PDHonline Course L155G (5 PDH)

Data Models and Data processing in GIS

Instructor: Steve Ramroop, Ph.D.

2020

PDH Online | PDH Center

5272 Meadow Estates Drive
Fairfax, VA 22030-6658
Phone: 703-988-0088
www.PDHonline.com

An Approved Continuing Education Provider

Slide 1

Lecture 5 Content

- **Geographic Information Systems (GIS)**

Data Processing And Management

This lecture is the continuation of the GIS topic identified in the course description which is Data Models, Data Structure and Data Management. The emphasis in this lecture is on the analysis capabilities of typical GIS software.

Slide 2

Lecture Content:

- Introduction
- Organizing Data for Analyses data layers
- GIS Data Manipulation And Analyses

This slide gives the topics covered in this lecture. The lecture gave a basic introduction into the need to properly organize GIS data layers in order to perform manipulation and analyses of the data for the intended GIS application.

Slide 3

Introduction

- GIS spatial analysis functions distinguishes it from other types of information systems.
- These functions utilizes the spatial and attribute data to answer questions about the real world.
- Note that GIS models are designed to mimic only the selected aspects of reality which depends upon the intended GIS application.
- Generally, the more factors that a model takes into account, the more complex the GIS application becomes, and the more expensive it is to use and maintain the GIS.
- A more complex model may or may not provide "better" answers; it depends on the questions being addressed and the analyzing capabilities of the GIS software being used.

Note that the focus of this slide is on the analysis capabilities of GIS software. But the capability depends upon: the model; the data sets; and the intended GIS application. The more complex a GIS model, the more will be the demands for higher: quality data; GIS processing capability; and computing power.

Slide 4

- **Models are used sometimes repeatedly to perform analyses that test alternative scenarios.**
- **To develop the best answers from the information available requires a systematic framing of the questions to be addressed.**
- **There is often an influencing data storage tendency to begin the questions to be answered by being guided by the data needed.**
- **The analysis functions available in GIS and the organization of the data influences the final usefulness of the end result.**

This slide gives some general statements about GIS: models; analysis; and answers.

Slide 5

Organizing Data for Analyses

- The organization of data sets are influenced by the intended use of the final GIS application result
- For example: it may be acceptable to store oil well locations on the same map (coverage) as the roads for the purpose of a hazard response in the oil industry but in another application such as a tourist map, then the oil well locations are not important
- Geographic Information are organized within a GIS so as to optimize the convenience and efficiency with which they can be used

This slide identifies the importance of organizing data. There are issues regarding the storage of data into separate data layers, and its relation to the efficiency of its data storage. The ideal situation is that point, lines, and polygons must be stored separately. For example, on a single data layer, electricity poles maybe added as graphics while the power lines are stored as features with topology definitions.

Slide 6

- **The organization of a company will influence the GIS application by the:**
  - **types of data to be used;**
  - **types of analyses to be performed; and**
  - **methods used to encode the data.**
- **Geographic Information is organized as a set of feature classes (e.g. roads, streams, landuse, etc.) or data layers.**

This slide gives two other points which influences the definition of GIS applications.

Slide 7

Data Layers

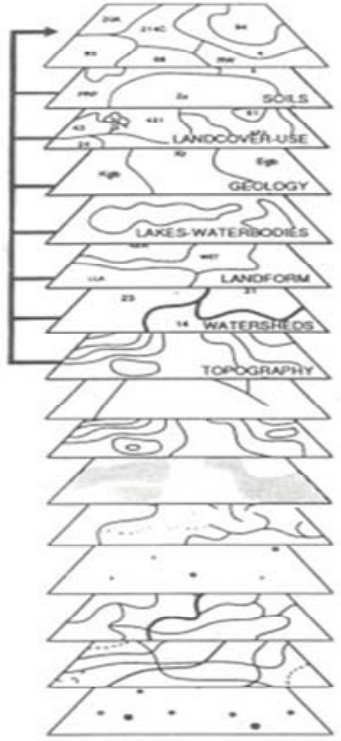
" A data layer consists of a set of logically related geographic features and their attributes"

- the features to be grouped in a single data layer are chosen for the convenience of the user
- organization principle may be to group similar feature types
- For example: roads and railways combined as a single transportation data layer; streams and lakes as a hydrology data layer
- The following two slide gives an example of data layers

This slide gives a definition of data layers. Some GIS vendors refer to data layers by different terminologies such as: data sets; feature classes.

Slide 8

GIS data layers commonly used in natural resource GIS applications



LAYER NAME:	ATTRIBUTES:
TERRAIN UNITS (polygons) Input Scale: 1:50,000 to 1:100,000	-Soil Types Component Texture Depth Slope Drainage Erosion pH Nitrogen Phosphorus Potassium -Landcover/Use Dominant Species Canopy Closure Stem Density Mid-Story DBH -Geology -Lakes and Waterbodies -Landform -Watershed Basin -Topography Type
FAULT (lines) Input Scale: 1:50,000 to 1:100,000	- Type - Name - Hazard
ELEVATION (lines and points) Input Scale: 1:50,000 to 1:100,000	- Elevation
SLOPE-ASPECT (polygons) Input Scale: 1:50,000 to 1:100,000 Derived from ELEVATIONS	- Slope - Aspect - Surface-area
STREAMS (lines) Input Scale: 1:50,000 to 1:100,000	- Name - Type - Width - Periodicity - Order
WELLS-GAUGING STATIONS (points) Input Scale: 1:50,000 to 1:100,000	- Well Number - Gauging Station Number - Springs - Basin Number
OWNERSHIP/ADMINISTRATIVE (polygons) Input Scale: 1:50,000 to 1:100,000	- Owner(s) - Township/Range - Section - County - Local Districts - State Districts - National Districts - Mineral Leases
TRANSPORTATION LINES (lines) (Roads, railroads, etc.) Input Scale: 1:50,000 to 1:100,000	- Name - Type - Width
SETTLEMENT/POINTS OF INTEREST (points) Input Scale: 1:50,000 to 1:100,000	- Type - Description (house, historic, archaeological)

L155 - GIS Data Models and Data Processing
Lecture 5
Dr. Steve Ramroop

8

This is an example of various data layers used in a natural resource GIS application. Notice the multitude of data layers and its attribute table definitions.

Slide 9

- **GIS Data Manipulation And Analyses**
  - **some common GIS processes the ability to perform the following:-**
    1) **Coordinate Transformations**
    2) **Map join / Edge Matching**
    3) **Windowing**
    4) **Coordinate Filtering**
    5) **Theissen Polygons**
    6) **Measurements**
    7) **Optimal Path Selections**
    8) **Digital Terrain Analyses**
    9) **Statistical Analyses**
    10) **Attribute Analysis Functions**
- **Each of these GIS processes are presented in the following slides.**

This slide identifies a few of the common analyzing capabilities that are typically available in GIS software. Each of the GIS analyzing capabilities is presented for the remaining of this lecture. Note that there is a multitude of analyzing capabilities available in GIS software. The more you work with GIS software the more one will learn about the various analyzing capabilities.

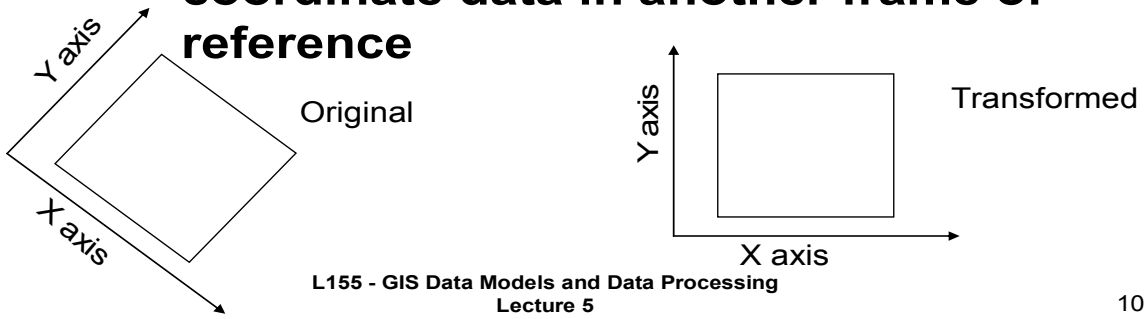
Slide 10

1) Coordinate Transformations

- **Mathematical expression or formula:**

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} a1 & a2 \\ a3 & a4 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} AX \\ AY \end{bmatrix}$$

- **It is used to convert coordinate data within one frame of reference to coordinate data in another frame of reference**



All GIS software must be able to transform from one coordinate system into another coordinate system. There is a rotation and a translation of the coordinate systems. These slide shows the general form of the equation that has the rotation and translation matrix from one coordinate system to another. Such an equation is coded into the analyzing capability of the GIS software.

Slide 11

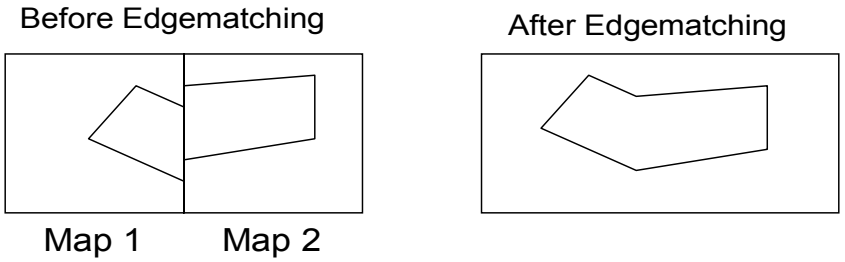
- **Used for a variety of applications**
  - **Changing one map projection to another map projection**
  - **Changing digitizer coordinates to UTM**
  - **Removal of distortion in your documents**
  - **Changing scale, rotation and shift**
- **Transformation can be:**
  - **Conformal**
    - » **Scale, rotation & shift are maintained**
  - **Affine**
    - » **Scale, rotation, shift with stretch are maintained**
  - **Polynomial**
    - » **Higher order transformation**

Transformations of the coordinate systems are a common analyzing capability in GIS software. The reasons from using transformations are identified on this slide. There are three general types of transformation (Conformal, Affine, Polynomial), and their functions differ (as shown on this slide).

Slide 12

2) Map Join/Edge match

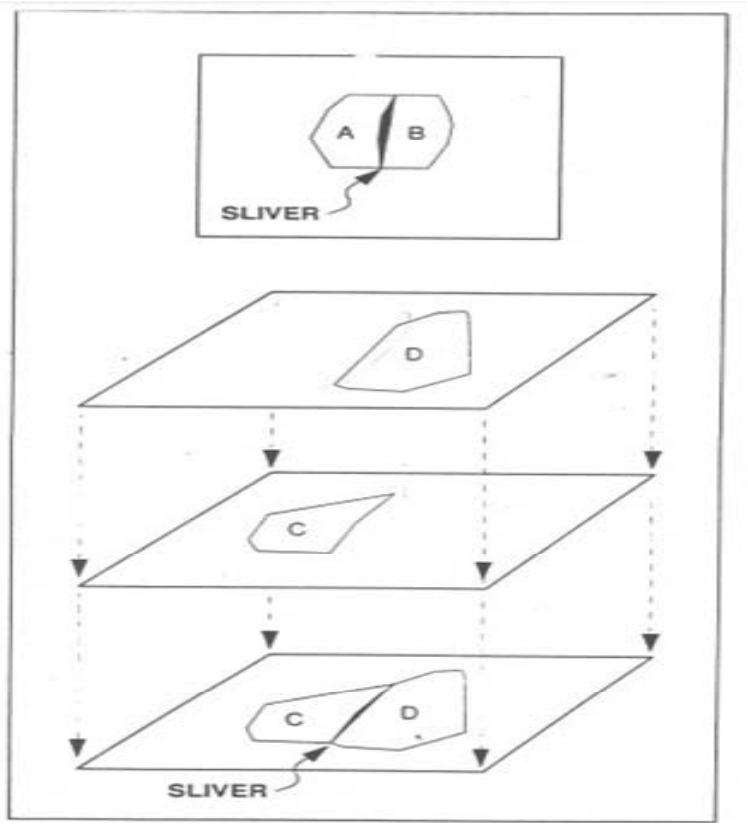
- **This is the process of removing inconsistencies at the edges of digital maps from compiled map data so that features match across map sheets**
- **Edge match, map join and merge all constitute to one function:**
  - **Edge match can be considered as a tie**
  - **Map join is putting them together**
  - **Merge is a global joining of a group of maps**



This slide gives another common analysis function on all GIS software. This is the ability to join multiple neighboring maps of a map series.

Slide 13

This figure shows the error called "slivers". It occurs when feature boundaries are digitized twice which leaves an unnoticeable polygon.



L155 - GIS Data Models and Data Processing
Lecture 5
Dr. Steve Ramroop

13

When maps are merged there are usually the unwanted polygons at the merged boundaries that were digitized more than once. These errors are called sliver polygons and they must be removed.

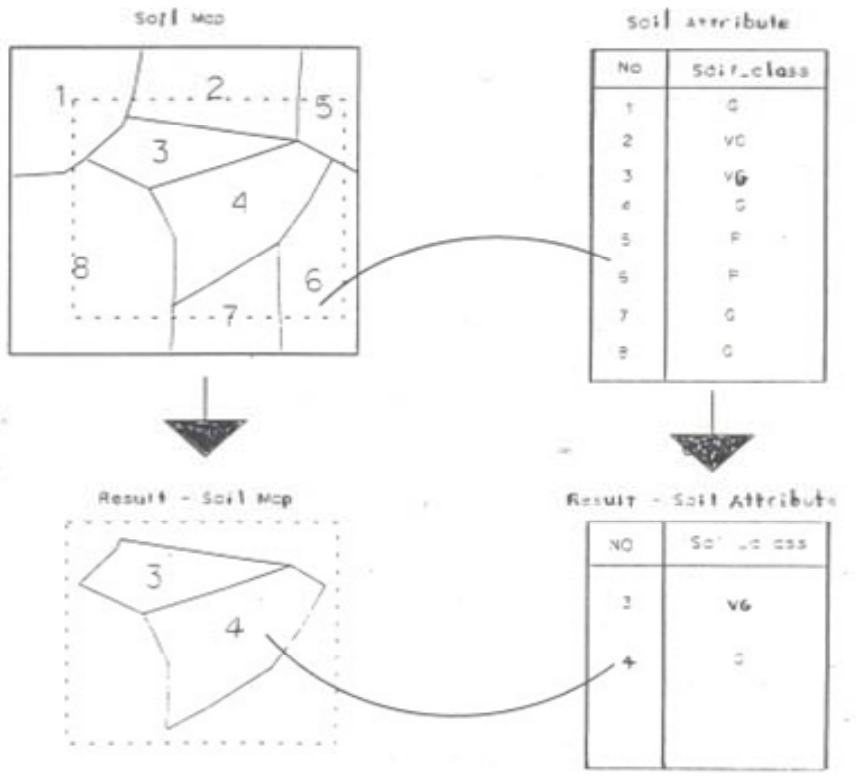
Slide 14

3) Windowing

- **A function that allows the user to define a specific geographic area which is used to delimit (select or define) a piece of map area**
- **The process involves maintaining the attributes with the map features**
- **Map features and associated attributes can retrieved using windows for display, analysis and manipulation**
- **The following slide shows a window example**

This is a common GIS analysis capability which is available in all GIS software. That is the ability to define a selection window to select the spatial data and at the same time the attribute data. Windowing can also be the ability to define a window to zoom into more details on the digital map.

Slide 15

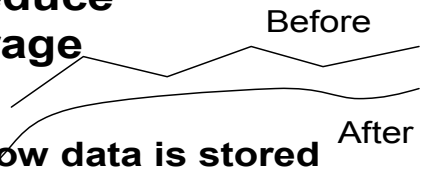


This slide gives an example of defining a window and the results obtained. Notice that the results show only the polygons that fall within the window (shown as the dotted line) and its associated attributes in the attribute table.

Slide 16

4) Coordinate filtering

- **The process of weeding out superfluous coordinates**
- **Used to generalize maps and reduce requirements for computer storage**
- **Other terms used are:**
  - **Data compression – based upon how data is stored**
  - **Line thinning – that is removing unwanted points**
- **Criteria:**
  - **Distance interval – based upon distance traveled**
  - **Time interval – based upon time lapsed**
  - **Angles – based upon features angles of orientation**
  - **Threshold (2nd difference-gradient and reflection point)**



Another GIS analyzing capability called Coordinate filtering is shown on this slide. This is the ability to smooth the vector data collected usually when collected using a digitizing table and at the same time minimize the storage of the data set. Filtering can be based on the distance traveled, time traveled, angles, and by setting thresholds.

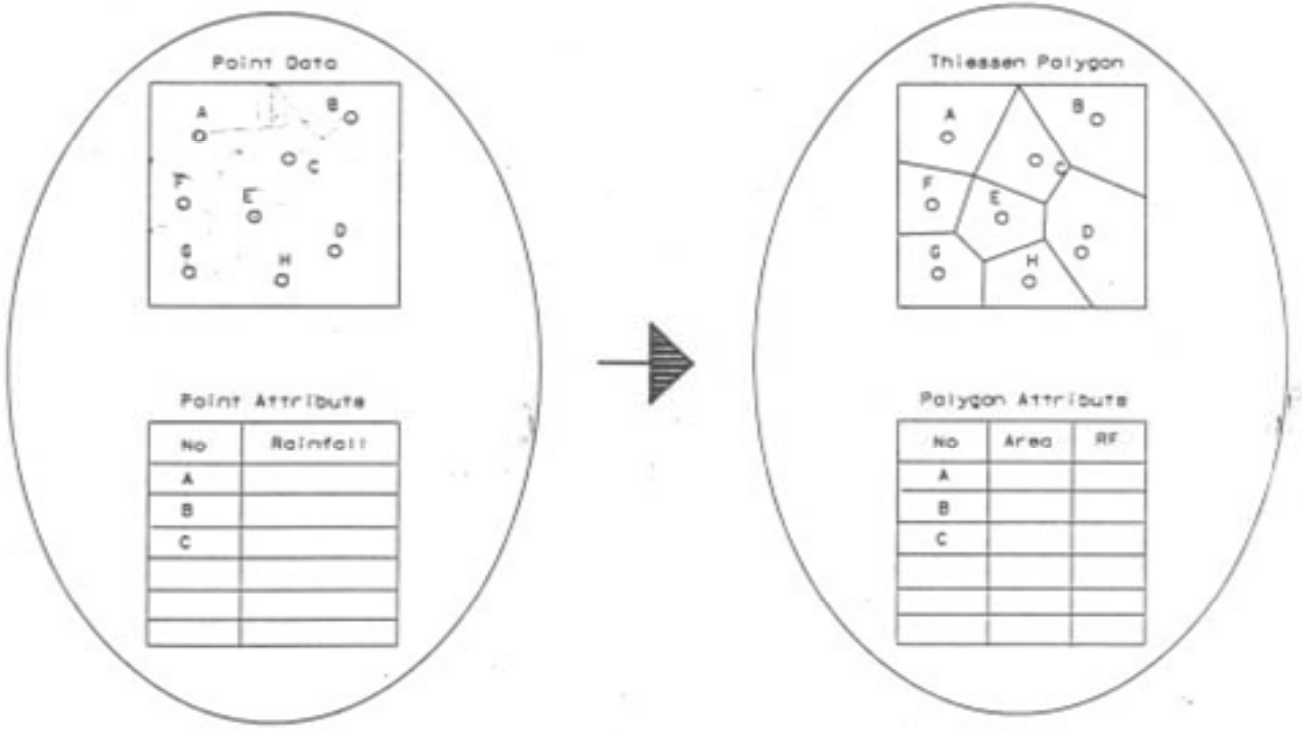
Slide 17

5) Thiessen polygon

- Creation of polygons or areas around randomly spaced point location
- It is carried out by dividing equally the distance between paired points, then generating perpendicular lines, to these midway points which are then extended to intersect and form areas
- The process is used for creating polygons, mapping and analyzing qualitative data where continuous coverage map is desired, but where contouring is not appropriate

All GIS software has the ability to generate theissen polygons. This involves the creation of polygons from points containing attribute data. This is the creation of a continuous polygon data layer.

Slide 18



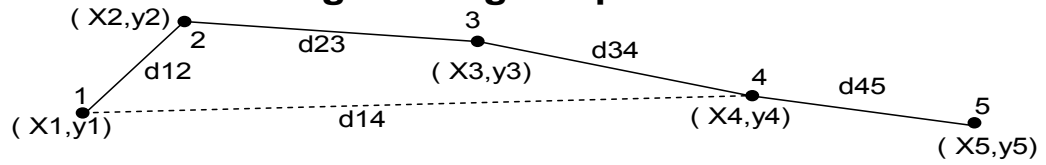
This slide shows the generation of theissen polygons from point data.

Slide 19

5) Measurements

a) Distance calculation

- **Distance can either be the distance between successive coordinate pairs forming a straight line or following an irregular path**



Distances are calculated

$$d12 = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$

$$d23 = \sqrt{(x3 - x2)^2 + (y3 - y2)^2}$$

$$d34 = \sqrt{(x4 - x3)^2 + (y4 - y3)^2}$$

The ability to measure distances is an important analyzing capability in all GIS software. This slide gives the basic understanding of how the distances are computed—that is, either along a given line or straight line distances. The formulae are coded into the GIS software analyzing capability.

Slide 20

- **Total distance along the path (1-2-3-4) is (d12+d23+d34)**
- **Straight distance between (1) and (4) is given by:**

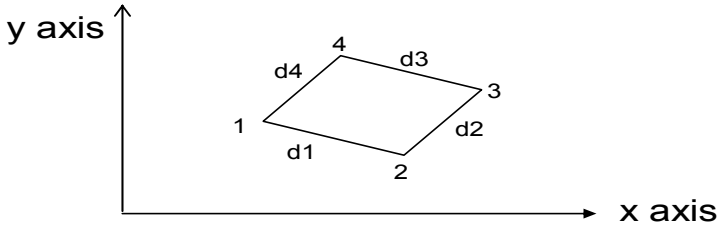
$$d34 = \sqrt{(x4 - x3)^2 + (y4 - y3)^2}$$

This slide shows the formula for straight line distances are coded into the GIS software analyzing capability.

Slide 21

b) Area / Perimeter calculation

- **if the coordinates of points which form the polygon are known, then the area can be calculated using the formula given below**



Area = ½ {(Sun of parallel sides X Height)}

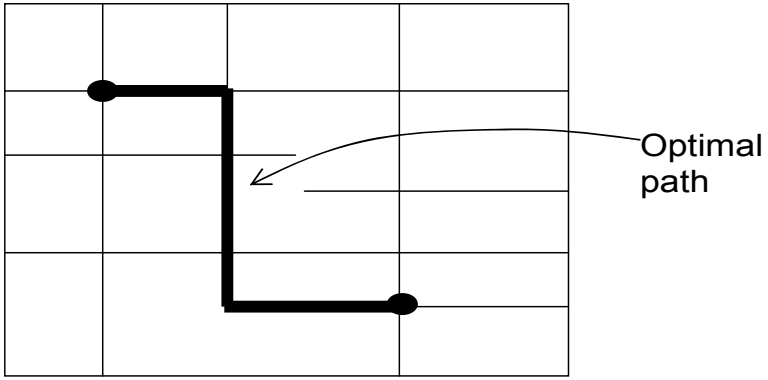
Perimeter = d1 +d2 +d3 + d4

Another common measurement in all GIS software is the ability to compute the area and perimeter in a polygon data layer. Area is typically computed using the area trapezium rule. Such a rule is coded into the GIS software analyzing capability.

Slide 22

7) Optimal path selection

- **This is the process of using a network database of road and road intersections to select the optimal path based on a specified time and distance relationship.**



By specifying origins and destinations points on a road network database, the GIS software will be able to delineate the shortest path. This is typically called network analysis. The shortest path can also take into consideration multiple attribute information such as road width, speed limit, time of day, and such like.

Slide 23

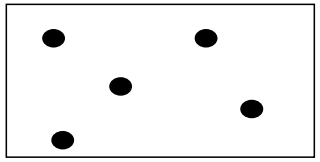
8) Digital terrain analysis

a) Spatial interpolation

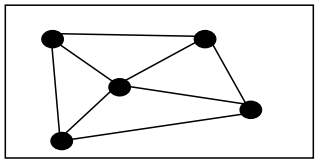
- To interpolate heights between point of known elevation for the purpose of establishing contours or elevation values between known points
- This is done using randomly distributed point data to derive elevation into a regular grid or triangulated surface or facets

Digital terrain analysis refers to many other types of analysis capabilities. The first type is the spatial interpolation capability. This is the ability to perform analysis to create a continuous surface, for example from spot heights and contour lines. The continuous surface is called a Digital Elevation Model (DEM).

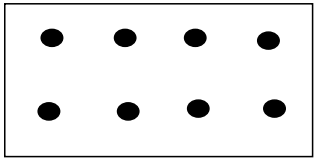
Random point



Triangular network



Regular grid



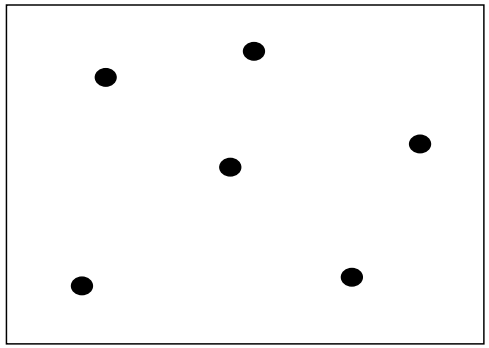
- **The most common methods of interpolations (which uses the general trend in the neighborhood) are:**
  - **Linear interpolation**
  - **Windowing (filtering)**

This slide identifies a number of data sources which can be used in the creation of a DEM. The methods of interpolations can be a:

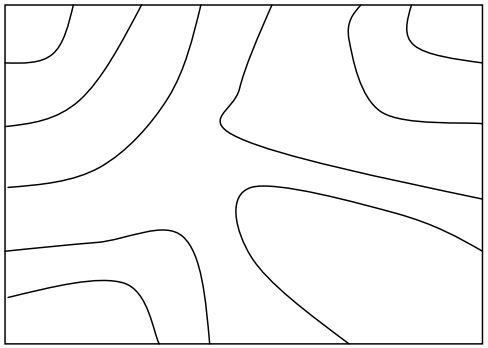
- Linear Interpolation:
  - This involves the approximation of neighboring values from know values.
- Windowing:
  - Involves the use of filters that takes into consideration the neighboring values in order to apply values to unknown points. It travels from one pixel to the other until the entire raster data set is interpolated.

b) Contouring

Random points



Contour lines



c) Slope, gradient and aspect

- These are different views of a surface
  - Slope addresses the steepness of the area
  - Gradient is the maximum slope
  - Aspect relates to the direction in which the gradient is oriented relative to north

The second and third types of the spatial interpolation capabilities are shown in this slide:

Contouring:

- The ability to generate contour lines is another analysis capability from random points, or from a continuous surface.

Slope, gradient and aspect:

- Another analysis capability is the ability to prepare maps of slope, gradient, and aspect by using a DEM.

Slide 26

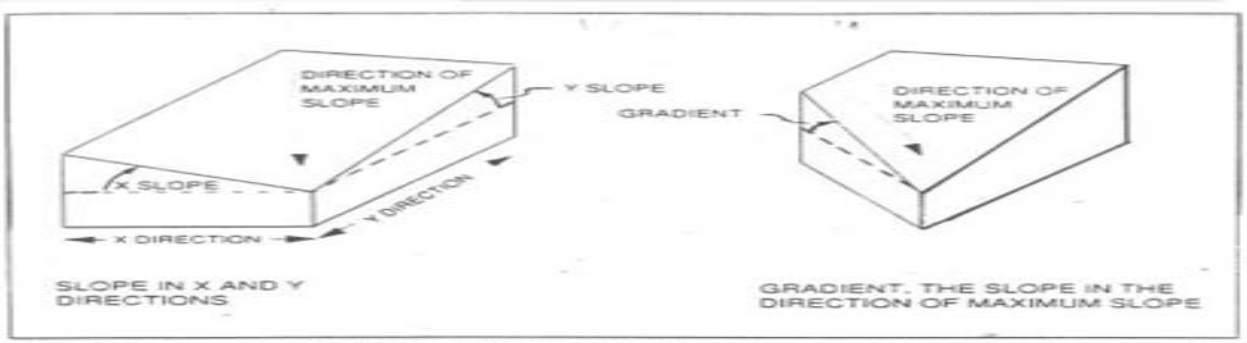


Figure 7.22 Measurement of Slope. The slope at a point is the angle measured from the horizontal to a plane tangent to the surface at that point. The value of the slope will depend on the direction in which it is measured. Slope is commonly measured in the direction of the coordinate axes e.g. in the X-direction and Y-direction. The slope measured in the direction at which it is a maximum is termed the **gradient**.

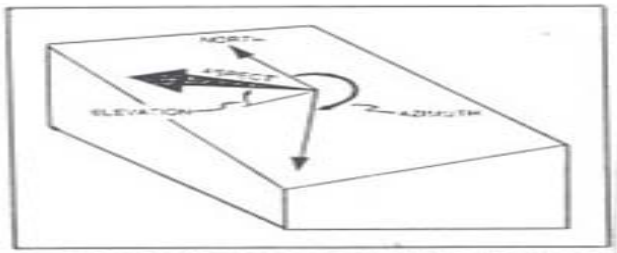


Figure 7.23 Measurement of Aspect. There are two components in the measurement of aspect. The horizontal angle or azimuth is the aspect direction. It is the angle formed by moving clockwise from north to the direction of maximum slope. The vertical angle or elevation angle is measured from the horizontal to a line drawn perpendicular to the surface.

This slide shows slope, gradient, and aspect.

Slide 27

d) Sun intensity

- It is calculated based on angle of slope, direction of slope, and location of the sun relative to that slope.
- Sun estimate and angle are calculated based on time and hour of day and comprises of the angle of incidence of the sun.

e) Watershed analysis

- The use of a digital elevation model (DEM) to determine the water flow and other hydrological characteristics.

f) Viewsheds

- Using a DEM, areas of visibility from any given observation point can be identified.

g) Cross-sections

- Also called ground profiles.
- It is the sectional view of the terrain topography.

h) 3D Viewing

- This is the generation of three dimensional views of the data using a DEM.

This slide identifies some other digital terrain analysis capabilities that make use of a DEM.

Slide 28

9) Statistical analysis

- **This concerns with the aggregation of data according to well defined statistical procedures in order to provide the user with a better perspective on the collection of data**
- **It includes a number of well defined functions:**
  - **Mean**
  - **Standard deviation**
  - **Regression and correlation**
  - **Hypothesis testing …..**
  - **etc….**
- **Data is viewed as a collection rather than as individual entities**
- **It is important to understand what is being done by the functions in order to be able to properly interpret the result of the analysis**

Another GIS software analysis capability is through statistical analysis. Most GIS software has the ability of calculate mean, standard deviation, and other simple statistical analysis. The GIS software has the ability to integrate with other third party statistical software such as SPSS.

Slide 29

10) Attribute analysis functions

a) Attribute editing functions:

- Add : that is to add new attributes
- Update : to make changes to existing data
- Delete : to remove data

b) Attribute query functions

- Based upon the operators used to build the queries that are used to query databases
  - Examples : AND, OR, >, <, =, ≥, ≤, ≠, etc.
- Queries are typically based using the *"Structured Query Language (SQL)"*

The ability to add, update and delete attribute data is a must in all GIS software. In addition, it is also important to be able to build queries using Structured Query Language (SQL), which queries the spatial and attribute databases.

… The End …