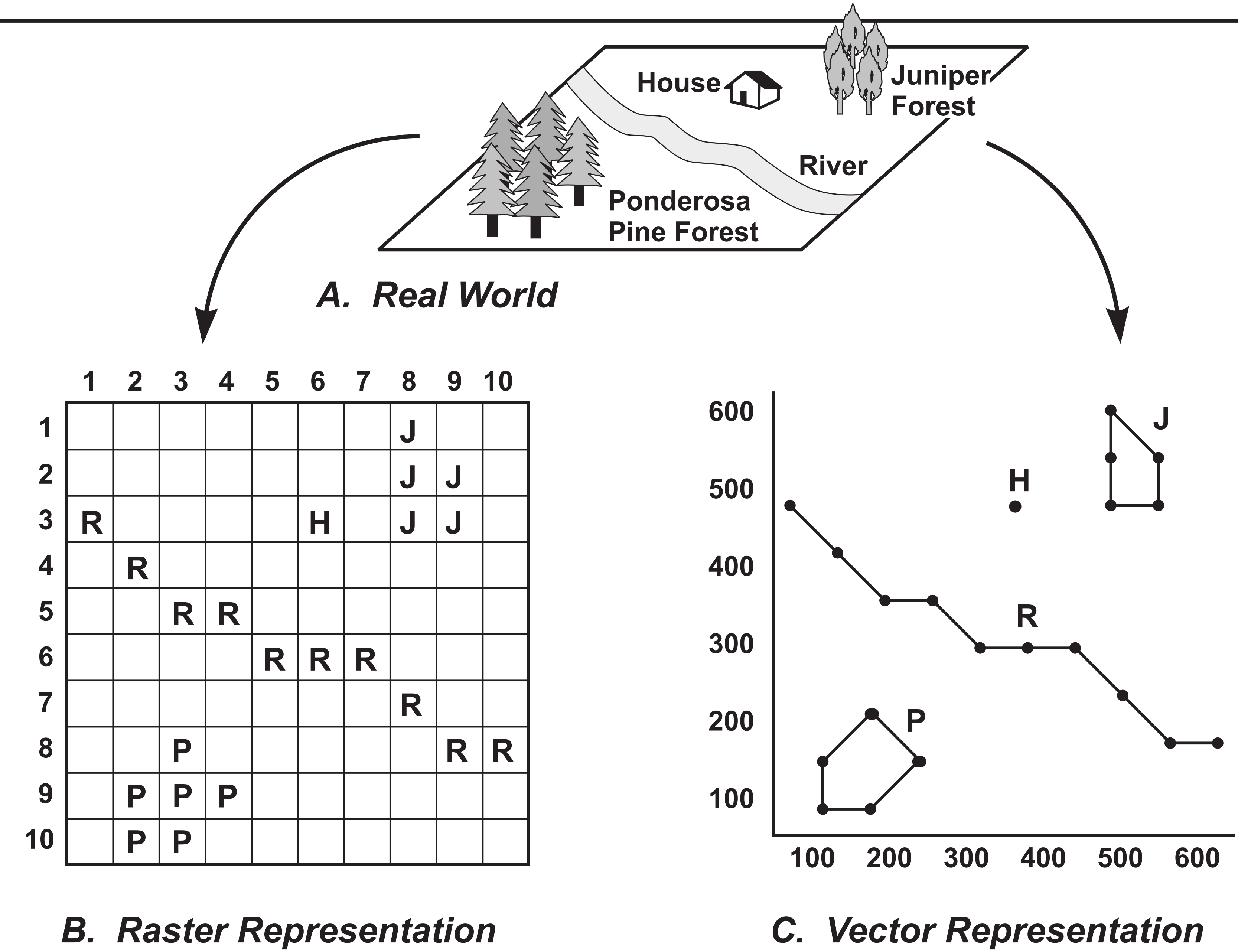
1-1

What Is GIS?
Geographic Information System

DEFINITION: **Any sequence of interrelated functions that achieves the input, storage, processing, and subsequent generation of spatial data.**

FIGURE

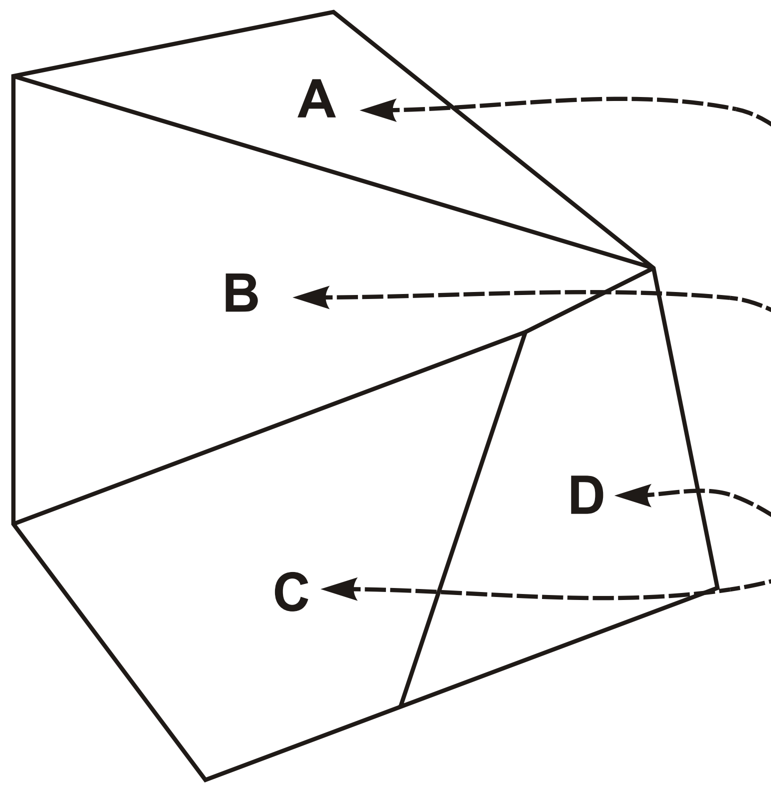
1-2



Raster and Vector Data Models

FIGURE
1-3

NSFGIS\TOPIC1\OH0103.CDR



ID	*Type*	*Age*	*Acres*
A	P. Pine	30	12
B	P. Pine	15	25
C	D. Fir	20	20
D	L. Pine	5	14

Linking Spatial and Attribute Data

FIGURE 1-4

NSFGIS\TOPIC1\OH0104.CDR

Land Unit Layer - Data File

ID	*Area*	*Veg.*	*Soil*
A	56	RH	1
B	78	RP	3
C	112	RH	4
D	42	ASP	2

Soils Database

Soil	*S.I.*	*% O.M.*
1	79	1.2
2	102	6.4
3	56	4.1
4	84	0.7

Linking Databases

FIGURE 1-5

NSFGIS\TOPIC1\OH0105.CDR

Elements of Topology

In the *Vector* World:

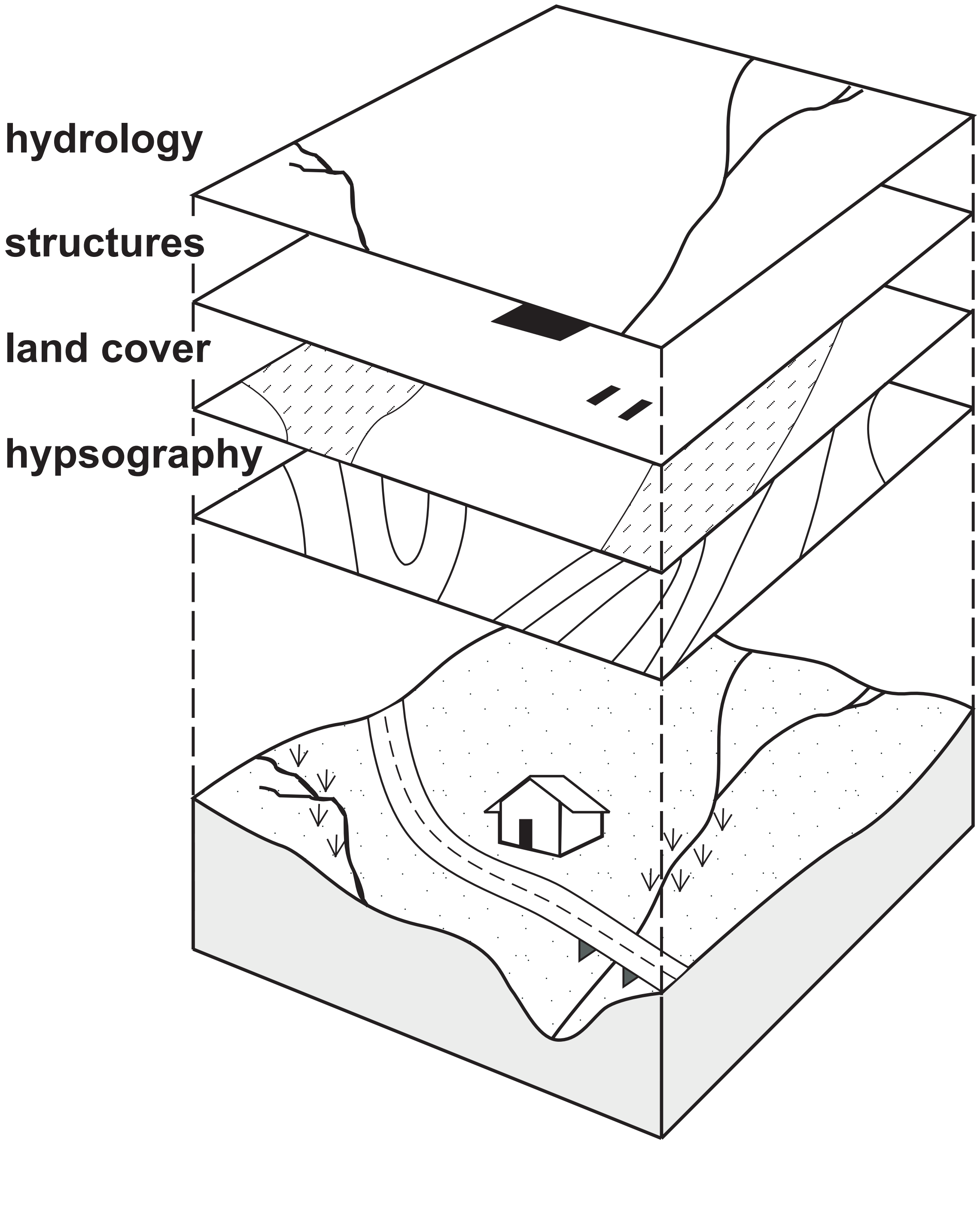
A. Coordinate systems
B. Tics
C. Discrete points
D. Nodes
E. Vertices
F. Arcs
G. Polygons
H. Annotation
I. Coverage

In the *Raster* World:

A. Cell
B. Whole cell
C. Partial cell
D. Lines
E. Areas
F. Surfaces

FIGURE

1-6



The Overlay Concept

FIGURE

1-7

Computer-Aided Mapping:
New Ways of Considering Maps

A. Graphic elements versus drawn lines

B. Data links

C. Flexibility

D. Mapping in real size / real world coordinates

FIGURE

1-8

Digital vs. Hand Drawn Maps

Advantages:

A. Lower costs

B. Faster production

C. Accuracy

D. Flexibility in output

Disadvantages:

A. Not as automated as once thought

B. Computers do not ensure high quality

FIGURE

1-9

GIS vs.
Computer Cartography

1. Data Stores

2. Data Indexes

3. Data Analysis Tools

4. Data Display Tools

FIGURE

1-10

Objectives

A. Define Geographic Information System.

B. Outline origins of the field.

C. Discuss GIS terminology and concepts.

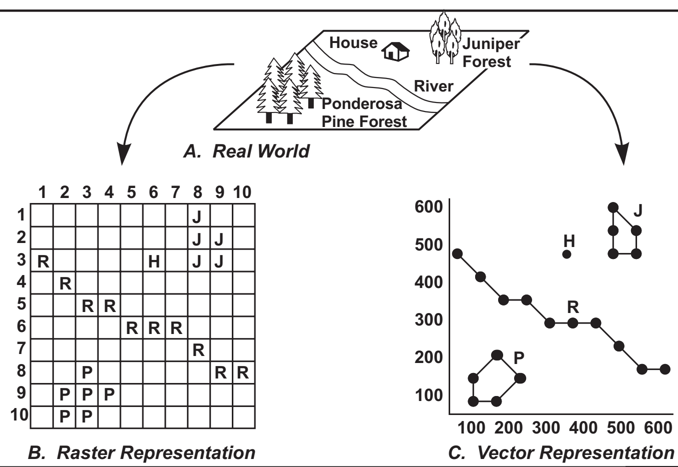
D. Compare the differences between GIS, CADD, and Automated Mapping.

FIGURE 1-1

What Is GIS? Geographic Information System

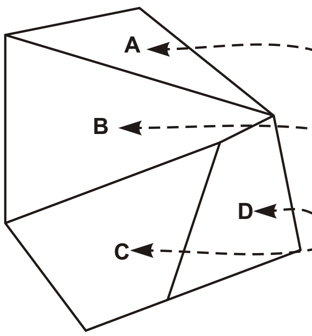
DEFINITION: Any sequence of interrelated functions that achieves the input, storage, processing, and subsequent generation of spatial data.

FIGURE 1-2



Raster and Vector Data Models

FIGURE 1-3



ID	Type	Age	Acres
A	P. Pine	30	12
B	P. Pine	15	25
C	D. Fir	20	20
D	L. Pine	5	14

Linking Spatial and Attribute Data

FIGURE 1-4

Land Unit Layer - Data File

ID	Area	Veg.	Soil
A	56	RH	1
B	78	RP	3
C	112	RH	4
D	42	ASP	2

Soils Database

Soil	S.I.	% O.M.
1	79	1.2
2	102	6.4
3	56	4.1
4	84	0.7

Linking Databases

FIGURE 1-5

Elements of Topology

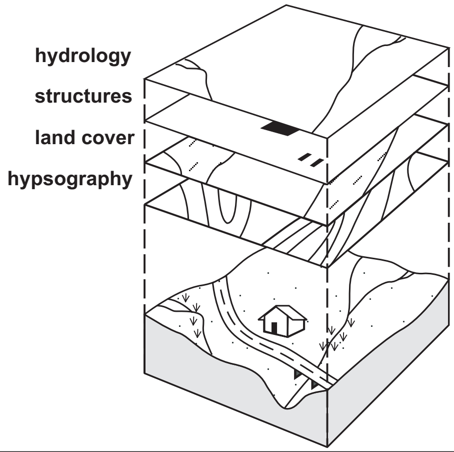
In the *Vector* World:

A. Coordinate systems
B. Tics
C. Discrete points
D. Nodes
E. Vertices
F. Arcs
G. Polygons
H. Annotation
I. Coverage

In the *Raster* World:

A. Cell
B. Whole cell
C. Partial cell
D. Lines
E. Areas
F. Surfaces

FIGURE 1-6



The Overlay Concept

FIGURE 1-7

Computer-Aided Mapping: New Ways of Considering Maps

A. Graphic elements versus drawn lines

B. Data links

C. Flexibility

D. Mapping in real size / real world coordinates

FIGURE 1-8

Digital vs. Hand Drawn Maps

Advantages:

A. Lower costs
B. Faster production
C. Accuracy
D. Flexibility in output

Disadvantages:

A. Not as automated as once thought
B. Computers do not ensure high quality

FIGURE 1-9

GIS vs.
Computer Cartography

1. Data Stores

2. Data Indexes

3. Data Analysis Tools

4. Data Display Tools

FIGURE
1-10

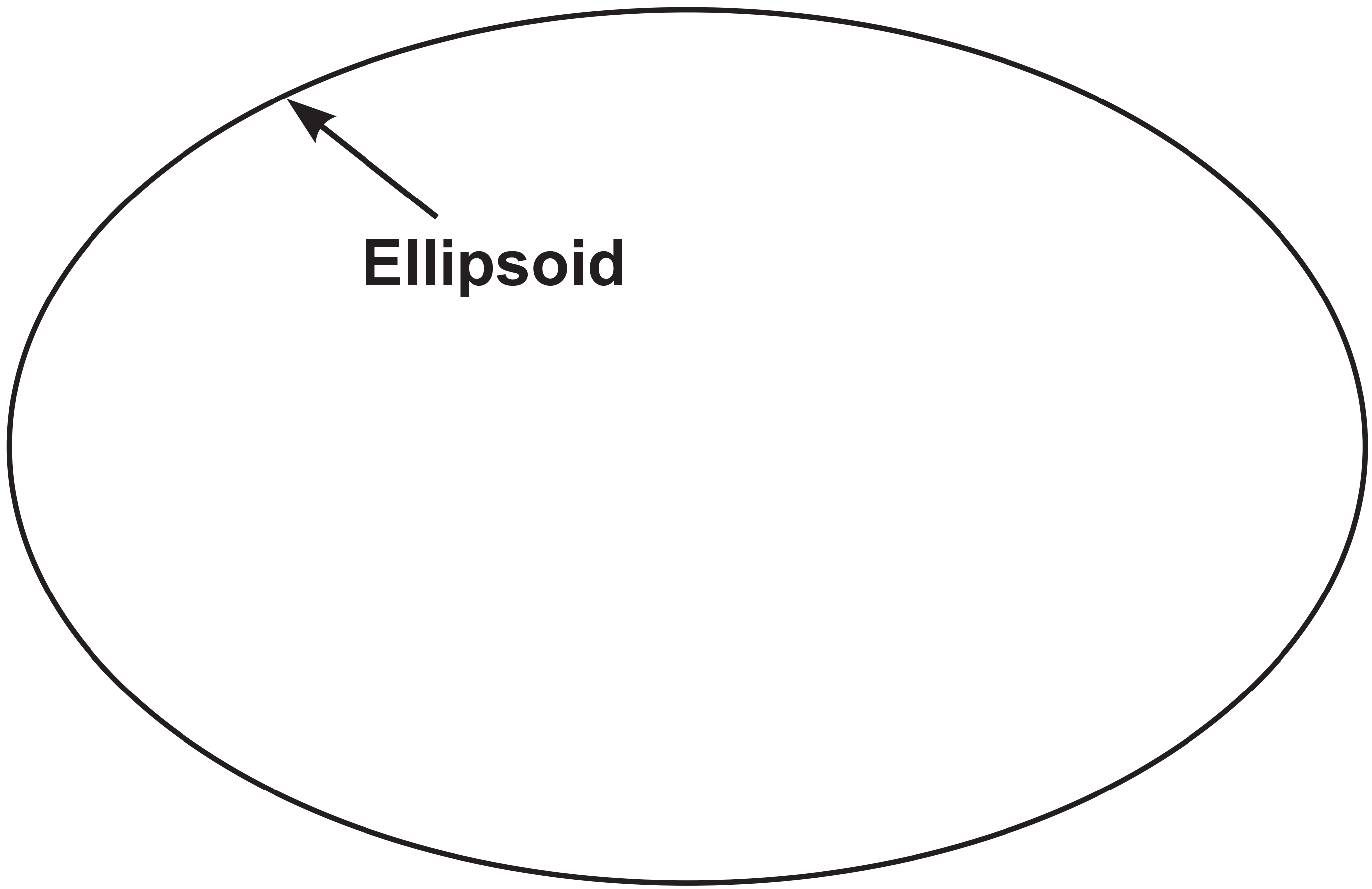
Objectives

Understand the basic concepts of geography pertaining to:

1. Geodesy
2. Map projections
3. Map based grid systems

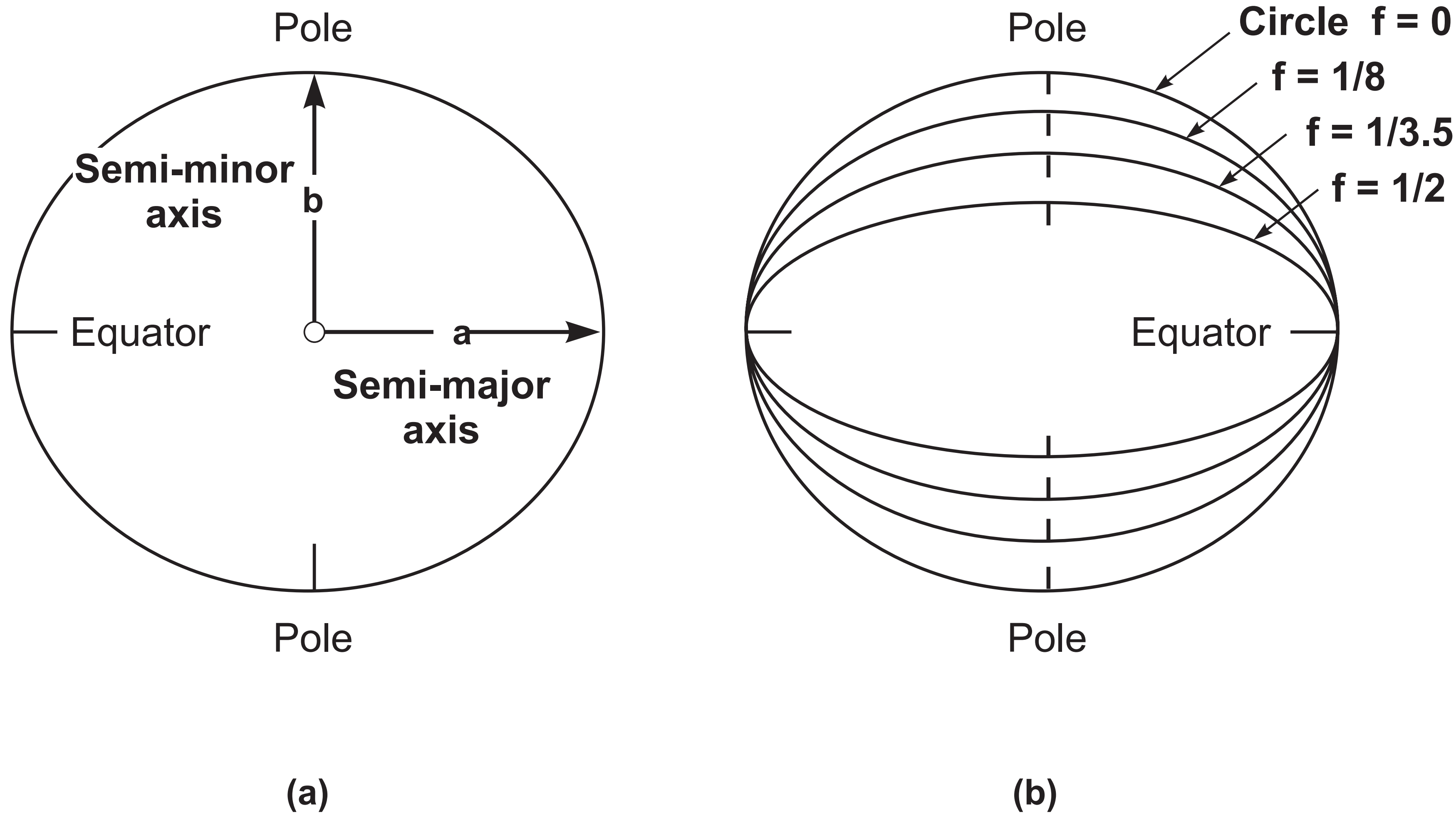
FIGURE

2-1



Ellipsoid

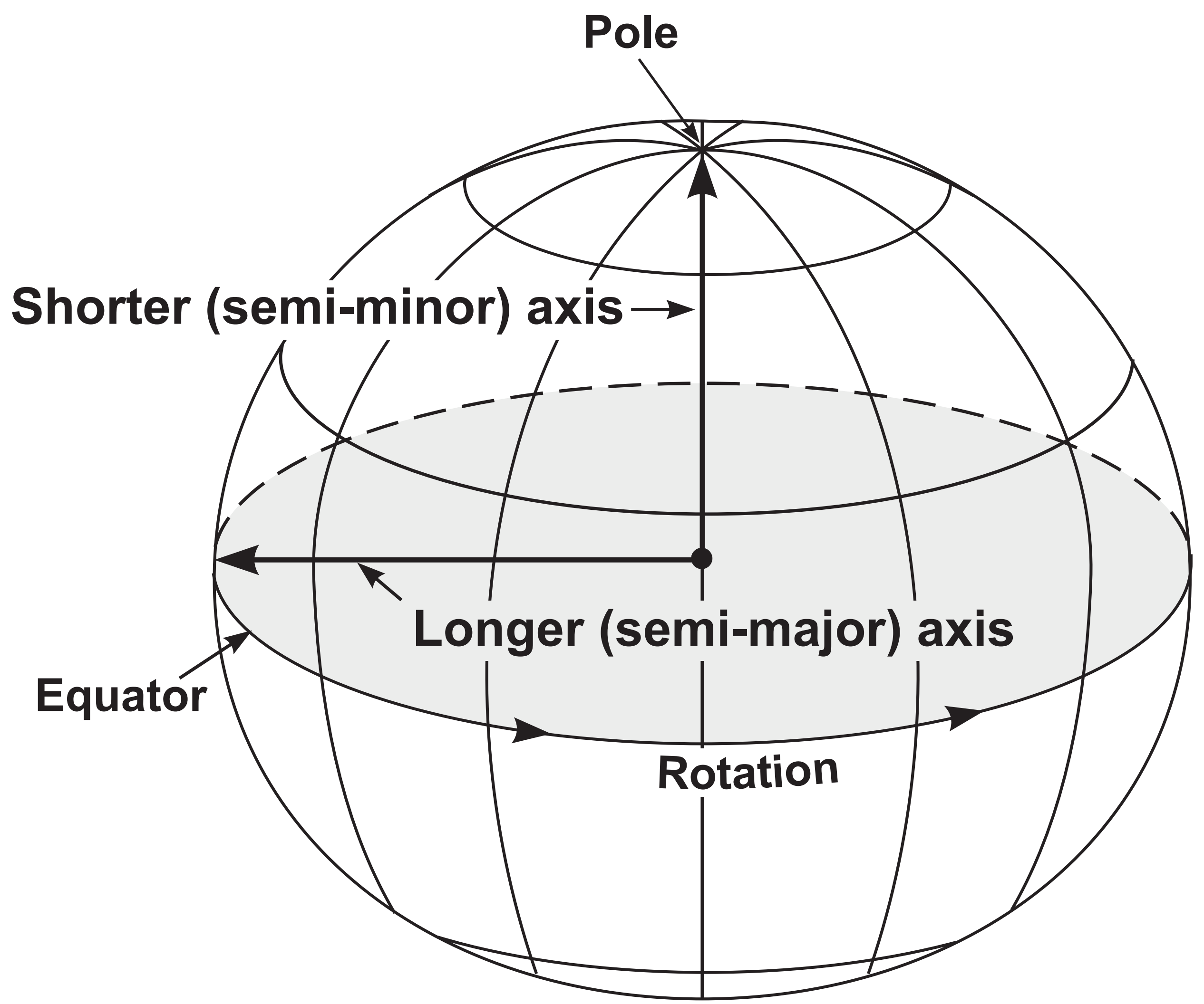
FIGURE 2-2



Ellipsoid Basics and Variations

FIGURE 2-3

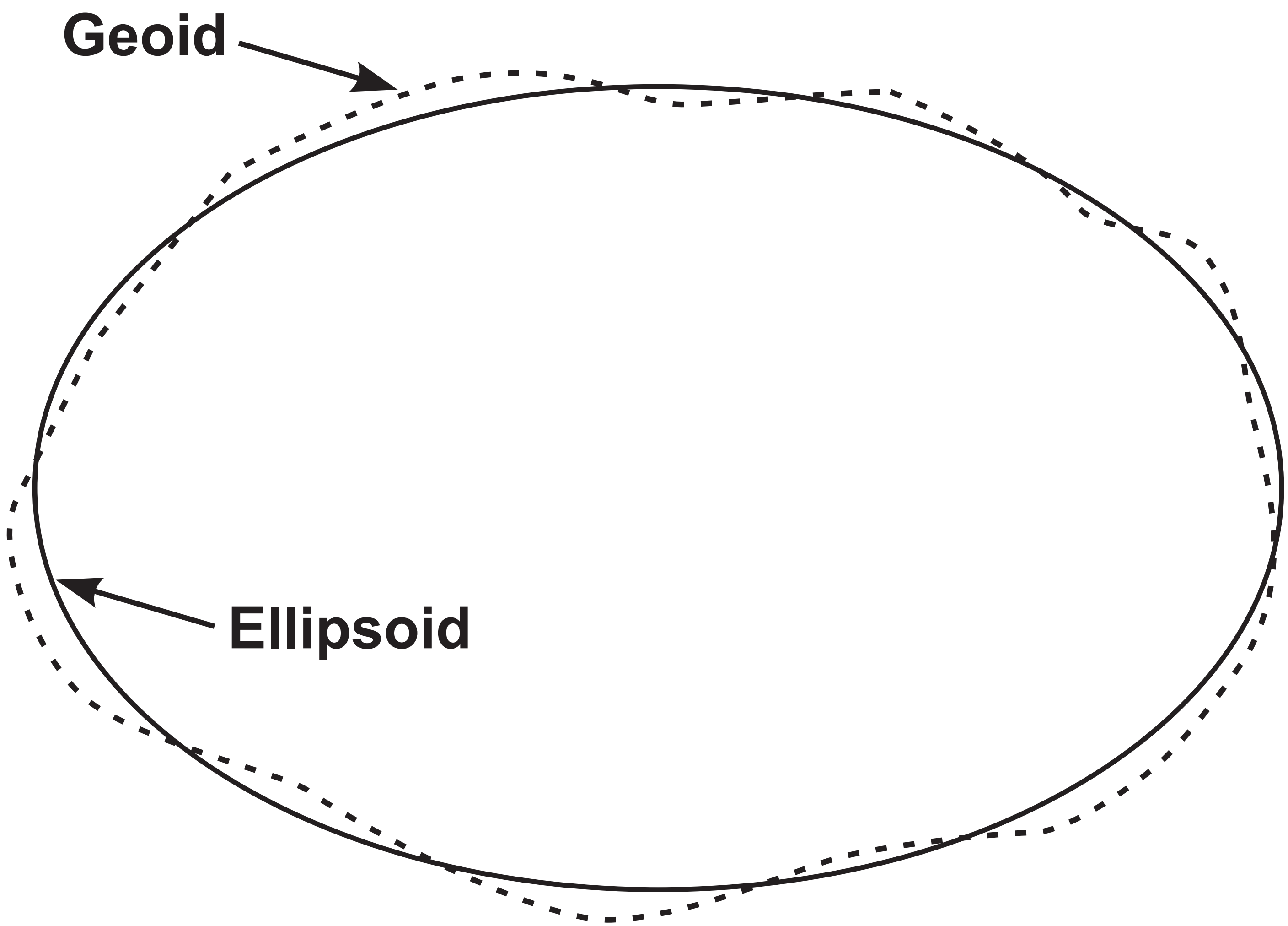
NSFGIS\TOPIC2\OH0203.CDR



Earth in Ellipsoid Form

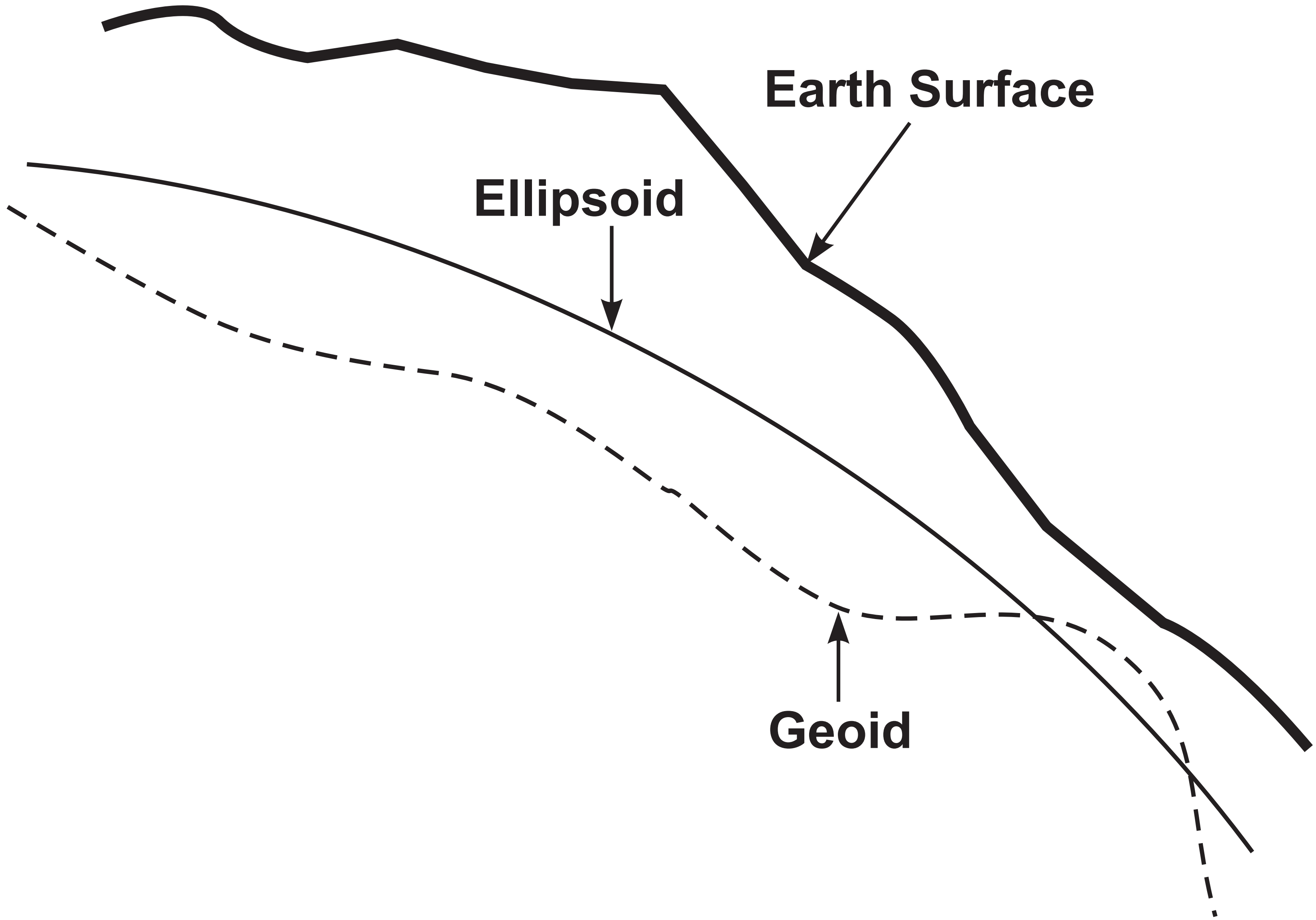
FIGURE
2-4

NSFGIS\TOPIC2\OH0204.CDR



Geoid vs. Ellipsoid

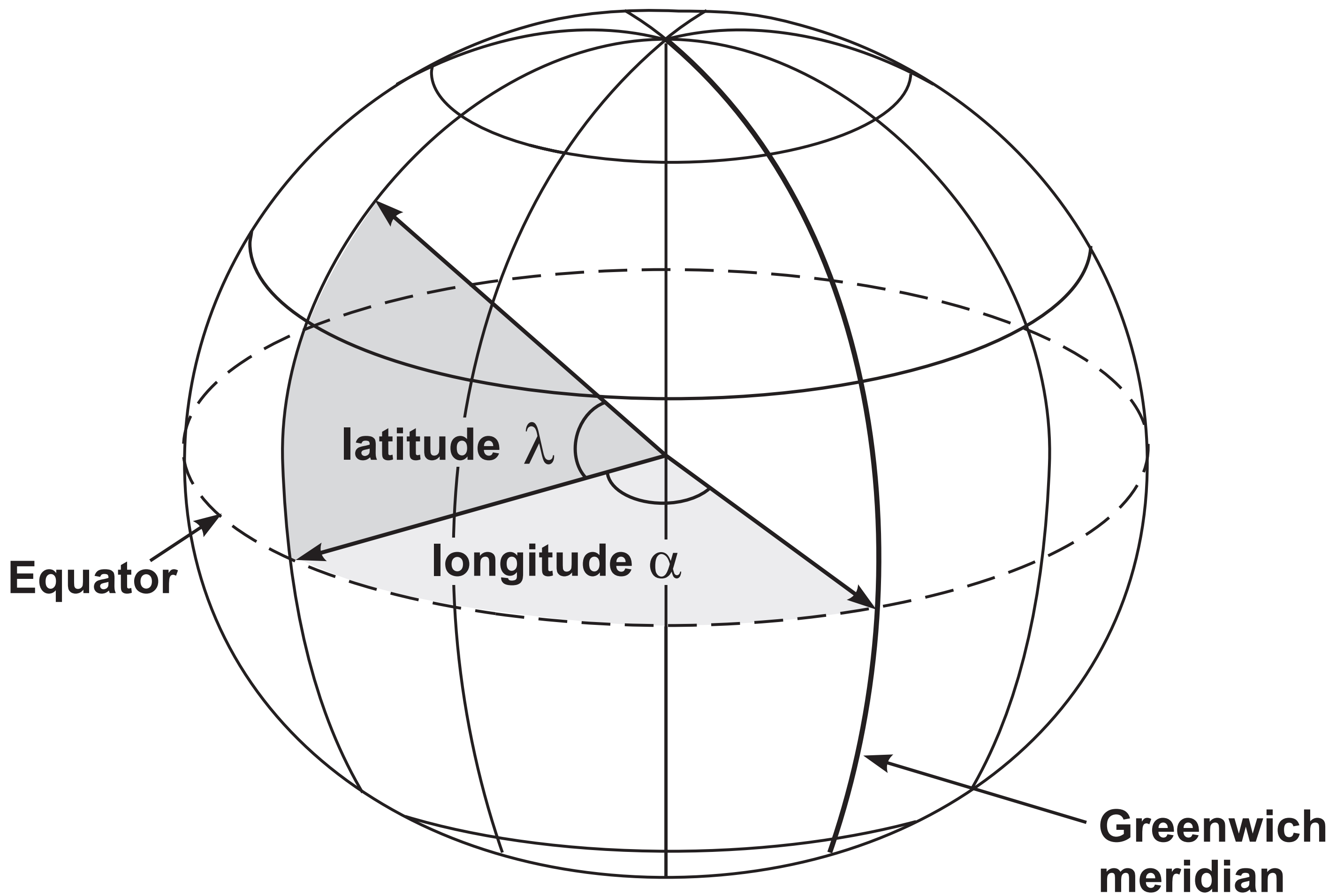
FIGURE 2-5



Ellipsoid, Geoid and Earth Surface

FIGURE **2-6**

NSFGIS\TOPIC2\OH0206.CDR



The Graticule: Latitude and Longitude

FIGURE 2-7

Since Earth's radius is not constant, Earth's surface distance contained in 1° of latitude or longitude in not constant.

Latitude Changes Little:

At Equator, 1 Degree = 110.5 km (68.7 mi)

At Poles, 1 Degree = 111.7 km (69.4 mi)

For Longitude:

At Equator, 1 Degree = 111.3 km (69.2 mi)

At Poles, 1 Degree = 0 km (0 mi)

FIGURE

2-8

A Map Projection Is:

A systematic rendering of points on the earth ellipsoid to points on a flat sheet.

Think of it as passing rays of light from some point through the globe and onto the map surface.

Why do we need to do this?

1. **To produce 2-dimensional maps.**
2. **To have a convenient cartesian coordinate system.**

FIGURE

2-9

Desirable Properties of Projections

A. Equivalence

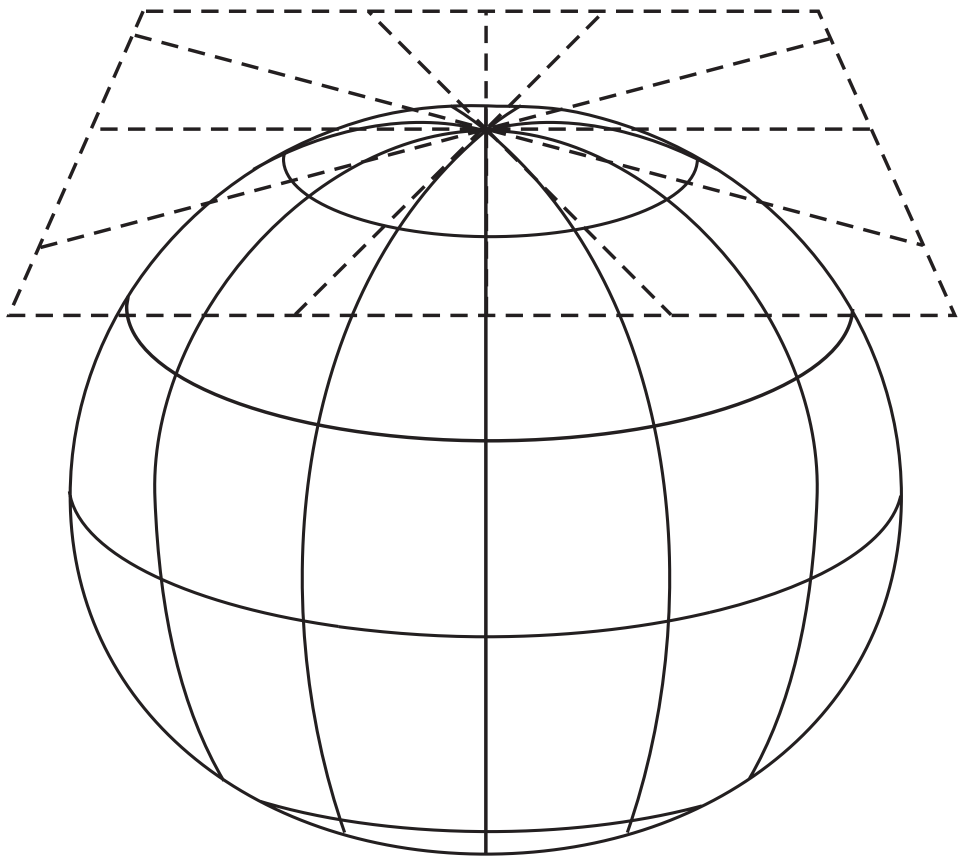
B. Conformality

C. Azimuthality

D. Equidistance

FIGURE

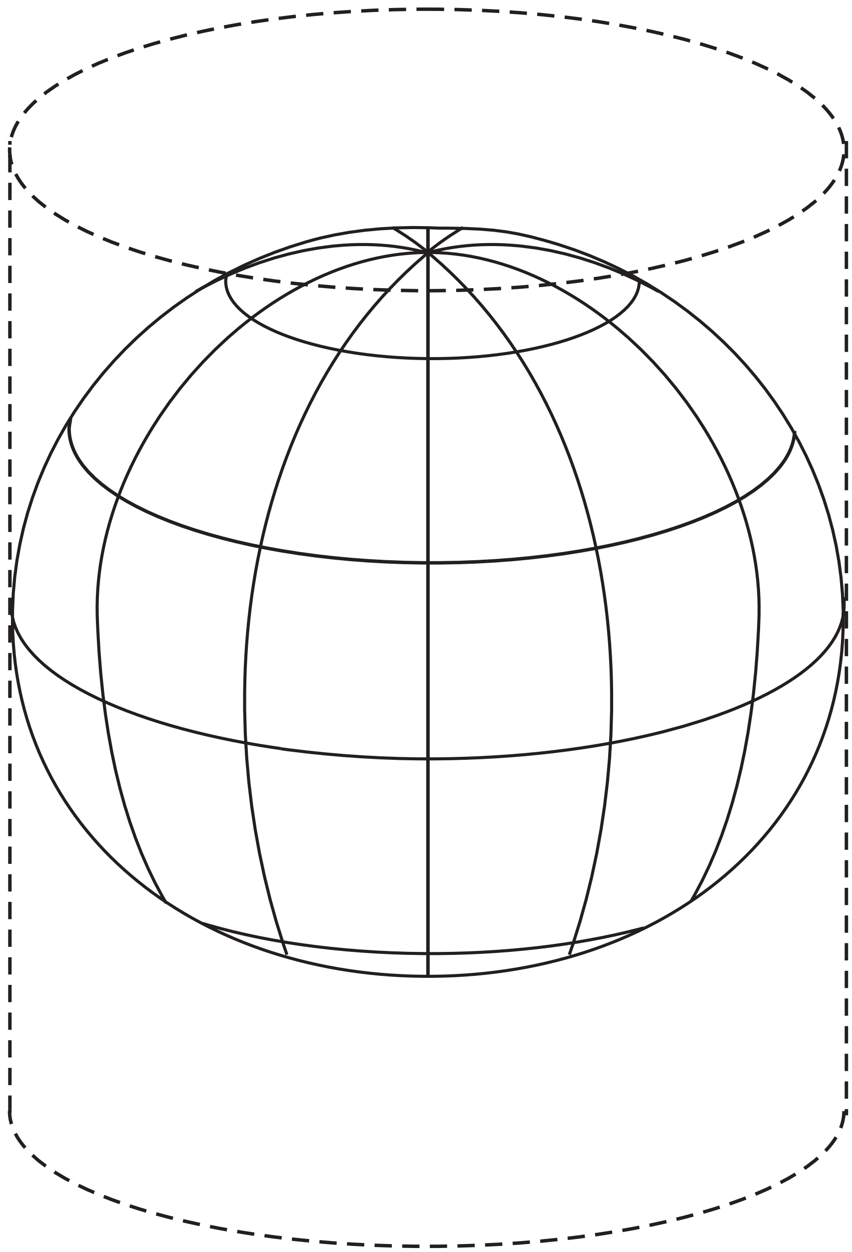
2-10



Polar Plane Projection

FIGURE

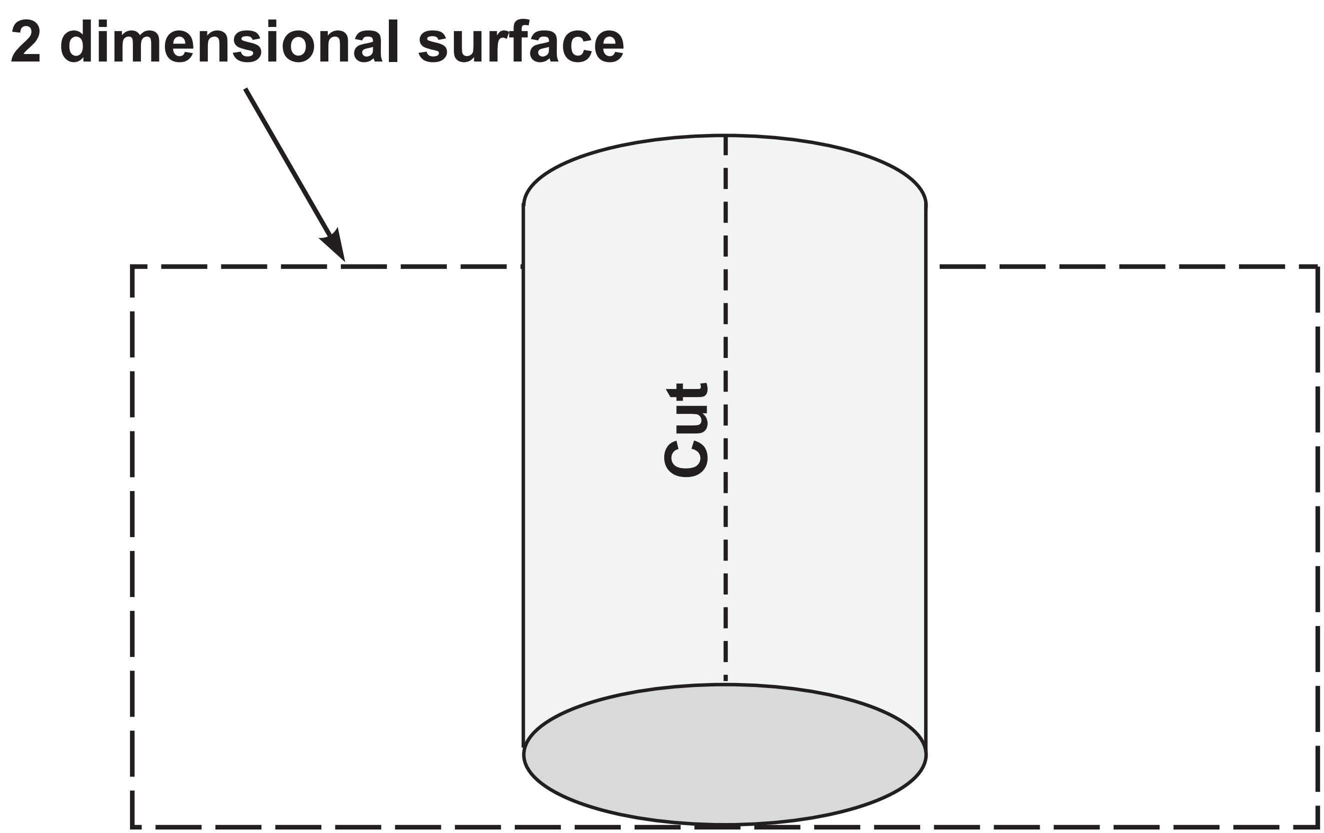
2-11



Cylindrical Projection

FIGURE

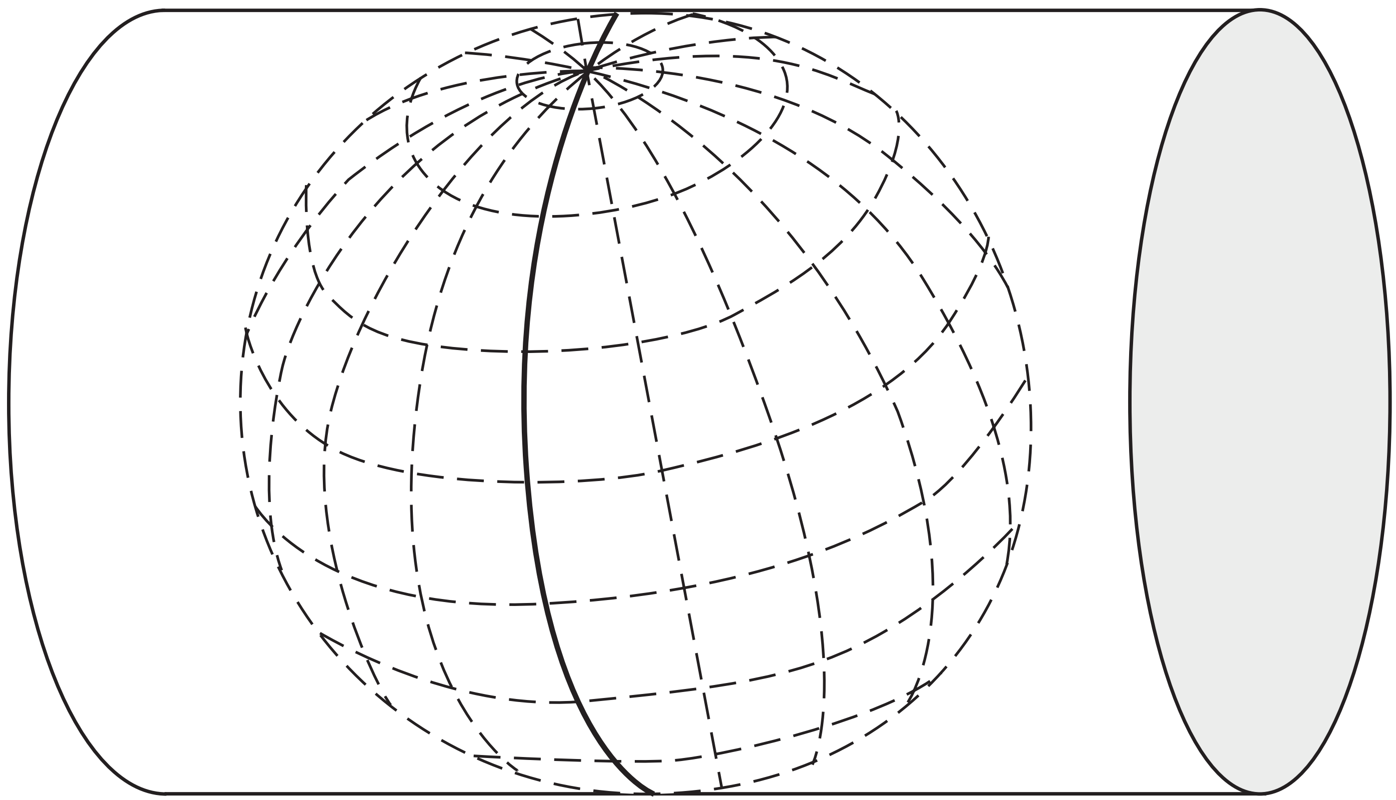
2-12



Cylinder Developed into Flat Surface

FIGURE 2-13

NSFGIS\TOPIC2\OH0213.CDR

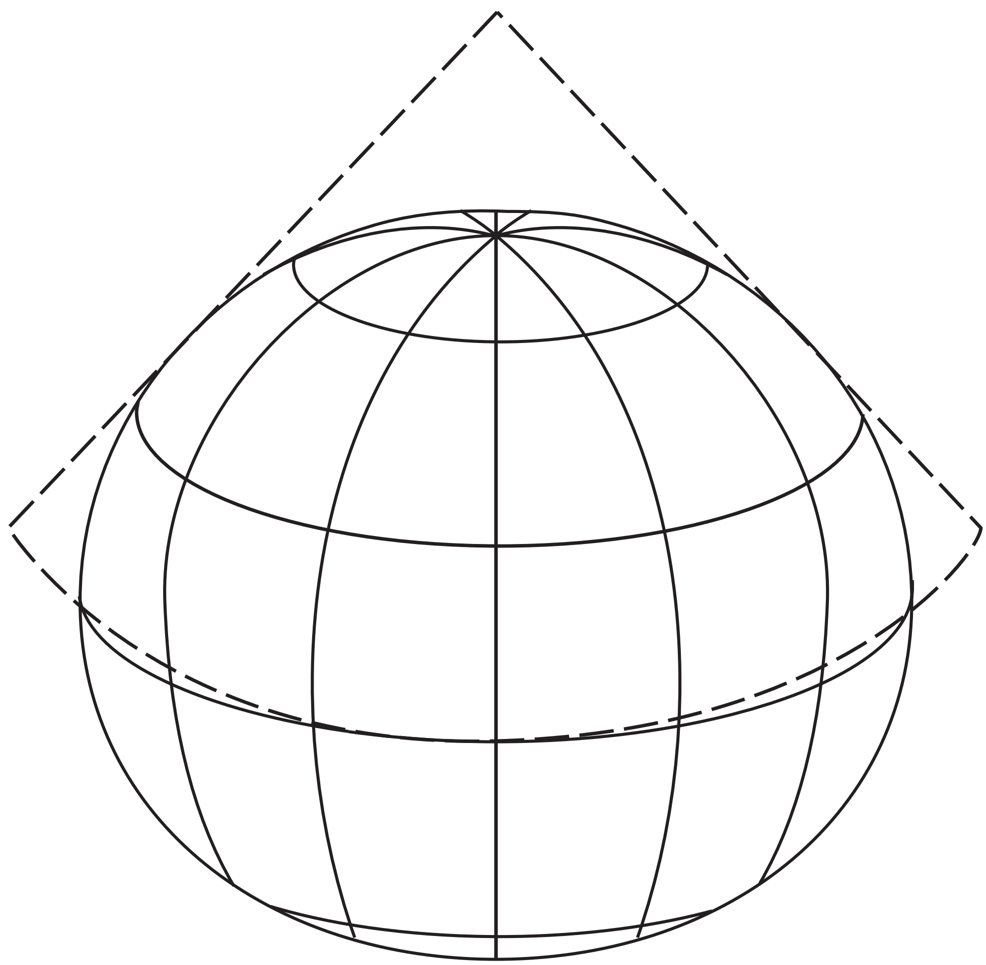


The transverse Mercator projection is projected onto a cylinder that is tangent along a meridian.

FIGURE

2-14

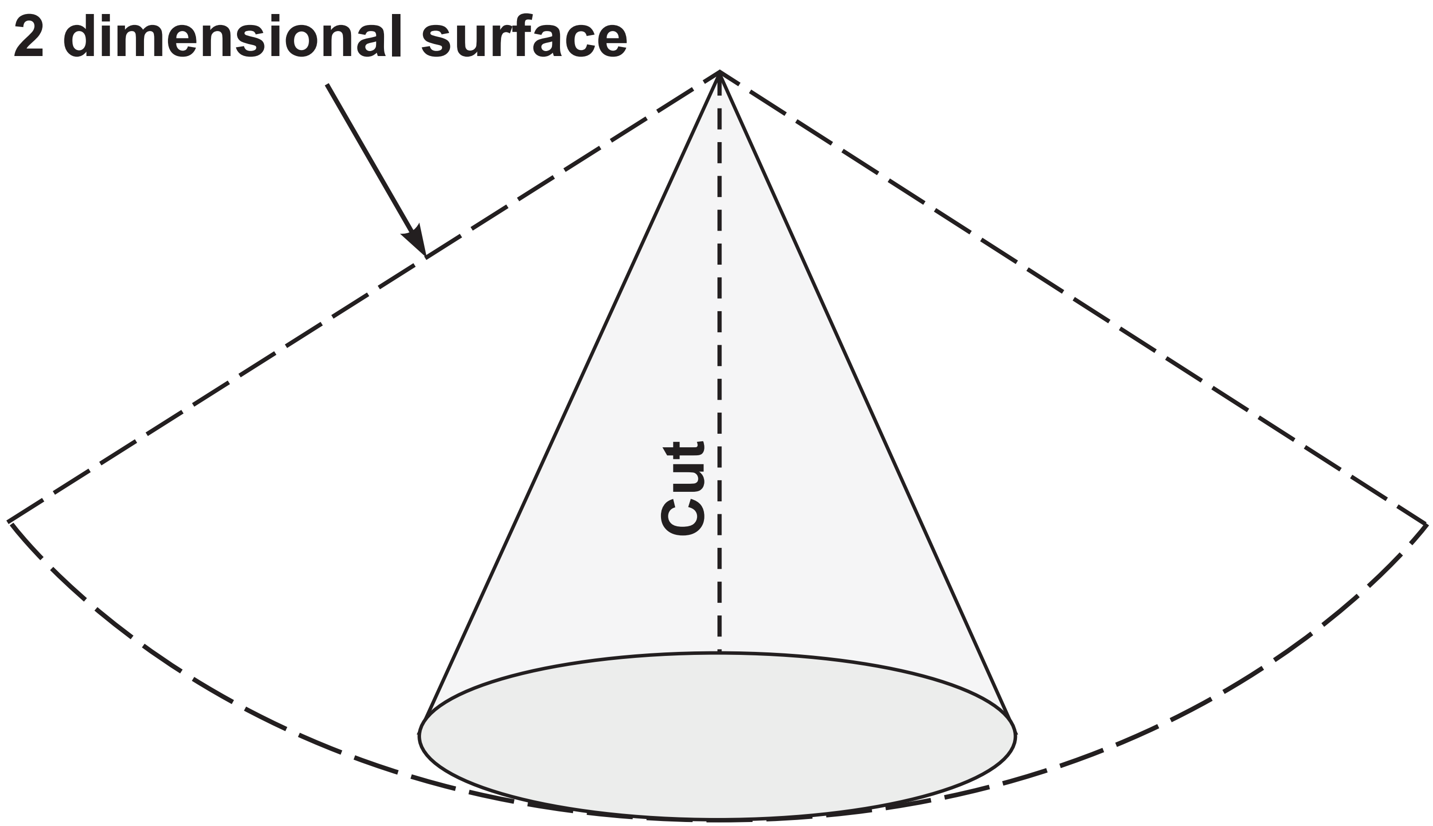
NSFGIS\TOPIC2\OH0214.CDR



Conic Projection

FIGURE

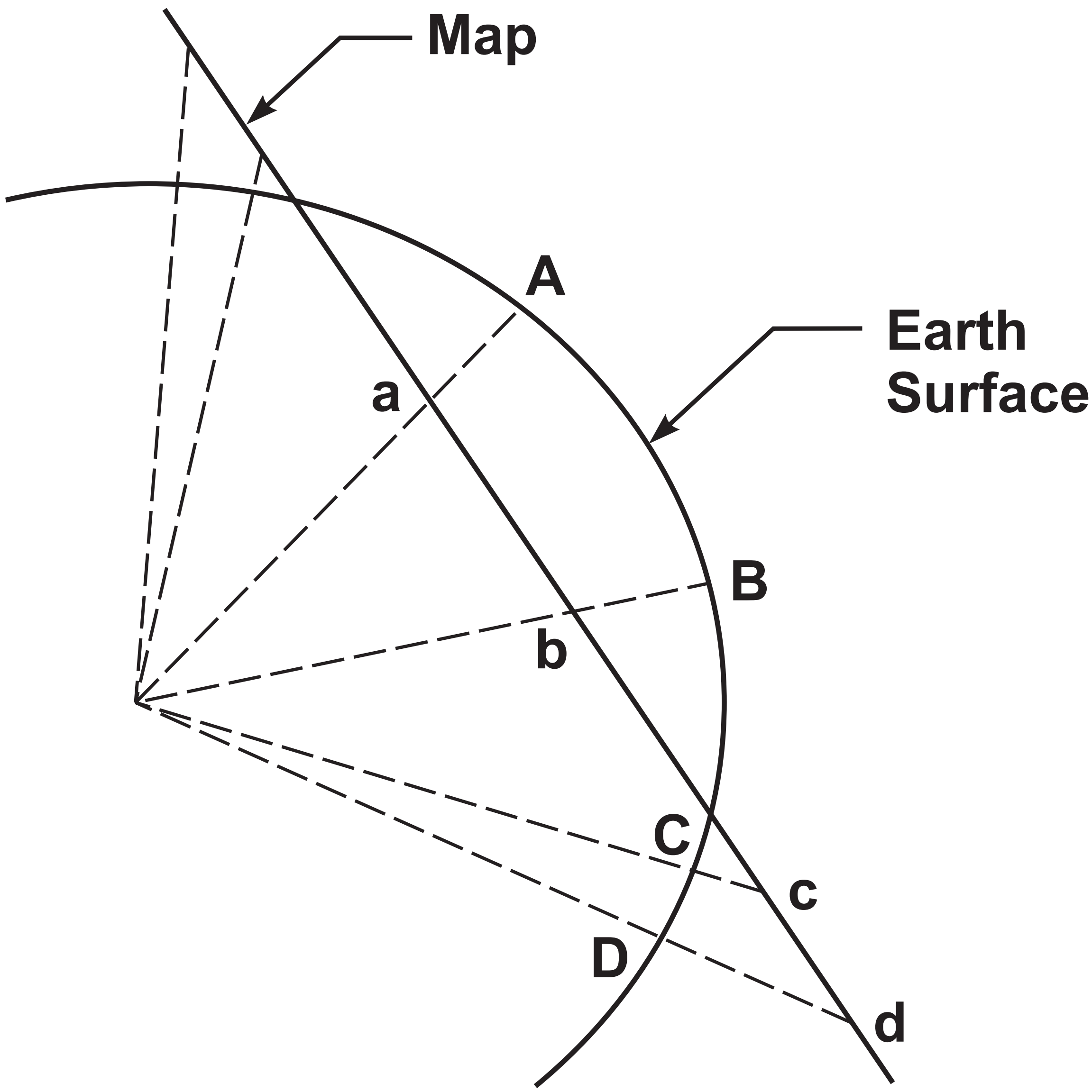
2-15



Cone Developed into Flat Surface

FIGURE 2-16

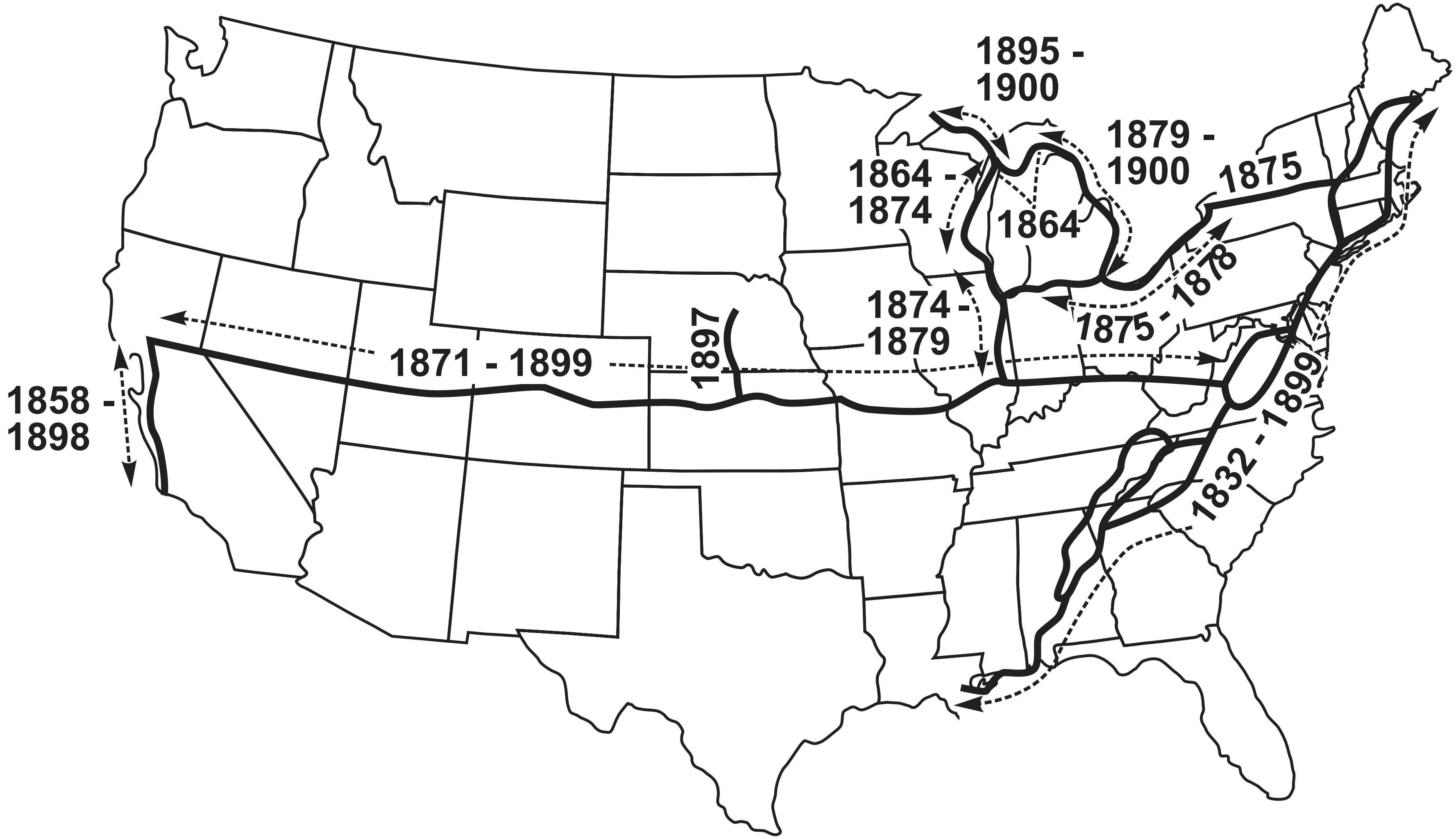
NSFGIS\TOPIC2\OH0216.CDR



Projection Distortion

FIGURE

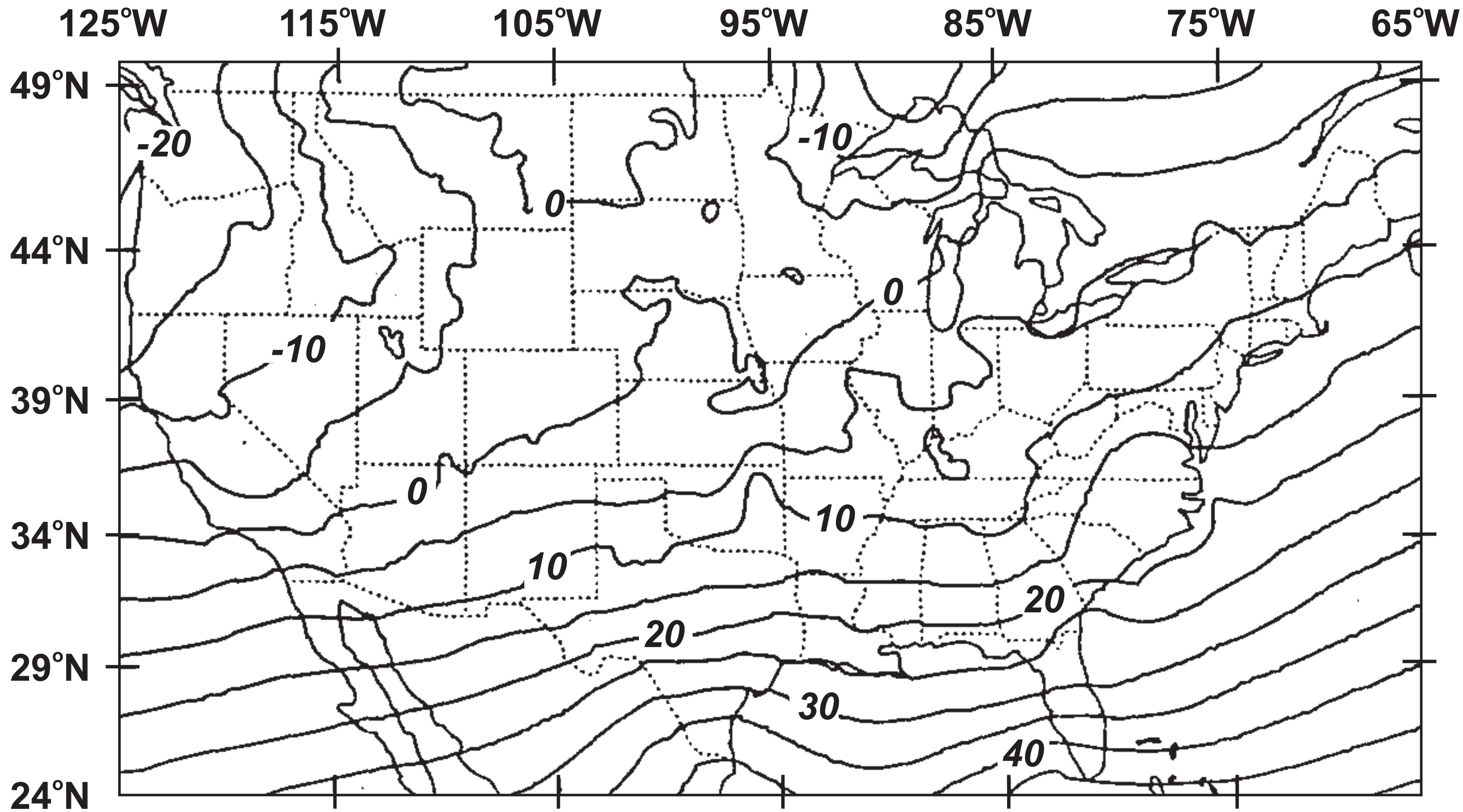
2-17



Original Horizontal Control

FIGURE
2-18

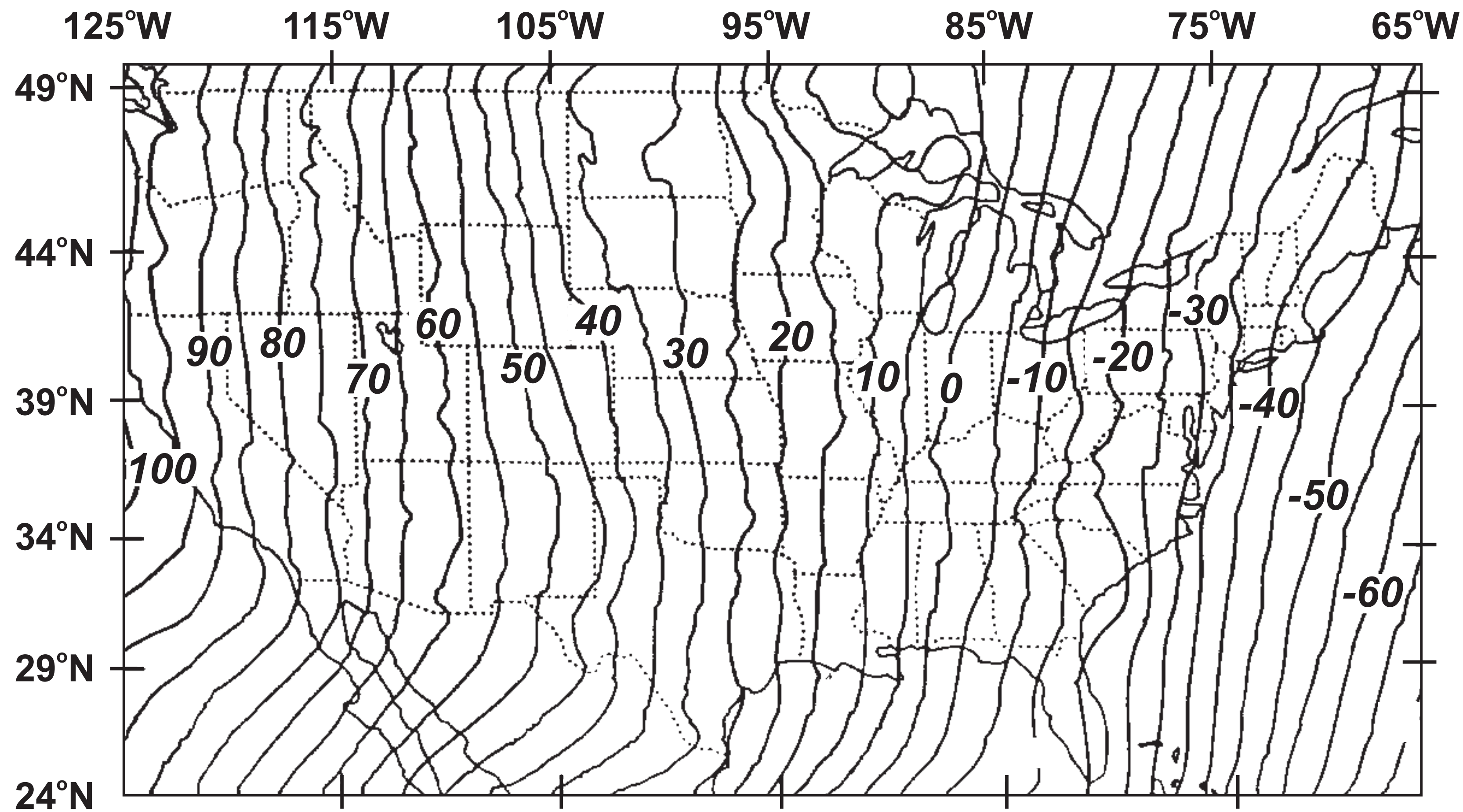
NSFGIS\TOPIC2\OH0218.CDR



Approximate Shift in Latitude from NAD27 to NAD83 (in meters)

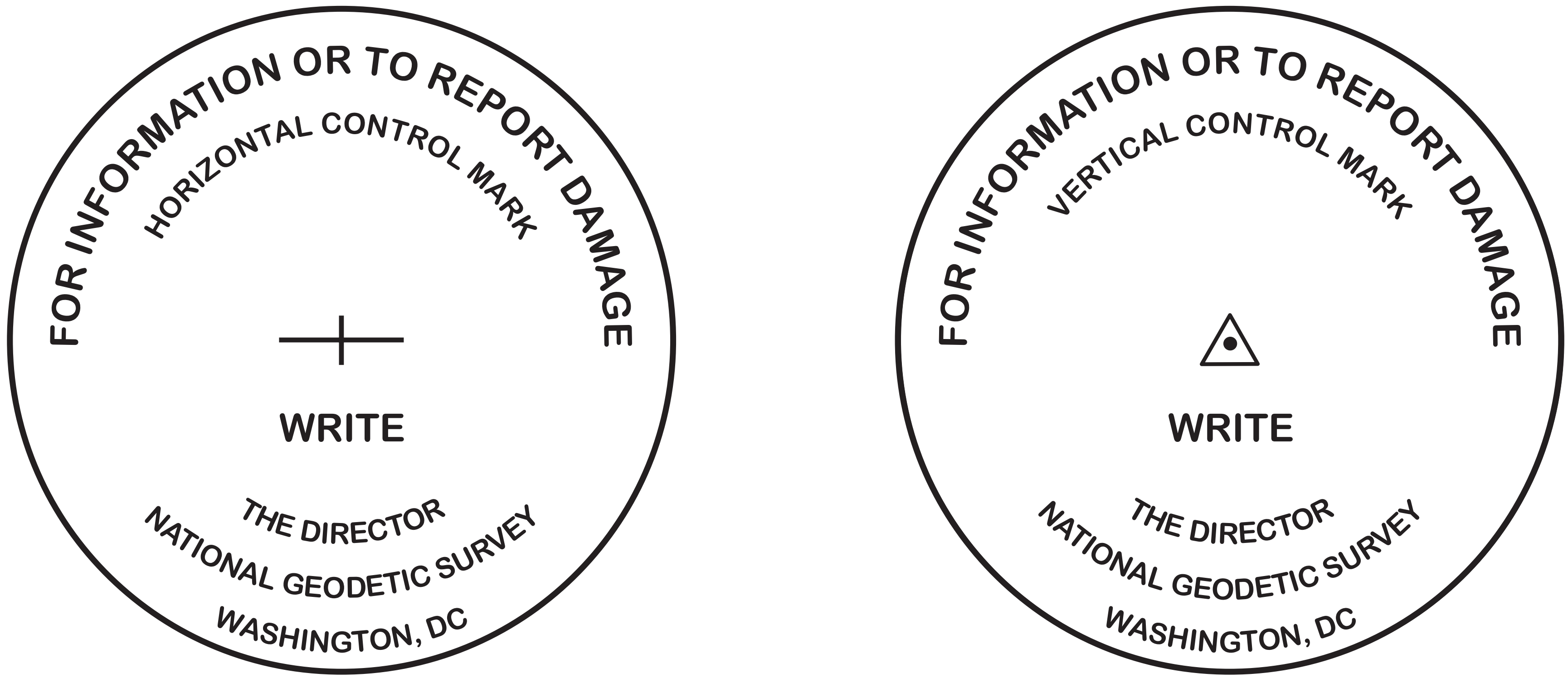
FIGURE 2-21

NSFGIS\TOPIC2\OH0221.CDR



Approximate Shift in Longitude NAD27 to NAD83 (in meters)

FIGURE 2-22



Horizontal

Vertical

Horizontal and Vertical Control Station Markers

FIGURE 2-23

NSFGIS\TOPIC2\OH0223.CDR

Horizontal Control Accuracy Standards

1. First Order

    a. 1 in 100,000

2. Second Order

    a. Class I   1 in 50,000

    b. Class II   1 in 20,000

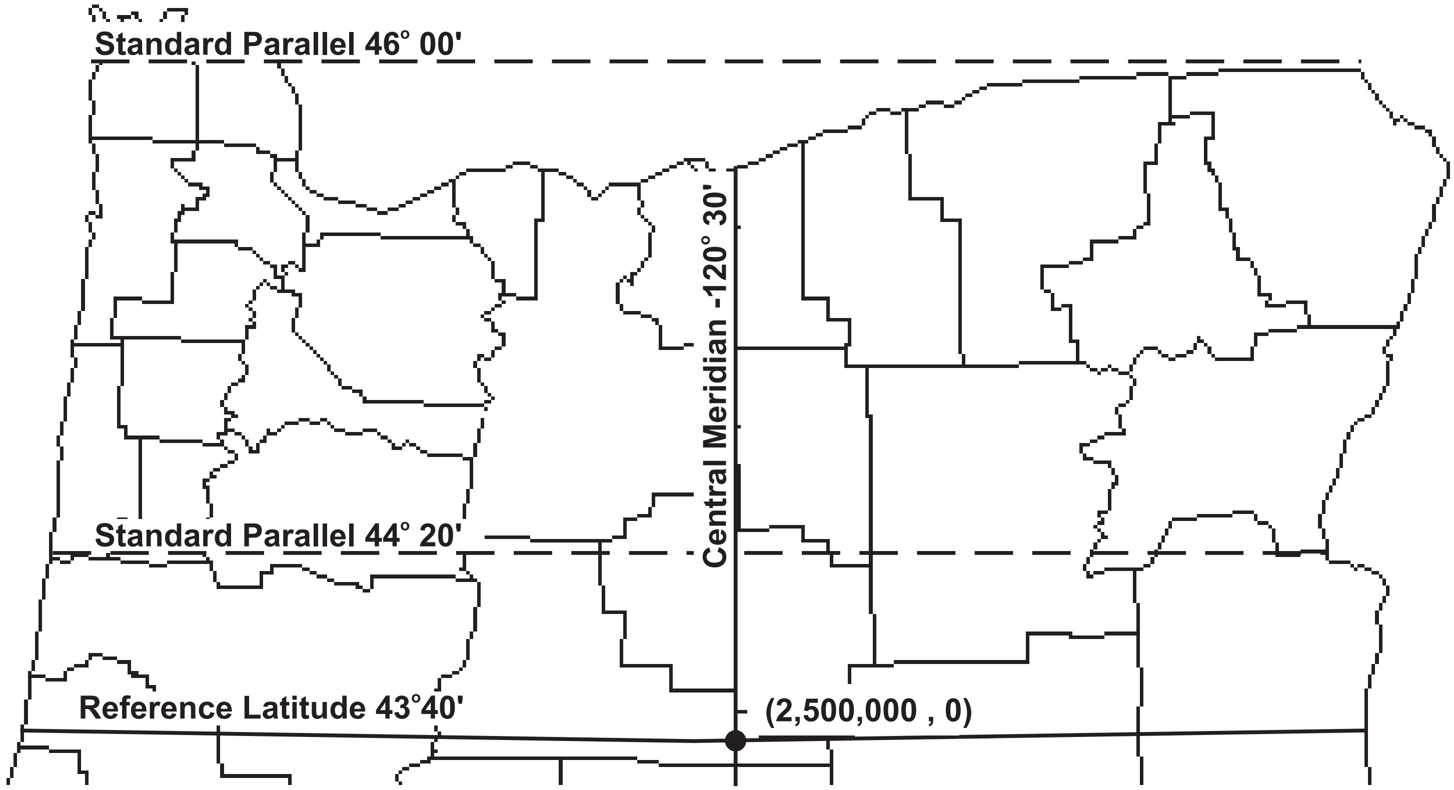
3. Third Order

    a. Class I   1 in 10,000

    b. Class II   1 in 5,000

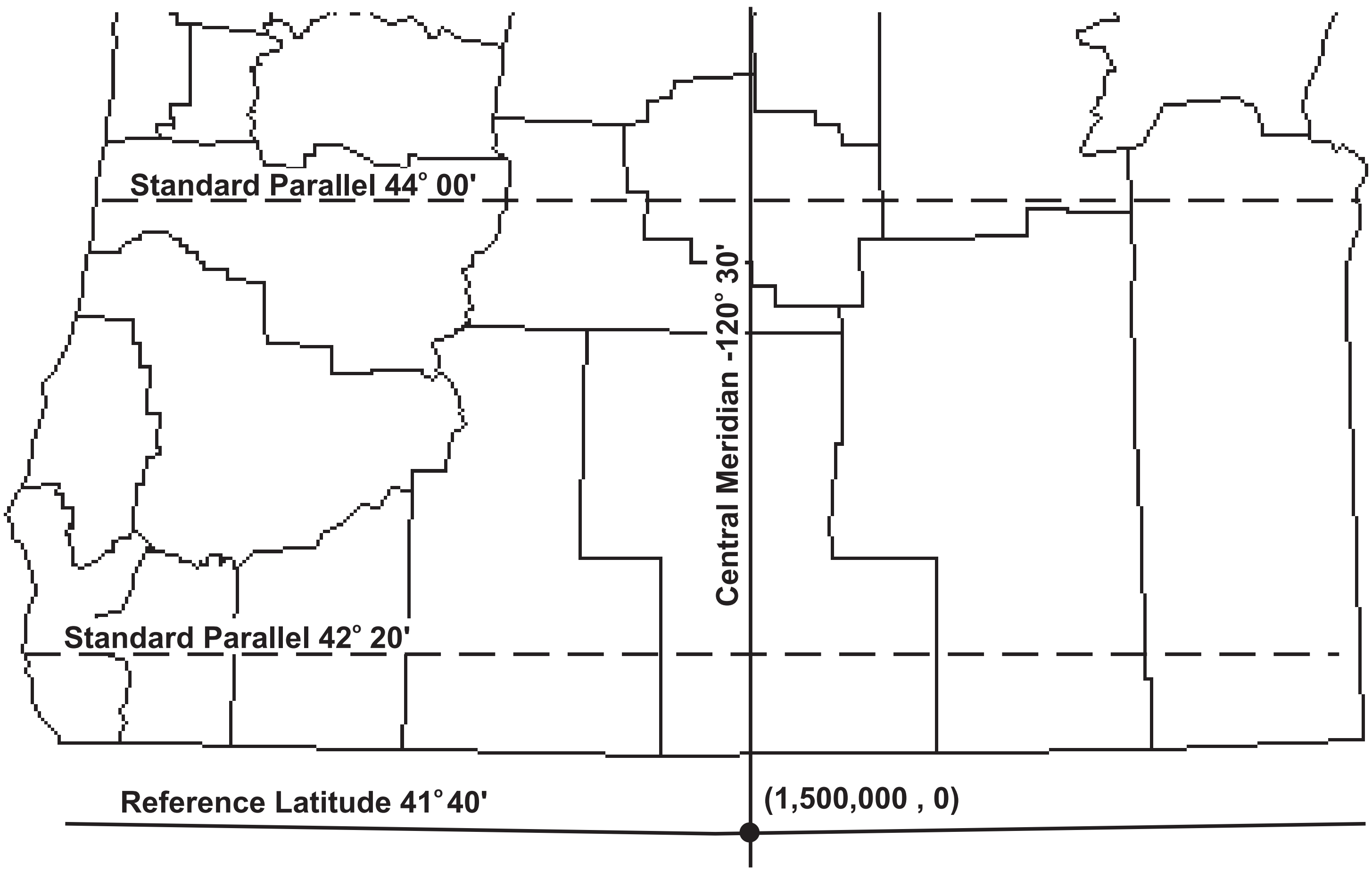
FIGURE

2-24



Oregon North State Plane Coordinate System

FIGURE 2-25



Oregon South State Plane Coordinate System

FIGURE 2-26

NSFGIS\TOPIC2\OH0226.CDR

The Two Types of Projections Most Commonly Used Are:

1. **Transverse Mercator**

   Used for states longer in north-south direction (eg. Illinois)

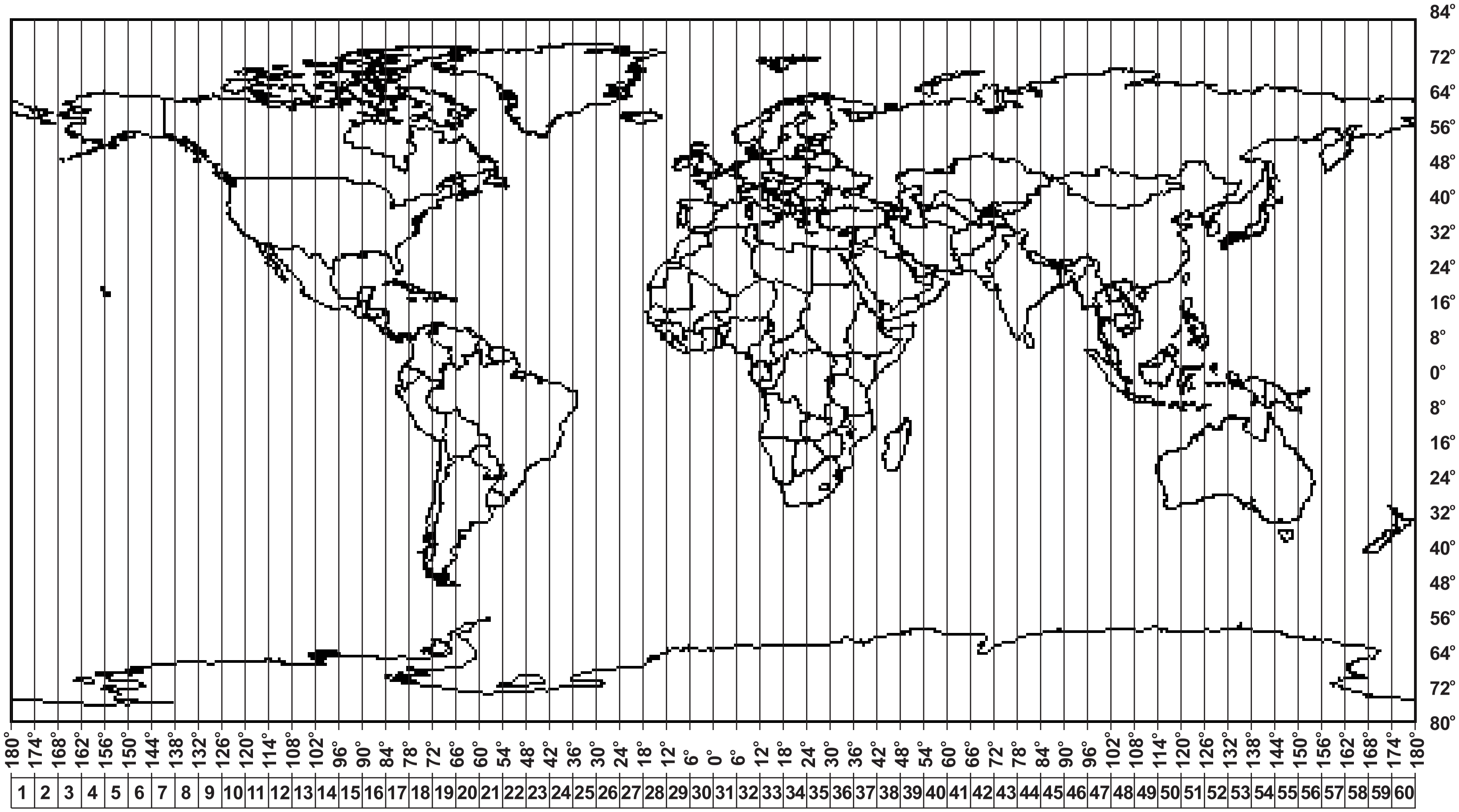
2. **Lambert Conformal Conic**

   Used for states longer in east-west direction (eg. Washington, Oregon)

Both are used to define a standard coordinate set for each state, known as the State Plane Coordinate System (SPCC)

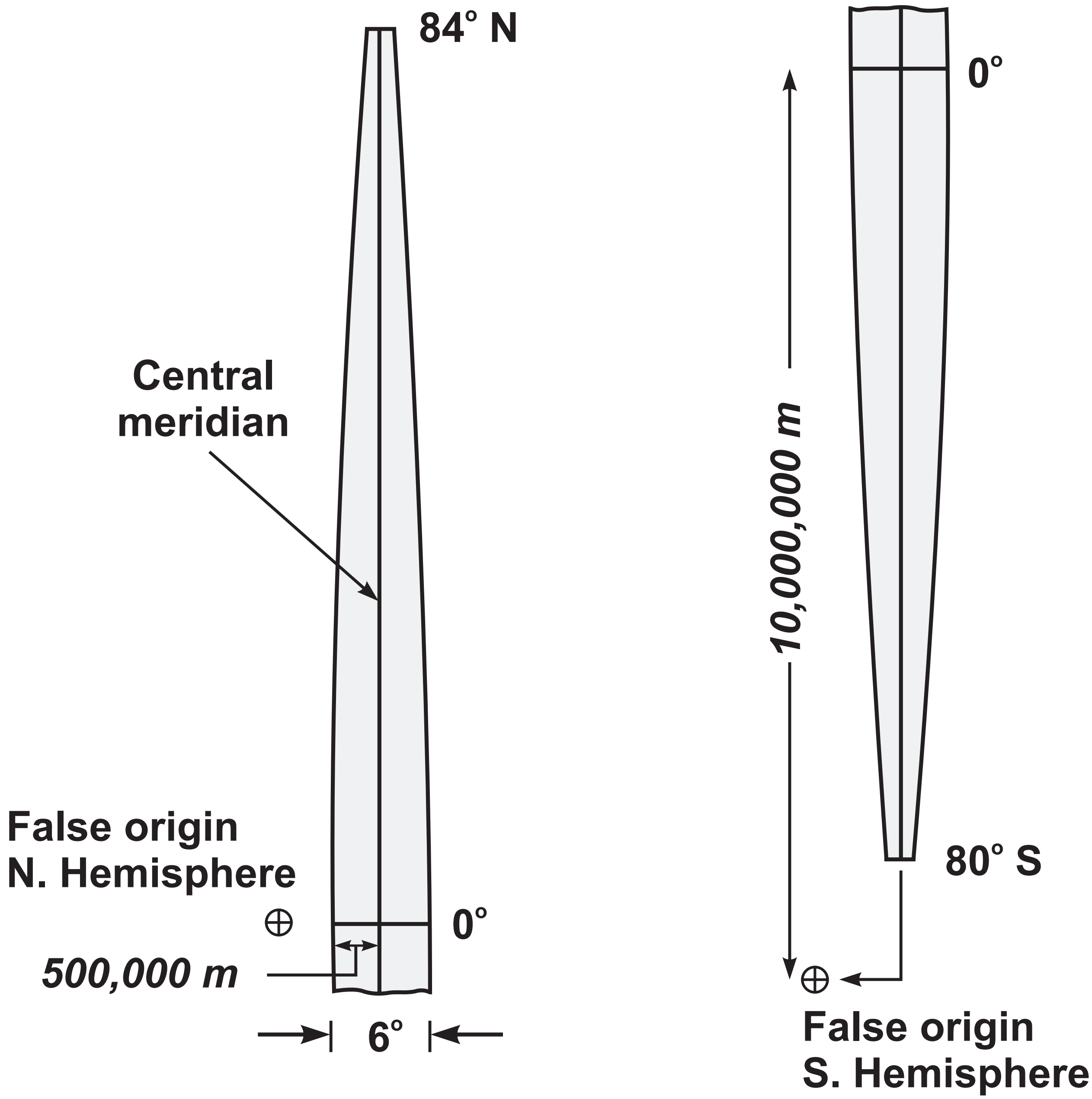
FIGURE

2-27



Universal Transverse Mercator (UTM) grid zone

FIGURE
2-28



UTM Zone

FIGURE

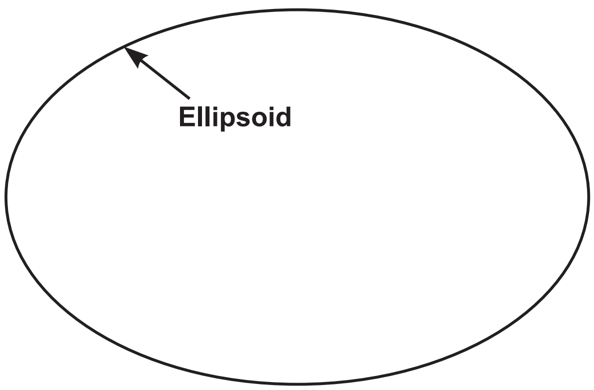
2-29

Objectives

Understand the basic concepts of geography pertaining to:

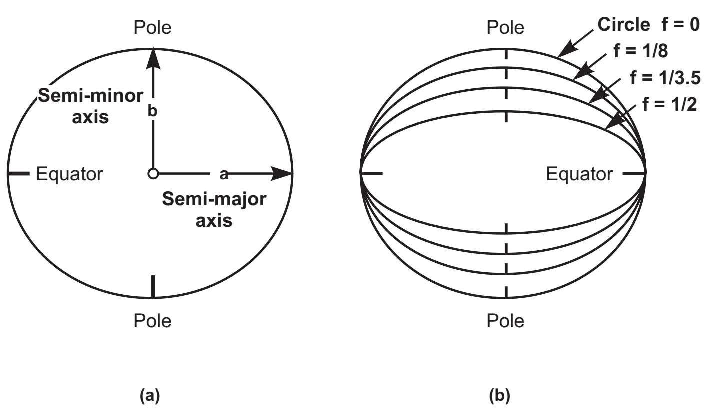
1. **Geodesy**
2. **Map projections**
3. **Map based grid systems**

FIGURE 2-1



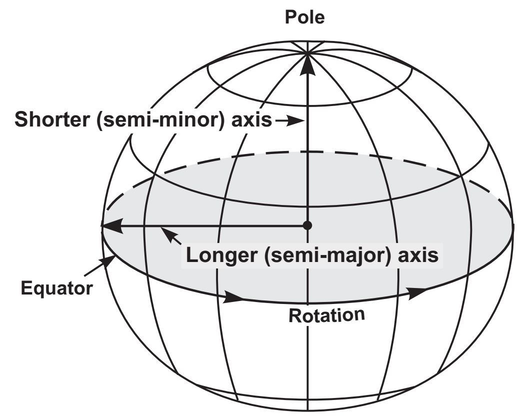
Ellipsoid

FIGURE 2-2

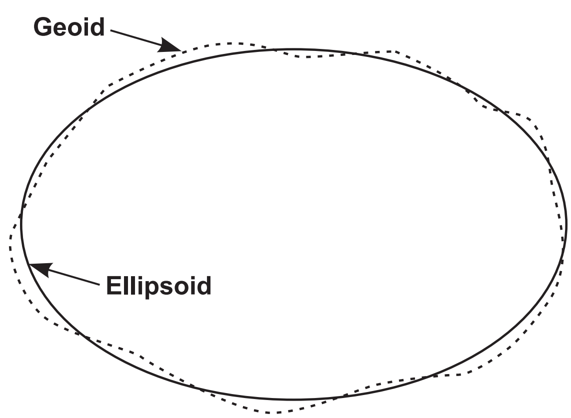


Ellipsoid Basics and Variations

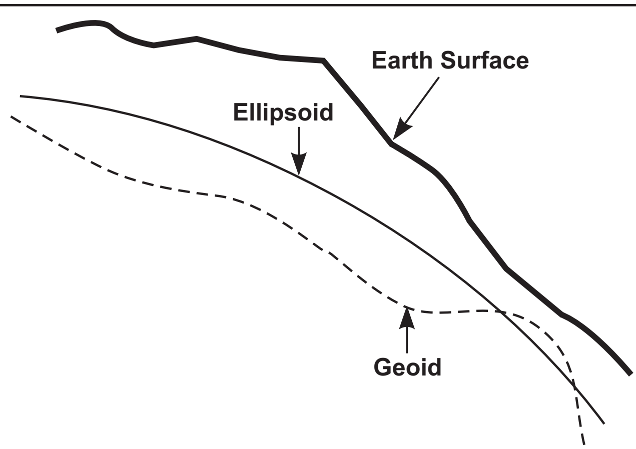
FIGURE 2-3



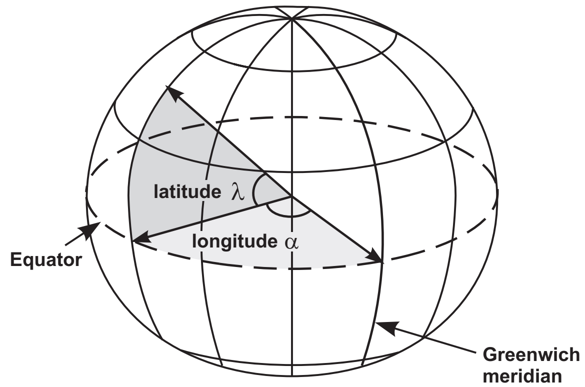
Earth in Ellipsoid Form — FIGURE 2-4



Geoid vs. Ellipsoid — FIGURE 2-5



Ellipsoid, Geoid and Earth Surface — FIGURE 2-6



The Graticule: Latitude and Longitude

FIGURE 2-7

Since Earth's radius is not constant, Earth's surface distance contained in 1° of latitude or longitude in not constant.

Latitude Changes Little:

At Equator, 1 Degree = 110.5 km (68.7 mi)

At Poles, 1 Degree = 111.7 km (69.4 mi)

For Longitude:

At Equator, 1 Degree = 111.3 km (69.2 mi)

At Poles, 1 Degree = 0 km (0 mi)

FIGURE 2-8

A Map Projection Is:

A systematic rendering of points on the earth ellipsoid to points on a flat sheet.

Think of it as passing rays of light from some point through the globe and onto the map surface.

Why do we need to do this?

1. To produce 2-dimensional maps.
2. To have a convenient cartesian coordinate system.

FIGURE 2-9

Desirable Properties of Projections

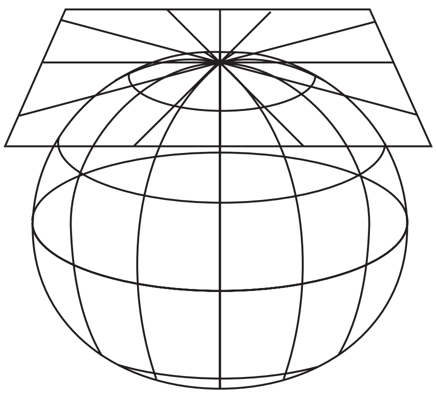
A. Equivalence

B. Conformality

C. Azimuthality

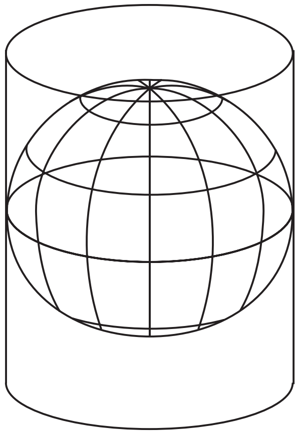
D. Equidistance

FIGURE 2-10



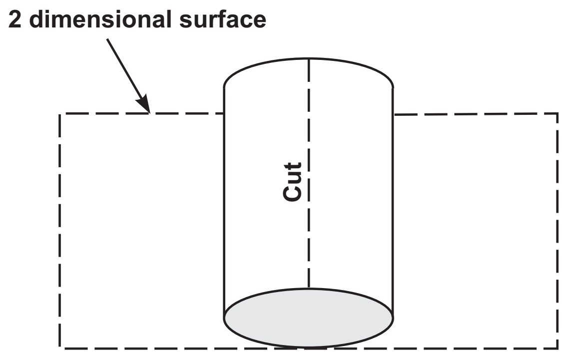
Polar Plane Projection

FIGURE 2-11



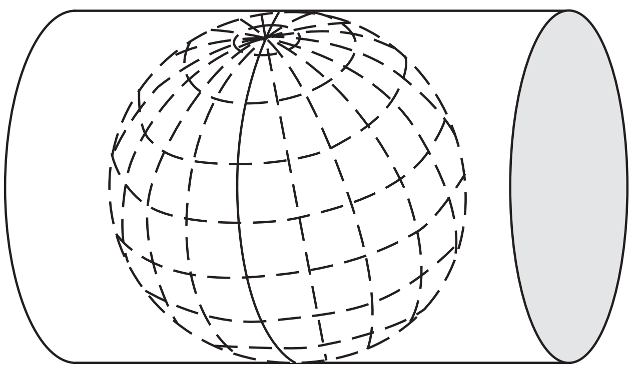
Cylindrical Projection

FIGURE 2-12



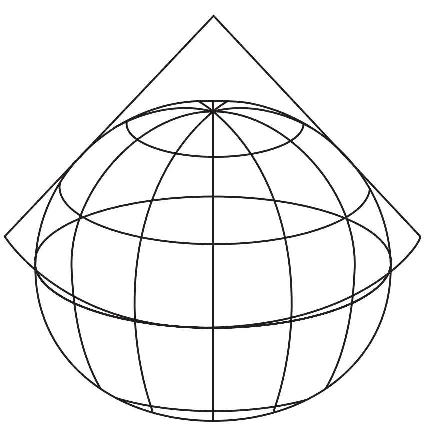
Cylinder Developed into Flat Surface

FIGURE
2-13



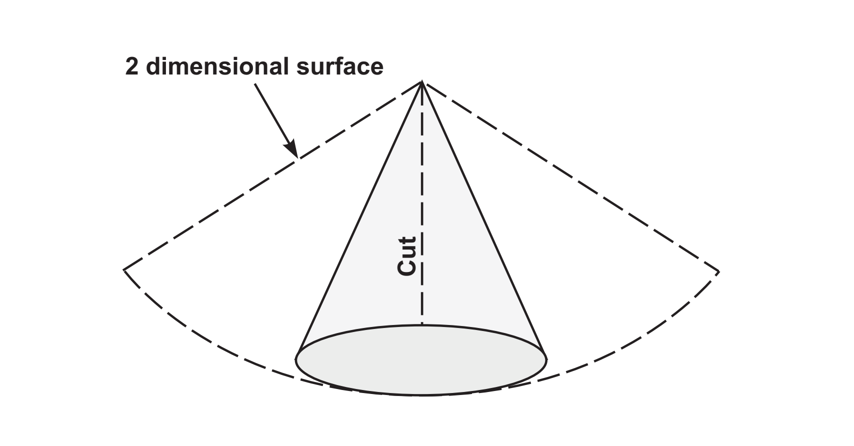
The transverse Mercator projection is projected
onto a cylinder that is tangent along a meridian.

FIGURE
2-14

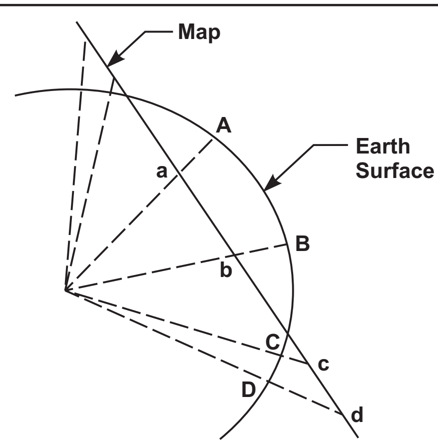


Conic Projection

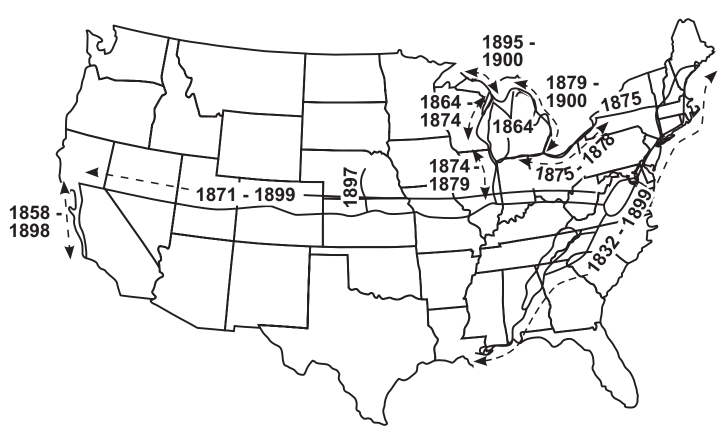
FIGURE
2-15



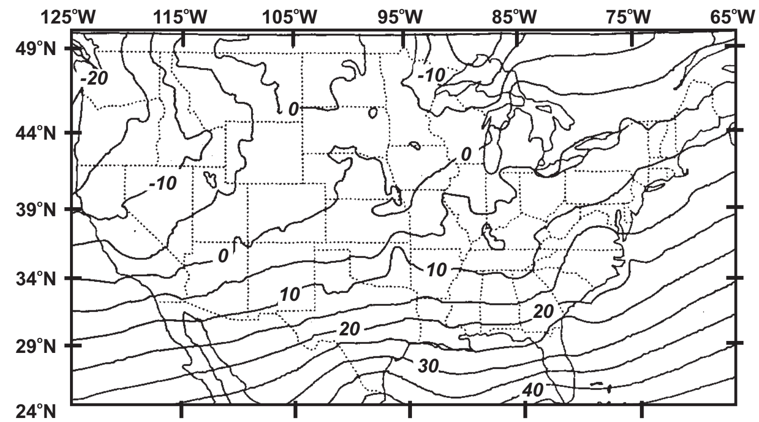
Cone Developed into Flat Surface	FIGURE 2-16



Projection Distortion	FIGURE 2-17

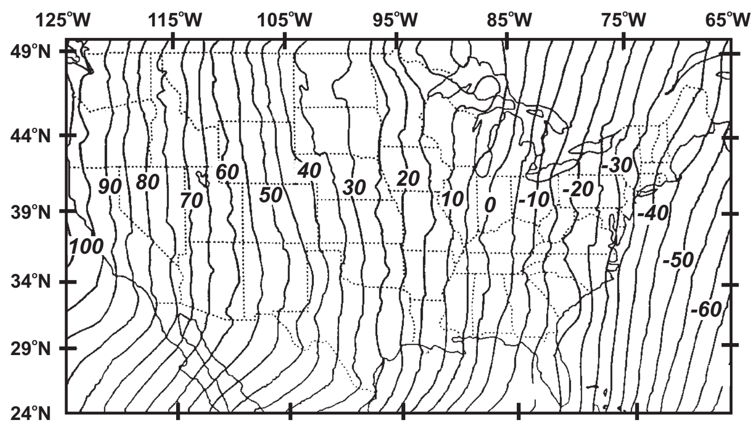


Original Horizontal Control	FIGURE 2-18



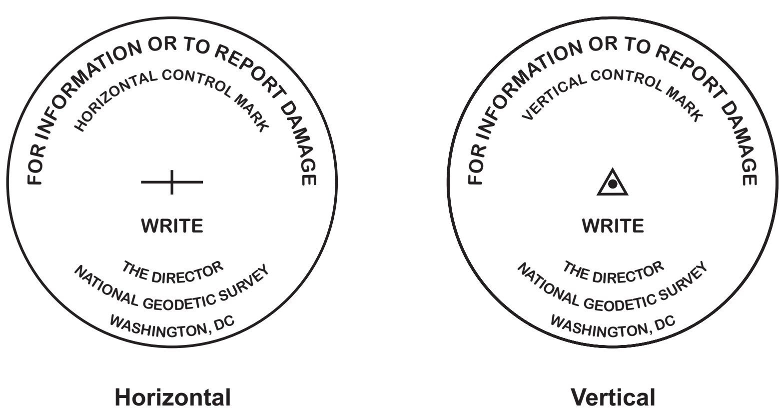
Approximate Shift in Latitude from NAD27 to NAD83 (in meters)

FIGURE 2-21



Approximate Shift in Longitude NAD27 to NAD83 (in meters)

FIGURE 2-22



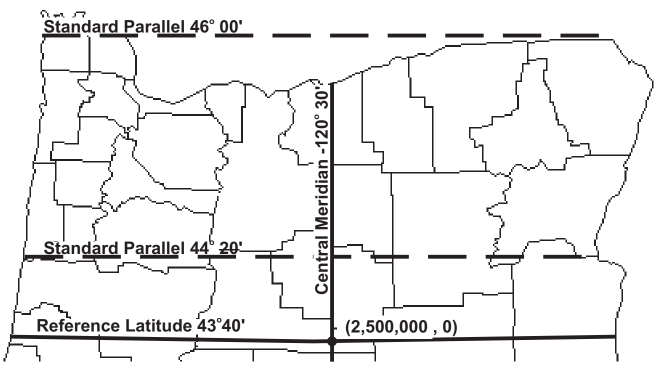
Horizontal and Vertical Control Station Markers

FIGURE 2-23

Horizontal Control Accuracy Standards

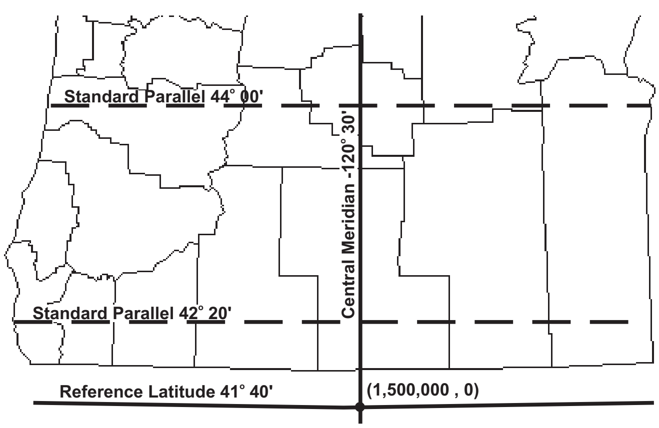
1. First Order
   a. 1 in 100,000
2. Second Order
   a. Class I 1 in 50,000
   b. Class II 1 in 20,000
3. Third Order
   a. Class I 1 in 10,000
   b. Class II 1 in 5,000

FIGURE 2-24



Oregon North State Plane Coordinate System

FIGURE 2-25



Oregon South State Plane Coordinate System

FIGURE 2-26

The Two Types of Projections Most Commonly Used Are:

1. **Transverse Mercator**

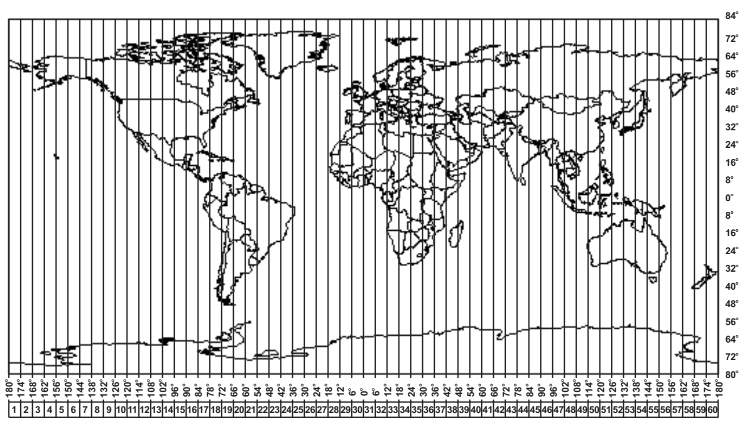
   Used for states longer in north-south direction (eg. Illinois)

2. **Lambert Conformal Conic**

   Used for states longer in east-west direction (eg. Washington, Oregon)

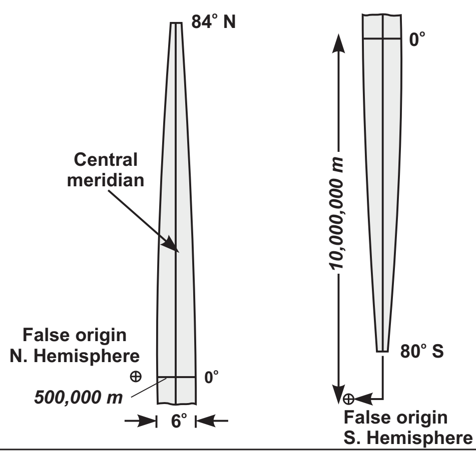
Both are used to define a standard coordinate set for each state, known as the State Plane Coordinate System (SPCC)

FIGURE 2-27



Universal Transverse Mercator (UTM) grid zone

FIGURE 2-28



UTM Zone

FIGURE 2-29

Objectives

A. Become familiar with various hardware associated with GIS

B. Become familiar with various software components of GIS

FIGURE

3-1

Basic GIS Hardware

A. Input

B. Central Processing Unit (CPU)

C. Volatile Storage

D. Non-volatile storage

E. Archival

F. Backup

G. Output

FIGURE

3-2

Categories of GIS Hardware

A. Personal Computers (PC's)

B. Workstations

C. Mainframes

FIGURE

3-3

Functional Elements of a Geographic Information System

1. Data Acquisition

2. Preprocessing

3. Data Management

4. Manipulation and Analysis

5. Product Generation

FIGURE

3-4

Objectives

A. Become familiar with various hardware associated with GIS

B. Become familiar with various software components of GIS

FIGURE 3-1

Basic GIS Hardware

A. Input

B. Central Processing Unit (CPU)

C. Volatile Storage

D. Non-volatile storage

E. Archival

F. Backup

G. Output

FIGURE 3-2

Categories of GIS Hardware

A. Personal Computers (PC's)

B. Workstations

C. Mainframes

FIGURE 3-3

Functional Elements of a Geographic Information System

1. Data Acquisition
2. Preprocessing
3. Data Management
4. Manipulation and Analysis
5. Product Generation

FIGURE 3-4

Objectives

A. Become familiar with the concept of modeling reality in a Geographic Information System.

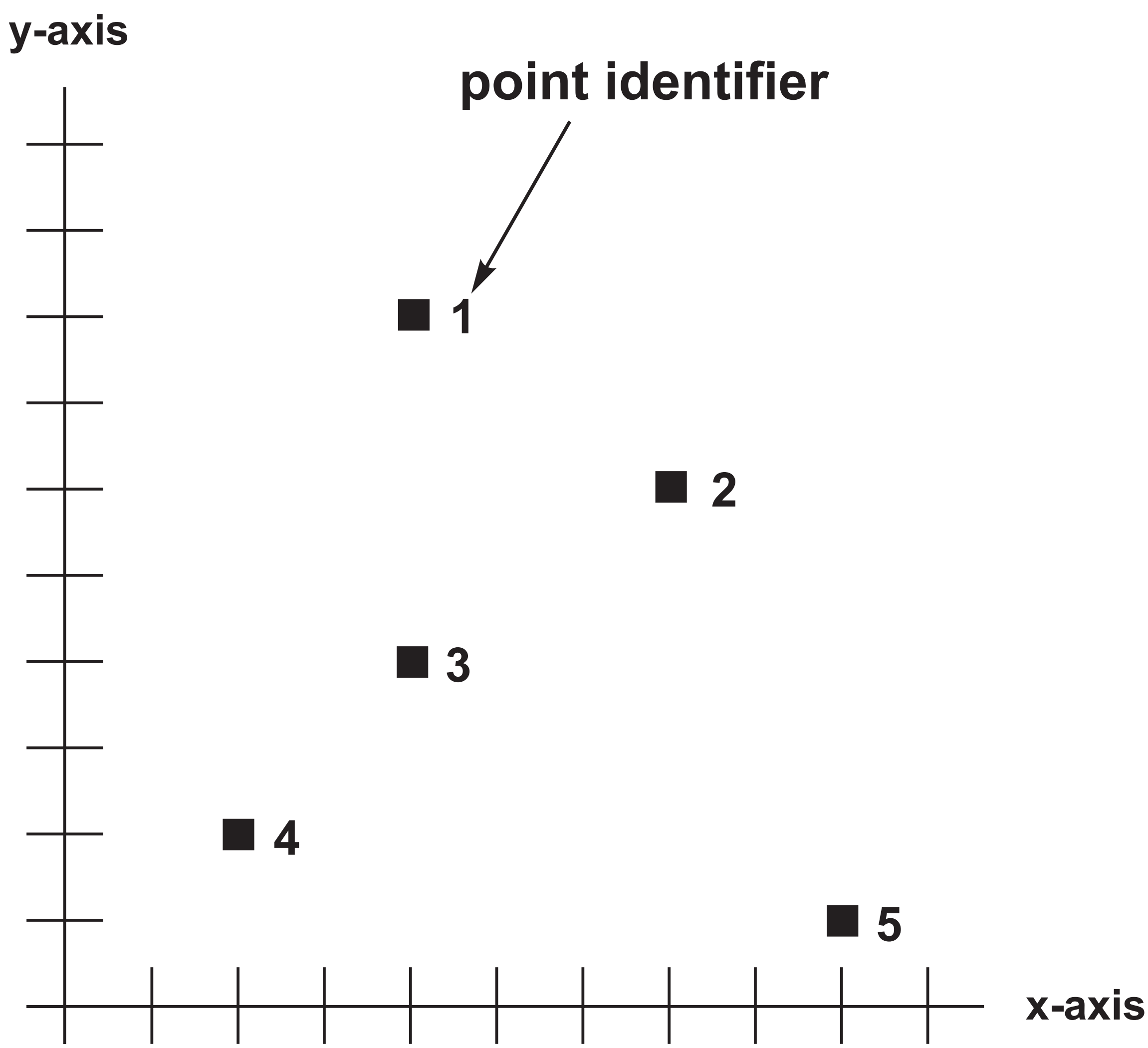
B. Understand various options for data models underlying GIS.

C. Understand how data structures contribute to GIS functionality.

D. Understand how file structures may ultimately affect GIS software efficiency.

FIGURE

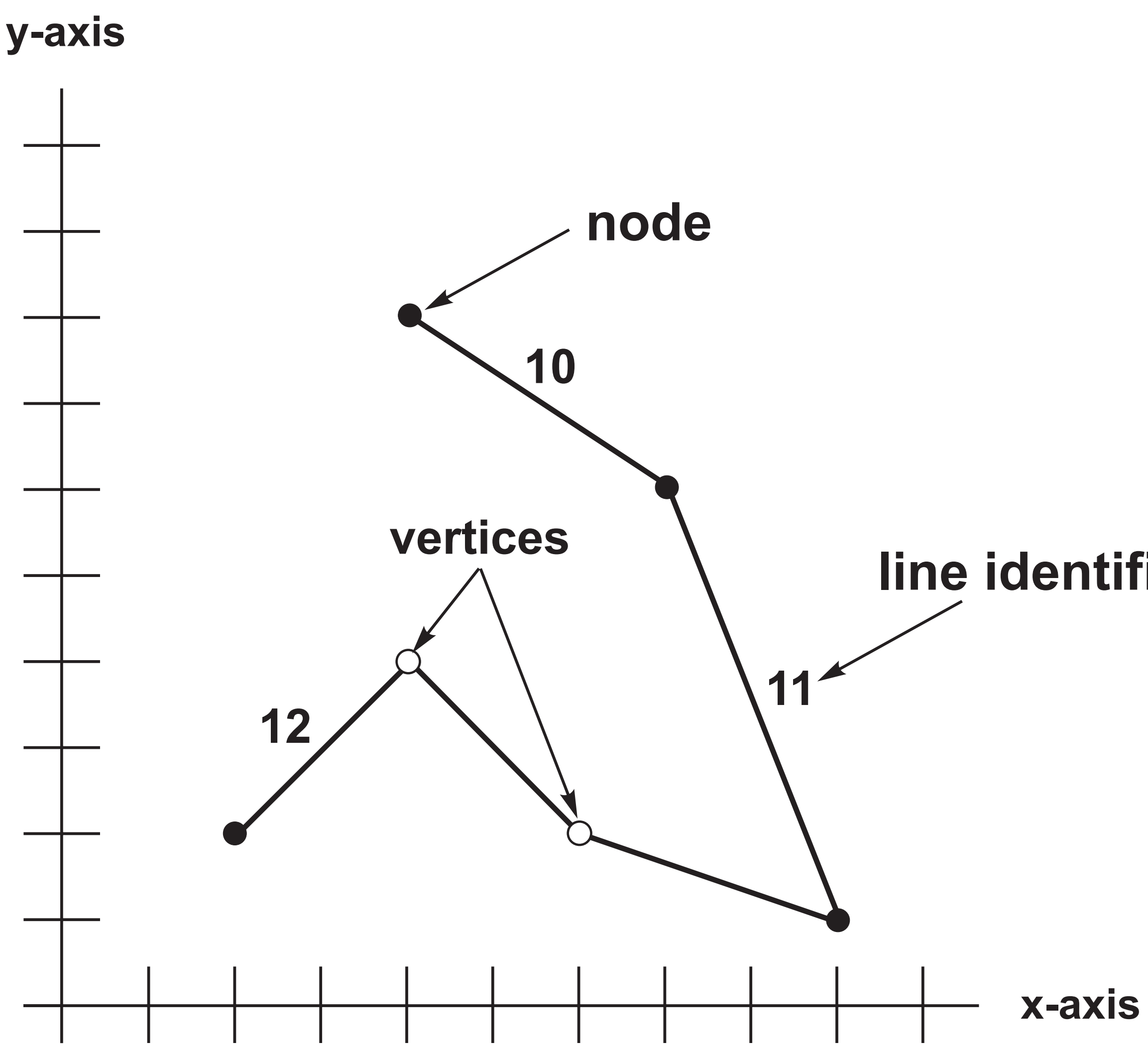
4-1



Points

FIGURE

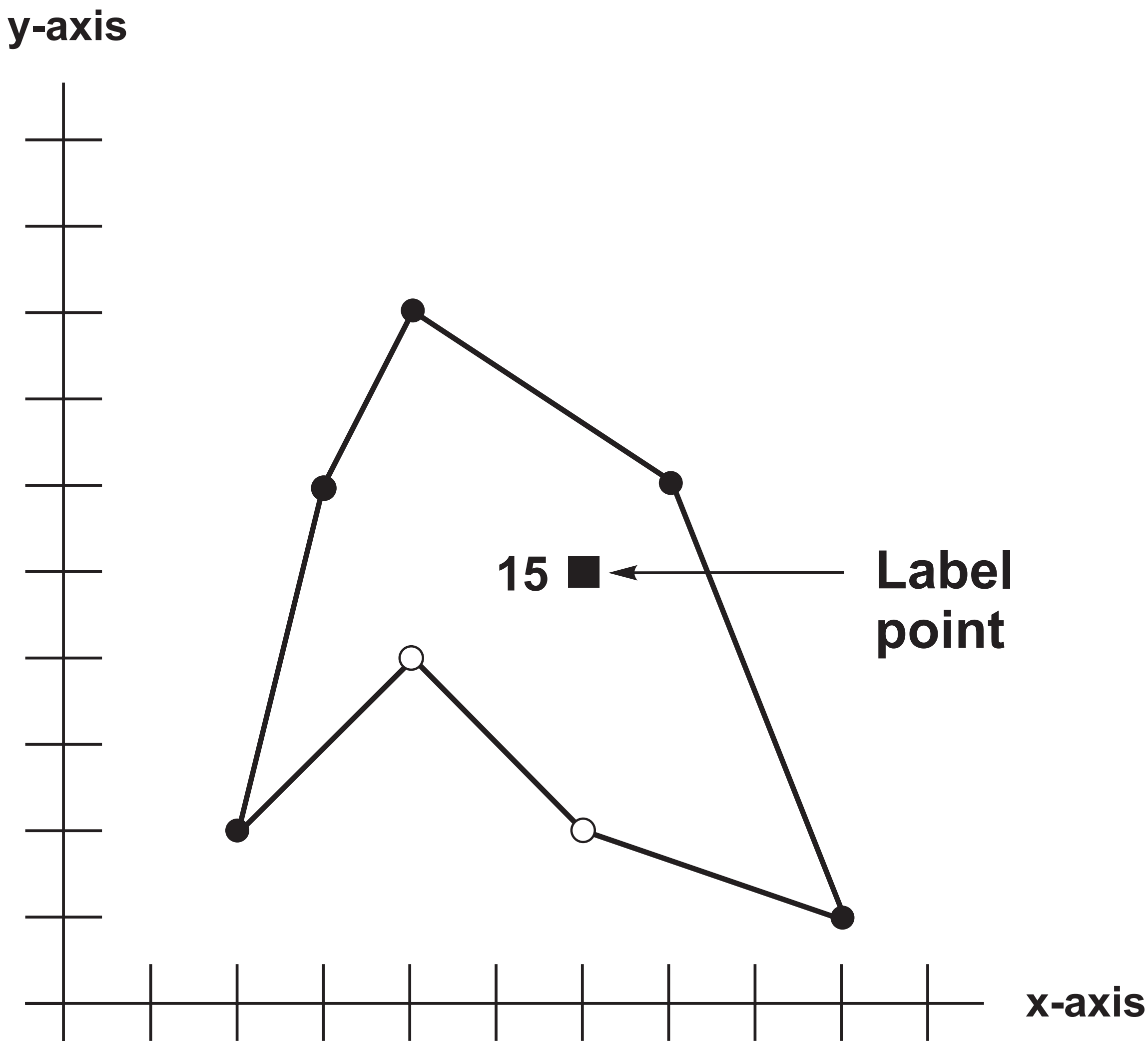
4-2



Line

FIGURE

4-3

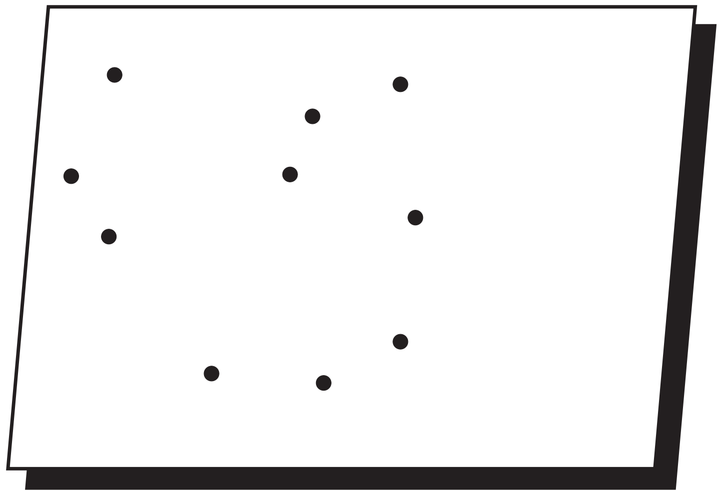


Polygon

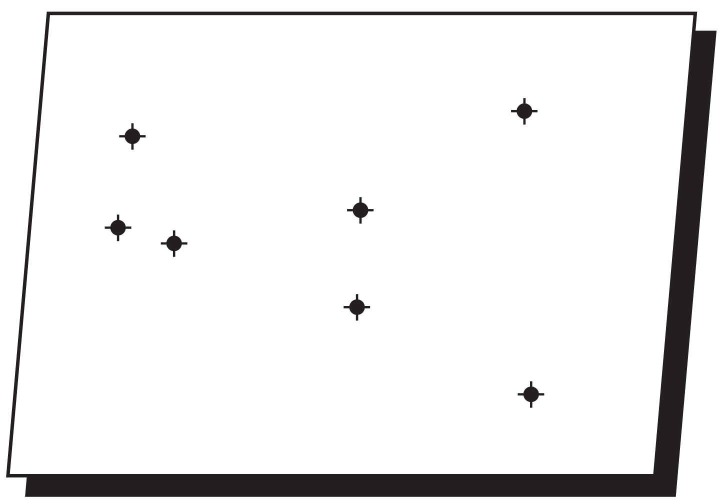
FIGURE

4-4

Wells



Springs

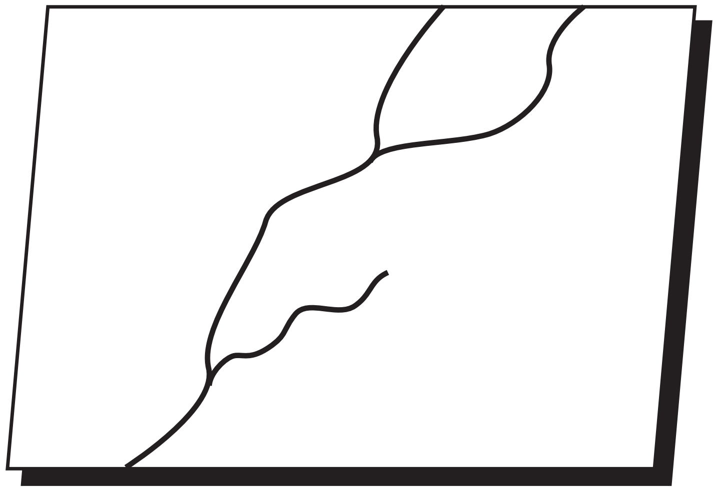


Point Features

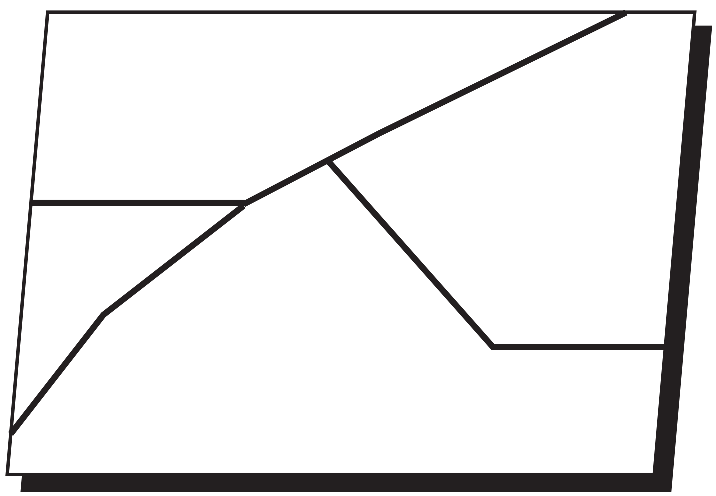
FIGURE

4-5

Streams



Roads

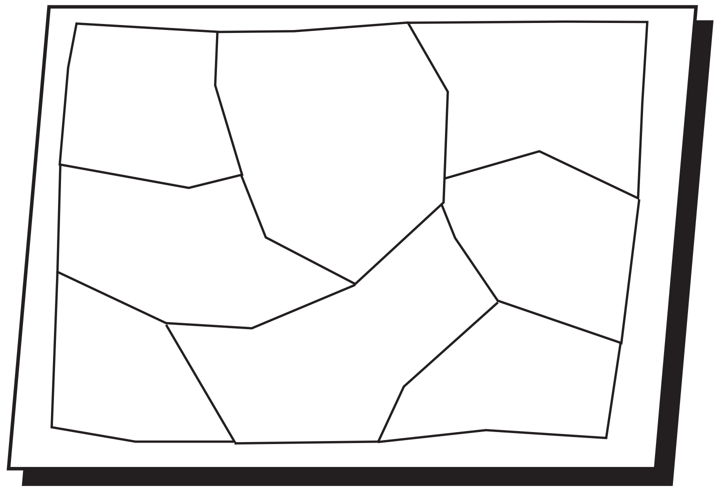


Line Features

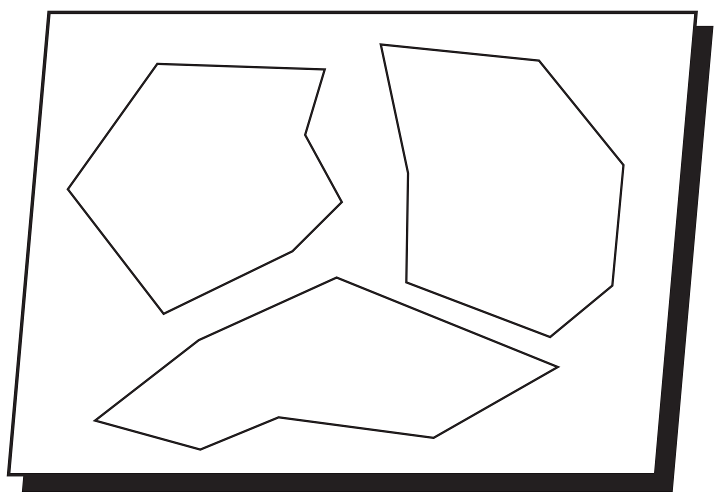
FIGURE

4-6

Plant Communities



Wildlife Habitats



Polygon Features

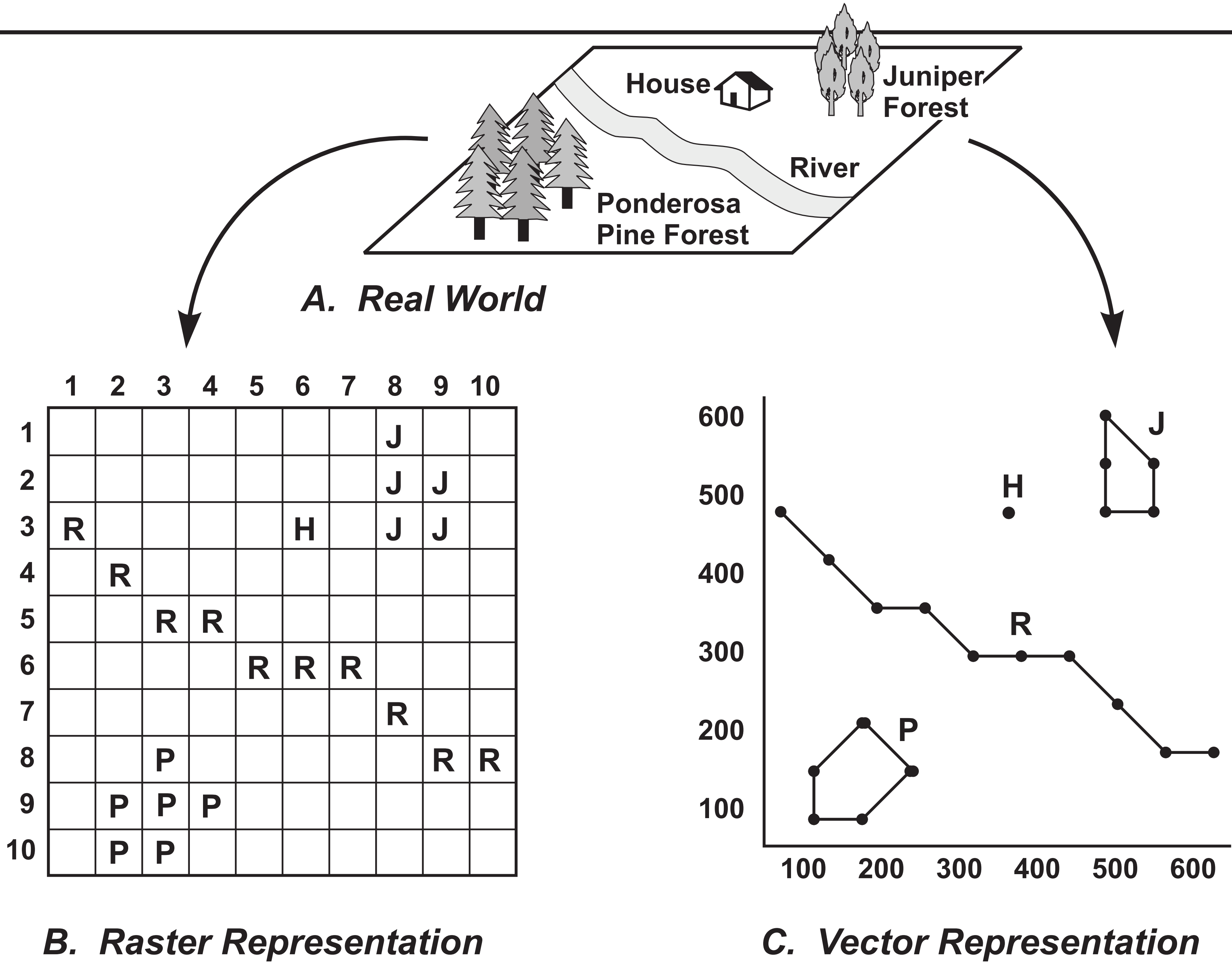
FIGURE

4-7

COCC Campus Map

FIGURE

4-8

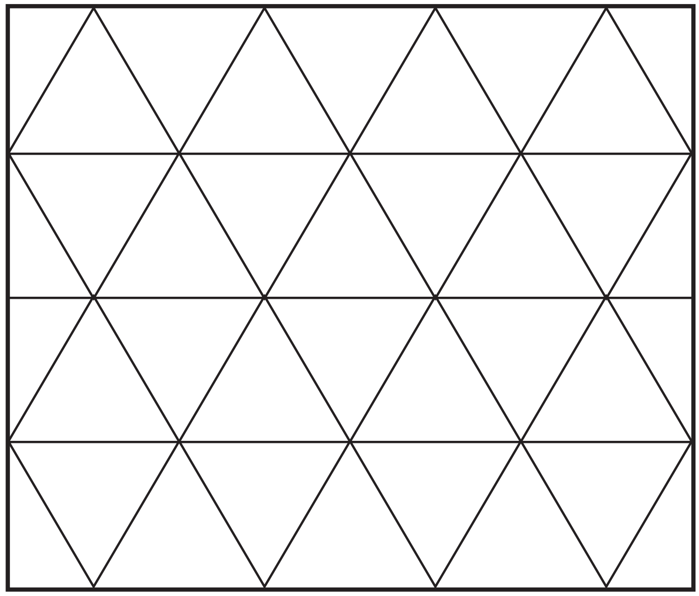


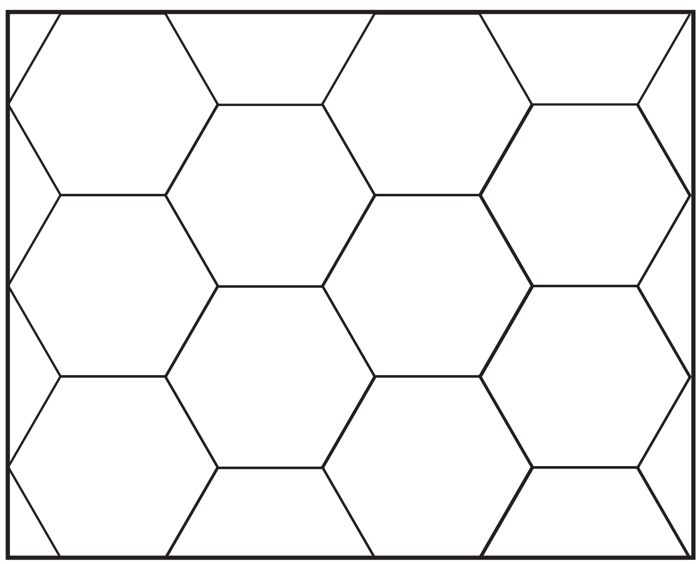
Raster and Vector Data Models

FIGURE

4-9

NSFGIS\TOPIC4\OH0409.CDR

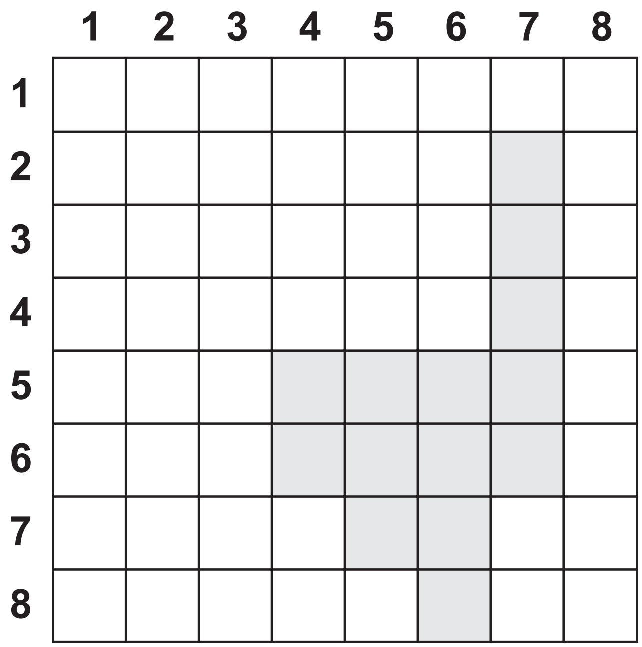




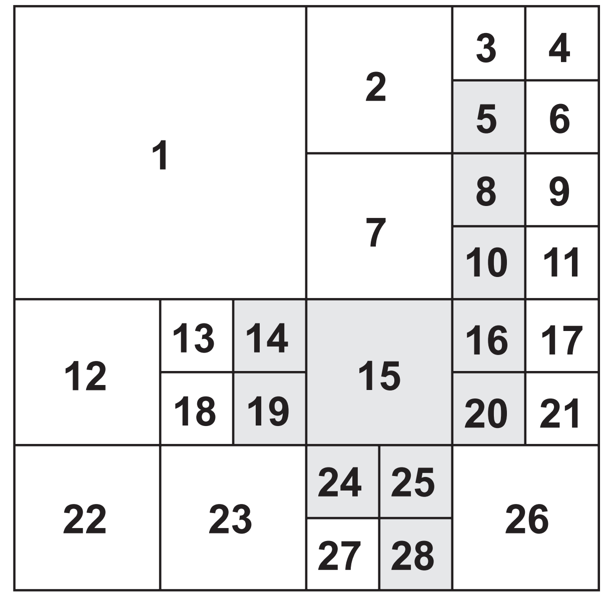
Grid Styles

FIGURE

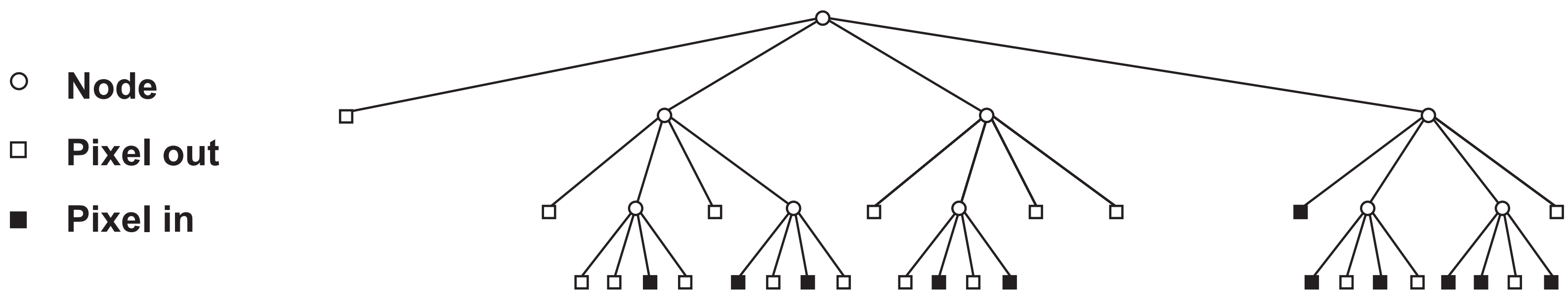
4-10



Simple region on a raster grid.



A quadtree representation of a simple region.



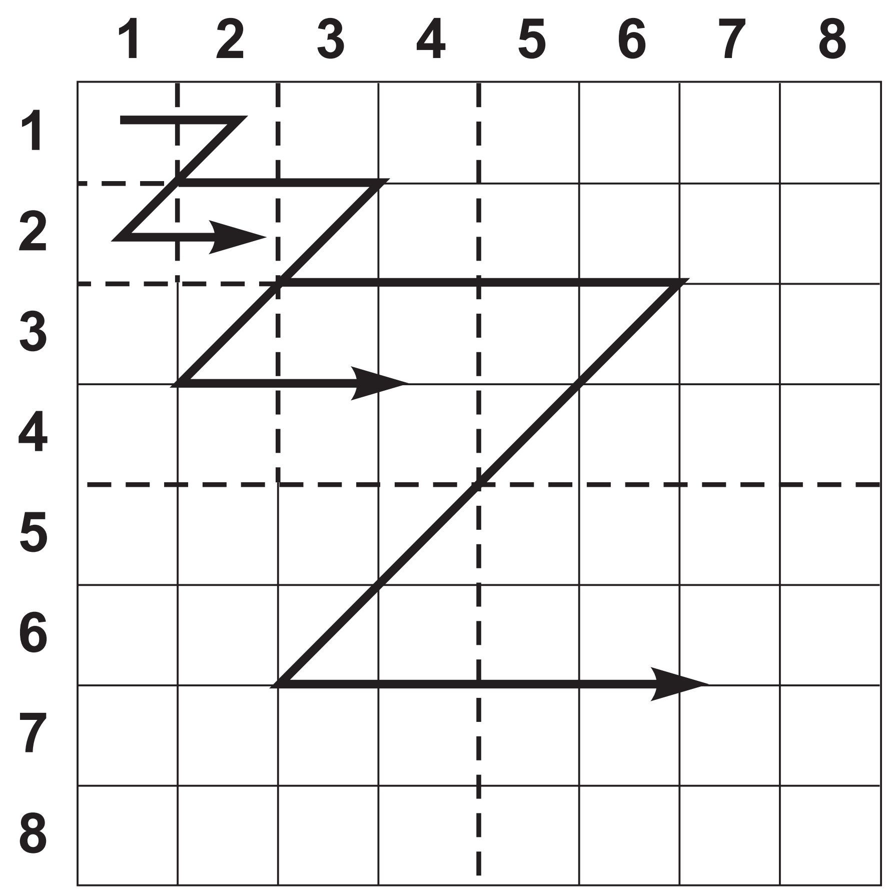
Quadtree structure of a simple region.

Quadtree

FIGURE

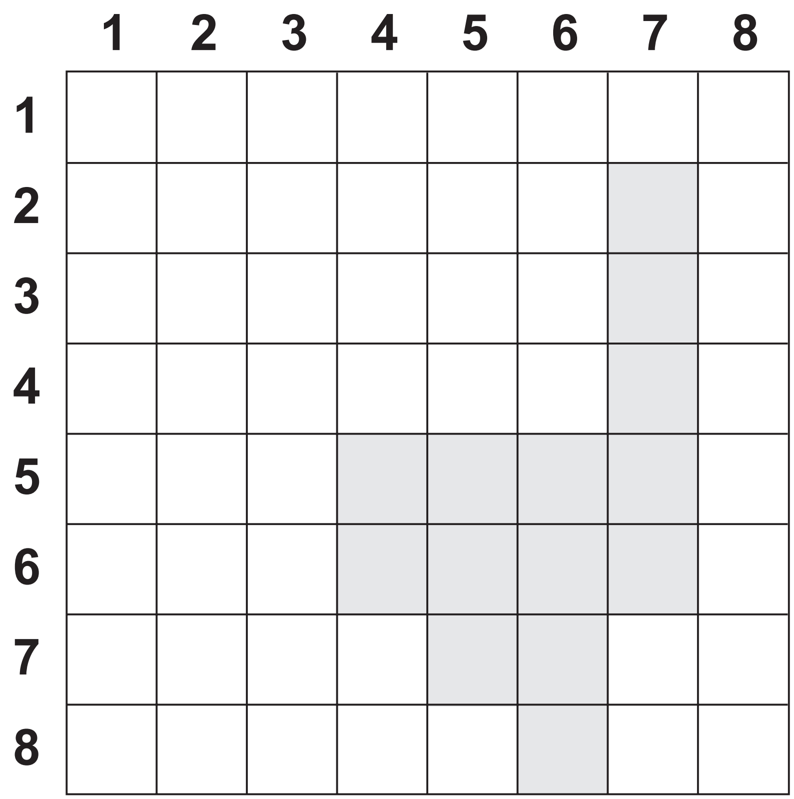
4-11

NSFGIS\TOPIC4\OH0411.CDR



Z Search Pattern of a Quadtree

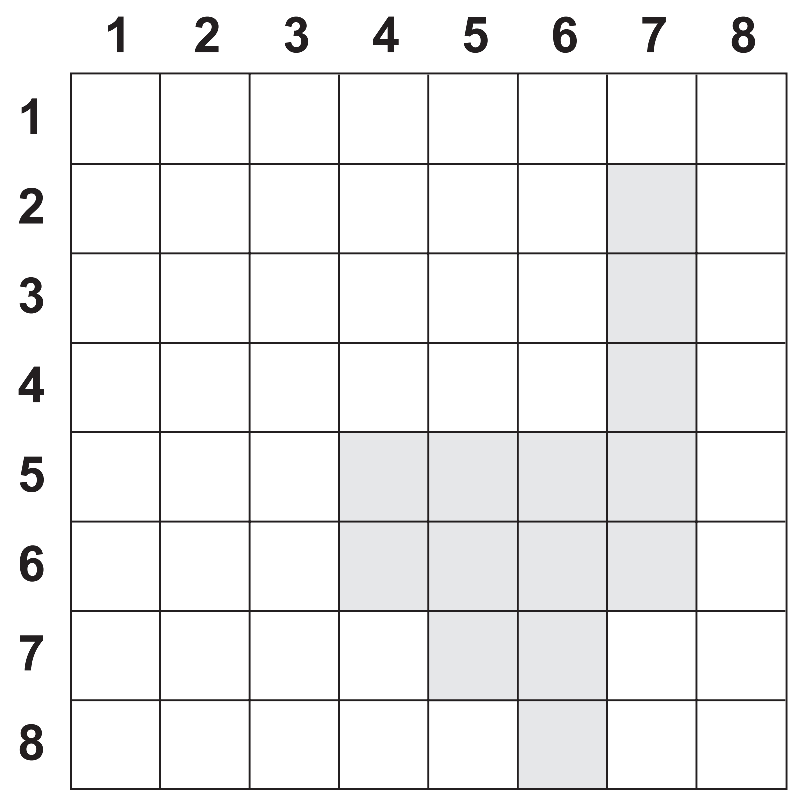
FIGURE 4-12



$3^5, 2, 3^2, 2, 1, 2, 1, 2, 1^2, 0^3, 1^3, 0$

Simple Region with Chain Codes

FIGURE 4-13

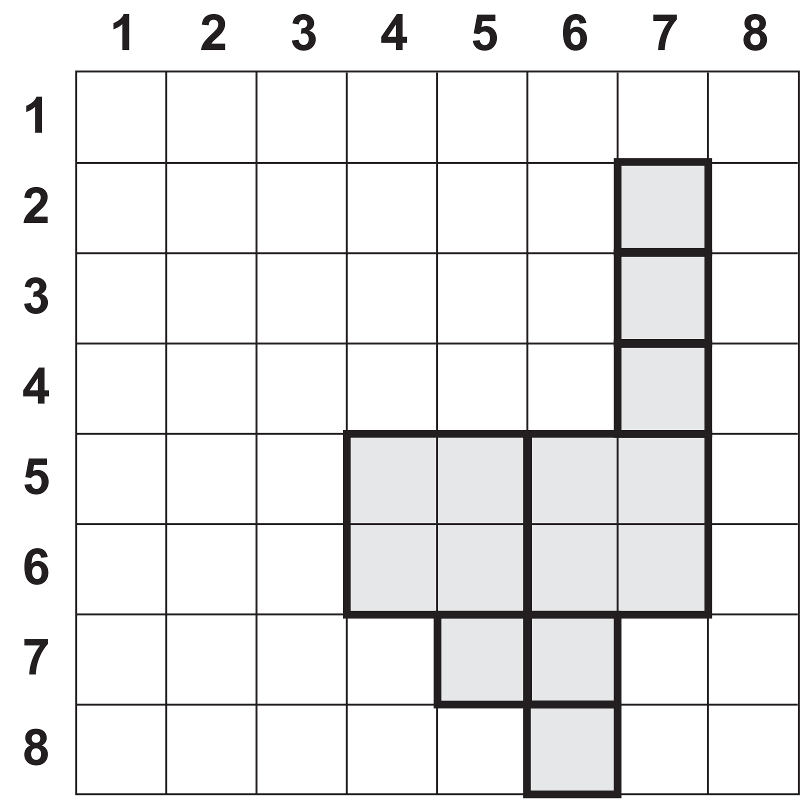


Row 2: 7,7
Row 3: 7,7
Row 4: 7,7
Row 5: 4,7
Row 6: 4,7
Row 7: 5,6
Row 8: 6,6

Simple Region with Run-Length Codes

FIGURE

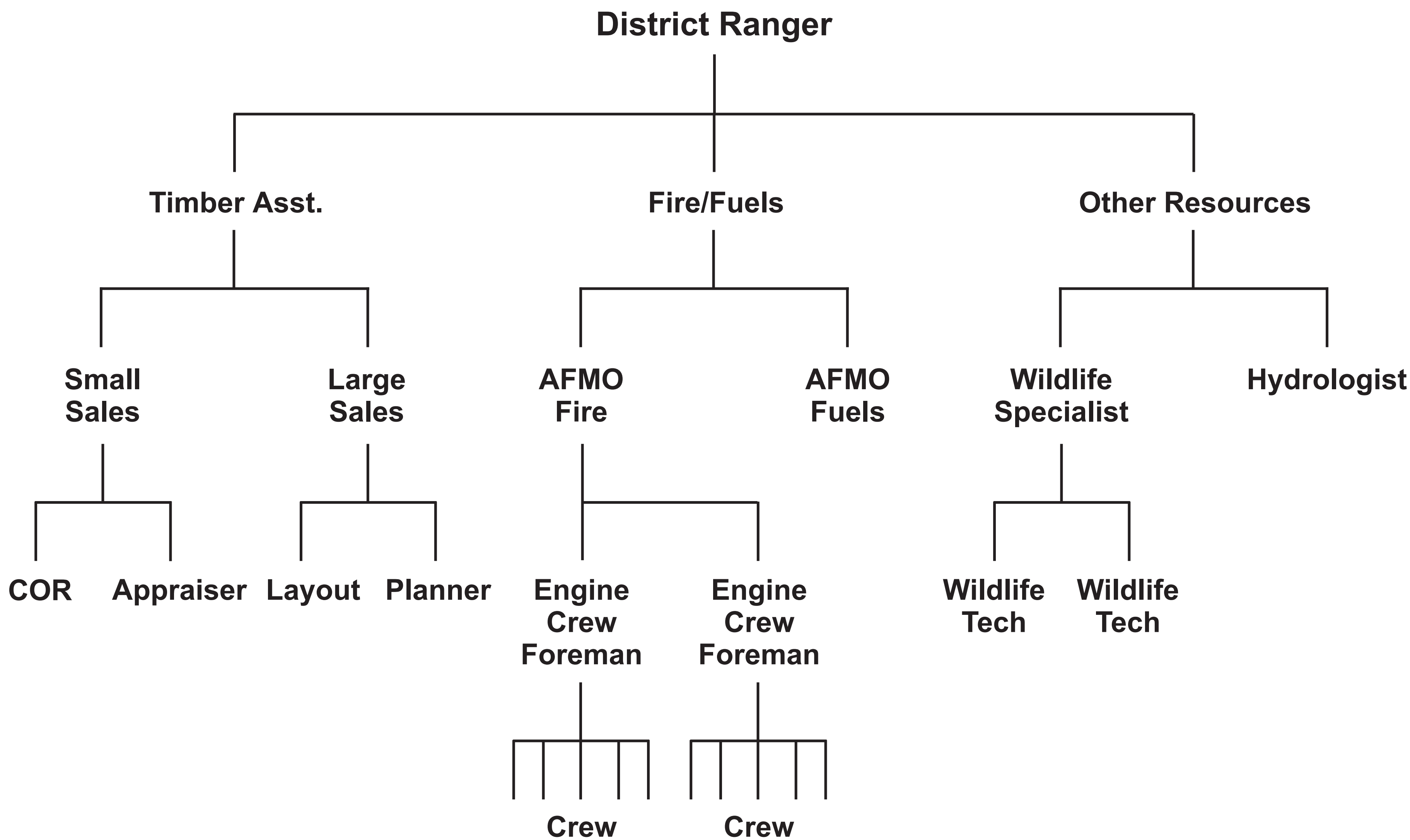
4-14



6, 1-squares + 2, 4-squares

Simple Region in Block Code

FIGURE 4-15



Hierarchical Structure

FIGURE 4-16

NSFGIS\TOPIC4\OH0416.CDR

Land Unit Layer - Data File

ID	*Area*	*Veg.*	*Soil*
A	56	RH	1
B	78	RP	3
C	112	RH	4
D	42	ASP	2

Soils Database

Soil	*S.I.*	*% O.M.*
1	79	1.2
2	102	6.4
3	56	4.1
4	84	0.7

Linking Databases

FIGURE 4-17

Example: a shopping list

- eggs
- buns
- milk
- cheese
- sugar
- apples
- taco shells

Simple List

FIGURE

4-18

Soils_id	*Area*	*Perimeter*	*Drainage*	*Soiltype*
1	2.447072e+009	323524.400	Fair	Sand/Loam
2	3.626703e+008	82463.410	Good	Gravel
3	2.279584+009	404514.900	Fair	Loam
4	2.498233e+003	278962.800	Good	Sand
5	8.018180e+008	137461.500	Poor	Clay
6	2.370808e+009	211413.900	Good	Sand
7	1.494432e+009	224793.500	Poor	Clay
8	4.003240e+009	327335.800	Fair	Sand/Loam

Ordered List

FIGURE 4-19

NSFGIS\TOPIC4\OH0419.CDR

Soiltype	Soils_id
Clay	5
Clay	7
Gravel	2
Loam	3
Sand	4
Sand	6
Sand/Loam	1
Sand/Loam	8

Soils_id	Area	Perimeter	Drainage	Soiltype
1	2.447072e+009	323524.400	Fair	Sand/Loam
2	3.626703e+008	82463.410	Good	Gravel
3	2.279584+009	404514.900	Fair	Loam
4	2.498233e+003	278962.800	Good	Sand
5	8.018180e+008	137461.500	Poor	Clay
6	2.370808e+009	211413.900	Good	Sand
7	1.494432e+009	224793.500	Poor	Clay
8	4.003240e+009	327335.800	Fair	Sand/Loam

Indexed List

FIGURE 4-20

NSFGIS\TOPIC4\OH0420.CDR

Objectives

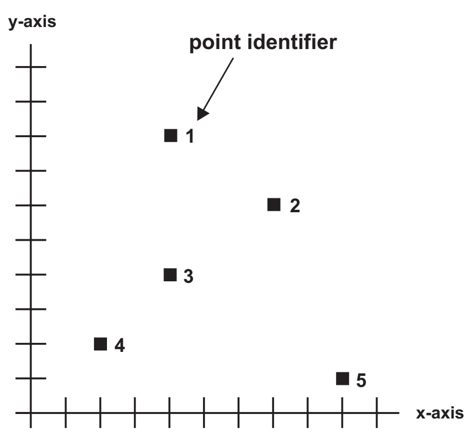
A. Become familiar with the concept of modeling reality in a Geographic Information System.

B. Understand various options for data models underlying GIS.

C. Understand how data structures contribute to GIS functionality.

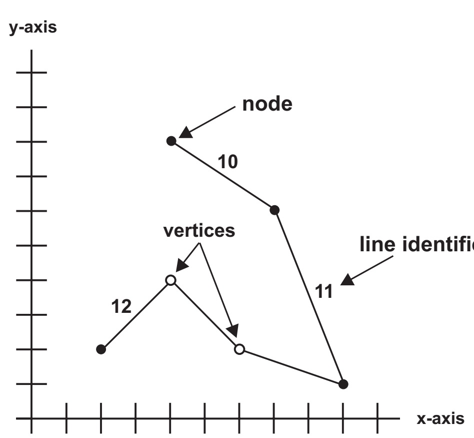
D. Understand how file structures may ultimately affect GIS software efficiency.

FIGURE 4-1



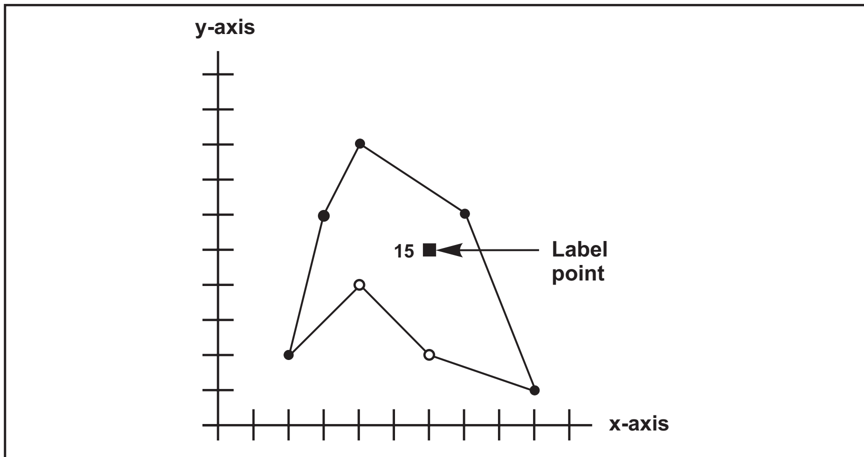
Points

FIGURE 4-2



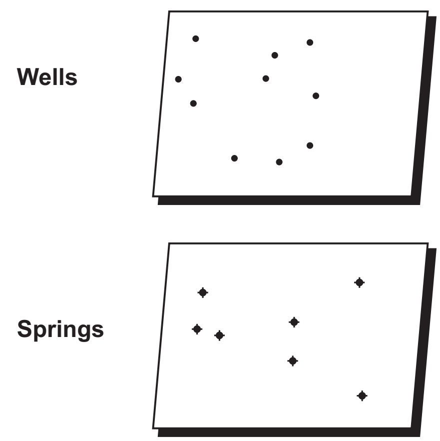
Line

FIGURE 4-3



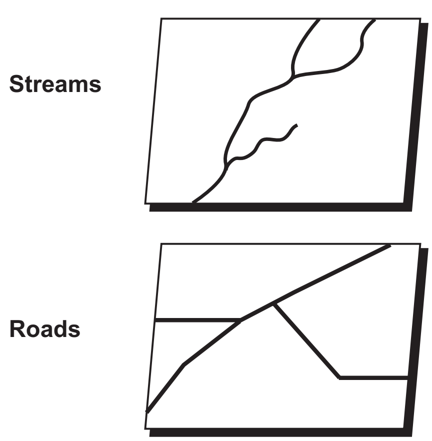
Polygon

FIGURE 4-4



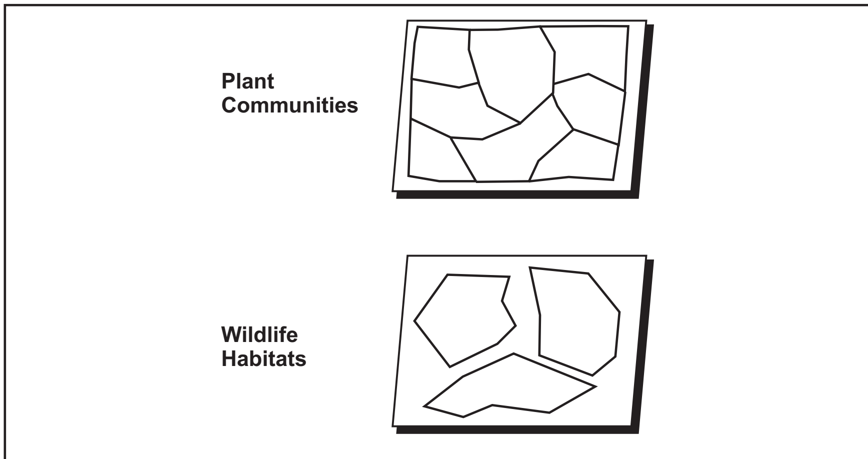
Point Features

FIGURE 4-5

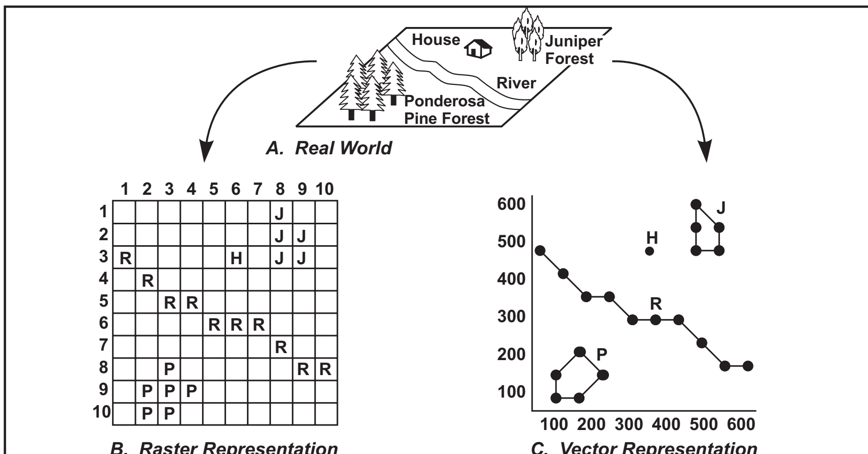


Line Features

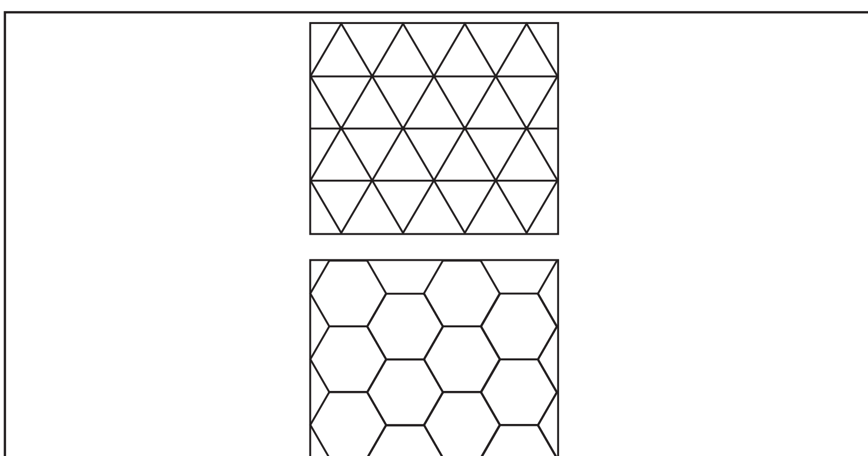
FIGURE 4-6



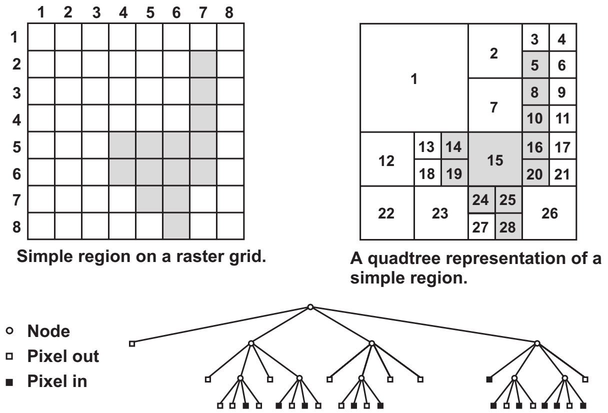
Polygon Features — FIGURE 4-7



Raster and Vector Data Models — FIGURE 4-9



Grid Styles — FIGURE 4-10

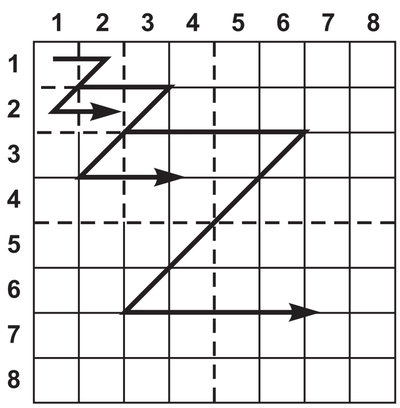


Simple region on a raster grid.

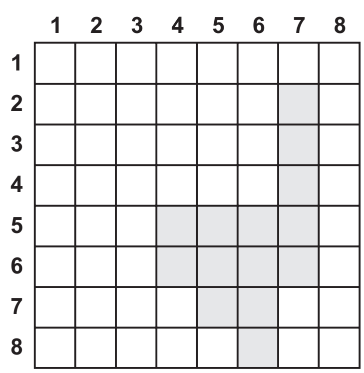
A quadtree representation of a simple region.

Quadtree structure of a simple region.

Quadtree — FIGURE 4-11

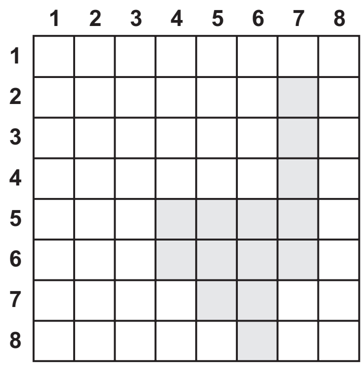


Z Search Pattern of a Quadtree — FIGURE 4-12



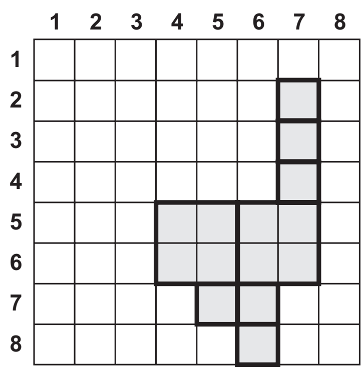
$3^5, 2, 3^2, 2, 1, 2, 1, 2, 1^2, 0^3, 1^3, 0$

Simple Region with Chain Codes — FIGURE 4-13



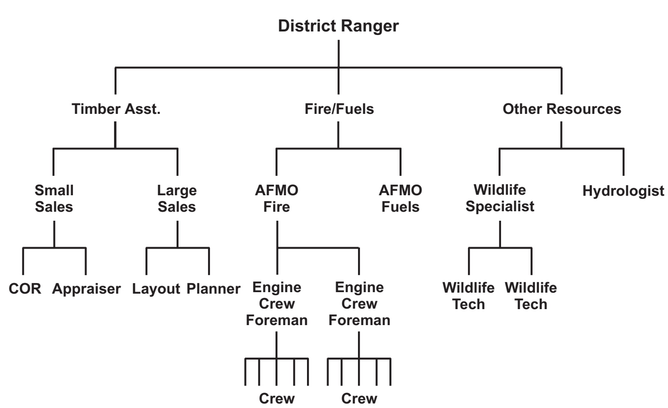
Row 2: 7,7
Row 3: 7,7
Row 4: 7,7
Row 5: 4,7
Row 6: 4,7
Row 7: 5,6
Row 8: 6,6

Simple Region with Run-Length Codes — FIGURE 4-14



6, 1-squares + 2, 4-squares

Simple Region in Block Code — FIGURE 4-15



Hierarchical Structure — FIGURE 4-16

Land Unit Layer - Data File

ID	*Area*	*Veg.*	*Soil*
A	56	RH	1
B	78	RP	3
C	112	RH	4
D	42	ASP	2

Soils Database

Soil	*S.I.*	*% O.M.*
1	79	1.2
2	102	6.4
3	56	4.1
4	84	0.7

Linking Databases

FIGURE 4-17

Example: a shopping list

- eggs
- buns
- milk
- cheese
- sugar
- apples
- taco shells

Simple List

FIGURE 4-18

Soils_id	*Area*	*Perimeter*	*Drainage*	*Soiltype*
1	2.447072e+009	323524.400	Fair	Sand/Loam
2	3.626703e+008	82463.410	Good	Gravel
3	2.279584+009	404514.900	Fair	Loam
4	2.498233e+003	278962.800	Good	Sand
5	8.018180e+008	137461.500	Poor	Clay
6	2.370808e+009	211413.900	Good	Sand
7	1.494432e+009	224793.500	Poor	Clay
8	4.003240e+009	327335.800	Fair	Sand/Loam

Ordered List

FIGURE 4-19

Soiltype	Soils_id
Clay	5
Clay	7
Gravel	2
Loam	3
Sand	4
Sand	6
Sand/Loam	1
Sand/Loam	8

Soils_id	Area	Perimeter	Drainage	Soiltype
1	2.447072e+009	323524.400	Fair	Sand/Loam
2	3.626703e+008	82463.410	Good	Gravel
3	2.279584+009	404514.900	Fair	Loam
4	2.498233e+003	278962.800	Good	Sand
5	8.018180e+008	137461.500	Poor	Clay
6	2.370808e+009	211413.900	Good	Sand
7	1.494432e+009	224793.500	Poor	Clay
8	4.003240e+009	327335.800	Fair	Sand/Loam

Indexed List

FIGURE 4-20

Objectives

A. Be able to discuss the raster data model.

B. Be able to discuss the vector data model.

C. Understand some of the various vector data models that have been used over time to store polygons.

D. Understand the differences between raster and vector data models.

E. Understand the situations when one of the models might be better to use.

FIGURE

5-1

We now use computers because they are:

1. Fast

2. Accurate

3. Cheap

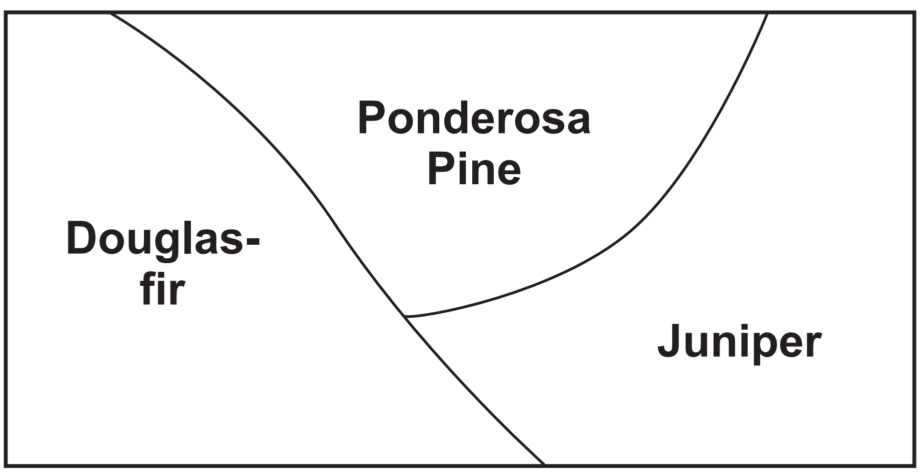
4. Flexible

FIGURE

5-2

D	D	P	P	P	P	P	P	J	J
D	D	D	P	P	P	P	P	J	J
D	D	D	D	P	P	P	J	J	J
D	D	D	D	D	J	J	J	J	J
D	D	D	D	D	D	J	J	J	J

Raster

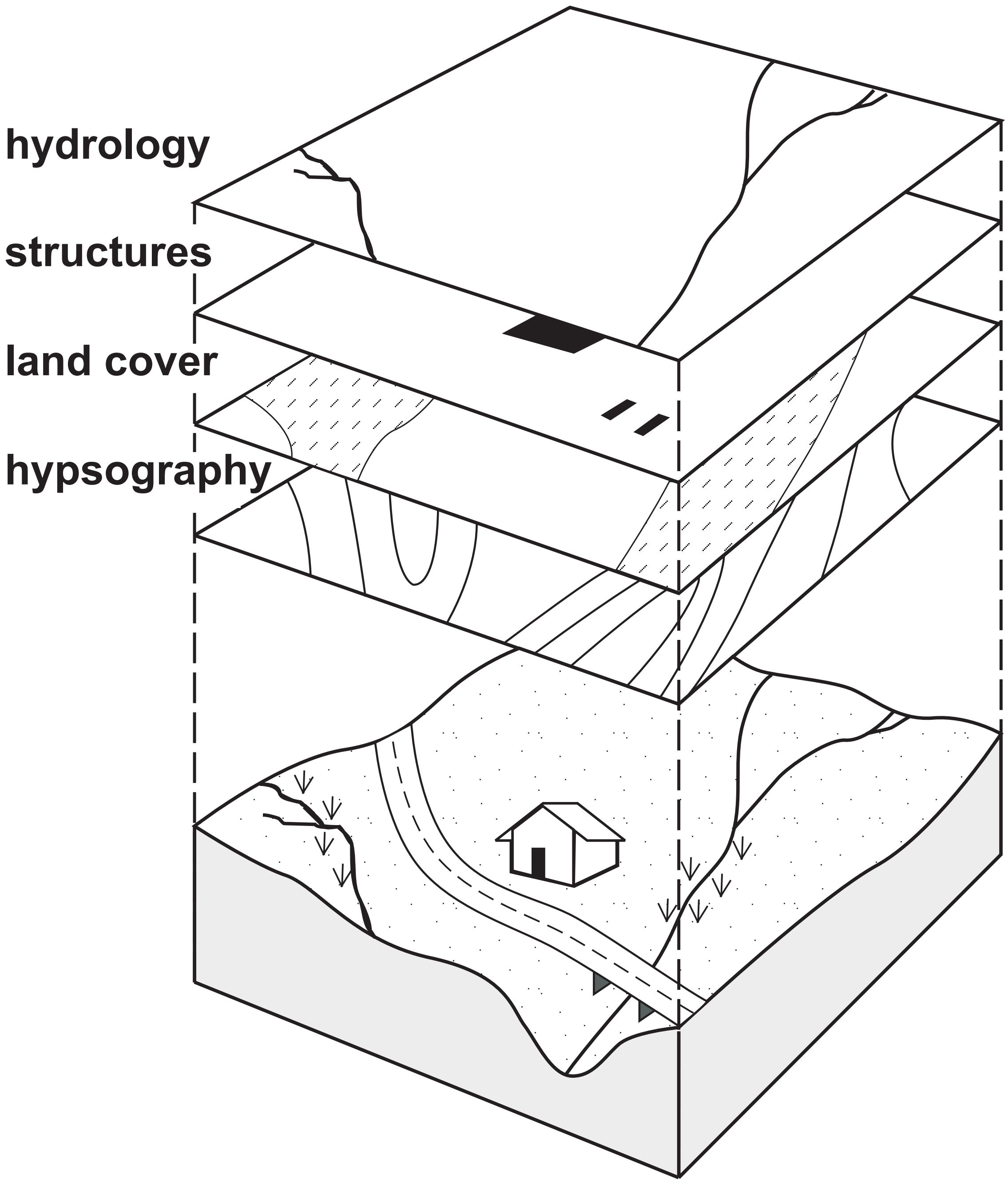


Vector

Major GIS Data Models

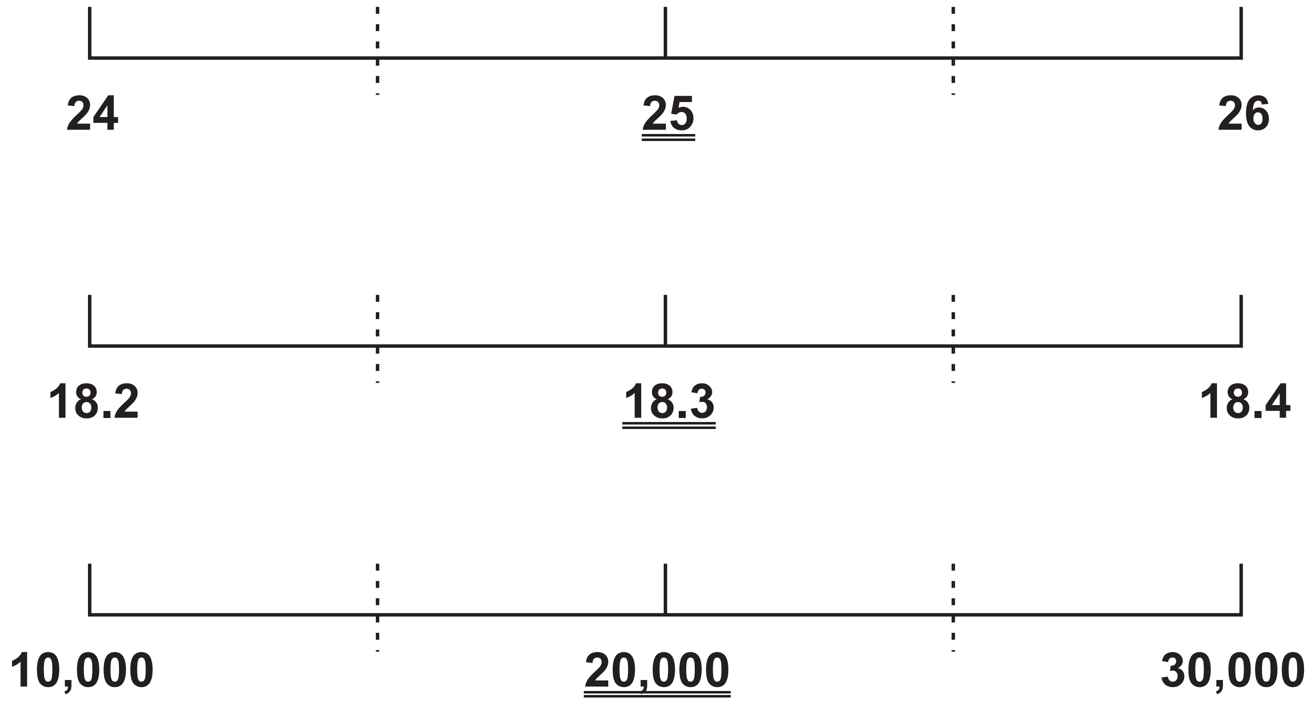
FIGURE

5-3



Vector Data Model

FIGURE
5-4



Number	Accurate to	Precise to	Uncertainty
25	2 digits	unit	± .5
18.3	3 digits	tenths	± .05
20,000	1 digit	ten thousands	± 5,000

Precision and Accuracy

FIGURE 5-5

NSFGIS\TOPIC5\OH0505.CDR

Types of files that have been used over time for storing polygons:

Location List

Polygon A, points:

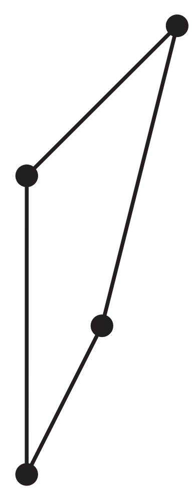
5 , 10

6 , 12

7 , 16

5 , 14

5 , 10



FIGURE

5-6

Types of files that have been used over time for storing polygons:

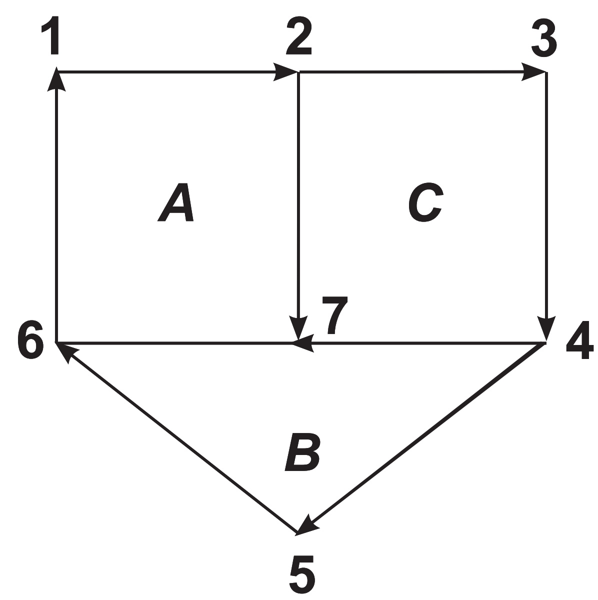
Points Dictionary

point #	x-coordinate	y-coordinate
1	23560	13254
2	25551	15173
3	15786	16552

polygon	points
A	1,2,5,7,11,16,1
B	1,9,10,11,12,13,14,4,3,2,1

FIGURE

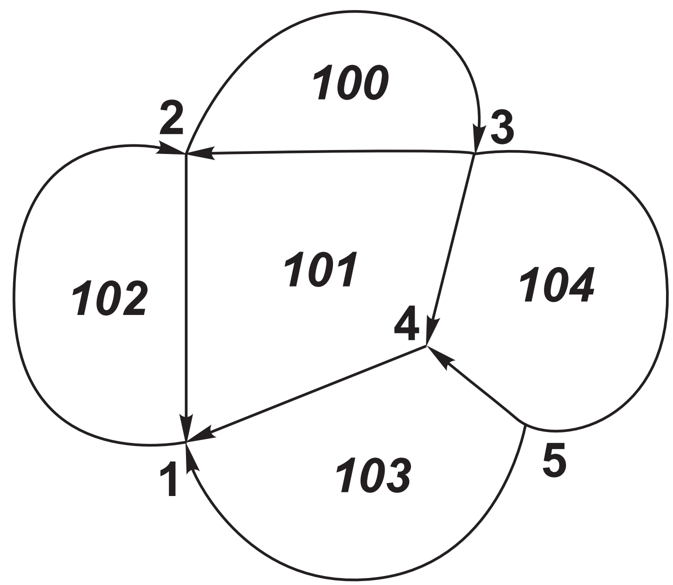
5-7



from node	to node	L poly	R poly	x from	y from	x to	y to
1	2	0	A				
2	7	C	A				
2	3	0	C				
3	4	0	C				

Topological Structure, DIME File

FIGURE 5-8

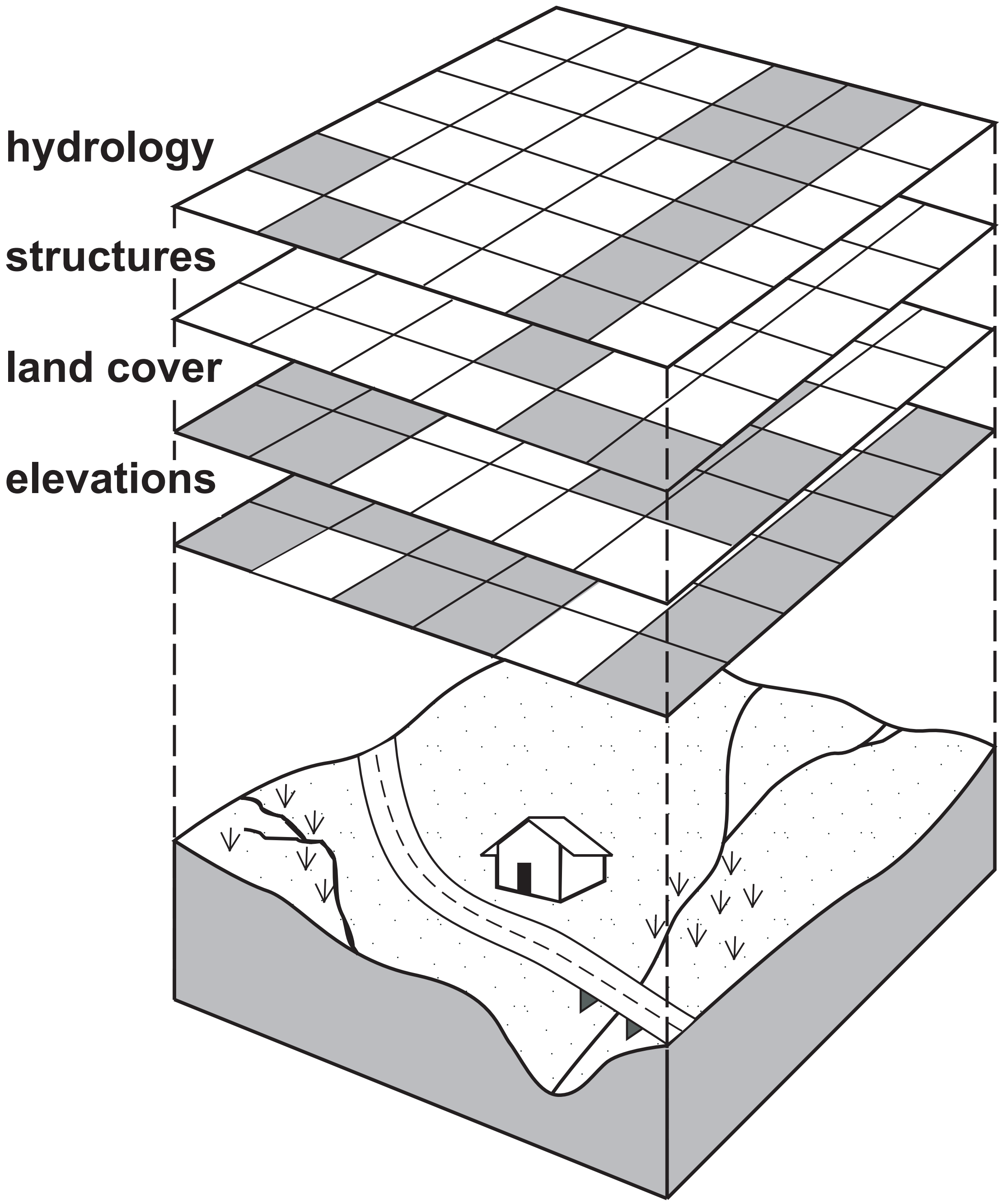


from node	to node	L poly	R poly	# points	coordinates
1	2	0	102	14	X_0, Y_0 — X_{13}, Y_{13}
2	1	101	102	6	X_0, Y_0 — X_5, Y_5
5	1	103	0	27	X_0, Y_0 — X_{26}, Y_{26}
4	1	103	101	4	X_0, Y_0 — X_3, Y_3

Topological Structure, Chain File

FIGURE 5-9

NSFGIS\TOPIC5\OH0509.CDR



FIGURE

Raster Data Model

5-10

Vector	Raster
1. low storage requirement	1. high storage requirement
2. store by object	2. store by space
3. potential for better accuracy	3. neighborhood structure makes certain analysis more functional
4. aesthetic appeal	4. ease of map overlay
5. analysis may be slow and complex	5. ease of statistics on general spatial characteristics
6. some objects are vector by definition	6. appealing to remote sensing community
	7. fixed resolution
	8. normally a cheaper system

Raster / Vector Comparison

FIGURE 5-11

Objectives

A. Be able to discuss the raster data model.

B. Be able to discuss the vector data model.

C. Understand some of the various vector data models that have been used over time to store polygons.

D. Understand the differences between raster and vector data models.

E. Understand the situations when one of the models might be better to use.

FIGURE 5-1

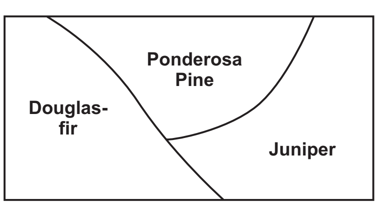
We now use computers because they are:

1. Fast
2. Accurate
3. Cheap
4. Flexible

FIGURE 5-2

D	D	P	P	P	P	P	P	J	J
D	D	D	P	P	P	P	P	J	J
D	D	D	D	P	P	P	J	J	J
D	D	D	D	D	J	J	J	J	J
D	D	D	D	D	D	J	J	J	J

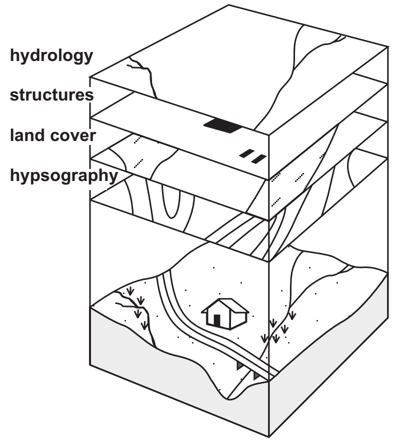
Raster



Vector

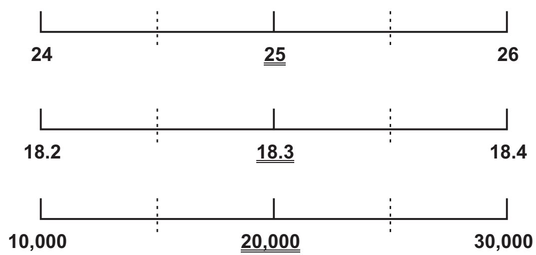
Major GIS Data Models

FIGURE 5-3



Vector Data Model

FIGURE 5-4



Number	Accurate to	Precise to	Uncertainty
25	2 digits	unit	± .5
18.3	3 digits	tenths	± .05
20,000	1 digit	ten thousands	± 5,000

Precision and Accuracy

FIGURE 5-5

Location List

Polygon A, points:

5 , 10
6 , 12
7 , 16
5 , 14
5 , 10

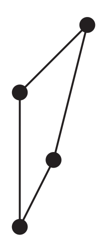


FIGURE 5-6

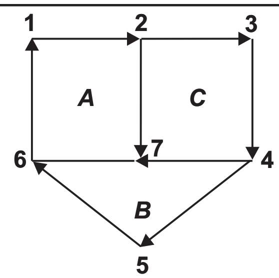
Types of files that have been used over time for storing polygons:

Points Dictionary

point #	x-coordinate	y-coordinate
1	23560	13254
2	25551	15173
3	15786	16552

polygon	points
A	1,2,5,7,11,16,1
B	1,9,10,11,12,13,14,4,3,2,1

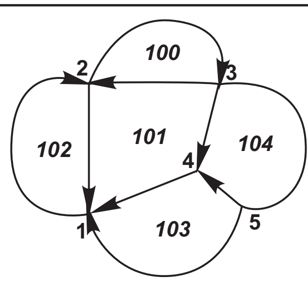
FIGURE 5-7



from node	to node	L poly	R poly	x from	y from	x to	y to
1	2	0	A				
2	7	C	A				
2	3	0	C				
3	4	0	C				

Topological Structure, DIME File

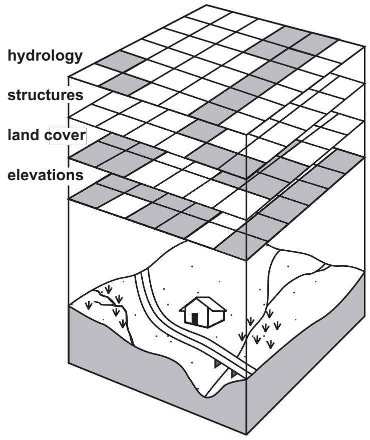
FIGURE 5-8



from node	to node	L poly	R poly	# points	coordinates
1	2	0	102	14	X_0, Y_0 — X_{13}, Y_{13}
2	1	101	102	6	X_0, Y_0 — X_5, Y_5
5	1	103	0	27	X_0, Y_0 — X_{26}, Y_{26}
4	1	103	101	4	X_0, Y_0 — X_3, Y_3

Topological Structure, Chain File

FIGURE 5-9



Raster Data Model — FIGURE 5-10

Vector	Raster
1. low storage requirement	1. high storage requirement
2. store by object	2. store by space
3. potential for better accuracy	3. neighborhood structure makes certain analysis more functional
4. aesthetic appeal	4. ease of map overlay
5. analysis may be slow and complex	5. ease of statistics on general spatial characteristics
6. some objects are vector by definition	6. appealing to remote sensing community
	7. fixed resolution
	8. normally a cheaper system

Raster/Vector Comparison — FIGURE 5-11

Objectives

A. Understand the basic components and analysis associated with Networks.

B. Become familiar with the concept of "fuzzy" logic and its implication on GIS.

FIGURE

6-1

Components of a Network

A. Nodes

B. Links

C. Turns

D. Attributes

1. Impedance
2. Stops
3. Centers
4. Resource demand
5. Routes
6. Barriers

FIGURE

6-2

Analysis in Networks

1. Optimal Routing

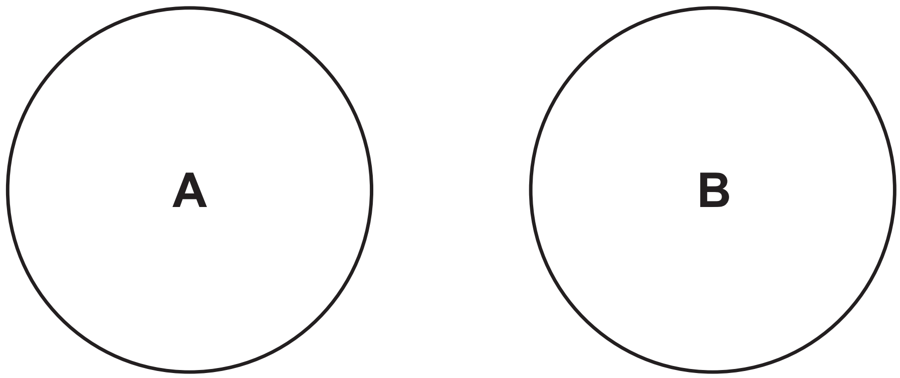
2. Allocation

3. Tracing

FIGURE

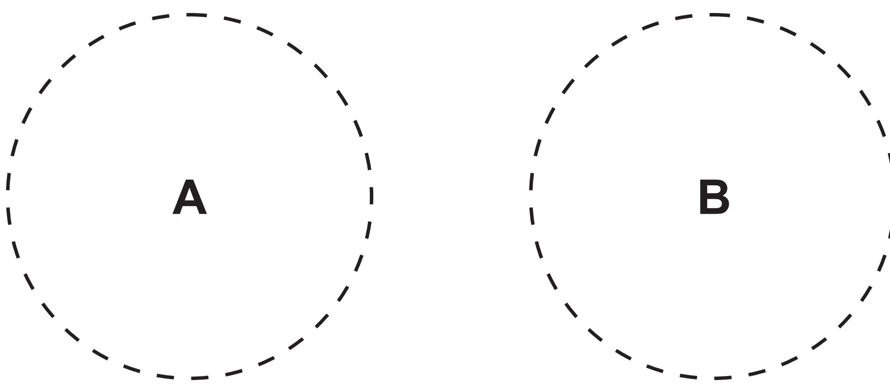
6-3

Traditional Set Theory



Mutually Exclusive Boundaries

Fuzzy Set Theory



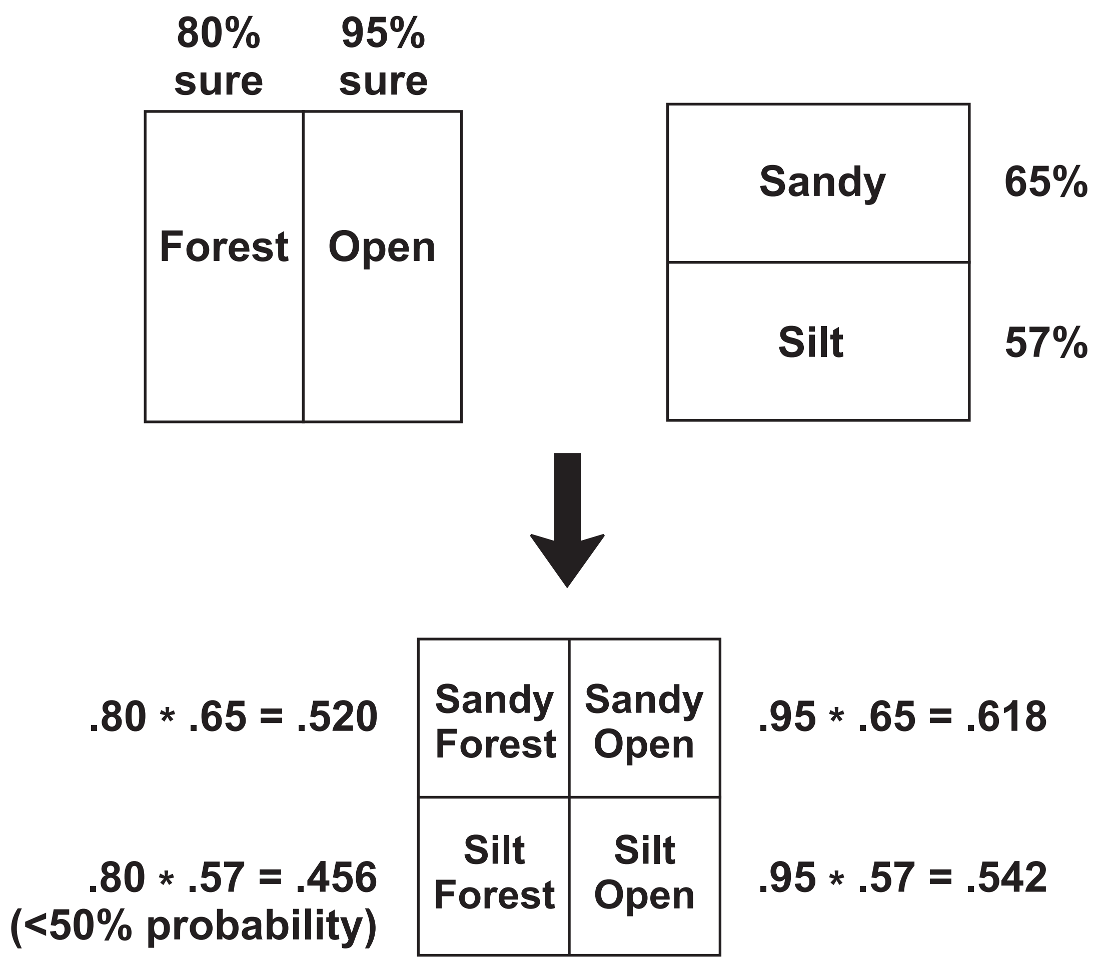
Non-absolute Boundaries

Fuzzy Logic

FIGURE

6-4

Example: Claiming independence of categories



Error Propagation

FIGURE 6-5

NSFGIS\TOPIC6\OH0605.CDR

Example:
Interval / ratio data in overlay process

A. Addition

	$\boxed{50}$	+	$\boxed{65}$	=	$\boxed{115}$
absolute error	± 5		± 5		± 10
relative error	10%		7.6%		8%

B. Subtraction

	$\boxed{65}$	-	$\boxed{50}$	=	$\boxed{15}$
absolute error	± 5		± 5		± 10
relative error	7.6%		10%		67%

C. Multiplication

	$\boxed{65}$	*	$\boxed{50}$	=	$\boxed{3250}$
absolute error	± 5		± 5		+600 -550
relative error	7.6%		10%		+18.5% -16.9%

FIGURE

6-6

Objectives

A. Understand the basic components and analysis associated with Networks.

B. Become familiar with the concept of "fuzzy" logic and its implication on GIS.

FIGURE 6-1

Components of a Network

A. Nodes

B. Links

C. Turns

D. Attributes

1. Impedance
2. Stops
3. Centers
4. Resource demand
5. Routes
6. Barriers

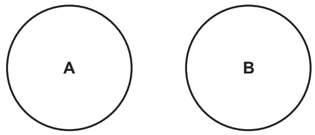
FIGURE 6-2

Analysis in Networks

1. Optimal Routing
2. Allocation
3. Tracing

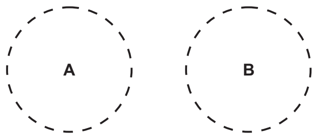
FIGURE 6-3

Traditional Set Theory



Mutually Exclusive Boundaries

Fuzzy Set Theory

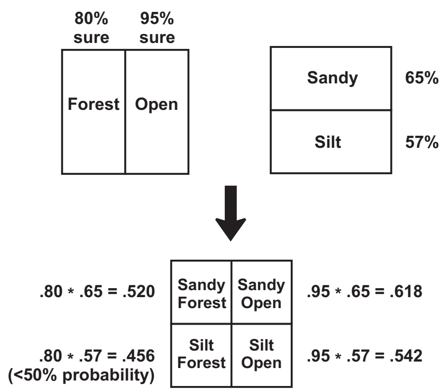


Non-absolute Boundaries

Fuzzy Logic

FIGURE 6-4

Example: Claiming independence of categories



Error Propagation

FIGURE 6-5

Example: Interval / ratio data in overlay process

A. Addition

	50	+	65	=	115
absolute error	± 5		± 5		± 10
relative error	10%		7.6%		8%

B. Subtraction

	65	-	50	=	15
absolute error	± 5		± 5		± 10
relative error	7.6%		10%		67%

C. Multiplication

	65	*	50	=	3250
absolute error	± 5		± 5		+600 -550
relative error	7.6%		10%		+18.5% -16.9%

FIGURE 6-6

Objectives

A. Understand classifications of data.

B. Identify data measurement levels.

C. Understand types of spatial variation.

D. List and discuss error and accuracy issues.

FIGURE

7-1

We usually observe conditions, or properties of features one of three ways:

1. **spatial**

2. **thematic**

3. **temporal**

FIGURE

7-2

Spatial Variation

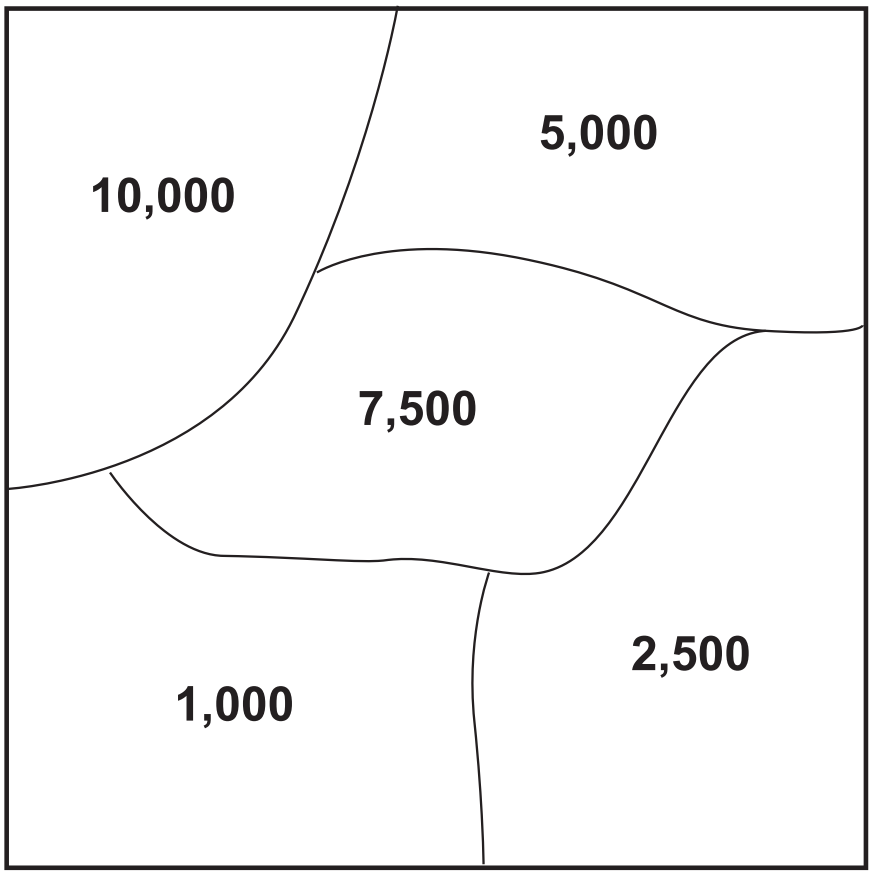
1. <u>Discrete</u> - homogeneity within an area, point, or along a line

2. <u>Continuous</u> - variation is continuous in all directions

FIGURE

7-3

Bd. Ft. Volume per Acre

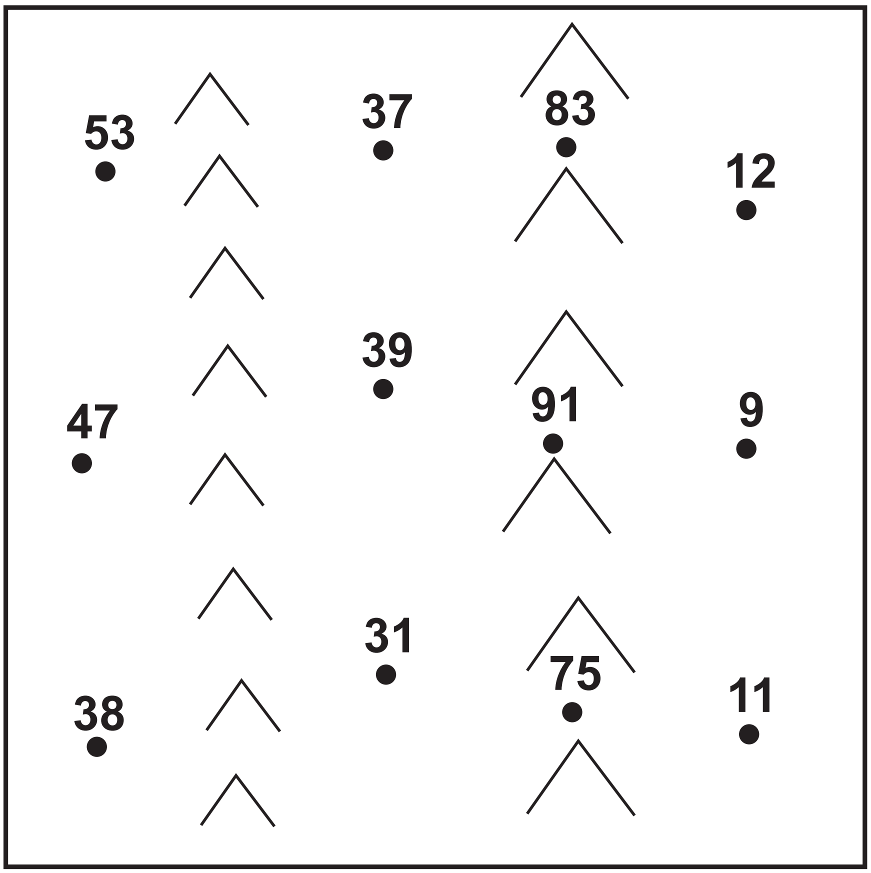


Discrete Variation

FIGURE

7-4

Ave. Rainfall per Year (inches)



Continuous Variation

FIGURE

7-5

Dimensions of Data

1. Points

2. Lines

3. Areas

4. Volumes

FIGURE

7-6

Measurement Levels

1. Nominal
2. Ordinal
3. Interval
4. Ratio

FIGURE

7-7

Reasons for Error with Cartographic Generalization

1. Selection (compilation)

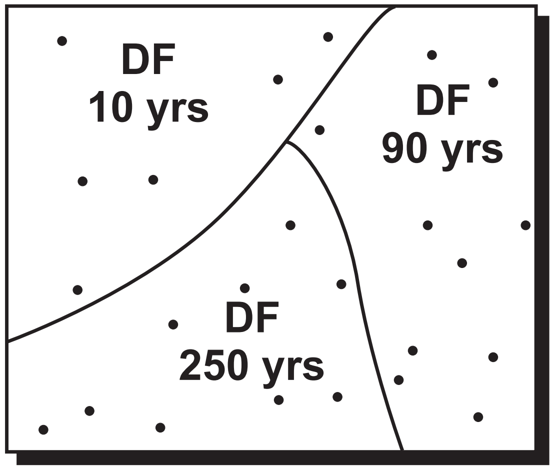
2. Simplification

3. Classification

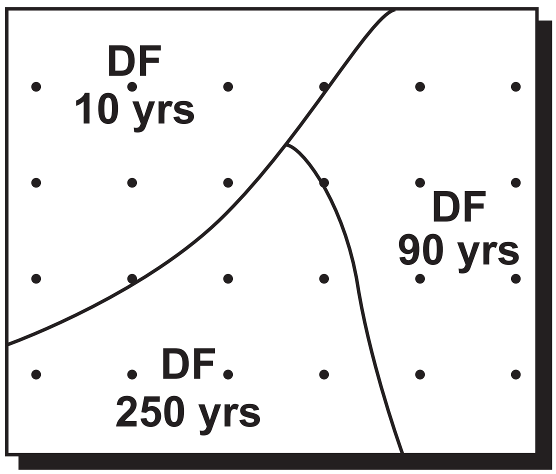
4. Symbolization

FIGURE

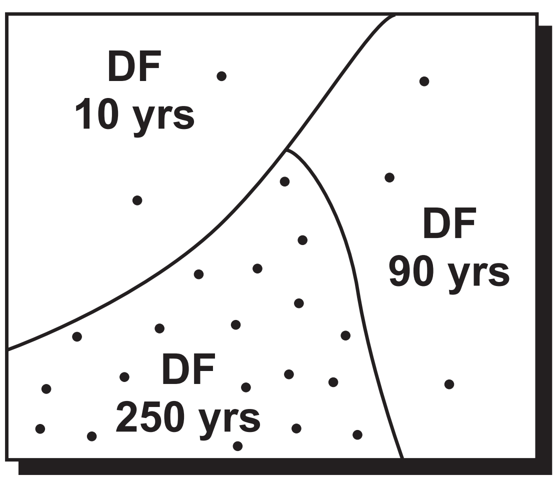
7-8



Random



Systematic



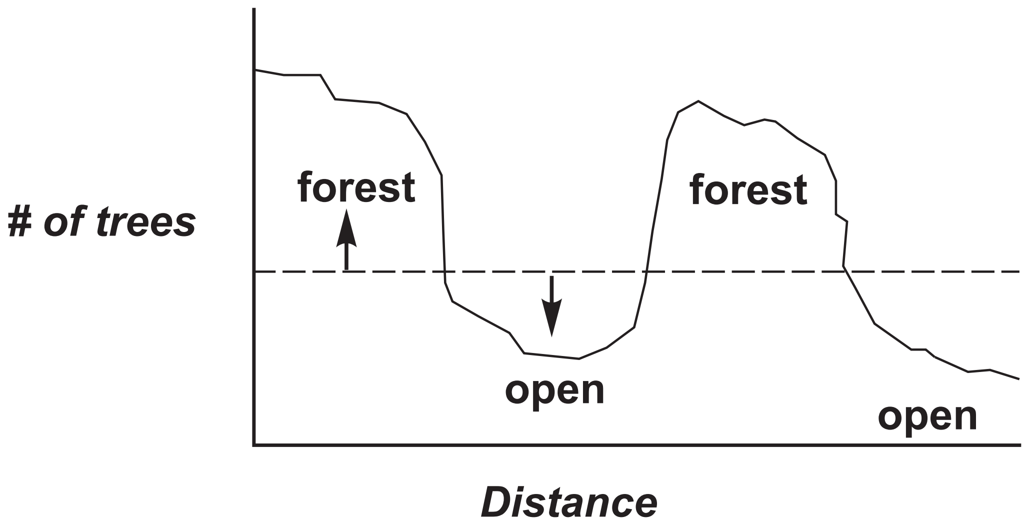
Stratified

Sampling Strategies

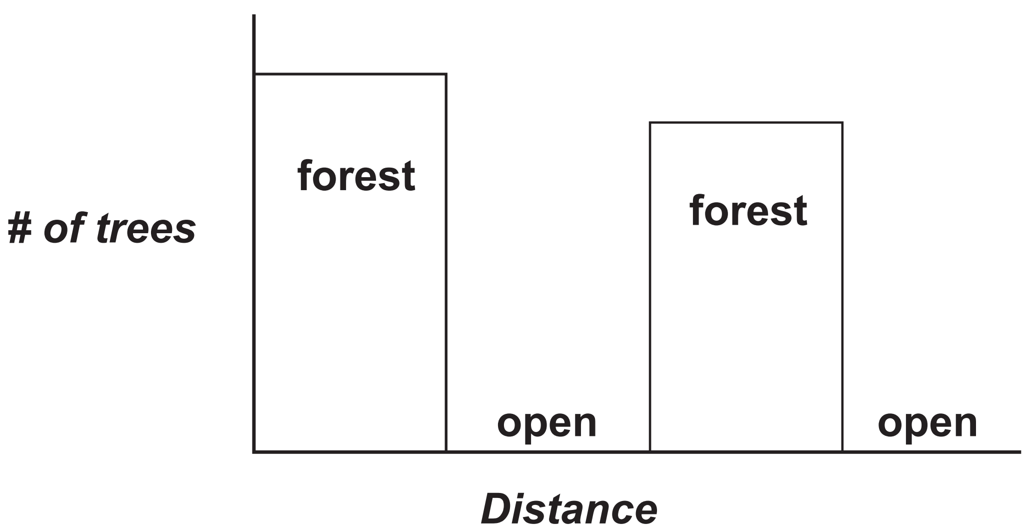
FIGURE

7-9

Transitional (continuous)



Abrupt (discrete)



Boundary Errors

FIGURE

7-10

Objectives

A. Understand classifications of data.

B. Identify data measurement levels.

C. Understand types of spatial variation.

D. List and discuss error and accuracy issues.

FIGURE 7-1

We usually observe conditions, or properties of features one of three ways:

1. spatial
2. thematic
3. temporal

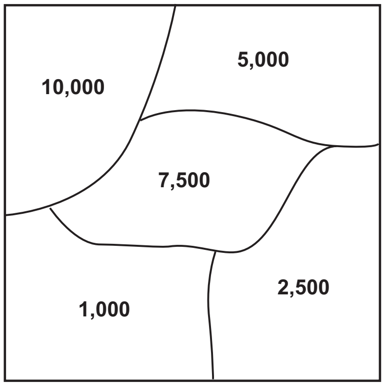
FIGURE 7-2

Spatial Variation

1. Discrete - homogeneity within an area, point, or along a line
2. Continuous - variation is continuous in all directions

FIGURE 7-3

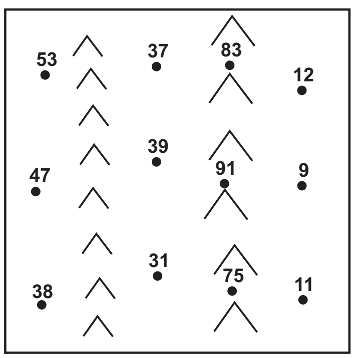
Bd. Ft. Volume per Acre



Discrete Variation

FIGURE 7-4

Ave. Rainfall per Year (inches)



Continuous Variation

FIGURE 7-5

Dimensions of Data

1. Points
2. Lines
3. Areas
4. Volumes

FIGURE 7-6

Measurement Levels

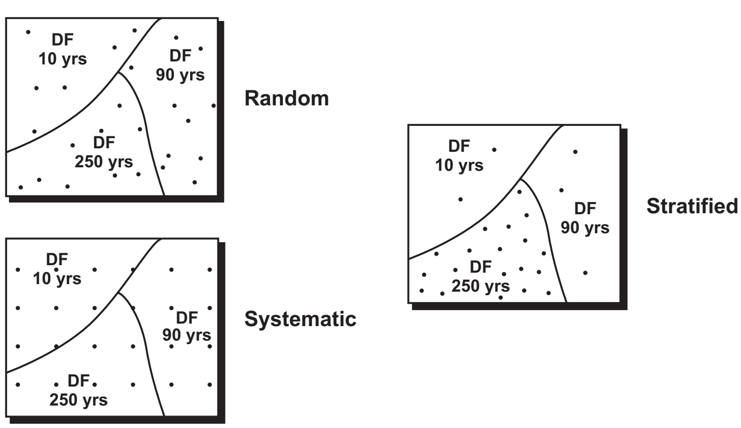
1. Nominal
2. Ordinal
3. Interval
4. Ratio

FIGURE 7-7

Reasons for Error with Cartographic Generalization

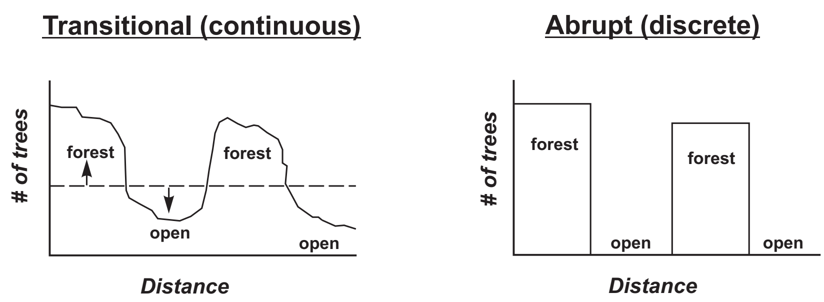
1. Selection (compilation)
2. Simplification
3. Classification
4. Symbolization

FIGURE 7-8



Sampling Strategies

FIGURE 7-9



Boundary Errors

FIGURE 7-10

Objectives

A. Be familiar with the various types of maps used to display information.

B. Understand the concept of Cartographic Modeling in GIS analysis.

C. Be able to name and explain the major analytical functions supported by GIS.

D. Understand Cartographic Neighborhoods in Raster GIS.

E. Be able to explain various examples of Raster analysis of neighborhoods.

FIGURE

8-1

Types of Maps

A. Reference Maps

B. Thematic maps

FIGURE

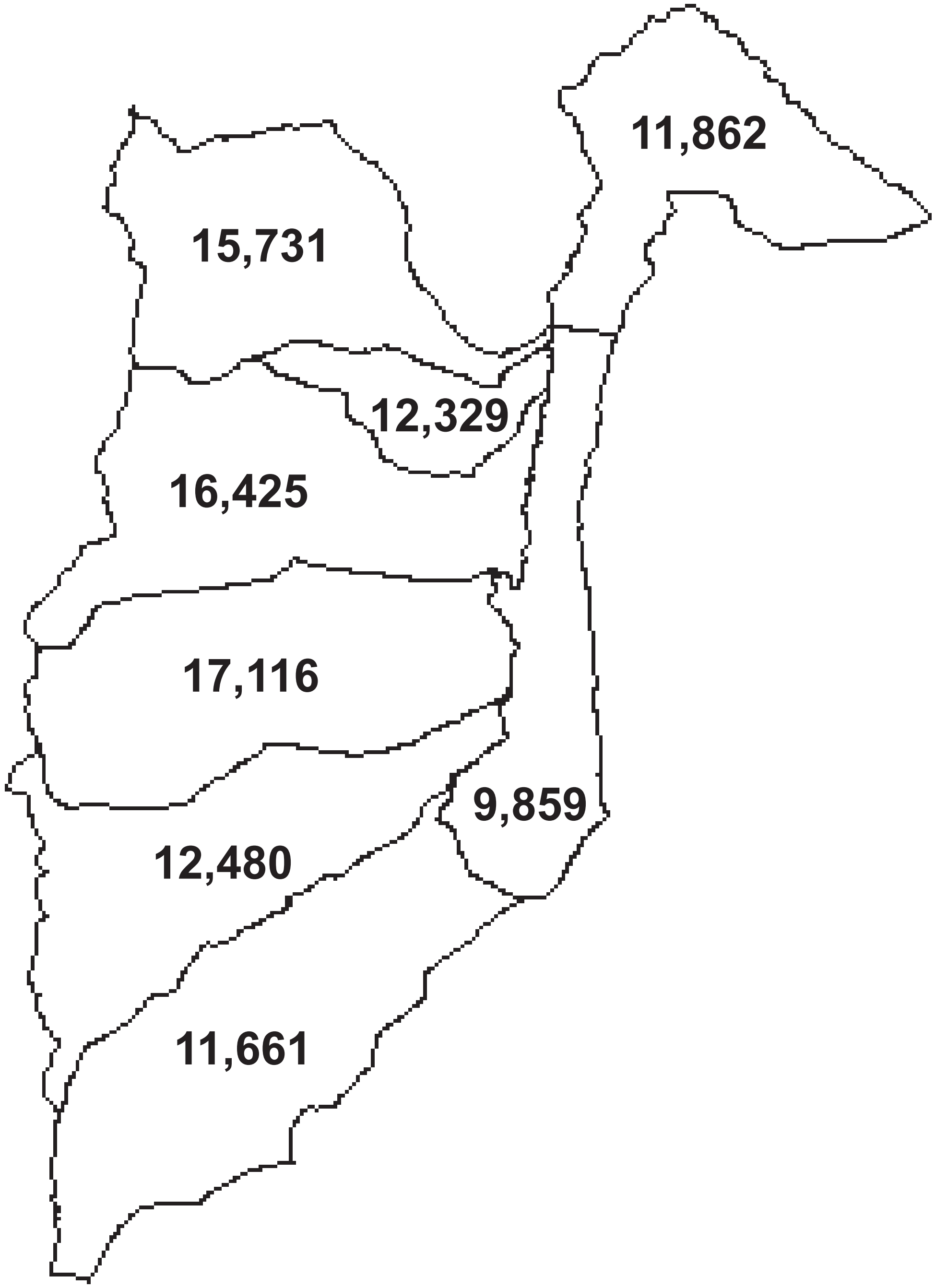
8-2

Types of Thematic Maps

1. choropleth map
2. isarithmic map
3. proximal map
4. trend surface map
5. residuals map

FIGURE

8-3

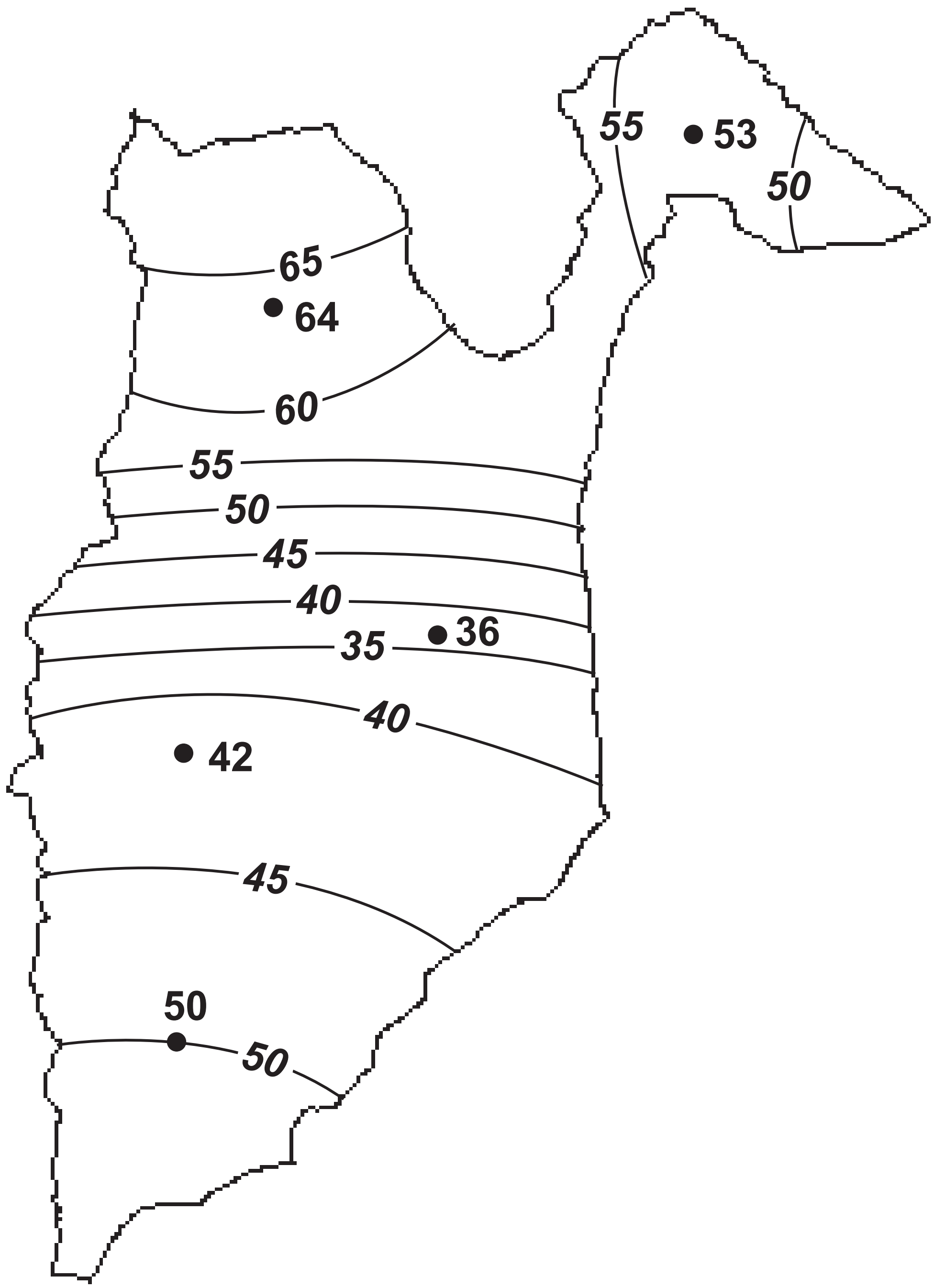


Choropleth Map

FIGURE

8-4

Ave. Rainfall per Year (inches)

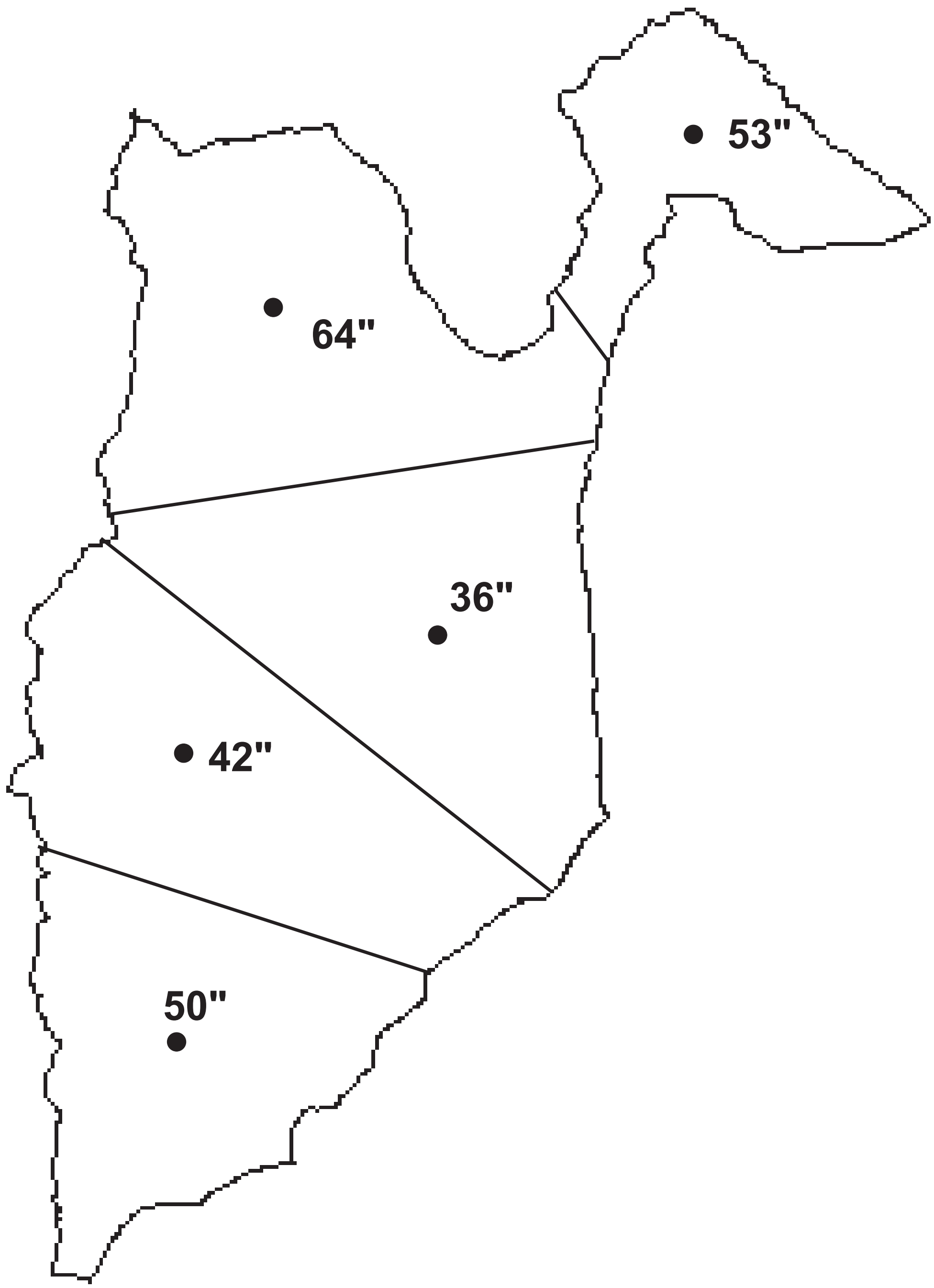


Isarithmic Map

FIGURE

8-5

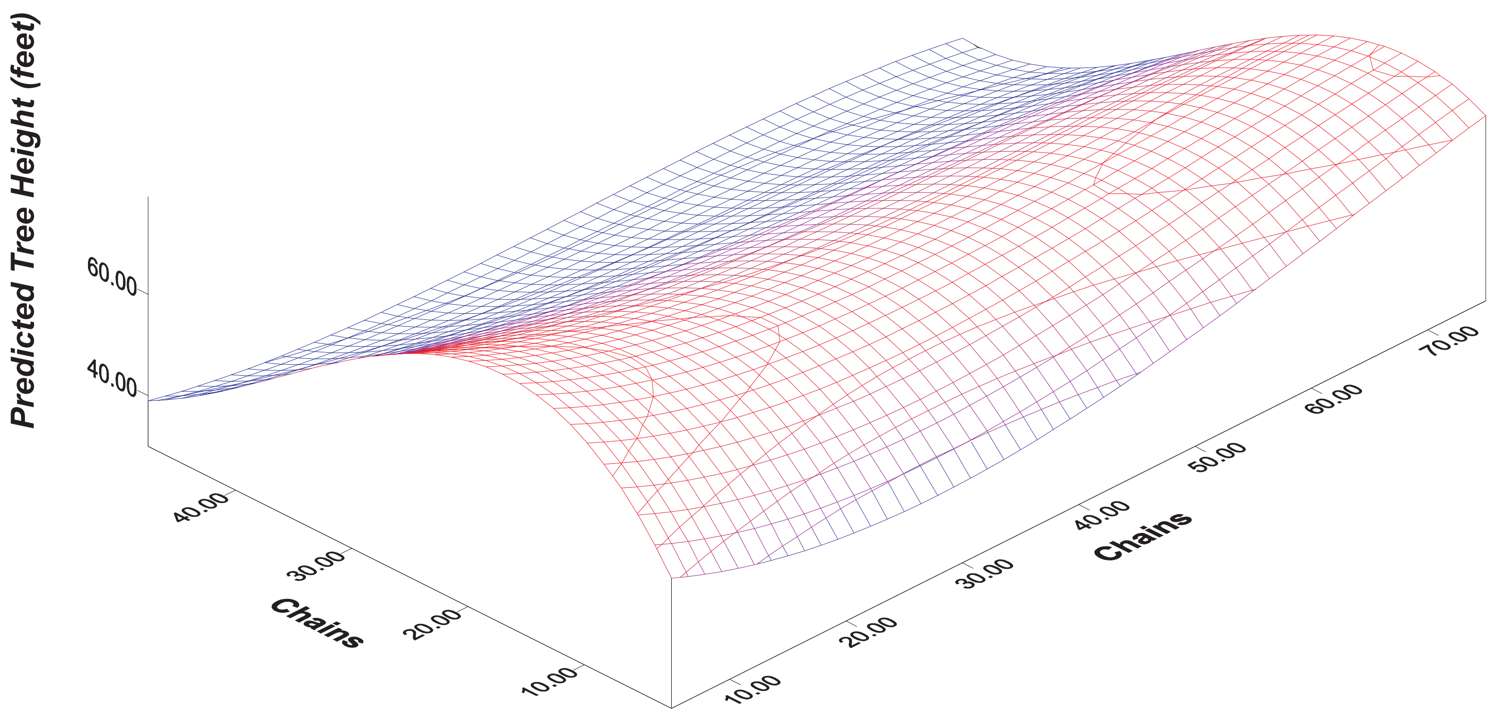
Ave. Rainfall per Year (inches)



Proximal Map

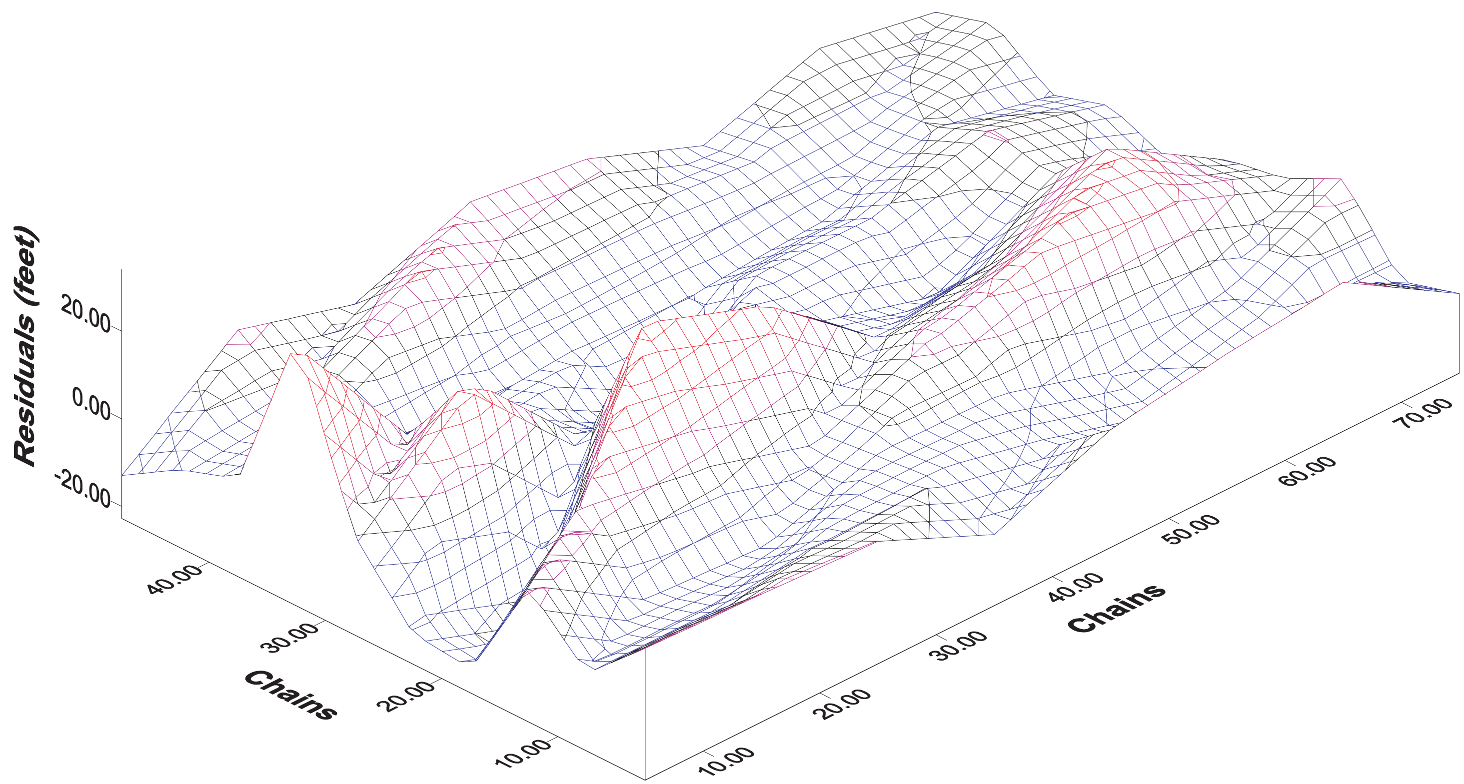
FIGURE

8-6



Third Order Trend Surface for Tree Heights

FIGURE
8-7



Residuals of Third Order Trend Surface for Tree Heights

FIGURE 8-8

Characteristics of Maps

1. Usually stylized, generalized, or abstracted.
2. Usually out of date.
3. Show only static situation.
4. Easy to use for certain questions.
5. Difficult for other questions.

FIGURE

8-9

Ways to Delineate Scale

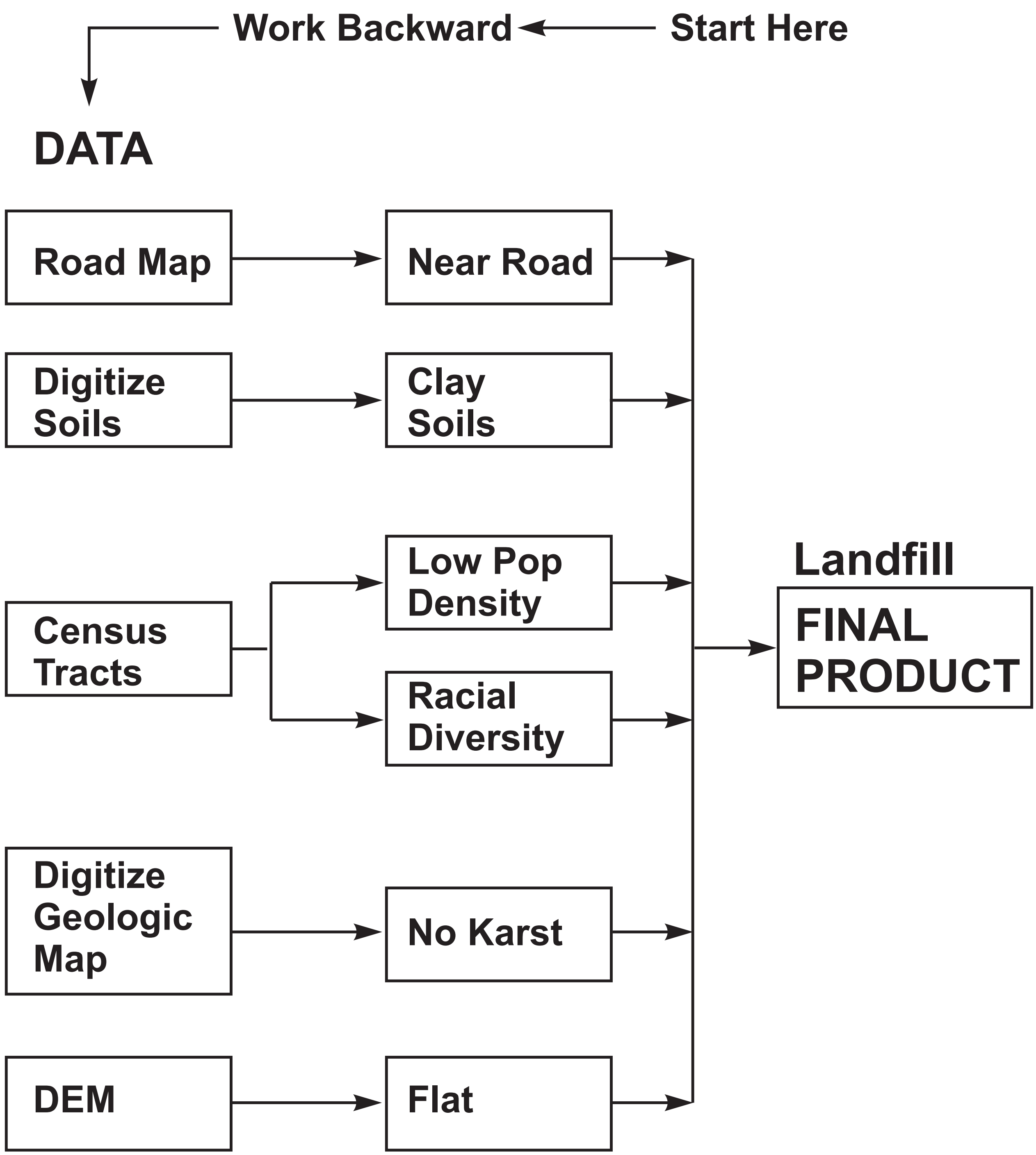
1. representative fraction (RF)

2. verbal (conversion scale)

3. graphic

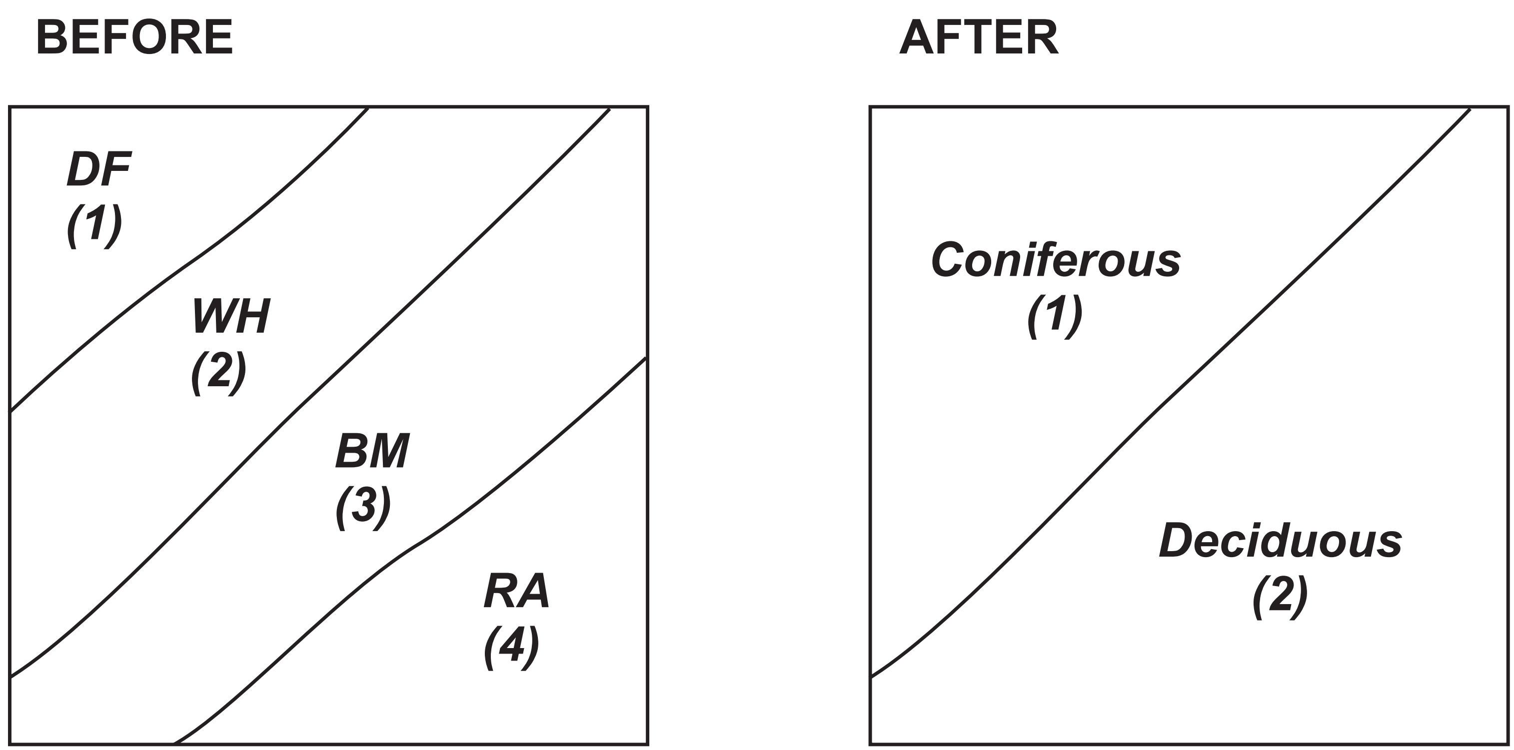
FIGURE

8-10



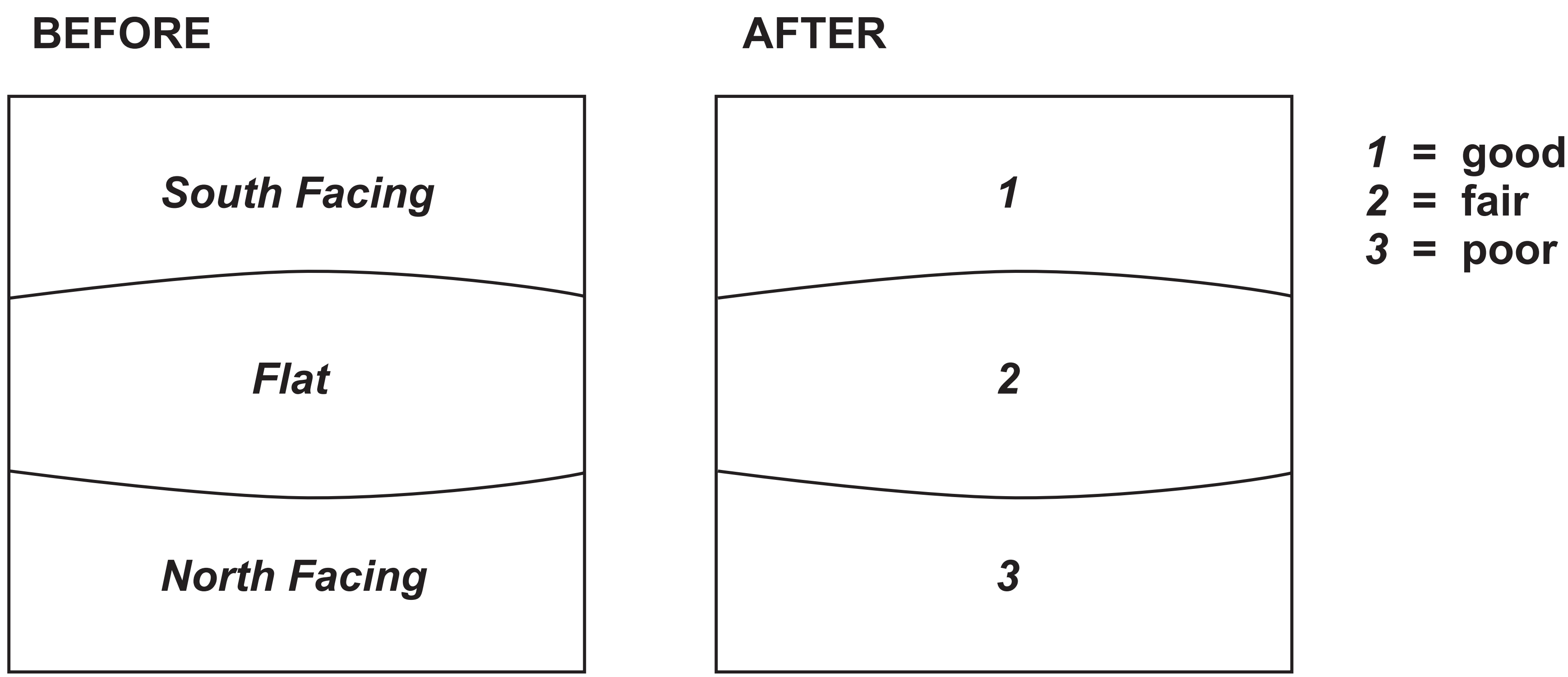
Cartographic Modeling

FIGURE 8-11



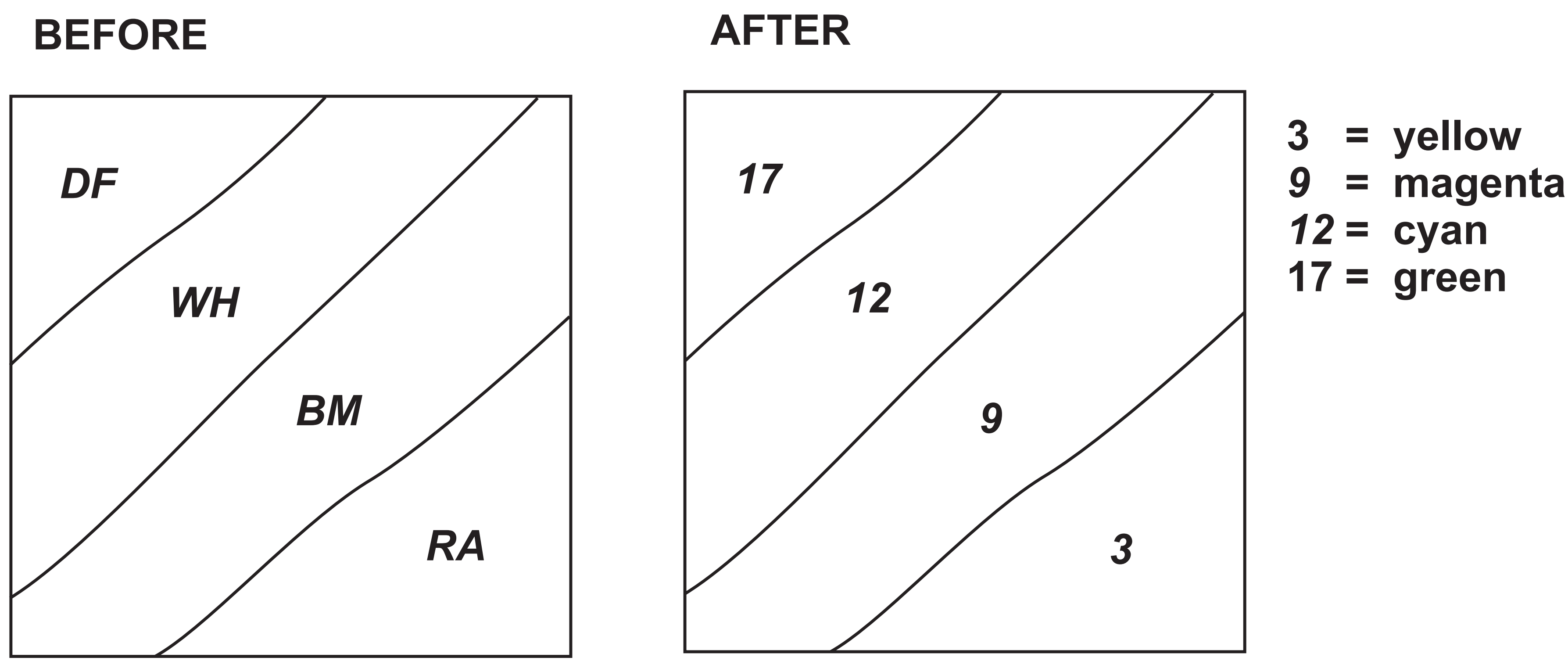
Reclass by Grouping

FIGURE 8-12



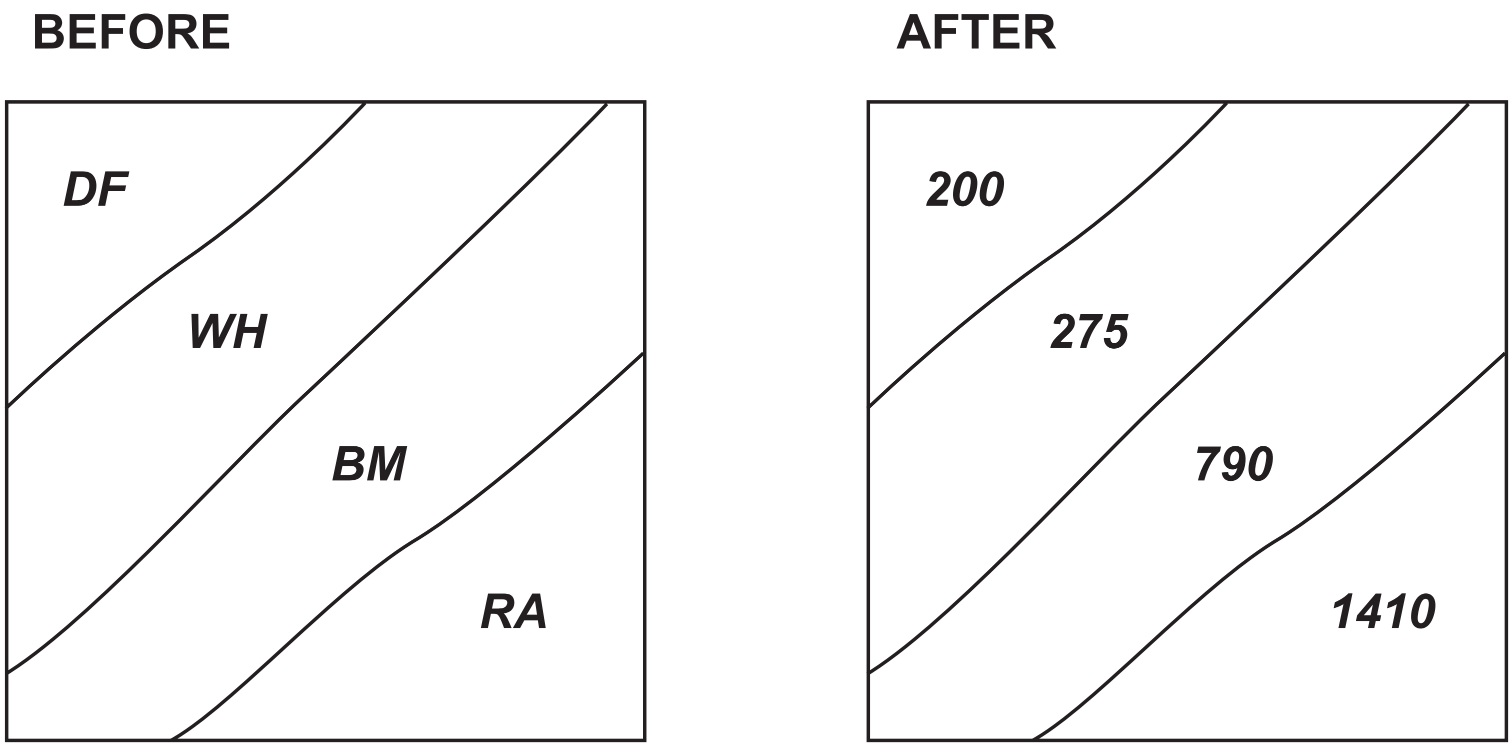
Reclass by Modeling

FIGURE 8-13



Reclass for Color Display

FIGURE 8-14



Reclass by Measurements

FIGURE 8-15

I-1, J-1	I-1, J	I-1, J+1
I, J-1	I, J kernel	I, J+1
I+1, J-1	I+1, J	I+1, J+1

Cartographic Neighborhoods

FIGURE 8-16

16	17	14
15	12	15
14	13	15

16+17+14+15+12+15+14+13+15 / 9 = mean (14)

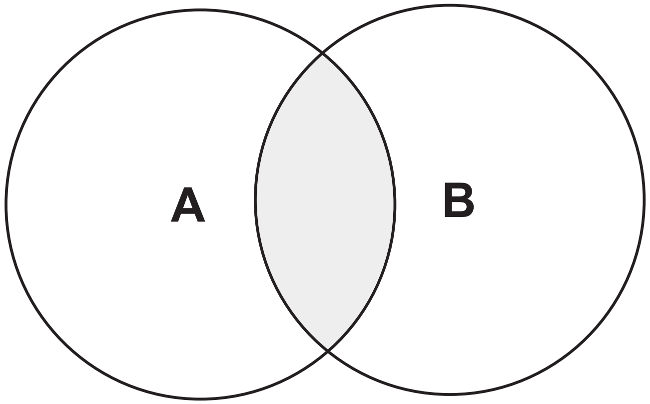
Filter

FIGURE 8-17

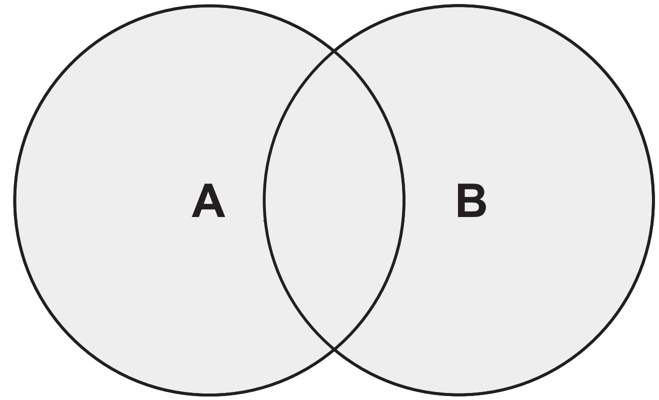
4355	4368	4312
4327	?	4315
4375	4382	4370

Slope

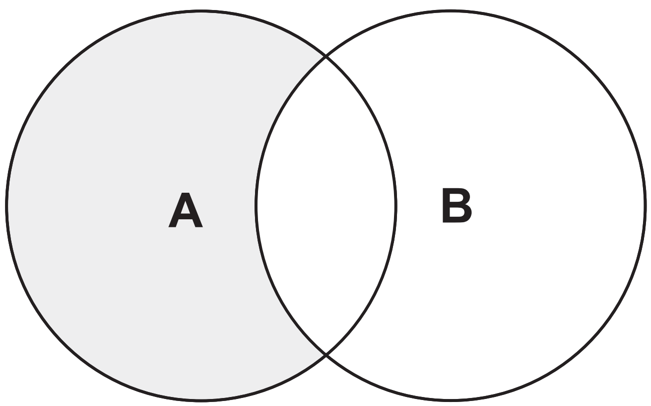
FIGURE 8-18



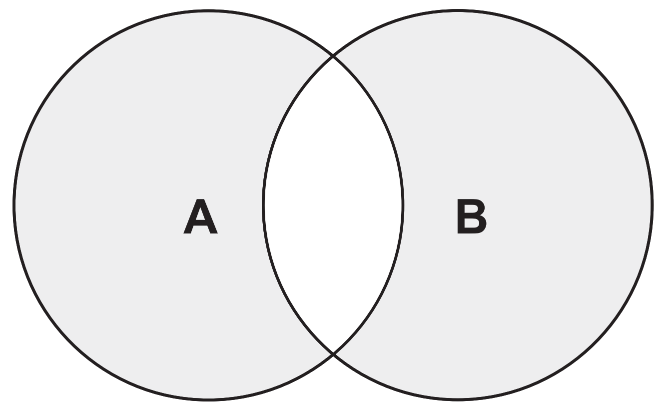
Boolean Operator <u>AND</u>



Boolean Operator <u>OR</u>



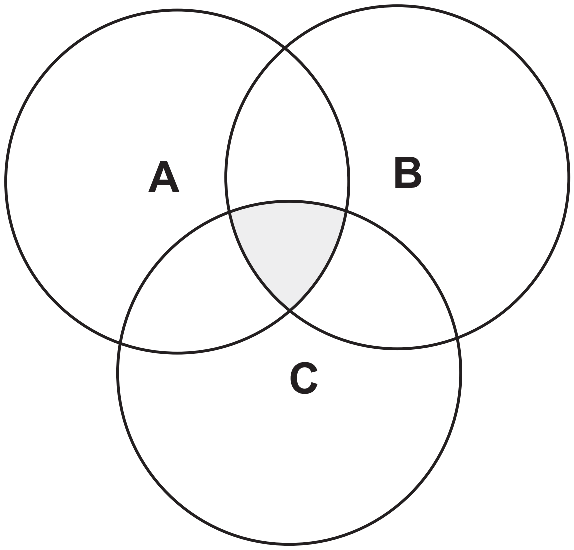
Boolean Operator <u>NOT</u>



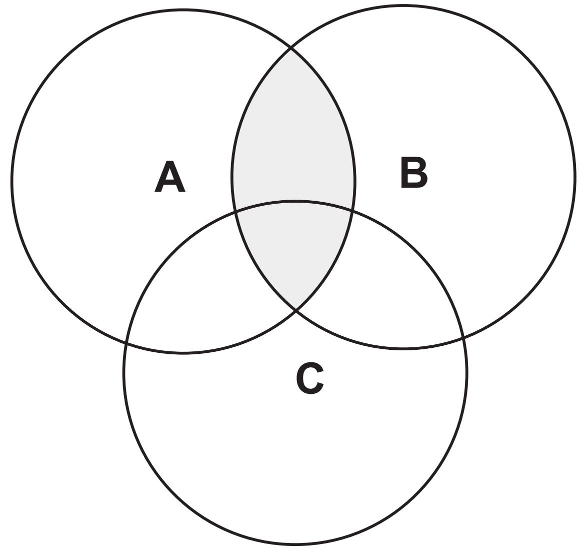
Boolean Operator <u>XOR</u>

Boolean Operators

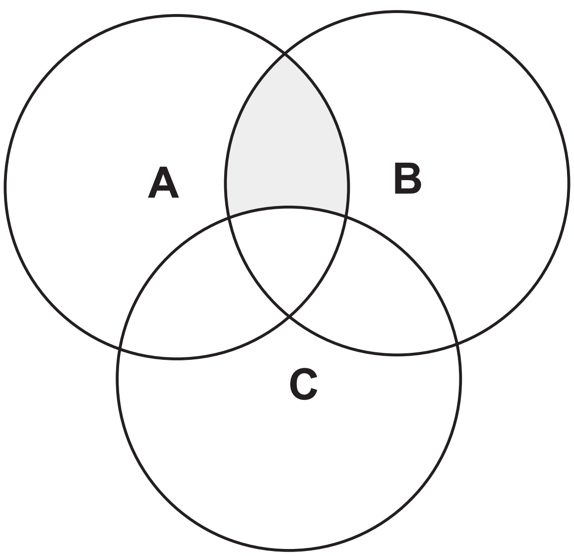
FIGURE 8-19



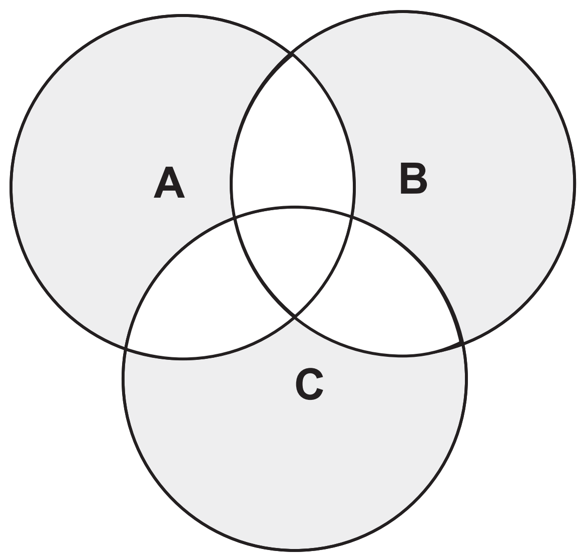
A and B and C



A and B



A and B not C

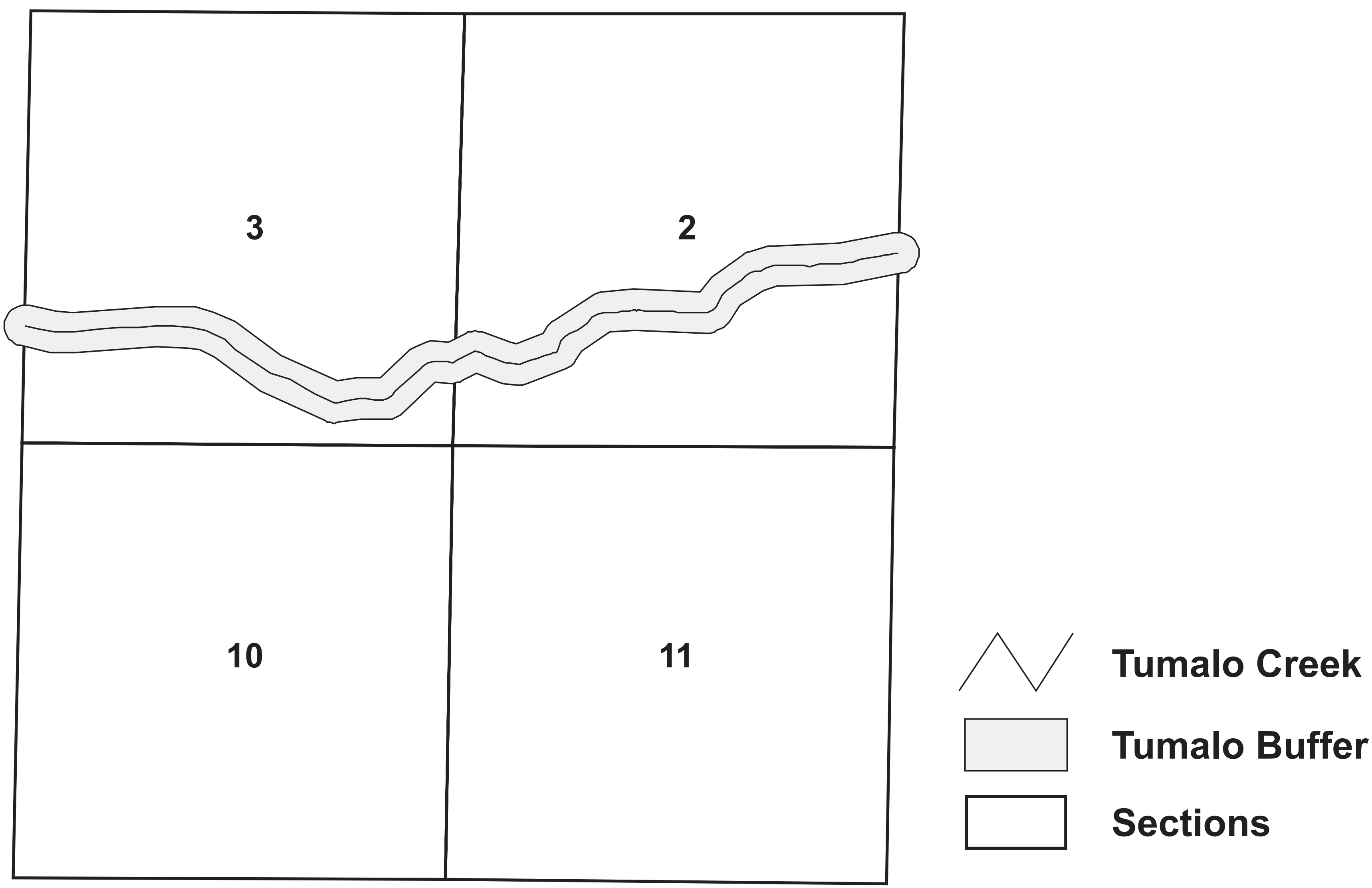


A xor B xor C

Examples of Boolean Operators

FIGURE

8-20

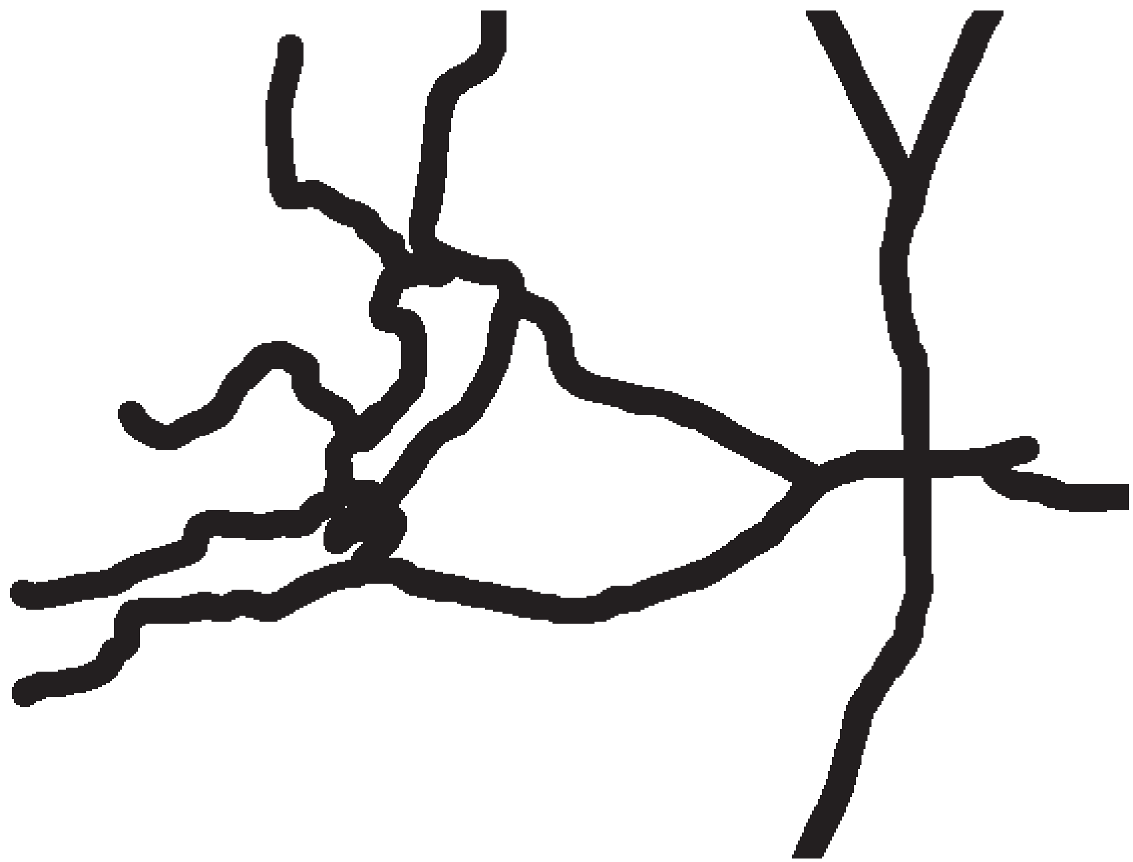


Vector Buffering / Line Buffer

Example: Stream Buffer

FIGURE 8-21

NSFGIS\TOPIC8\OH0821.CDR

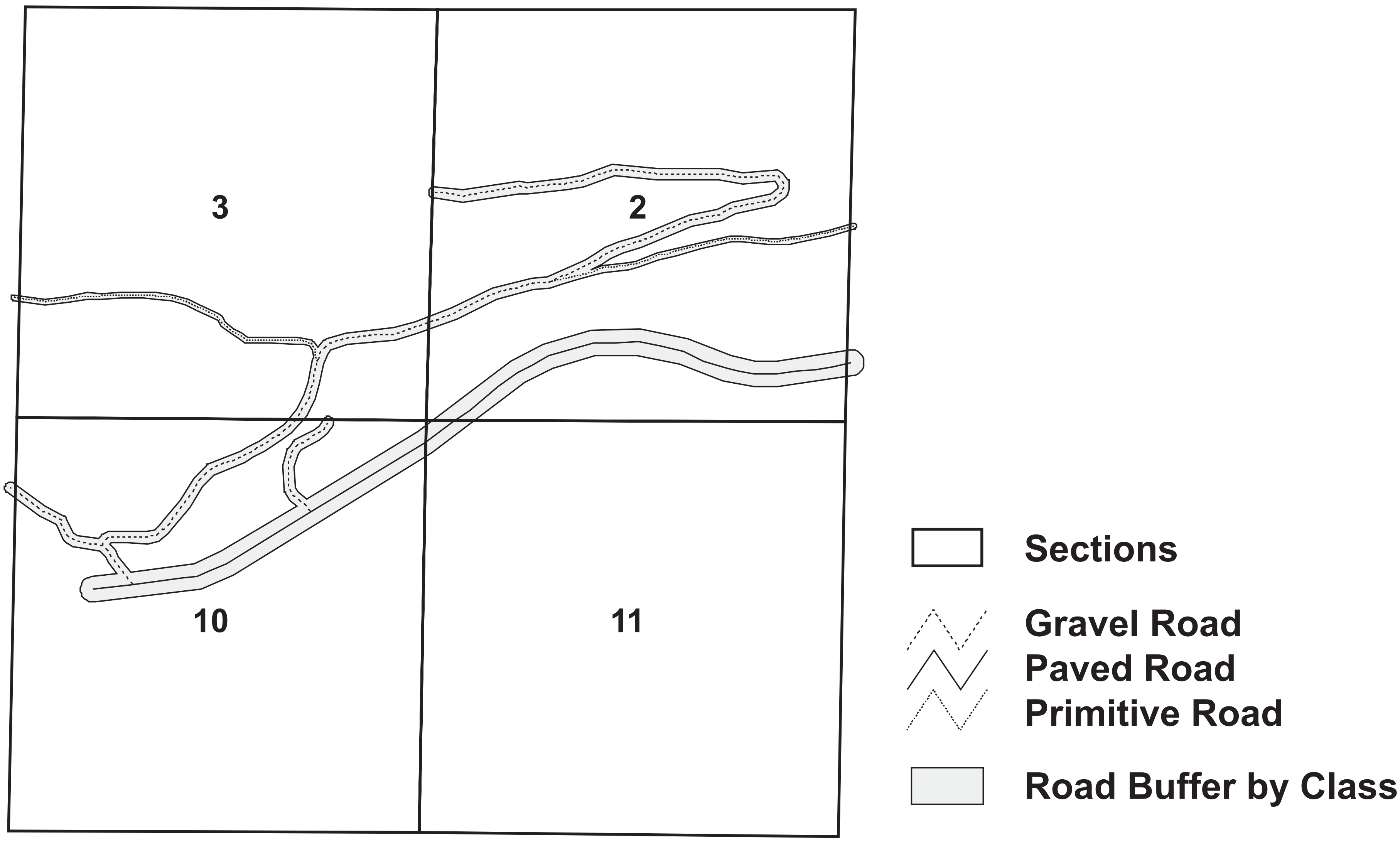


Raster Buffering

Example: Buffered Roads (200 meters)

FIGURE

8-22



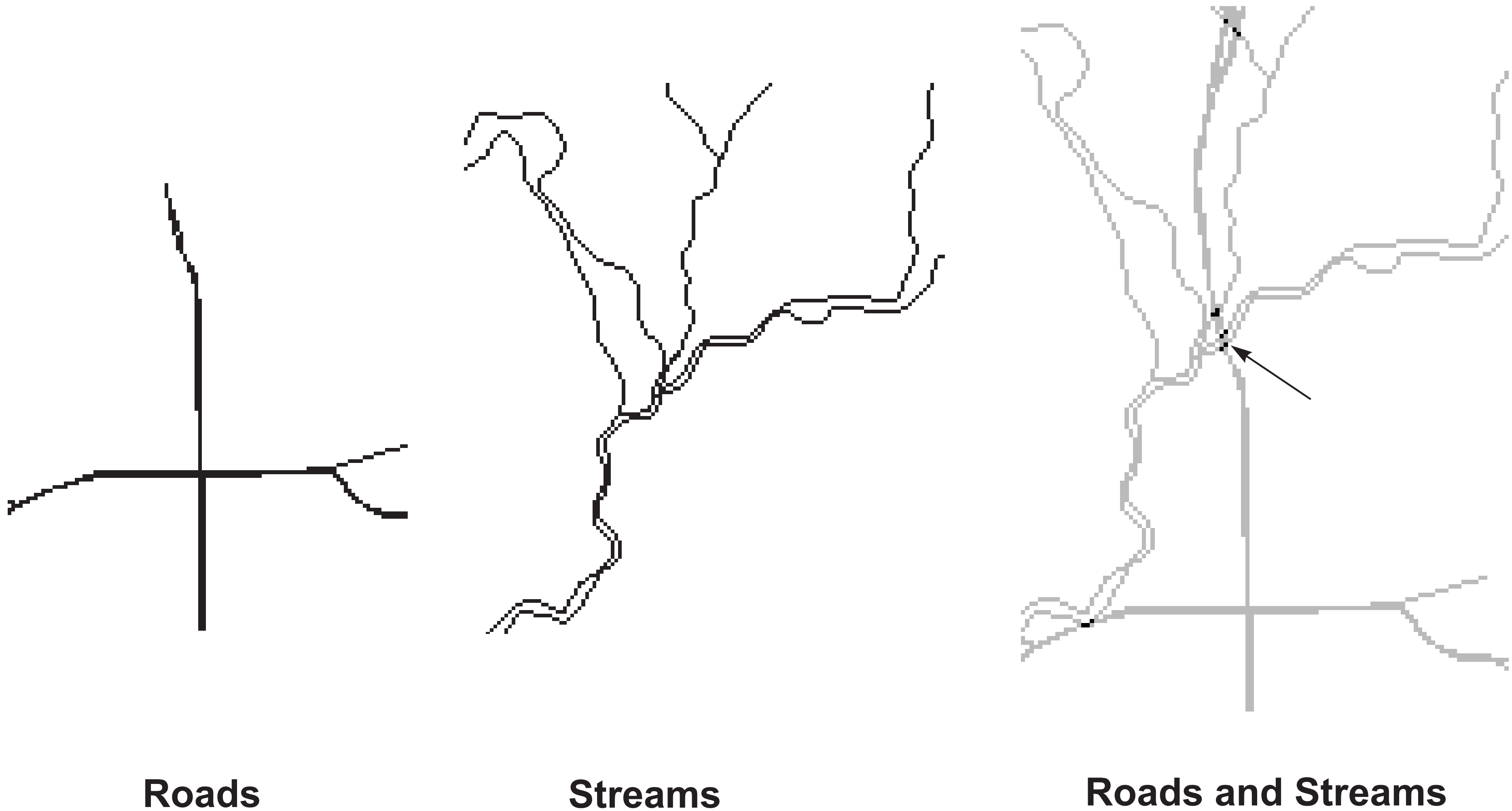
Constant and Variable Buffer

Example: Buffered Roads

FIGURE

8-23

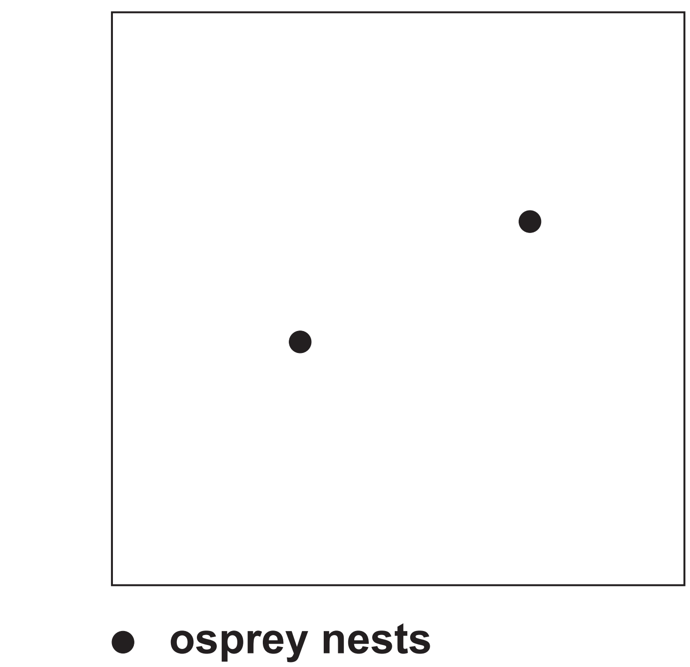
NSFGIS\TOPIC8\OH0823.CDR



Raster Overlay

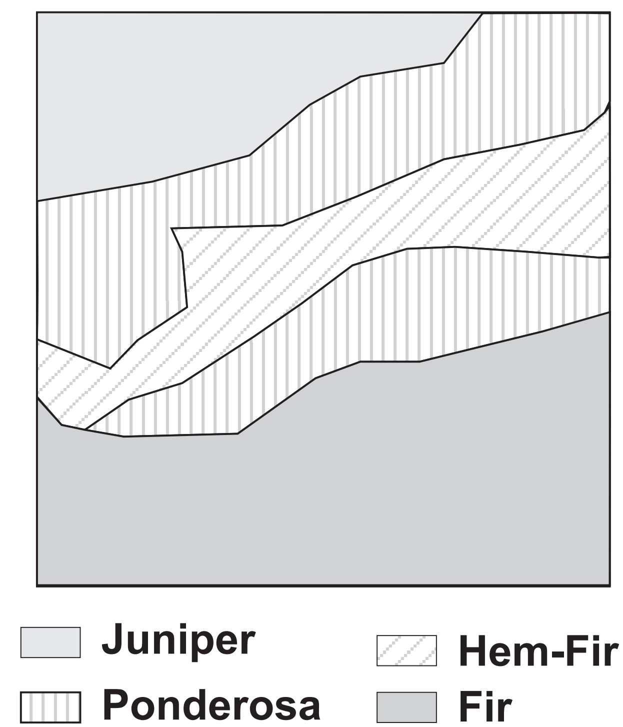
FIGURE 8-24

Raptor Nests



Area	Perimeter	Raptors_id	Raptor
0.000000	0.000000	5	osprey
0.000000	0.000000	11	osprey

Tree Cover Type



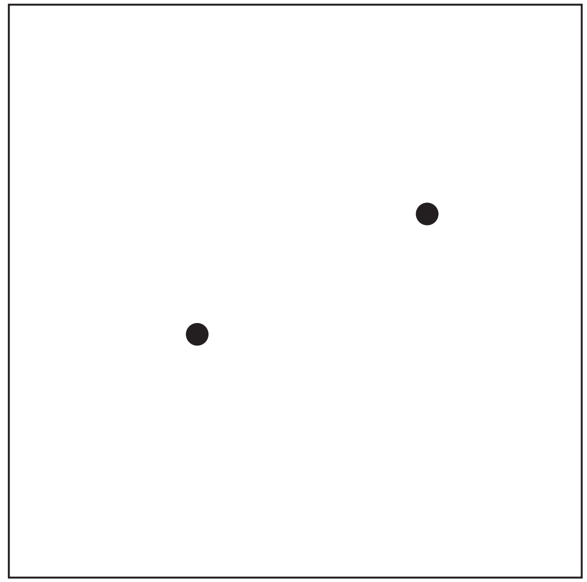
Area	Perimeter	Covtype_id	Covertype
2.507263e+009	250665.000000	3	JUNIPER
3.360276e+009	353346.900000	5	PONDEROSA
2.615034e+009	355404.800000	4	HEM-FIR
2.014542e+009	263314.100000	1	PONDEROSA
5.760761e+009	370030.700000	2	FIR

Vector Overlay - Points

FIGURE 8-25

NSFGIS\TOPIC8\OH0825.CDR

Overlay Raptor Nests by Cover Type



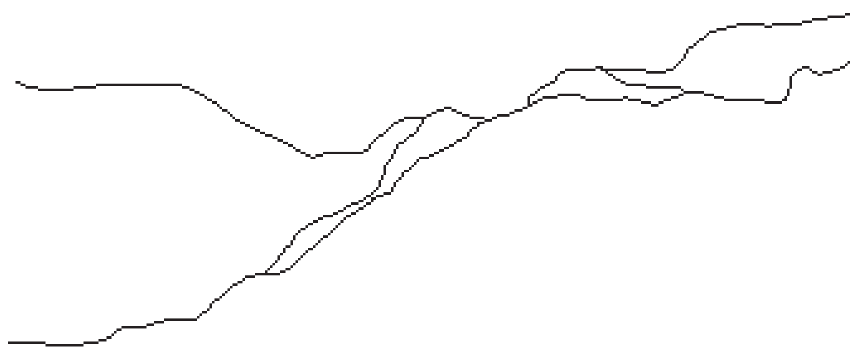
● osprey nests

Rap_cov_id	*Raptors_id*	*Raptor*	*Covtype_id*	*Covertype*
7	11	osprey	4	HEM-FIR
8	5	osprey	4	HEM-FIR

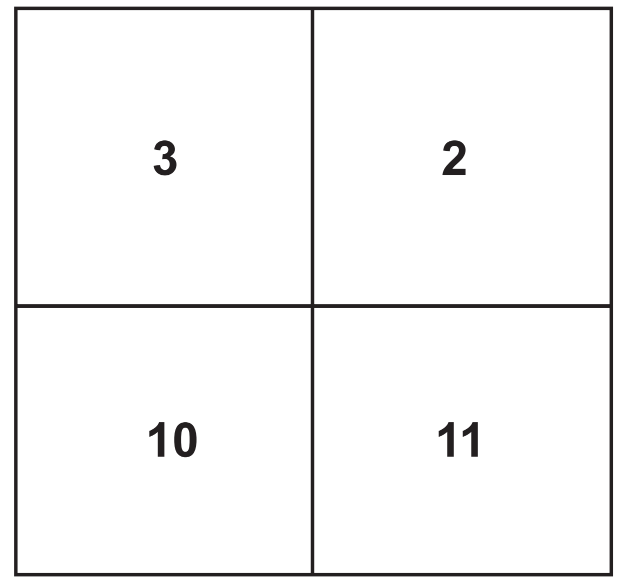
Vector Overlay - Points

FIGURE 8-26

Streams



Sections



Fnode_	Tnode_	Lpoly_	Rpoly_	Length	Streams_id	Str_nam
10	8	0	0	84918.200000	1	beaver brook
9	7	0	0	35789.910000	2	crossover str.
4	6	0	0	84514.160000	3	tumalo creek
6	2	0	0	54846.000000	4	beaver brook
6	1	0	0	55203.010000	5	tumalo creek
3	5	0	0	13969.010000	6	deadend ditch

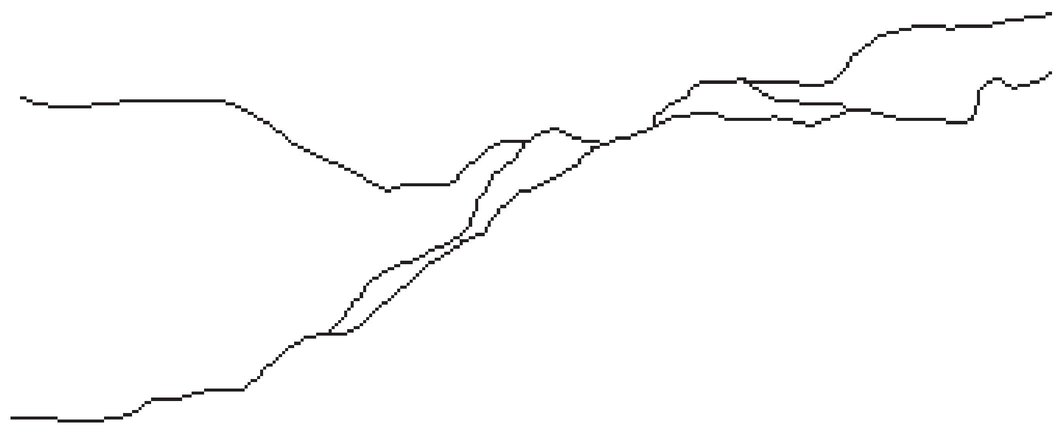
Area	Perimeter	Sections_i	Sec_num
4.000142e+009	253004.800000	2	3
4.098144e+009	256123.600000	4	2
4.026752e+009	253842.300000	1	10
4.132748e+009	257158.400000	3	11

Vector Overlay - Lines

FIGURE **8-27**

NSFGIS\TOPIC8\OH0827.CDR

Overlay Streams by Section



Fnode_	Tnode_	Lpoly_	Rpoly_	Length	Str_sec_id	Str_nam	Area	Perimeter	Sec_num
3	1	3	3	41607.740000	5	tumalo creek	4.098144e+009	256123.600000	2
3	5	3	3	13969.010000	6	deadend ditch	4.098144e+009	256123.600000	2
6	2	3	3	54846.000000	4	beaver brook	4.098144e+009	256123.600000	2
6	3	3	3	13595.280000	5	tumalo creek	4.098144e+009	256123.600000	2
8	7	2	2	247.418700	3	tumalo creek	4.000142e+009	253004.800000	3
9	6	3	3	6062.662000	3	tumalo creek	4.098144e+009	256123.600000	2
7	9	3	3	10233.260000	3	tumalo creek	4.098144e+009	256123.600000	2
10	9	3	3	11394.270000	1	beaver brook	4.098144e+009	256123.600000	2
4	8	2	2	67970.820000	3	tumalo creek	4.000142e+009	253004.800000	3
11	8	2	2	15652.360000	2	crossover str.	4.000142e+009	253004.800000	3
12	10	2	2	10057.970000	1	beaver brook	4.000142e+009	253004.800000	3
13	12	4	4	21482.300000	1	beaver brook	4.026752e+009	253842.300000	10
13	11	4	4	20137.560000	2	crossover str.	4.026752e+009	253842.300000	10
14	13	4	4	41983.640000	1	beaver brook	4.026752e+009	253842.300000	10

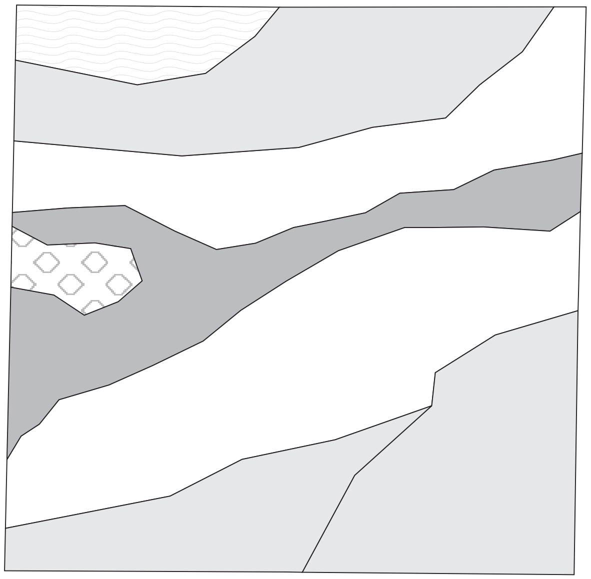
Vector Overlay - Lines

FIGURE

8-28

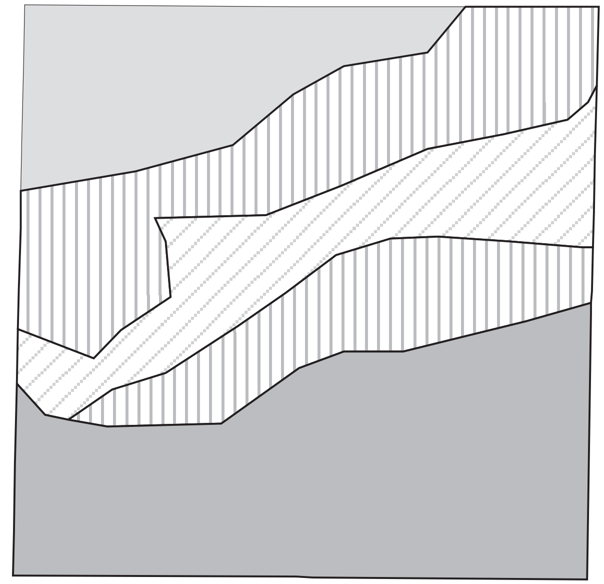
NSFGIS\TOPIC8\OH0828.CDR

Soils



Area	Perimeter	Soils_id	Drainage	Soiltype
8.018180e+008	137461.500000	5	POOR	CLAY
2.447072e+009	323524.400000	1	FAIR	SAND/LOAM
2.498233e+009	278962.800000	4	GOOD	SAND
2.279584e+009	404514.900000	3	FAIR	LOAM
4.003240e+009	327335.800000	8	FAIR	SAND/LOAM
3.626703e+008	82463.410000	2	GOOD	GRAVEL
2.370808e+009	211413.900000	6	GOOD	SAND
1.494432e+009	224793.500000	7	POOR	CLAY

Tree Cover Type

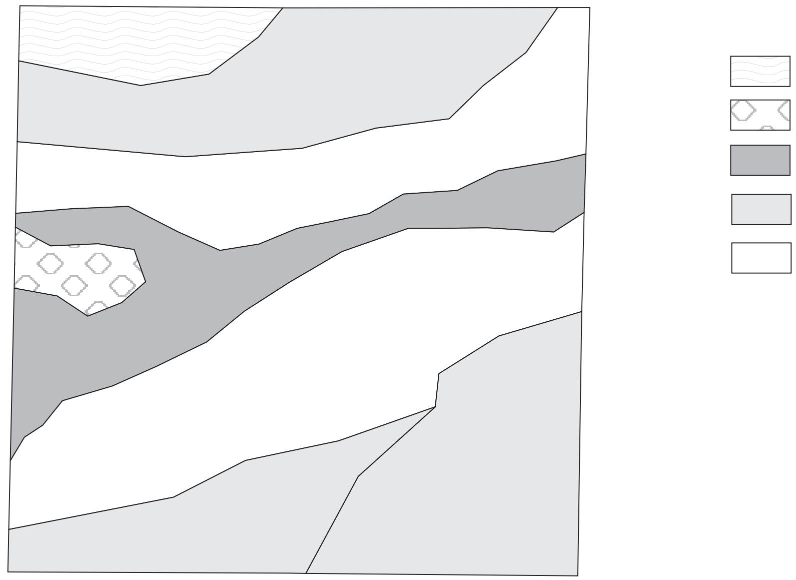


Area	Perimeter	Covtype_id	Covertype
2.507263e+009	250665.000000	3	JUNIPER
3.360276e+009	353346.900000	5	PONDEROSA
2.615034e+009	355404.800000	4	HEM-FIR
2.014542e+009	263314.100000	1	PONDEROSA
5.760761e+009	370030.700000	2	FIR

Vector Overlay - Polygons

FIGURE
8-29

Overlay Soils by Cover Type



Area	Perimeter	Soils_id	Drainage	Soiltype	Covtype_id	Covertype
8.018180e+008	137461.500000	5	POOR	CLAY	3	JUNIPER
1.441225e+009	300310.500000	1	FAIR	SAND/LOAM	5	PONDEROSA
7.850657e+008	224197.000000	1	FAIR	SAND/LOAM	4	HEM-FIR
2.207738e+008	86969.610000	1	FAIR	SAND/LOAM	3	JUNIPER
1.198349e+008	107061.200000	8	FAIR	SAND/LOAM	4	HEM-FIR
3.626344e+008	82455.700000	2	GOOD	GRAVEL	5	PONDEROSA
2.006870e+009	263115.700000	8	FAIR	SAND/LOAM	1	PONDEROSA
1.072813e+007	46361.310000	8	FAIR	SAND/LOAM	4	HEM-FIR
1.828748e+009	300424.700000	8	FAIR	SAND/LOAM	2	FIR
1.393650e+007	41721.320000	8	FAIR	SAND/LOAM	4	HEM-FIR
2.312254e+007	28246.360000	8	FAIR	SAND/LOAM	4	HEM-FIR
1.494432e+009	224793.500000	7	POOR	CLAY	2	FIR

Vector Overlay - Polygons

FIGURE 8-30

Objectives

A. Be familiar with the various types of maps used to display information.

B. Understand the concept of Cartographic Modeling in GIS analysis.

C. Be able to name and explain the major analytical functions supported by GIS.

D. Understand Cartographic Neighborhoods in Raster GIS.

E. Be able to explain various examples of Raster analysis of neighborhoods.

FIGURE 8-1

Types of Maps

A. Reference Maps

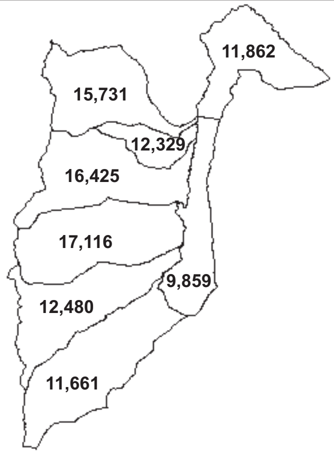
B. Thematic maps

FIGURE 8-2

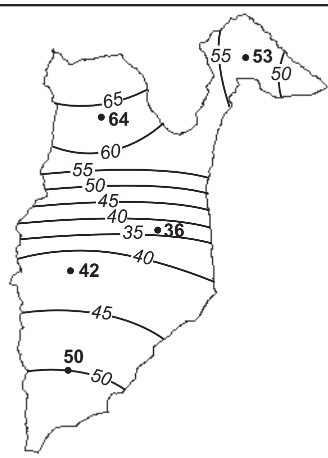
Types of Thematic Maps

1. choropleth map
2. isarithmic map
3. proximal map
4. trend surface map
5. residuals map

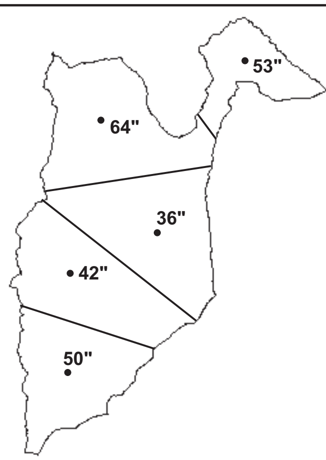
FIGURE 8-3



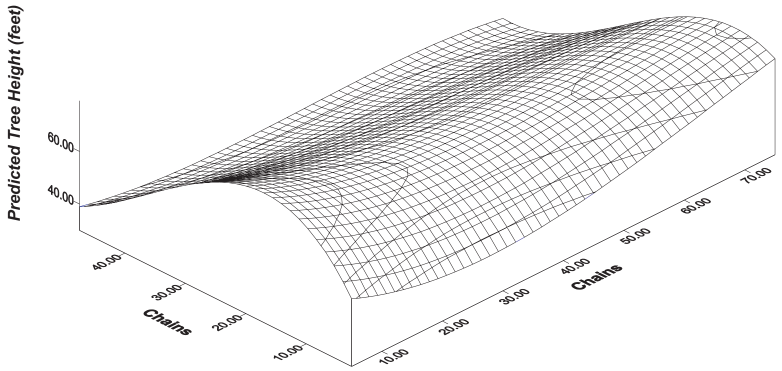
Choropleth Map | FIGURE 8-4



Isarithmic Map | FIGURE 8-5

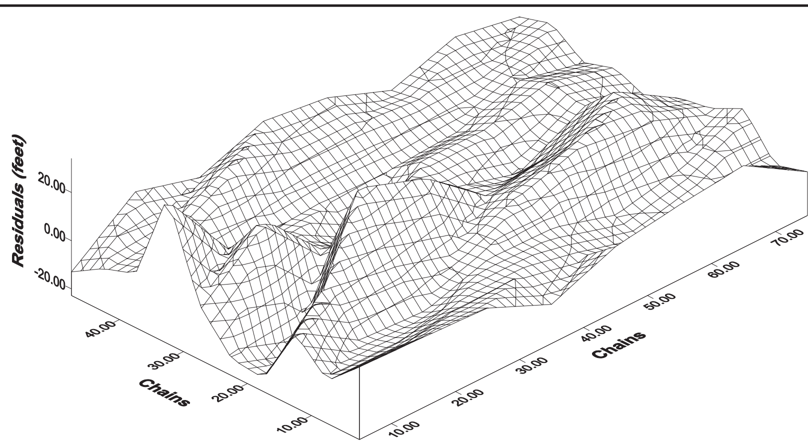


Proximal Map | FIGURE 8-6



Third Order Trend Surface for Tree Heights

FIGURE 8-7



Residuals of Third Order Trend Surface for Tree Heights

FIGURE 8-8

Characteristics of Maps

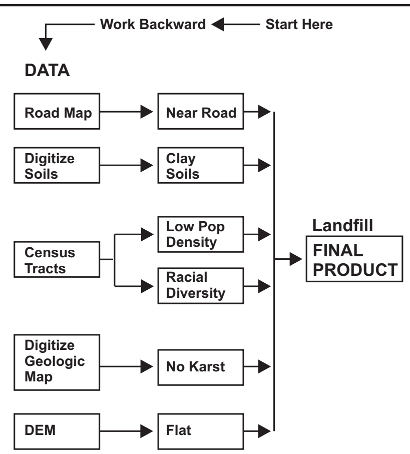
1. Usually stylized, generalized, or abstracted.
2. Usually out of date.
3. Show only static situation.
4. Easy to use for certain questions.
5. Difficult for other questions.

FIGURE 8-9

Ways to Delineate Scale

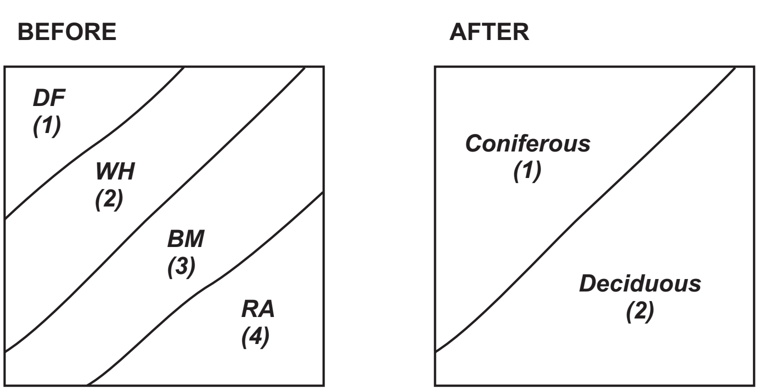
1. representative fraction (RF)
2. verbal (conversion scale)
3. graphic

FIGURE 8-10



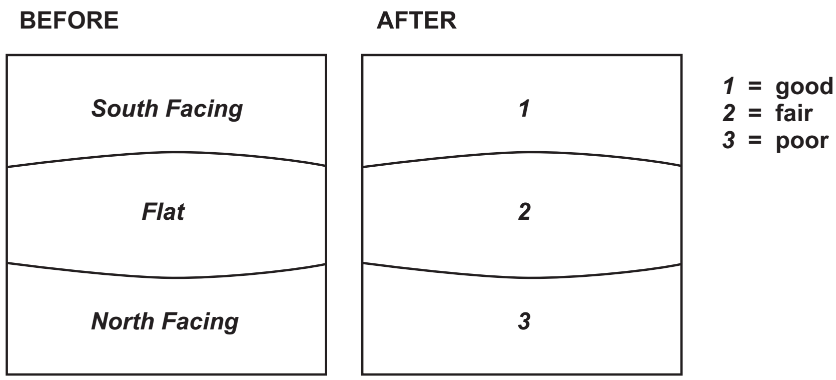
Cartographic Modeling

FIGURE 8-11



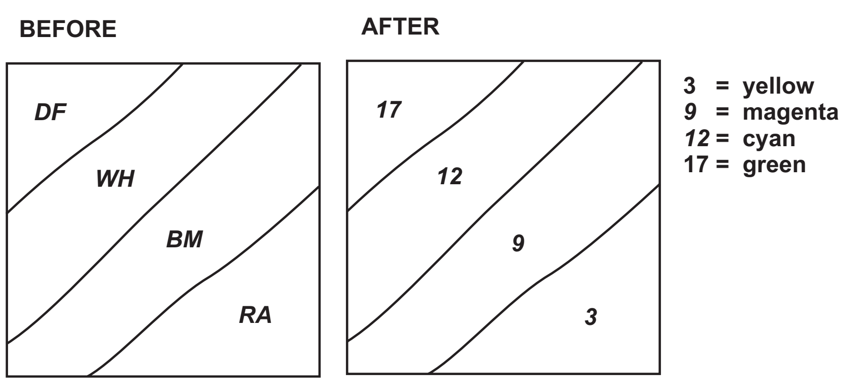
Reclass by Grouping

FIGURE 8-12



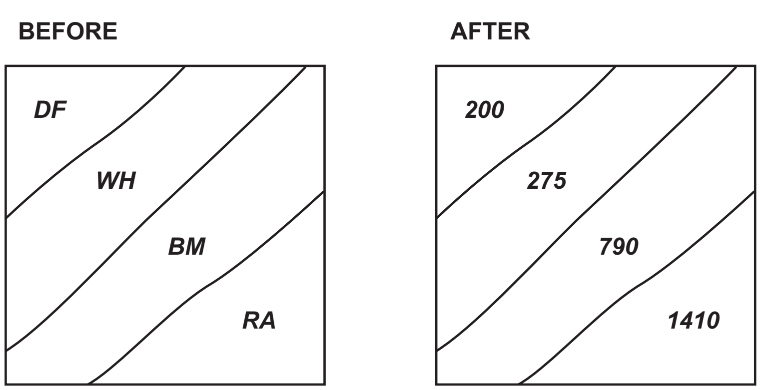
Reclass by Modeling

FIGURE 8-13



Reclass for Color Display

FIGURE 8-14



Reclass by Measurements

FIGURE 8-15

I-1, J-1	I-1, J	I-1, J+1
I, J-1	I, J kernel	I, J+1
I+1, J-1	I+1, J	I+1, J+1

Cartographic Neighborhoods FIGURE 8-16

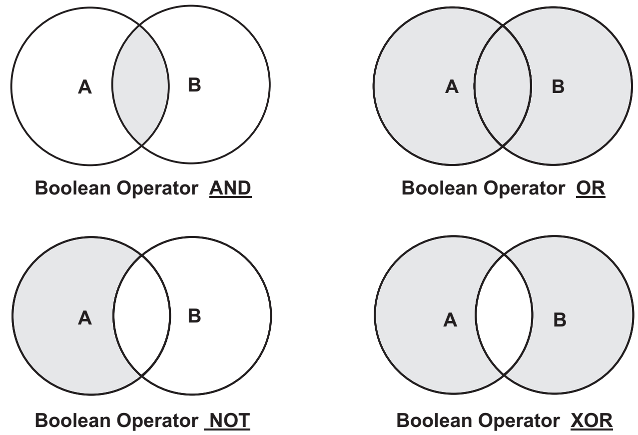
16	17	14
15	12	15
14	13	15

16+17+14+15+12+15+14+13+15 / 9 = mean (14)

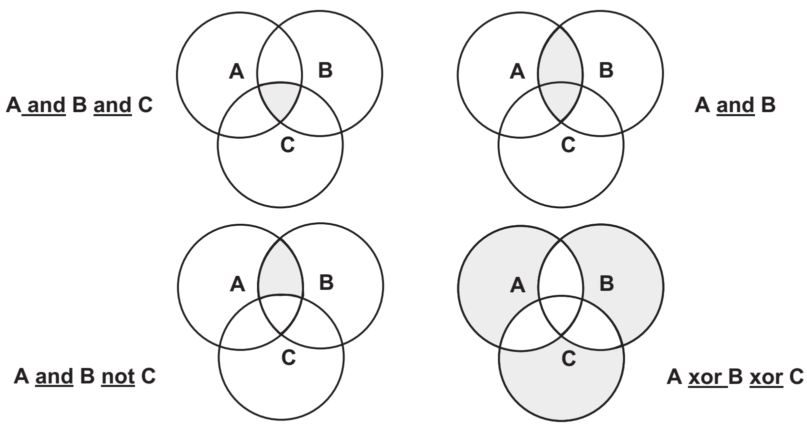
Filter FIGURE 8-17

4355	4368	4312
4327	?	4315
4375	4382	4370

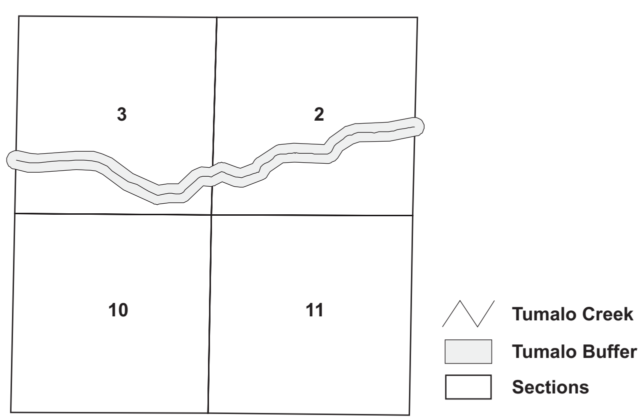
Slope FIGURE 8-18



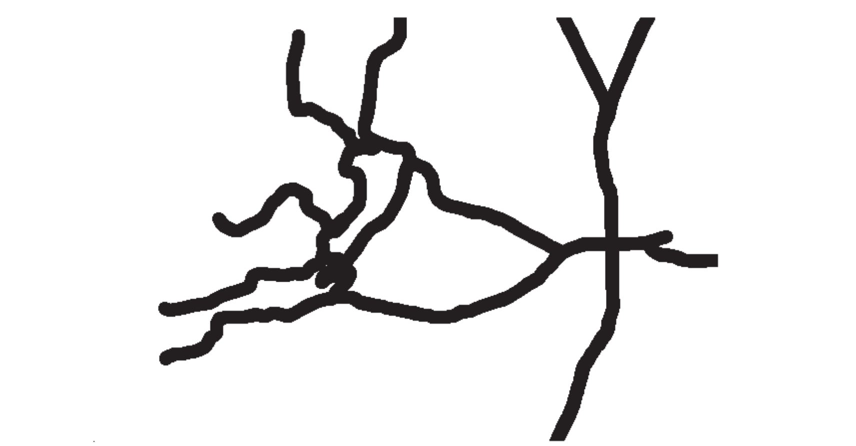
Boolean Operators FIGURE 8-19



Examples of Boolean Operators FIGURE 8-20

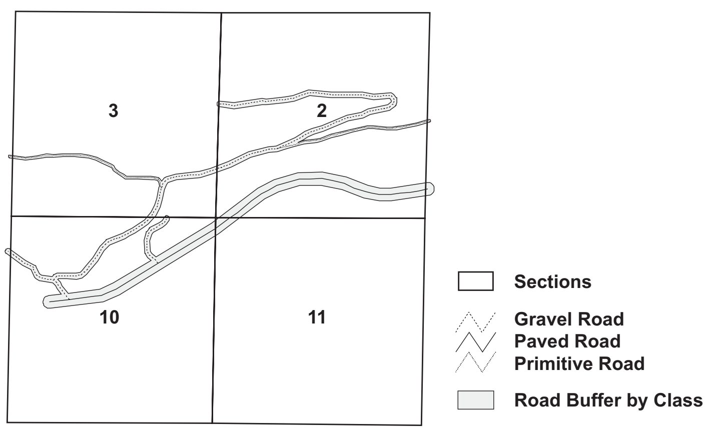


Vector Buffering / Line Buffer
Example: Stream Buffer FIGURE 8-21



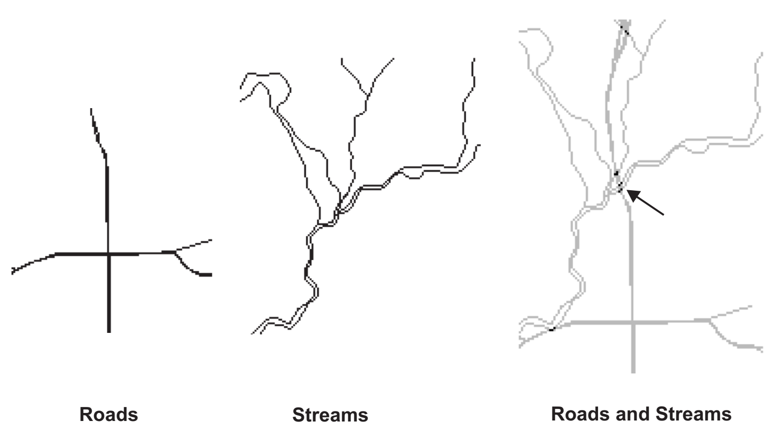
Raster Buffering
Example: Buffered Roads (200 meters)

FIGURE 8-22



Constant and Variable Buffer
Example: Buffered Roads

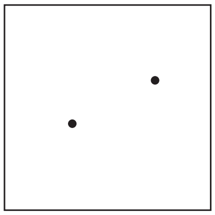
FIGURE 8-23



Raster Overlay

FIGURE 8-24

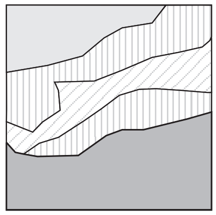
Raptor Nests



● osprey nests

Area	Perimeter	Raptors_id	Raptor
0.000000	0.000000	5	osprey
0.000000	0.000000	11	osprey

Tree Cover Type



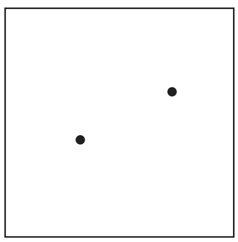
Juniper
Ponderosa
Hem-Fir
Fir

Area	Perimeter	Covtype_id	Covertype
2.507263e+009	250665.000000	3	JUNIPER
3.360276e+009	353346.900000	5	PONDEROSA
2.615034e+009	355404.800000	4	HEM-FIR
2.014542e+009	263314.100000	1	PONDEROSA
5.760761e+009	370030.700000	2	FIR

Vector Overlay - Points

FIGURE 8-25

Overlay Raptor Nests by Cover Type



● osprey nests

Rap_cov_id	Raptors_id	Raptor	Covtype_id	Covertype
7	11	osprey	4	HEM-FIR
8	5	osprey	4	HEM-FIR

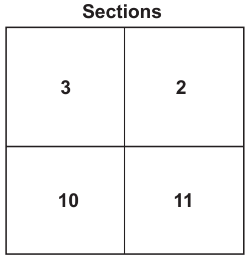
Vector Overlay - Points

FIGURE 8-26

Streams

Fnode_	Tnode_	Lpoly_	Rpoly_	Length	Streams_id	Str_nam
10	8	0	0	84918.200000	1	beaver brook
9	7	0	0	35789.910000	2	crossover str.
4	6	0	0	84514.160000	3	tumalo creek
6	2	0	0	54846.000000	4	beaver brook
6	1	0	0	55203.010000	5	tumalo creek
3	5	0	0	13969.010000	6	deadend ditch

Sections



Area	Perimeter	Sections_i	Sec_num
4.000142e+009	253004.800000	2	3
4.098144e+009	256123.600000	4	2
4.026752e+009	253842.300000	1	10
4.132748e+009	257158.400000	3	11

Vector Overlay - Lines

FIGURE 8-27

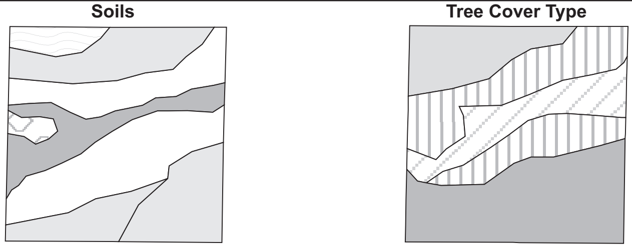
Overlay Streams by Section

Fnode_	Tnode_	Lpoly_	Rpoly_	Length	Str_sec_id	Str_nam	Area	Perimeter	Sec_num
3	1	3	3	41607.740000	5	tumalo creek	4.098144e+009	256123.600000	2
3	5	3	3	13969.010000	6	deadend ditch	4.098144e+009	256123.600000	2
6	2	3	3	54846.000000	4	beaver brook	4.098144e+009	256123.600000	2
6	3	3	3	13595.280000	5	tumalo creek	4.098144e+009	256123.600000	2
8	7	2	2	247.418700	3	tumalo creek	4.000142e+009	253004.800000	3
9	6	3	3	6062.662000	3	tumalo creek	4.098144e+009	256123.600000	2
7	9	3	3	10233.260000	3	tumalo creek	4.098144e+009	256123.600000	2
10	9	3	3	11394.270000	1	beaver brook	4.098144e+009	256123.600000	2
4	8	2	2	67970.820000	3	tumalo creek	4.000142e+009	253004.800000	3
11	8	2	2	15652.360000	2	crossover str.	4.000142e+009	253004.800000	3
12	10	2	2	10057.970000	1	beaver brook	4.000142e+009	253004.800000	3
13	12	4	4	21482.300000	1	beaver brook	4.026752e+009	253842.300000	10
13	11	4	4	20137.560000	2	crossover str.	4.026752e+009	253842.300000	10
14	13	4	4	41983.640000	1	beaver brook	4.026752e+009	253842.300000	10

Vector Overlay - Lines

FIGURE 8-28

Soils / Tree Cover Type



Clay
Gravel
Loam
Sand
Sand/Loam

Area	Perimeter	Soils_id	Drainage	Soiltype
8.018180e+008	137461.500000	5	POOR	CLAY
2.447072e+009	323524.400000	1	FAIR	SAND/LOAM
2.498233e+009	278962.800000	4	GOOD	SAND
2.279584e+009	404514.900000	3	FAIR	LOAM
4.003240e+009	327335.800000	8	FAIR	SAND/LOAM
3.626703e+009	82463.410000	2	GOOD	GRAVEL
2.370808e+009	211413.900000	6	GOOD	SAND
1.494432e+009	224793.500000	7	POOR	CLAY

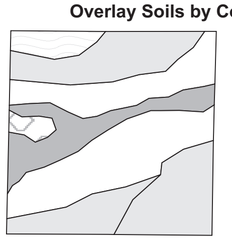
Juniper
Ponderosa
Hem-Fir
Fir

Area	Perimeter	Covtype_id	Covertype
2.507263e+009	250665.000000	3	JUNIPER
3.360276e+009	353346.900000	5	PONDEROSA
2.615034e+009	355404.800000	4	HEM-FIR
2.014542e+009	263314.100000	1	PONDEROSA
5.760761e+009	370030.700000	2	FIR

Vector Overlay - Polygons

FIGURE 8-29

Overlay Soils by Cover Type



Clay
Gravel
Loam
Sand
Sand/Loam

Area	Perimeter	Soils_id	Drainage	Soiltype	Covtype_id	Covertype
8.018180e+008	137461.500000	5	POOR	CLAY	3	JUNIPER
1.441225e+009	300310.500000	1	FAIR	SAND/LOAM	5	PONDEROSA
7.850657e+008	224197.000000	1	FAIR	SAND/LOAM	4	HEM-FIR
2.207738e+008	86969.610000	1	FAIR	SAND/LOAM	3	JUNIPER
1.198349e+008	107061.200000	8	FAIR	SAND/LOAM	4	HEM-FIR
3.626344e+008	82455.700000	2	GOOD	GRAVEL	5	PONDEROSA
2.006870e+009	263115.700000	8	FAIR	SAND/LOAM	1	PONDEROSA
1.072813e+007	46361.310000	8	FAIR	SAND/LOAM	4	HEM-FIR
1.828748e+009	300424.700000	8	FAIR	SAND/LOAM	2	FIR
1.393650e+007	41721.320000	8	FAIR	SAND/LOAM	4	HEM-FIR
2.312254e+007	28246.360000	8	FAIR	SAND/LOAM	4	HEM-FIR
1.494432e+009	224793.500000	7	POOR	CLAY	2	FIR

Vector Overlay - Polygons

FIGURE 8-30

Objectives

A. Become familiar with sources and types of attribute data.

B. Become familiar with sources and types of spatial data.

FIGURE

9-1

Sources of Attribute Data

1. US Census Bureau

2. Private sources

3. Remote sensed images

FIGURE

9-2

Sources of Spatial Data

1. DIMECO

2. World Data Bank I and II

3. DIME (Dual Independent Map Encoding)

4. TIGER (Topologically Integrated Geographic Encoding and Referencing)

FIGURE

9-3

US Geodata

A. Geographic Names Information System (GNIS)

B. Land Use / Land Cover

C. Planimetric Data

D. Elevation Data

FIGURE

9-4

Objectives

A. Become familiar with sources and types of attribute data.

B. Become familiar with sources and types of spatial data.

FIGURE 9-1

Sources of Attribute Data

1. US Census Bureau
2. Private sources
3. Remote sensed images

FIGURE 9-2

Sources of Spatial Data

1. DIMECO
2. World Data Bank I and II
3. DIME (Dual Independent Map Encoding)
4. TIGER (Topologically Integrated Geographic Encoding and Referencing)

FIGURE 9-3

US Geodata

A. Geographic Names Information System (GNIS)

B. Land Use / Land Cover

C. Planimetric Data

D. Elevation Data

FIGURE
9-4