GIS Analysis

DR. APARNA S. BHASKAR

Contents

SPATIAL DATA ANALYSIS

Spatial analysis is the process by which turn raw data into useful information.

The study of the locations, shapes of geographic features and the relationships between them.

Spatial analysis generally requires both non-spatial (attributes) and spatial information (location of objects).

The principal objective of spatial data analysis is to transform and combine data from diverse sources/disciplines into useful information, to improve one's understanding or to satisfy the requirements of objectives of decision makers.

Spatial analysis is how we understand our world mapping where things are, how they relate, what it all means, and what actions to take.

WHAT WE DO WITH SPATIAL ANALYSIS?

- Identify high crime area
- Selection of a best location for a new business
- Spread of disease
- Suitable site determination for garbage
- dumping
- Identifying the Bus routing (Shortest path)
- Flood vulnerable zones identification and etc..

What is spatial interpolation?

Spatial interpolation is the process of using points with known values to estimate values at other points.

These points with known values are called **known points, control points, sampled points, or observations.**

Spatial interpolation is a process of intelligent guesswork in which the investigator attempt to make a reasonable estimate of the value of a continuous field at places where the field has not actually been measured.

Spatial interpolation is therefore a means of creating surface data from sample points so that the surface data can be displayed as a 3-D surface or an isoline map and used for analysis and modeling.

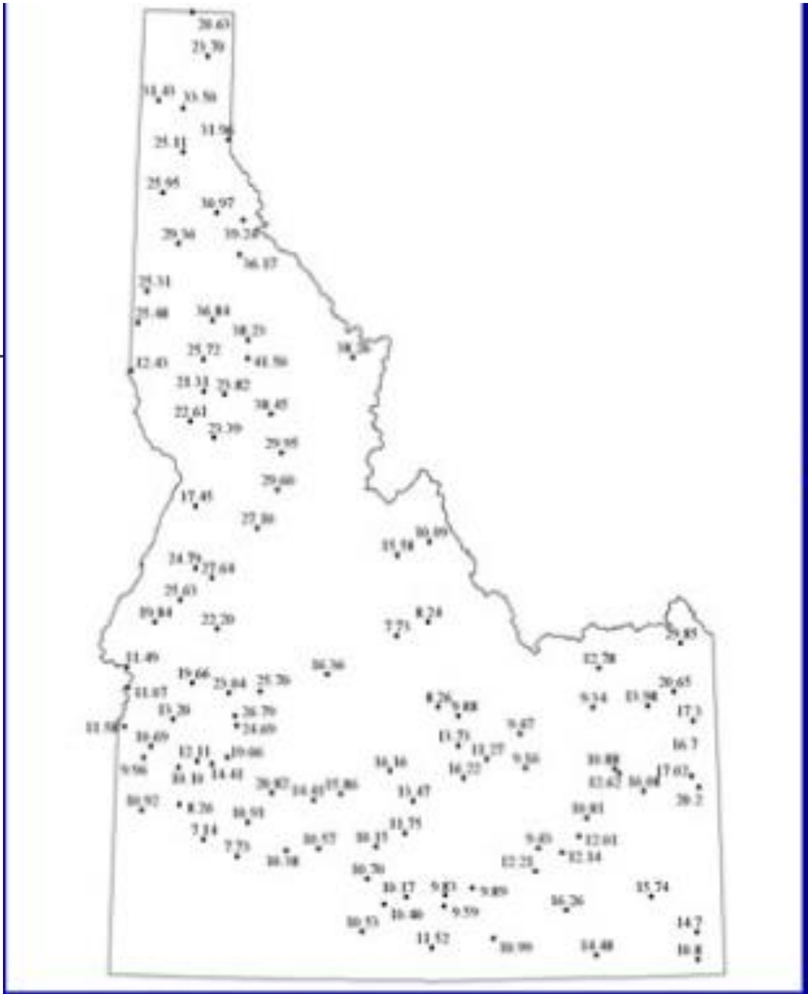
Applications

In estimating rainfall, temperature and other attributes at places where no direct measurements of these variable are available

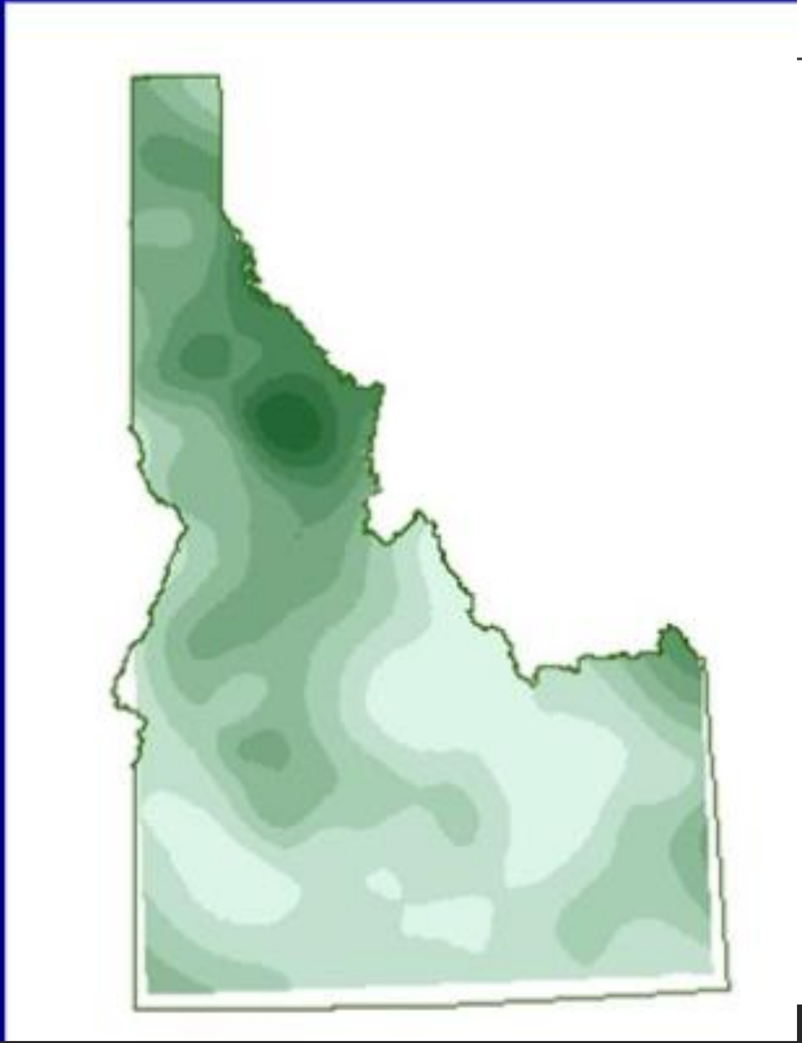
In estimating the elevation of the surface between the measured locations of DEM

In contouring when it is necessary to guess where to place contours between measured locations

In resampling raster, the operation that must take place whenever raster data must be transformed to another grid



A map of 105 weather stations in Idaho and their 30-year average annual precipitation values



Spatial interpolation

Type of spatial interpolation

1. Global interpolation and local interpolation
2. Exact interpolation and inexact interpolation
3. Deterministic interpolation and stochastic interpolation
4. Abrupt *vs* Smooth

Global interpolation:

Use every point available to estimate an unknown value.

Global interpolators determine a single function which is mapped across the whole region

- e.g. Trend surface

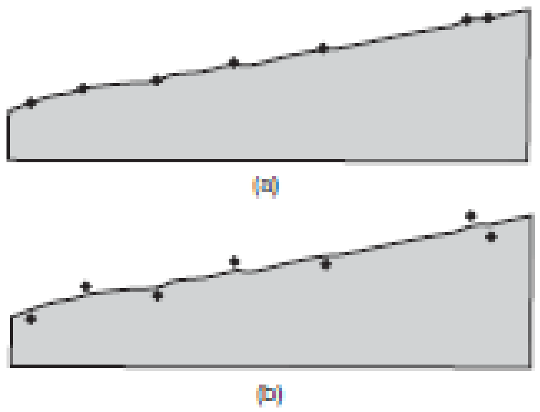
Local interpolation:

Uses sample of known points to estimate an unknown values.

Local interpolators apply an algorithm repeatedly to a small portion of the total set of points

- e.g. Inverse distance weighted

Global interpolation:	Local interpolation:
More number of control points are considered	Less number of control points are considered
Designed to capture general trend of surface	Designed to find short range variation
Less accuracy	More accuracy



Exact interpolation:

Predicts a value at the point location that is same as its known value.

Exact interpolators honor all data points

- e.g. Inverse distance weighted

Inexact interpolation:

Predicts a value at the point location that differs from its known value.

Approximate interpolators try to approach all data points

- e.g. Trend surface

Exact interpolation	Inexact interpolation
Generates a surface that passes through the control points	Generates a surface that slightly differs from the control points

Deterministic interpolation:

A deterministic interpolation method provides no assessment of errors with predicted values.

Stochastic interpolation:

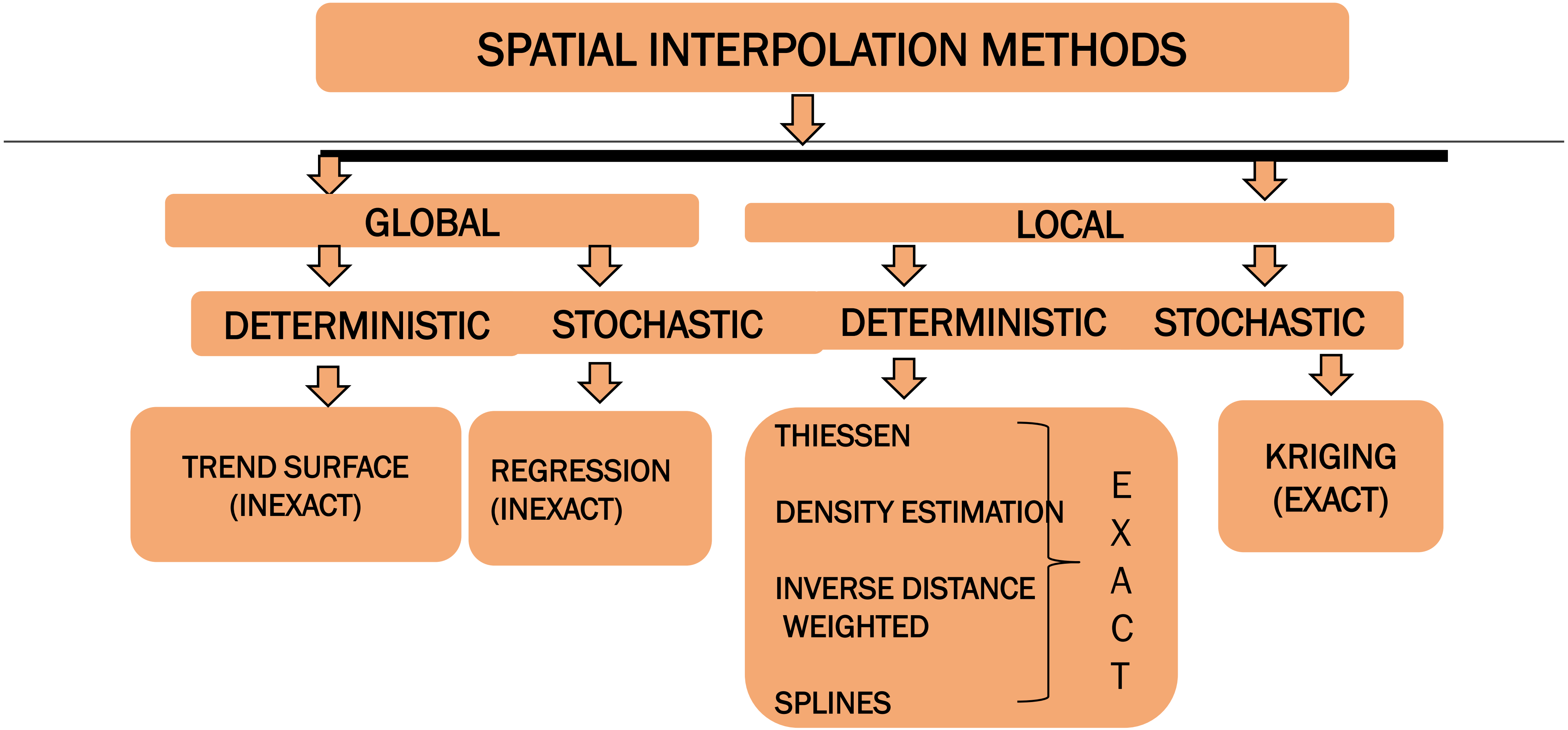
A stochastic interpolation method, offers assessment of prediction errors with estimated variances.

Gradual/Abrupt Interpolators

Gradual/smooth interpolators assume continuous and smooth behavior of data everywhere.

Abrupt interpolators allow for sudden changes in data due to boundaries or undefined derivatives.

CLASSIFICATION OF SPATIAL INTERPOLATION METHODS



GLOBAL METHODS

I. TREND SURFACE:

Trend surface analysis approximates points with known values with a polynomial equation.

A linear or first order trend surface uses the equation :

$$Z_{x,y} = b_o + b_1x + b_2y$$

Where, Z is a attribute value with function x and y coordinates

b_o , b_1 and b_2 are the coefficients estimated from known points

Example: precipitation

The trend surface model is computed by least square method, the ‘goodness of fit’ of model can be measured and tested.

If the distribution of most natural phenomenon is more complex than an incline lane surface from first order model.

Hence higher order trend surface is required to approximate more complex surfaces.

Example: hills and valleys [Third order model]

GIS package offers upto 12th order trend surface models

II. **REGRESSION MODEL:**

A regression model relates a dependent variable to a number of independent variables in a linear equation (an interpolator) which can be then used for prediction and estimation.

This model uses both spatial and nonspatial attributes for the interpolation.

$$\mathbf{Z = b_o + b_1x + b_2y + b_3e + b_4p}$$

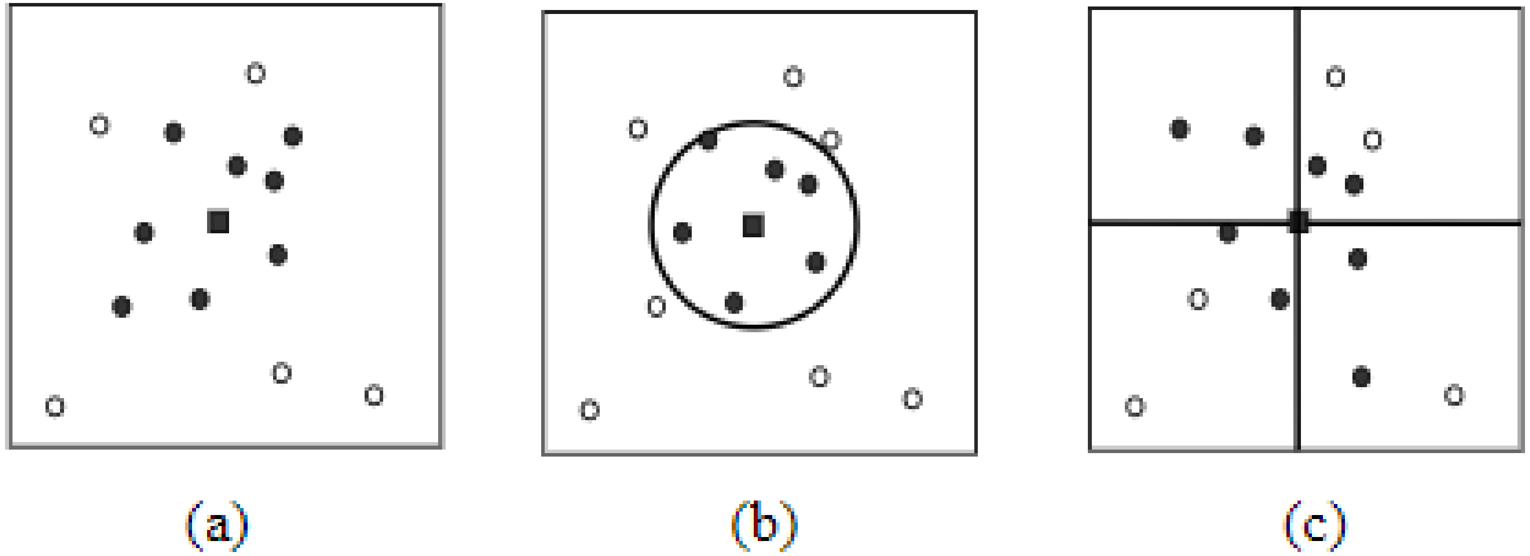
Where, Z is a dependent variable

b_o, b_1, b_2, b_3 and b_4 are coefficients and

x, y, e and p are the independent variables

Example: snow water equivalent

Local methods



Find the closest point to the point to be estimated

Points within the radius

Points within each quadrant

I. THIESSEN POLYGON:

Thiessen polygon interpolation also known as **Nearest neighbor interpolation**.

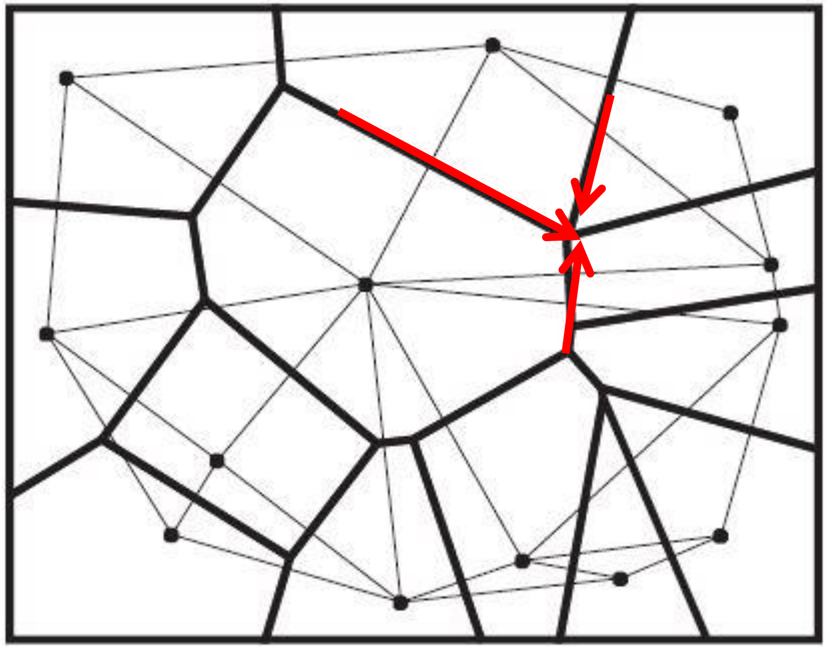
This is conceptually the simplest interpolation method, in the sense that the mathematical function used is the simple equality function, and only one point, the nearest point, is used to assign a value to an unknown location.

Thiessen polygon interpolation do not use an interpolator but require initial triangulation for connecting known points can form different set of triangles (Delauney triangulation).

After triangulation, Thiessen polygon can be easily constructed by connecting lines drawn perpendicular to the sides of each triangle at their mid points.

Example: Average areal precipitation

Thiessen polygon (thick) are interpolated from known points and Delaunay triangulation (thinner).



II. DENSITY ESTIMATION:

Density estimation measures cell densities in a raster by using a sample of known points.

There are two types of Density Estimation Methods:

1. **Simple density estimation method**
2. **Kernel density estimation method**

III. **INVERSE DISTANCE WEIGHTED INTERPOLATION:**

The IDW is simple and intuitive deterministic interpolation method based on principle that sample values closer to the prediction location have more influence on prediction value than sample values farther apart. Using higher power assigns more weight to closer points resulting in less smoother surface.

On the other hand, lower power assigns low weight to closer points resulting in smoother surface. Major disadvantage of IDW is “bull's eye” effect (higher values near observed location) and edgy surface.

Inverse distance weighting results in smooth interpolated surfaces. The values do not jump discontinuously at edges, as occurs with Thiessen polygons, and sometimes with fixed radius interpolation.

IV. Splines

- The mathematical equivalent of using a flexible ruler (called a spline)
- Can be used as an exact or approximate interpolator, depending upon the degrees of freedom granted (e.g. polynomial order)
- Best for smooth datasets, can cause wild fluctuations otherwise

Kriging

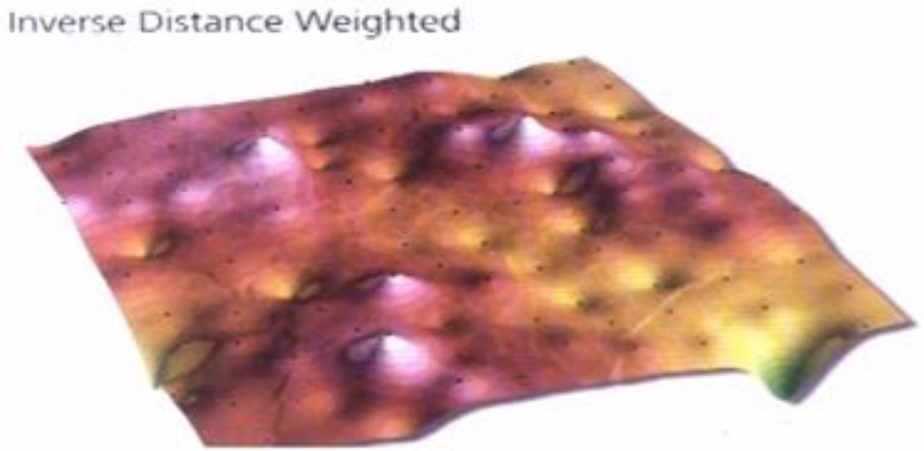
Developed by Georges Matheron, as the "theory of regionalized variables", and D.G. Krige as an optimal method of interpolation for use in the mining industry.

Kriging is the geostatistical method for spatial interpolation.

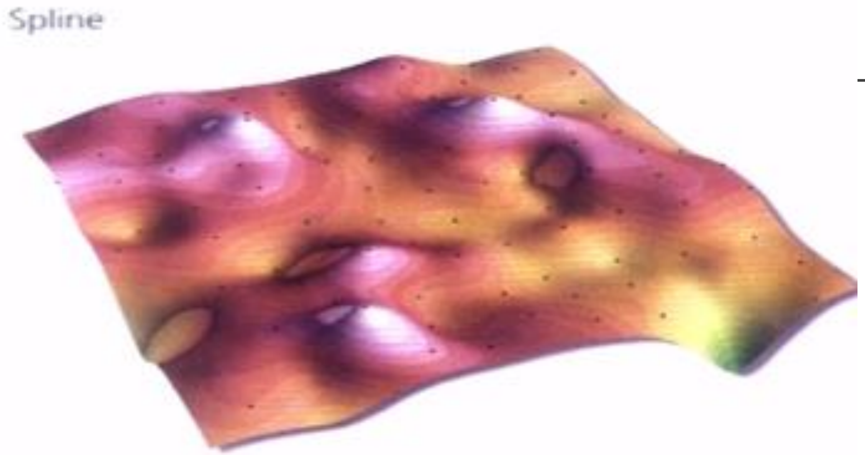
Basis of technique is the rate at which the variance between points changes over space

This is expressed in the variogram which shows how the average difference between values at points changes with distance between points

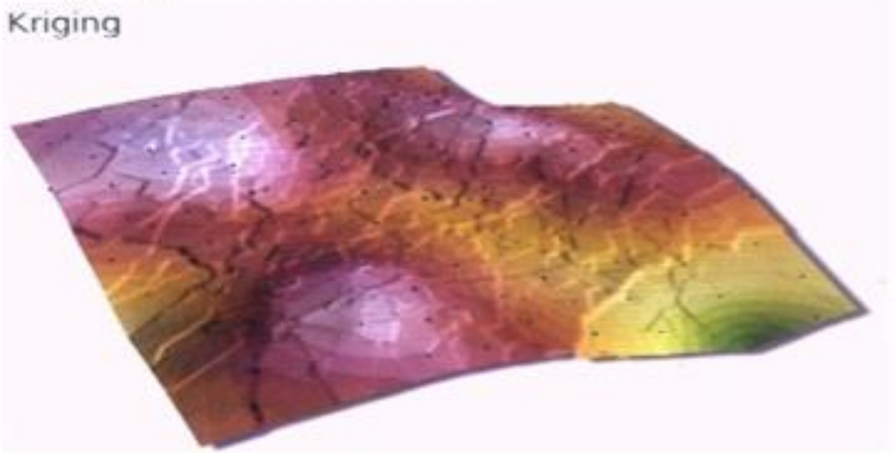
Inverse Distance Weighted (IDW) / Movir Average



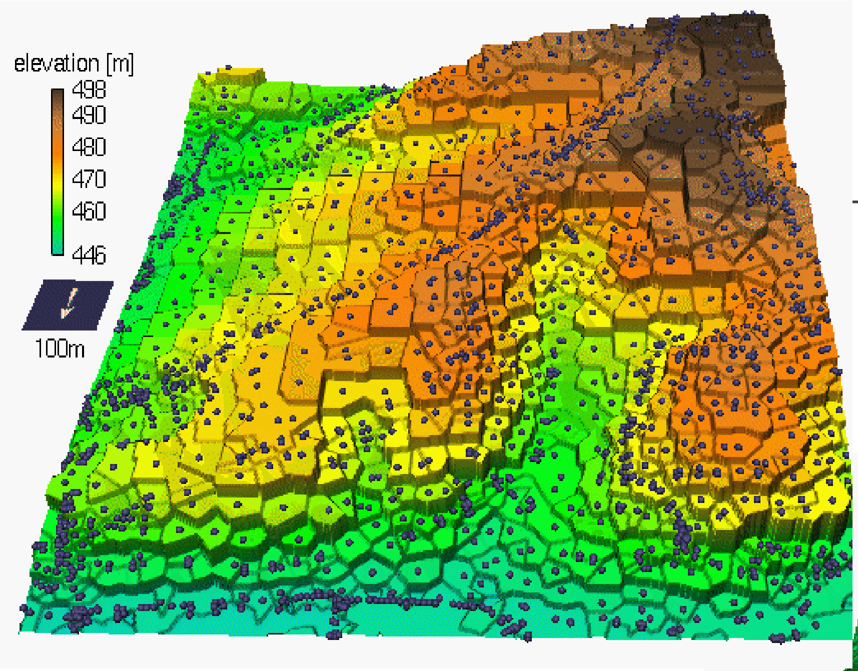
Spline / Moving Surface



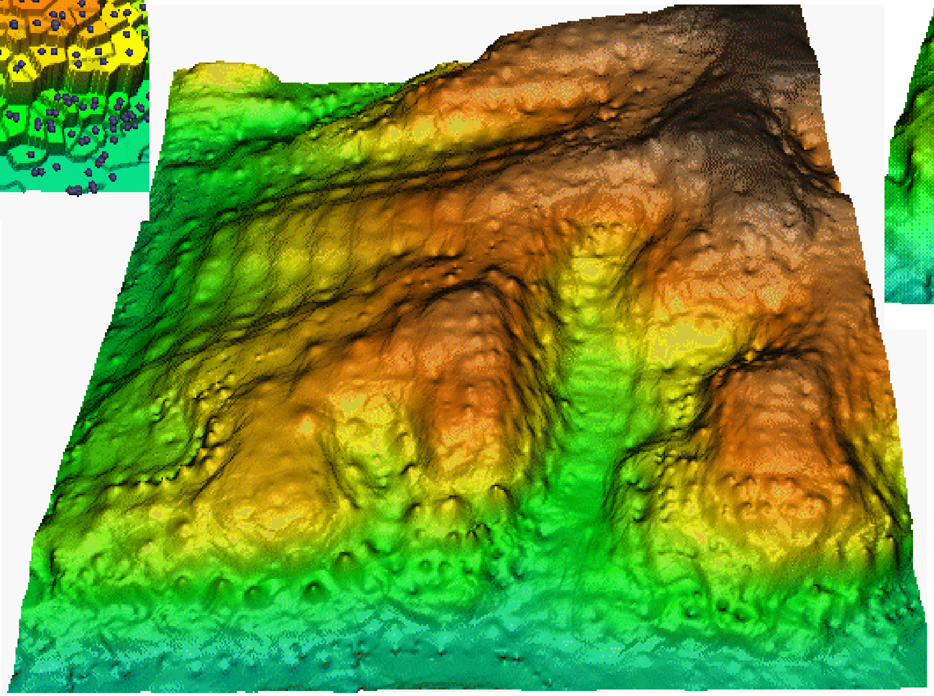
Kriging



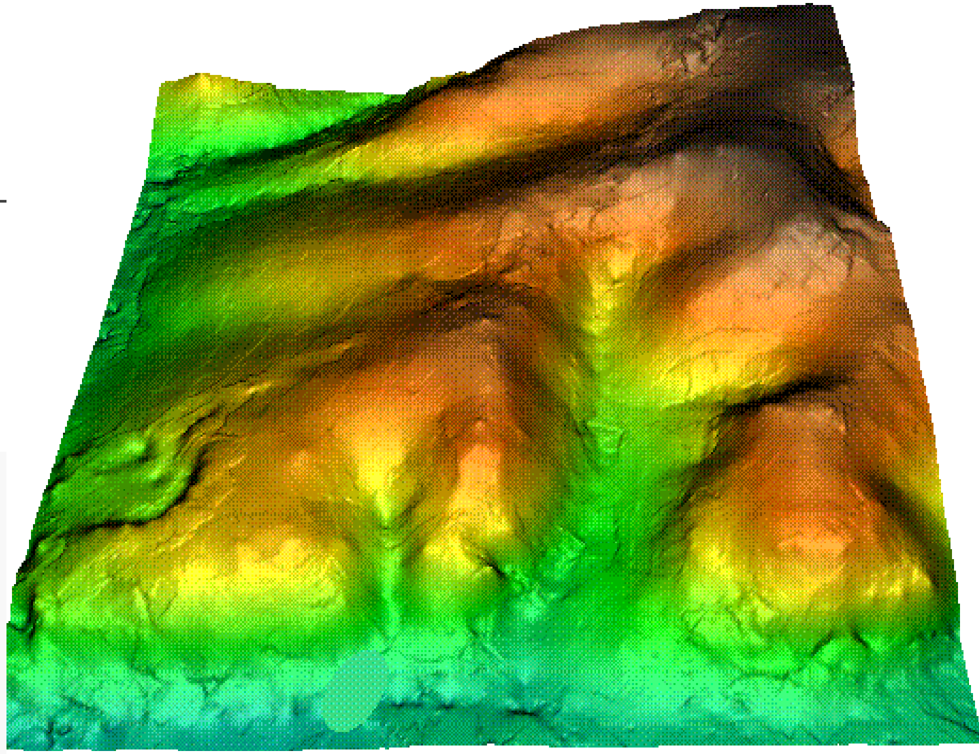
Theissen Polygon



Inverse Distance Weighting



Kriging



- Interpolation method depends upon
  - Character of data
  - Your assumptions of data behavior
- When possible, best way to compare methods is to
  1. try several methods
  2. make sure you understand theory
  3. refine best method

Reclassification Technique

The process of taking input cell values and replacing them with new output cell values

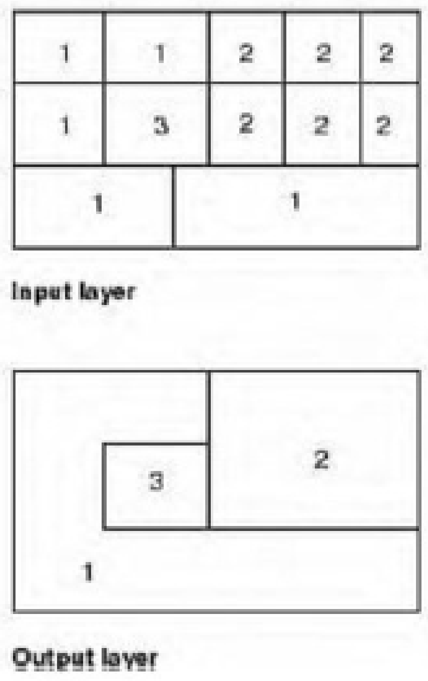
Reclassification is often used to simplify or change the interpretation of raster data by changing a single value to a new value, or grouping ranges of values into single values—for example, assigning a value of 1 to cells that have values of 1 to 50, 2 to cells that range from 51 to 100, and so on.

Reclassification serves three main purpose

- Creating a simplified raster: eg 1 for slopes 0-10%, 2 for 10-20%
- Creating new raster that contains a unique category or values: eg slopes of 10-20%, irrigated croplands
- Creating a new raster that shows the ranking of cell values in the input raster: eg 1 to 5, with 1 being the least suitable and 5 being the most suitable

RECLASSIFICATION

- Reclassification of attribute data by dissolving a part of the boundaries and merging into new reclassified polygons.
- Example: Population densities classified into classes such as "Sparsely populated" or "Over crowde



DATA RETRIEVAL

Data retrieval is the process of identifying and extracting data from database.

Data retrieval is feasible and easy if there is no ambiguity in the data and it should be presented in standard structure.

Data retrieval: 1. Locating, 2. Selecting by attributes 3. Buffering 4, Map overlay 5. Map algebra.

Data retrieval is done by applying the rules of Boolean logic to operate on the attributes.

DATA RETRIEVAL – SOURCES (Example)

Digital Library (Books, Journals)

Web Search engine (Flight ticket, Movie, Education)

USGS Data Library (Satellite image – Free of cost)

NRSC Data Center (Purchasing Indian satellite)

Etc..

BUFFER

A buffer is a zone with a width created around a spatial feature and is measured in units of distance from the feature.

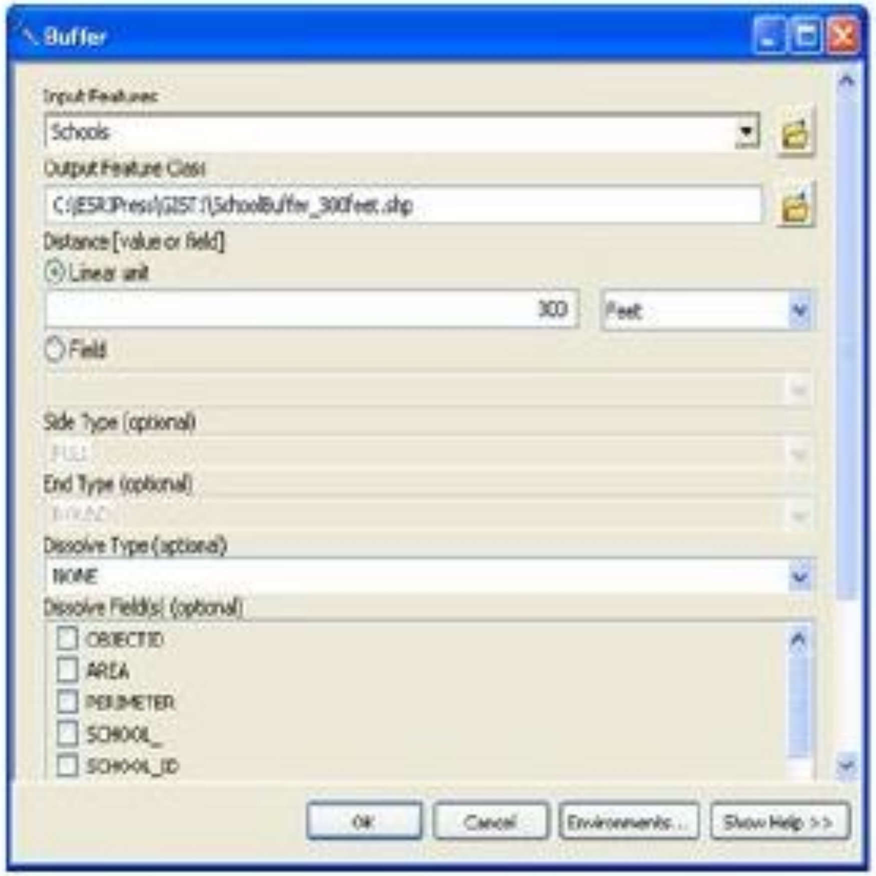
Buffering is used for neighborhood analysis which aims to evaluate the characteristics of the area surrounding the spatial feature.

Common examples of buffering include the identification of properties within a certain distance of an object,

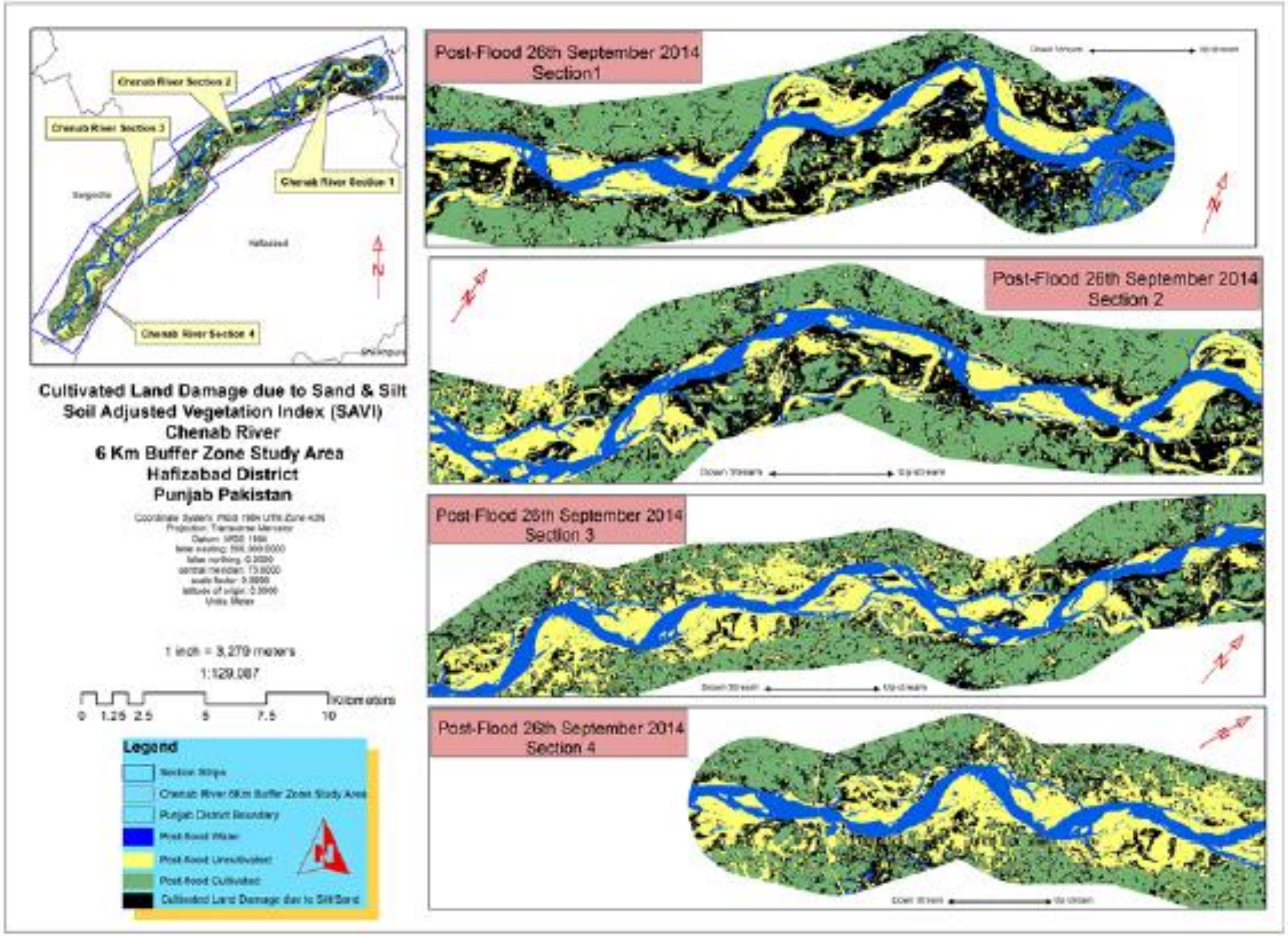
Delineation of areas around natural features where human activities are restricted, determination of areas affected by location etc

BUFFERING - POINT

- Crimes near schools



BUFFERING – AREA (FLOODING)



Query

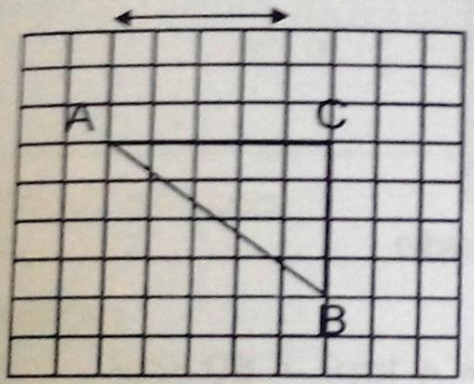
Query is a request to select features or records from a database is known as a Query.

Query is often written as a statement or logical expression. Analysing the data by querying it is Known as the Query Analysis. This is an immensely powerful feature that allows you to view a subset of the layer without selecting & exporting those features to a new layer each time. Three types of Queries are there (a) Interactive (b) Selection (SQL) (c) Spatial (Location) and Graphical

Spatial Measurement Methods

- Measurements allow to produce ratios of lengths to widths and of perimeters to areas
- The GIS user need to describe not only what objects are, how many objects exist and where they are, but also how large they are, how far apart and what the distance between them is like
- Method of measurements depend upon the type of data, data model and availability of software

In raster, measurements are done using Pythagorean geometry



AC = 5 grid cells × 2km/side

= 10 km

BC = 4 grid cells × 2 km/side

= 8 km

$$AB = \sqrt{AC^2 + BC^2}$$

$$= \sqrt{100 + 64}$$

$$= \sqrt{164}\ \text{km}$$

Fig 1: Length Calculation

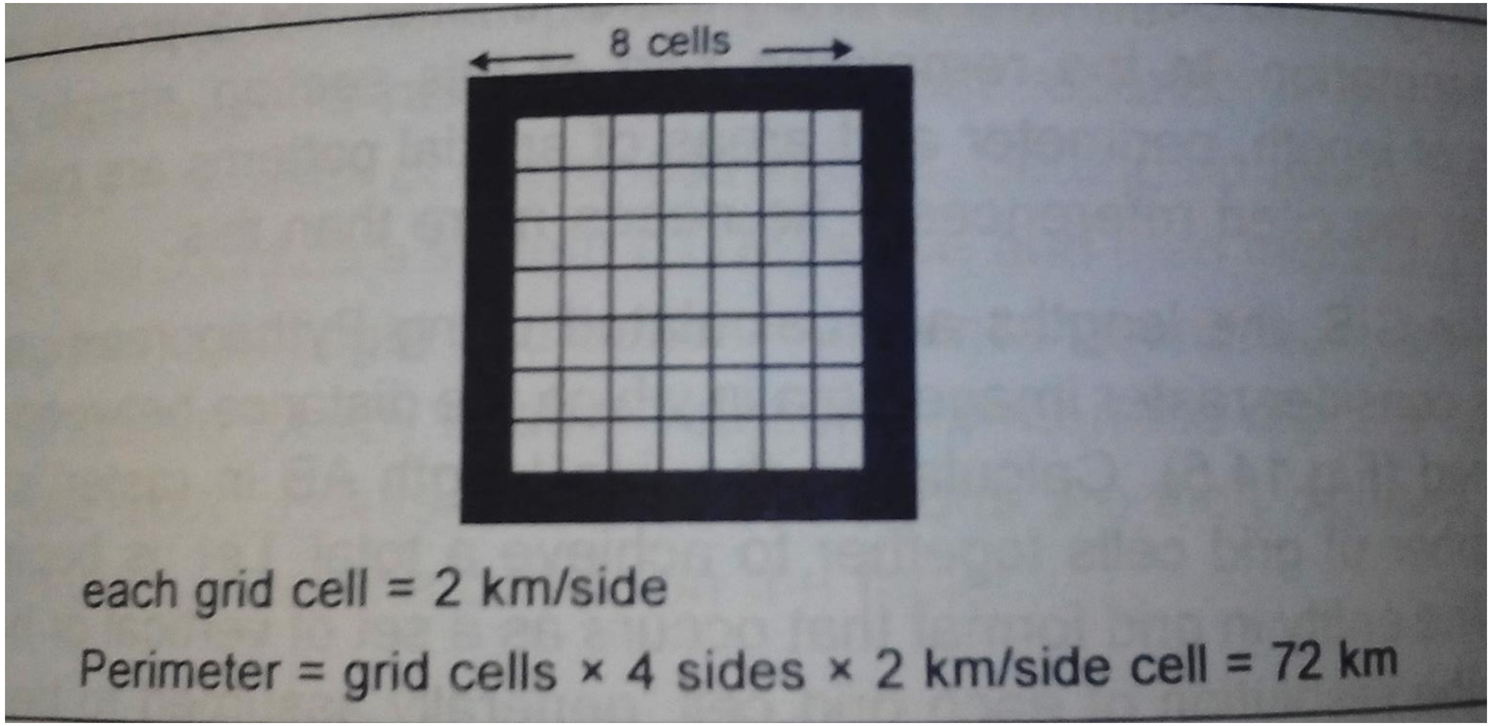


Fig 2: Perimeter Calculation

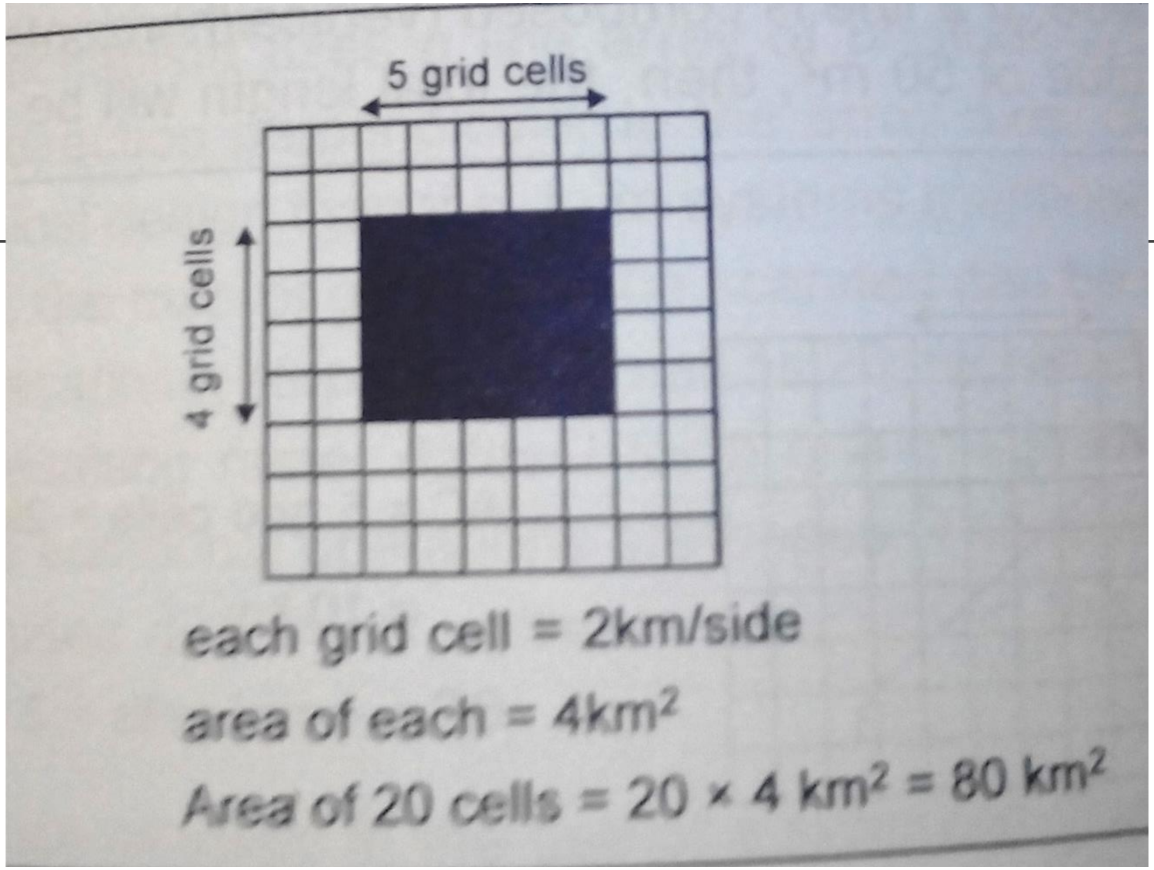


Fig 3: Area Calculation

In vector, measurements are also done using Pythagorean geometry

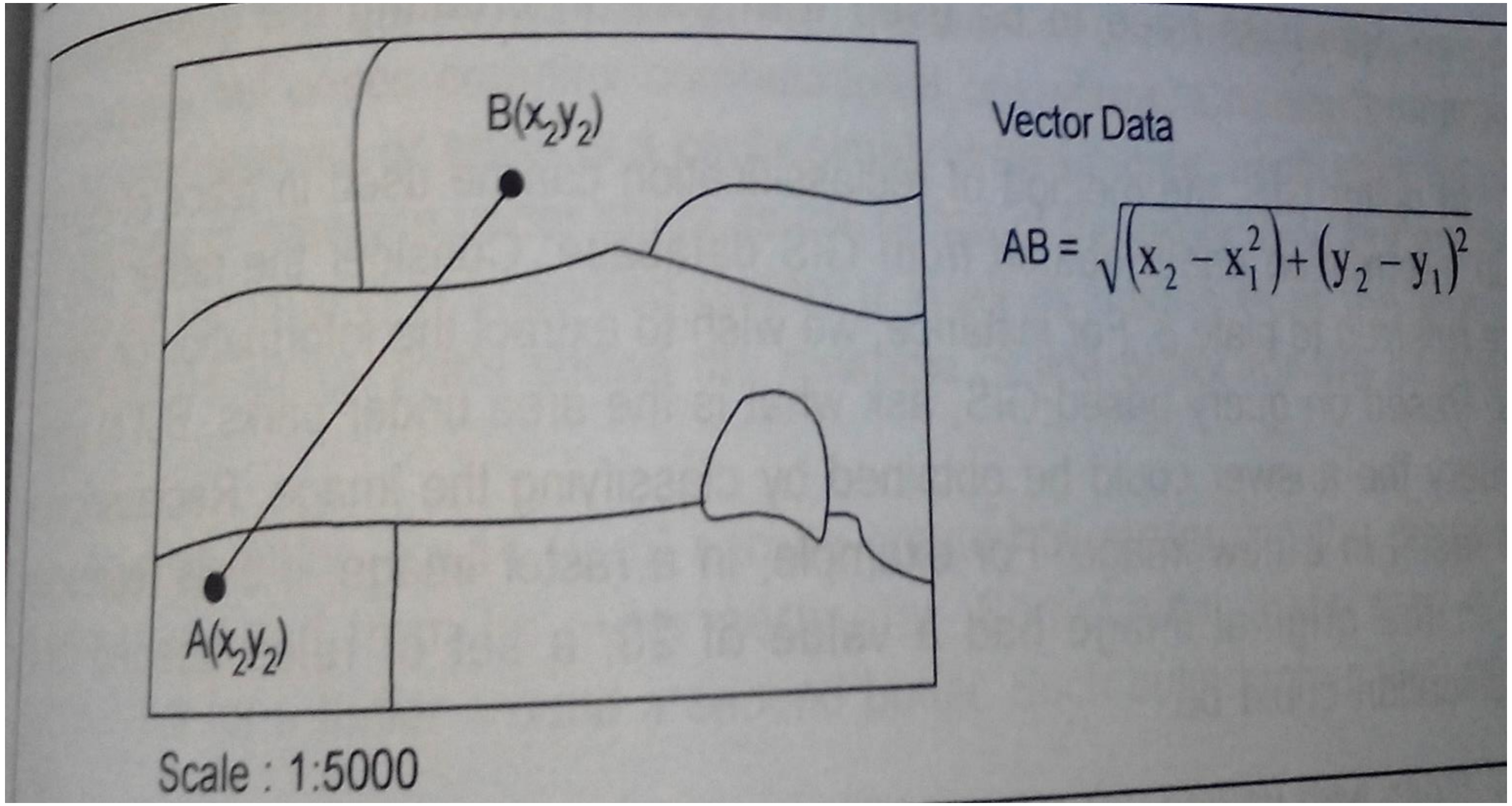


Fig 4: Vector Distance Calculation

OVERLAY

- Overlay that superimposes multiple data sets together for the purpose of identifying relationships between them.
- An overlay operation combines features and attributes from the input layers.

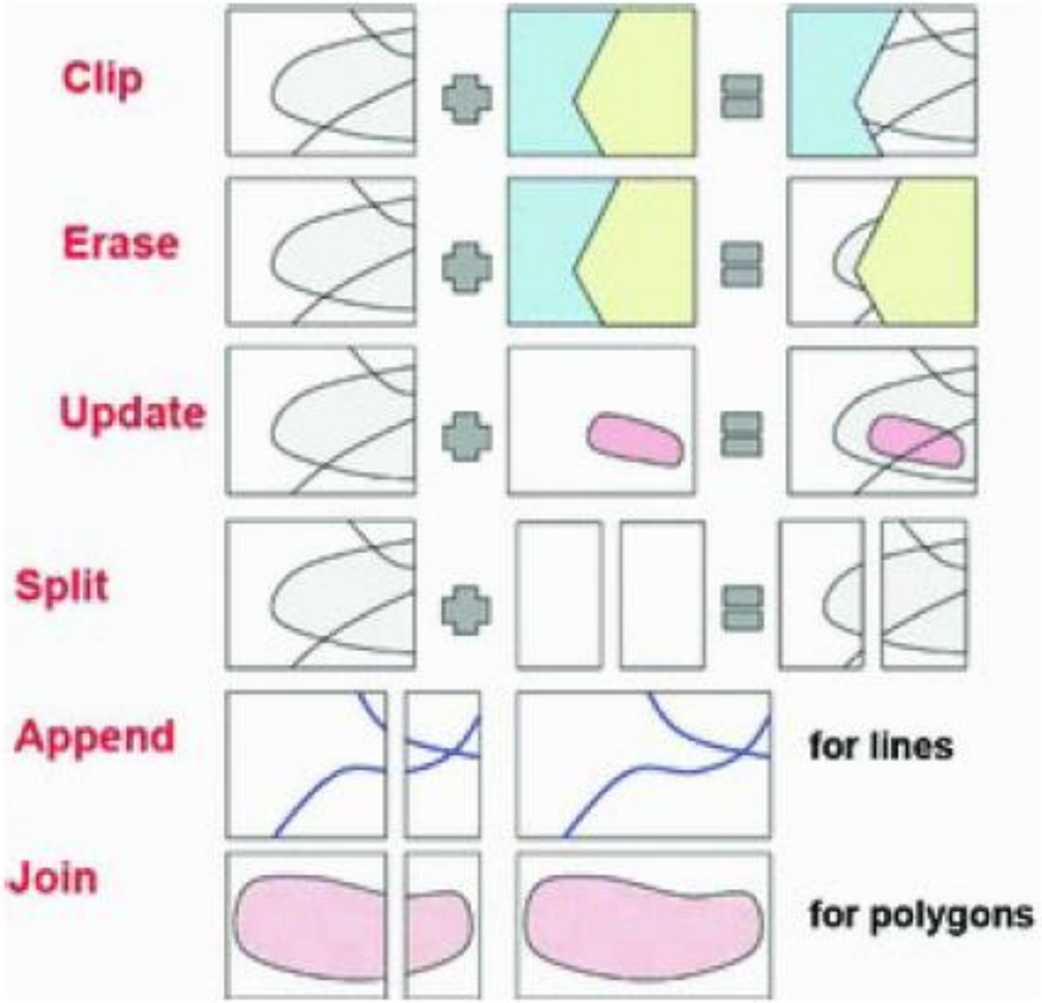
OVERLAY

- Overlaying of more than two layers, including rebuilding topology of the merged points, lines and polygons and operations on the merged attributes for suitability study, risk management and potential evaluation.

COVERAGE REBUILDING

- **Clip :** To identify and preserve features within the boundary of interest specified by users. It is called cookie cutter.
- **Erase:** To erase features inside the boundary while preserving features outside the boundary
- **Update :** To replace features within the boundary by cutting out the current polygons and pasting in the updated polygons.
- **Split :** To create new coverages by clipping geographic features with divided borders.
- **Append :** To merge the same feature classed of points and lines from the adjacent coverages.
- **Map Join:** To join the adjacent polygon features into a single coverage to rebuild to topology. It is called mosaicking.

COVERAGE REBUILDING



OVERLAY

- Steps in overlay operation

  1. Take two or more input layers

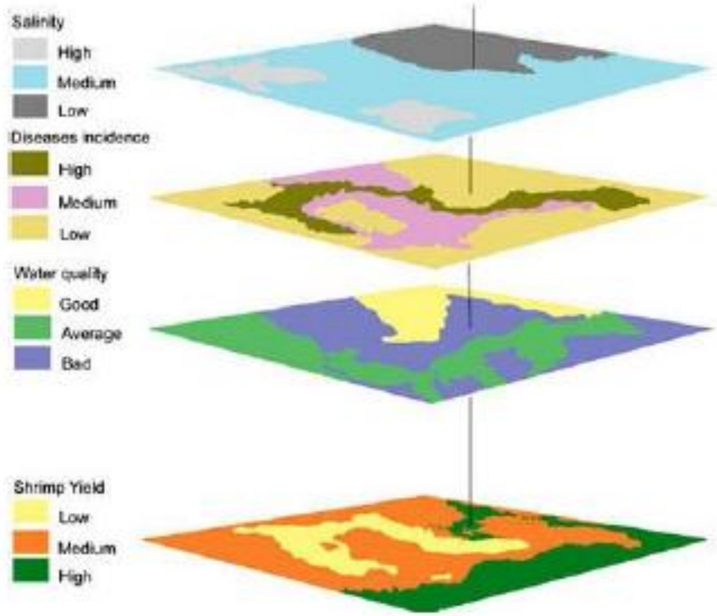
  2. Assume they are georeferenced in the same system

  3. Overlap in study area

The principle is to compare the characteristics of the same location in superimposed data layers, and to produce a new output value for each location.

OVERLAY

- Identify the shrimp yield
- Inputs required: Salinity, Disease incidence, Water quality



VECTOR OVERLAY

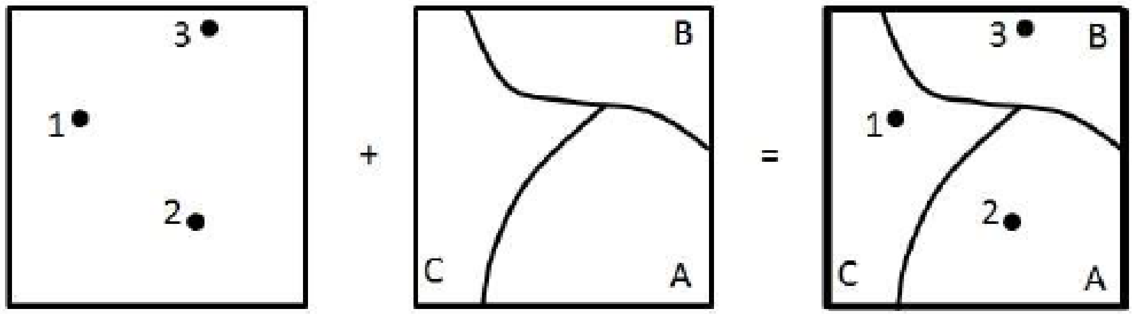
- Involves combining point, line, polygon geometry and associated attribute.
- Large attribute tables may result if overlay operations combine many layers
- Duplicate attribute field may also exist

TYPES OF VECTOR OVERLAY

- Point in polygon
- Line on polygon
- Polygon on polygon

POINT IN POLYGON OVERLAY

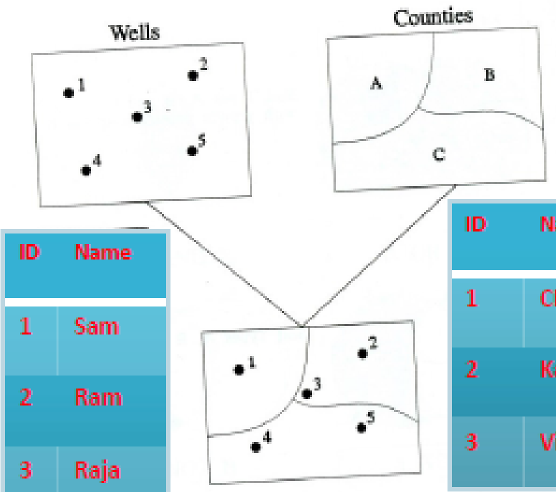
- Points are overlaid on polygon. Topology of point in polygon is " contained in" relationship.
- Point topology in the new data layer is a new attribute of polygon for each point.



Point Topology

Point	Polygon
1	C
2	A
3	B

POINT IN POLYGON OVERLAY



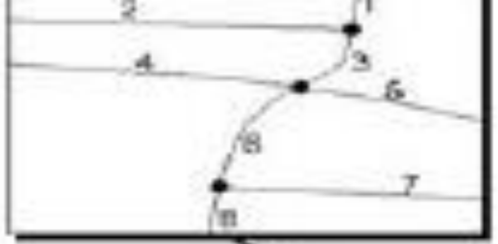
ID	Name
1	Sam
2	Ram
3	Raja
4	Pooja
5	Riya

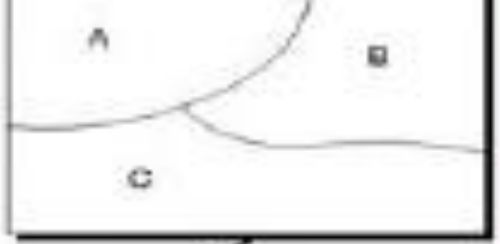
ID	Name
1	Chennai
2	Kancheepuram
3	Vilupuram

ID	Name

LINE – ON – POLYGON OVERLAY

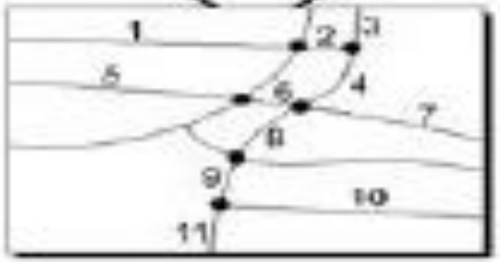
- Line – on – polygon algorithm overlays line objects on area objects and compute " is contained in" relationship.
- Lines are broken at each area object boundary to form line segments and new attributes created for each output line specifying the area it belongs to.





ID	Highway
1	35
2	22
3	35
4	60
5	60
6	35
7	82
8	35

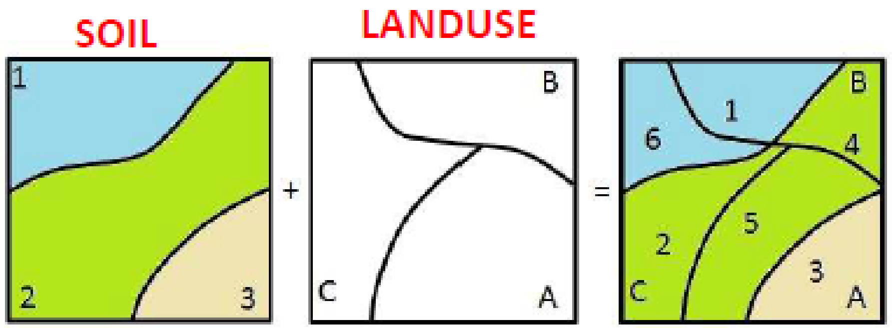
ID	Soil Type
A	Black
B	Brown
C	Red



ID	Original	Highway	Soil Type
1	2	22	Black
2	2	22	Brown
3	1	35	Brown
4	3	35	Brown
5	4	60	Black
6	4	60	Brown
7	5	60	Brown
8	6	35	Brown
9	6	35	Red
10	7	82	Red
11	8	35	Red

POLYGON-ON-POLYGON OVERLAY

- Two layers of area objectives are overlaid. The number of new polygons is usually larger than that of the original polygons.



Polygon Topology

ID	Polygon (A)	Polygon (B)
1	1	B
2	2	C
3	3	A
4	2	B
5	2	A
6	1	C

RASTER OVERLAY

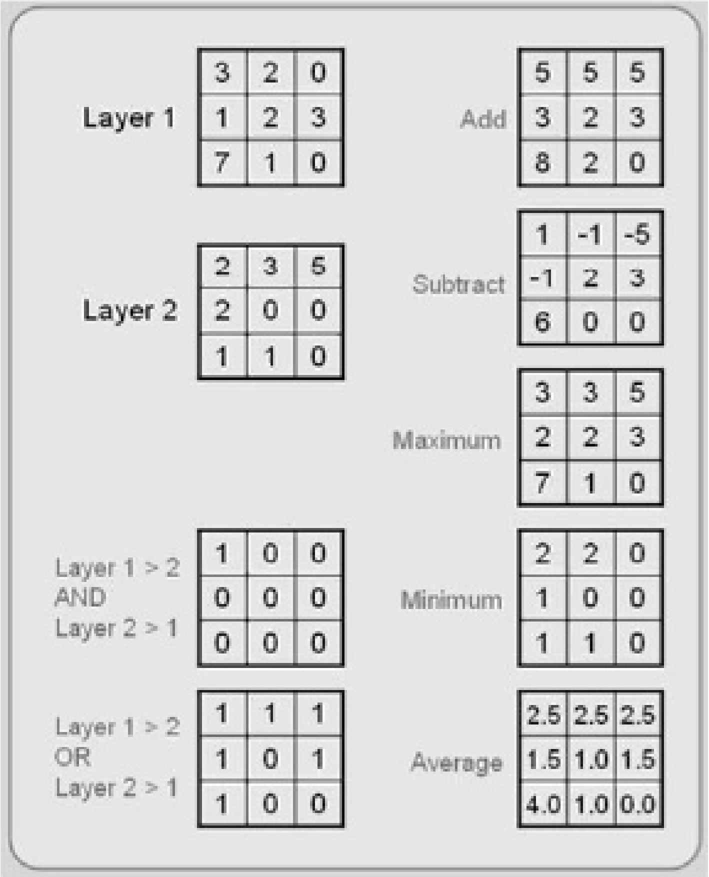
- Raster overlay can be performed by using map algebra or mathematics. Using map algebra, input layers may be added, subtracted, multiplied or divided to produce an output value.

RASTER OVERLAY

- Raster (Arithmetic) overlay: Raster overlay superimposes at least two input raster layers to produce an output layer. Each cell in the output layer is calculated from the corresponding pixels in the input layers.
- To do this, the layers must line up perfectly, they must have the same pixel resolution and spatial extent.
- Raster overlay is flexible, efficient, quick, and offers more overlay possibilities than vector overlay.

RASTER OVERLAY

- Raster overlay simply uses arithmetic operators to compute the corresponding cells of two or more input layers together, uses Boolean algebra like AND or OR to find the pixels that fit a particular query statement, or executes statistical tests like correlation and regression on the input layers.



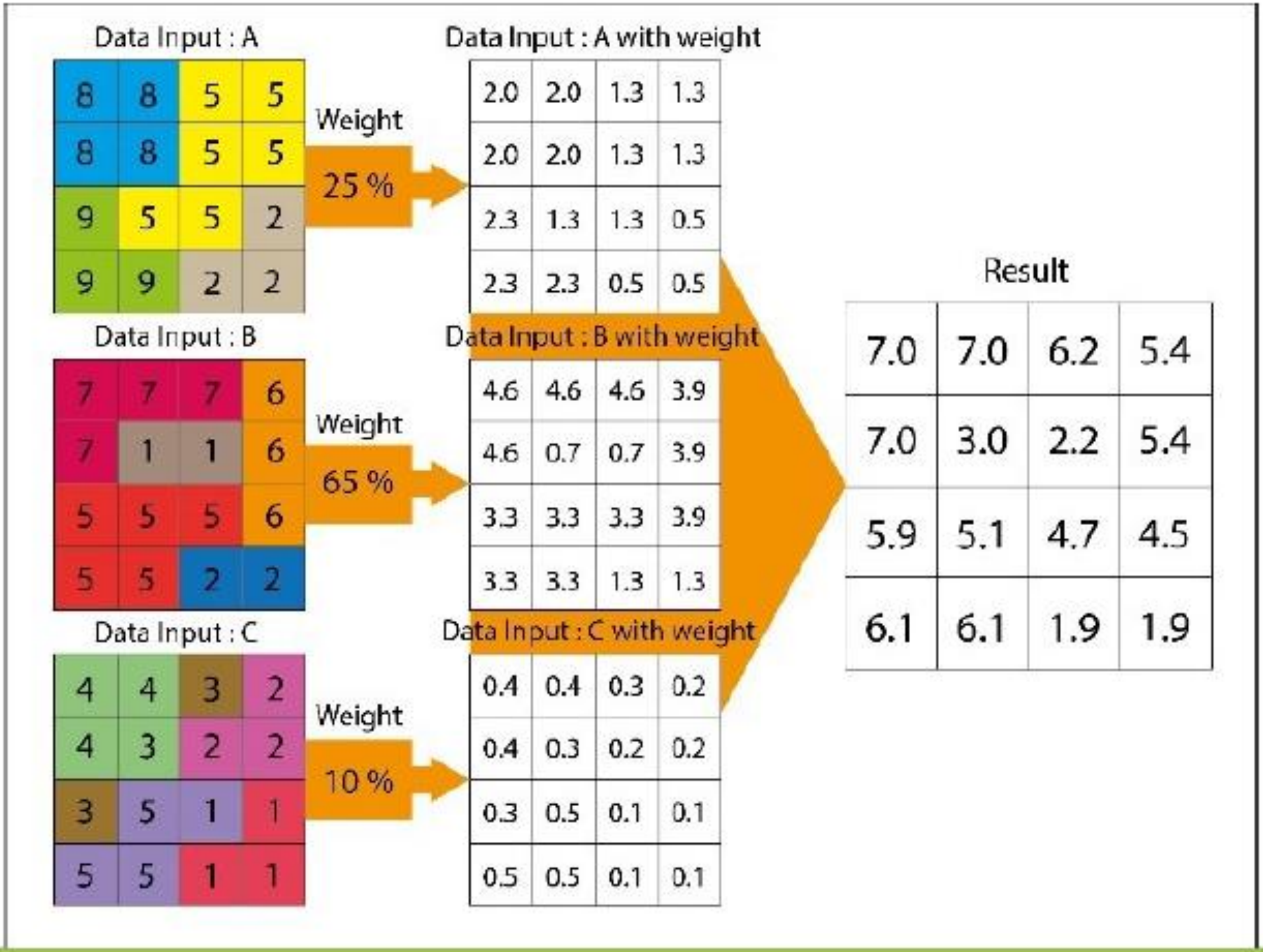
RASTER- WEIGHTED OVERLAY ANALYSIS

- Assigning a weight to each raster in the overlay process allows you to control the influence of different criteria in the suitability model.

RASTER-WEIGHTED OVERLAY ANALYSIS

- Steps: 1. Each raster layer is assigned a weight in the suitability analysis.
- 2. Values in the rasters are reclassified to a common suitability scale.
- 3. Raster layers are overlaid, multiplying each raster cell's suitability value by its layer weight and totaling the values to derive a suitability value.
- 4. These values are written to new cells in an output layer.

RASTER - WEIGHTED OVERLAY ANALYSIS



WEIGHTING BY RANKING

- Weighting the criteria by ranks in either ascending or descending order.
- Ascending means that the most important criterion is given rank 1, the second criterion rank 2 etc.
- When ranking in descending order, rank 1 is given to the least important criterion etc.

DIGITAL ELEVATION MODEL (DEM)

- DEM is a digital representation of 3- dimensional information (X, Y, Z) of the continuous topography of the bare earth surface in a particular reference coordinate system.
- A digital elevation model is a digital representation of ground surface topography or terrain.

DIGITAL ELEVATION MODEL

- A DEM is digital representation of topographic surface with the elevation or ground height above any geodetic datum. Following are widely used DEM in GIS.

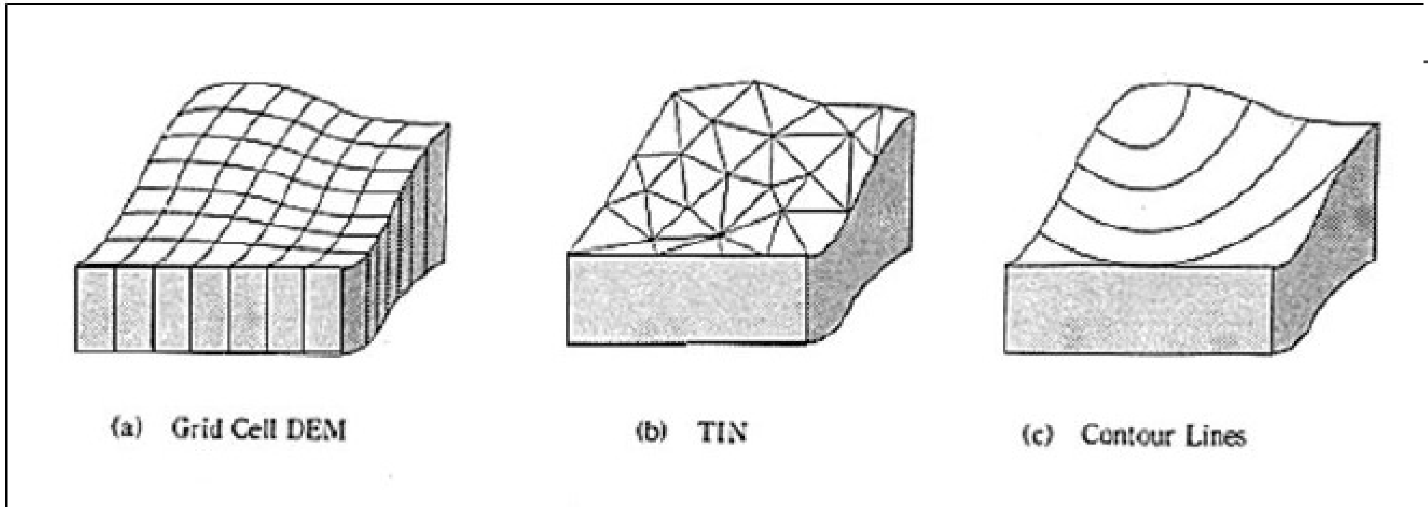
TYPES OF DEM

- DEMs are generated by using the elevation information from several points spaced at regular or irregular intervals.
- The elevation information may be obtained from different sources like field survey, topographic contours etc. DEMs use different structures to acquire or store the elevation information from various sources.

TYPES OF DEM

- Three main type of structures used are the following.
- a) Regular square grids
- b) Triangulated irregular networks (TIN)
- c) Contours

TYPES OF DEM



TYPES OF DEM

- Grid DEM : The result is a matrix whose indices are the coordinates and values are the elevation value at each point (raster representation)
- From this sample representation it is possible to get a representation of the relief.

REPRESENTATION OF DEM

- Grid DEM: They are based on the values of the elevation at the sampling points- one height per pixel (grid cell).
- The grid representation is the consequence of sampling elevation values in regular intervals of latitude and longitude.

DEM - PARAMETERS

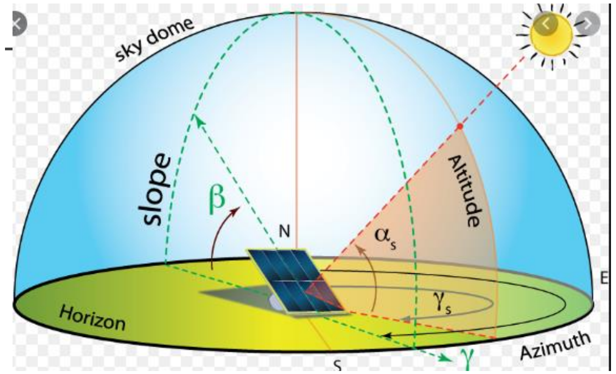
- A DEM (Digital Terrain Model) is digital representation of terrain features including elevation, slope, aspect, drainage and other terrain attributes.
- Usually a DTM is derived from a DEM or elevation data.
- Several terrain features including the following DTMs.
- 1. Slope and Aspect
- 2. Drainage network
- 3. Catchment area
- 4. Shading
- 5. Shadow
- 6. Slope stability

PARAMETERS DERIVED FROM DEM

- Slope – Displays the grade of steepness expressed in degrees or as percent slope. This image can reveal structural lineaments, fault scarps, fluvial terrace scarps, etc.
- 2. Aspect – Identifies the down-slope direction. Aspect images may enhance landforms such as fluvial networks, alluvial fans, faceted fault related scarps, etc.

PARAMETERS DERIVED FROM DEM

- 3. Shaded topographic relief or hill-shading – This image depicts relief by simulating the effect of the sun's illumination on the terrain.
- The direction and the altitude of the illumination can be changed in order to emphasize faults, lineaments, etc.
- This image is probably the most useful to display geological data related to landforms in terrains that show a close correlation between geology and topography.



PARAMETERS DERIVED FROM DEM

- 4. Flow direction – Shows the direction of flow by finding the direction of the steepest descent or maximum drop. This DEM derived surface depicts the drainage.
- 5. Basin – Function that uses a grid of flow direction (output of flow direction) to determine the contributing area.

DEM PARAMETERS

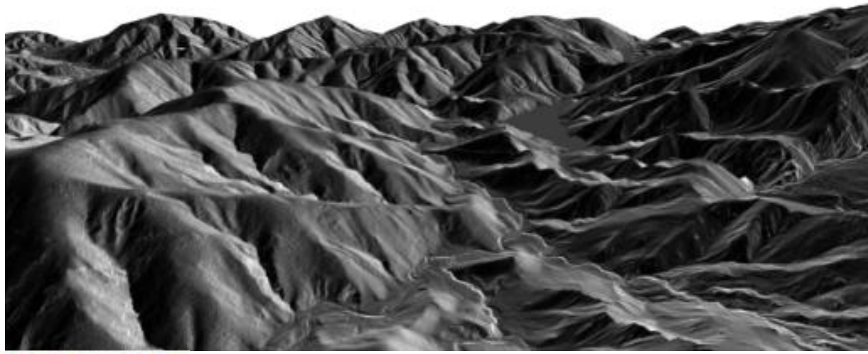
- Drainage basins : Drainage basin, also called Catchment Area, or Watershed, area from which all precipitation flows to a single stream or set of streams.
- Channel networks :

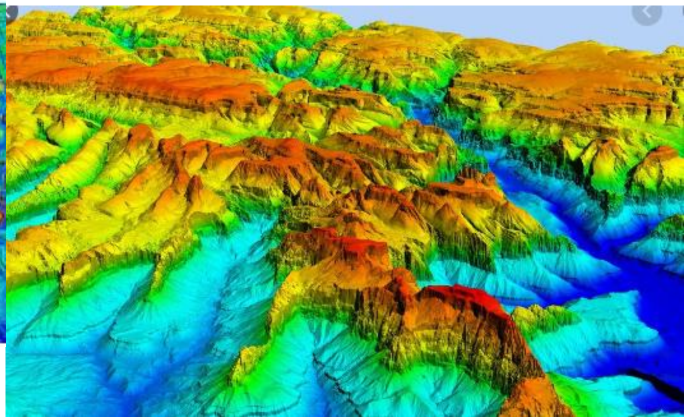
APPLICATION OF DEM

- Landslide probability
- Estimation of the volume of proposed reservoir
- Flood prone area mapping
- Hazard monitoring
- Natural resources exploration
- Agricultural management

MODELING SURFACE

- Surface is denoted here Earth's surface, Moon or asteroid created by using terrain's elevation data.
- Surface can be modeled by using
- DTM : Digital Terrain Model
- DEM: Digital Elevation Model
- DSM: Digital Surface Model





DIGITAL TERRAIN MODEL

- A Digital Terrain Model is a topographic model of the bare-earth terrain relief, that can be manipulated by computer programs.
- The data files contain the spatial elevation data of the terrain in a digital format which usually presented as a rectangular grid.
- Vegetation, buildings and other man-made (artificial) features are removed digitally- leaving just the underlying terrain.

DIGITAL TERRAIN MODEL

- DTM model is mostly related as raster data type, stored usually as a rectangular equal-spaced grid, with space of between 50 and 500 meters mostly presented in Geographic coordinate system.

DIGITAL SURFACE MODEL

- DSM : Surface model which captures the natural and built features on the Earth's surface.
- DSM's measure the height values of the first surface on the ground. This includes terrain features, buildings, vegetation and power lines etc. DSM's therefore provide a topographic model of the earth's surface. DSM's can be used to create 3D fly-through, support location-based systems and augmented simulated environments.

DIGITAL SURFACE MODEL

- DSM is generated using LIDAR (Light Detection and Ranging) system, which sends pulses of light to the ground and when the pulse of light bounces off/back its target and returns to the sensor, it gives the range of the Earth.
- LIDAR delivers a massive point cloud filled of varying elevation values (Height can come from the top of buildings, tree canopy, power lines, other built and natural features).
- DSM is useful in 3D modeling for telecommunications, urban planning and aviation (objects extrude from the earth, particularly useful in these application to identify obstructions).

What is expert system and its components?

Expert system, a computer program that uses artificial-intelligence methods to solve problems within a specialized domain that ordinarily requires human expertise.

An **expert system** generally consists of four **components**: a knowledge base, the search or inference **system**, a knowledge acquisition **system**, and the user interface or communication **system**. Knowledge **systems** solve difficult problems of the real world by performing inference processes on explicitly stated knowledge

The internal **structure** of an **expert system** can be considered to consist of three parts: the knowledge base ; the database; the rule interpreter. the set of productions; the set of facts held as working memory and a rule interpreter. The knowledge base holds the set of rules of inference that are used in reasoning.

Limitations of Expert System

- The response of the **expert system** may get wrong if the knowledge base contains the wrong information.
- Like a human being, it cannot produce a creative output for different scenarios.
- Its maintenance and development costs are very high.
- Knowledge acquisition for designing is much difficult.

What is the purpose of an expert system?

In artificial intelligence, an **expert system** is a computer **system** emulating the decision-making ability of a human **expert**. **Expert systems** are designed to solve complex problems by reasoning through bodies of knowledge, represented mainly as if–then rules rather than through conventional procedural code.