Bharathidasan University

Tiruchirappalli – 620 023, Tamil Nadu

M.Tech. Remote Sensing and GIS

Course : 24MTRS-05

GEOGRAPHIC INFORMATION SYSTEMS (GIS)

Unit-3 Raster and vector data Interpolation and Analyses

Dr. K. Palanivel

Professor, Department of Remote Sensing

24MTRS-05: GEOGRAPHIC INFORMATION SYSTEM --- 4 credits

UNIT I Basics of GIS - Definition - Evolution of GIS - Geospatial, spatial and non-spatial data - Components of GIS - Computer Hardware, Software Modules and Organisational Context of GIS - Data Structure in GIS - Types of Data (Points, Lines and Polygons) - Data sources - Ground and remote sensing survey - Data Structures (Raster data structures and Vector data Structures) – Database Structures.. 12 Hrs.

UNIT II GIS Data Input, Verification, Storage and Output - Spatial Data Input Processes and Devices (Sources of data, - Different Types of Data Entry methods, viz., Manual input, Run length code, Digitization, Automated Scanning, etc. – Vector to Raster conversion - Raster to Vector conversion - Input devices) - Entry of non-spatial data – Linking of Spatial & Non-spatial data – Element generation - Geodatabase creation - Data Verification (Errors of different types) - Correction (Rubber Sheet Transformation, Bilinear interpolation, Cubic Convolution, etc.) - GIS capabilities for Data correction – Data output (Types of Output, GIS Capabilities for output, Output devices) - File formats for raster and vector. 14 Hrs.

UNIT III Raster and vector data Interpolation and Analyses - Raster data analysis: Local, neighborhood and regional operations - Map algebra - Vector data analysis: Topological analysis, point-in-polygon, line-in-polygon, polygon-in-polygon - Proximity analysis: Buffering, Thiessen polygon - Non topological analysis: Attribute data retrieval and analysis – Concepts of SQL – ODBC - Basic Principles of Interpolation – Methods of Interpolation - Global Methods of Interpolation, Local Interpolation (Trend Surface Analysis) – Local Interpolation (Splines) - Optimal Interpolation (Kriging). 12 Hrs.

UNIT IV Surface Analysis, 3D Visualization and Network Modelling - Need for Three Dimensional Models - Methods of DEM & TIN - Products of DTM (Contour Maps, Shaded Relief Map, Maps Related to Slopes, Line Sight Maps, Drainage Analysis, Volume Estimation etc.) Usefulness of DEM/DTM. Overlay analysis, Capabilities (Point Operations, Regional Operations, Neighbourhood Operations) - Networking and Dynamic Segmentation – Applications, Minimum Distance Model, Maximum Covering Model (P-median model), Urban Transportation Planning Model. 14 Hrs.

UNIT V Data Classification & Advanced GIS - Principles – Types of Classification (Exogenous, Arbitrary, Idiographic & Serial) – Multivariate Analysis. Artificial Intelligence - Expert Systems - Object Oriented GIS - Web based GIS: Definition, merits - Architecture - Map server - Spatial data infrastructure - Spatial data standards - Free and open source - Proprietary GIS software. 12 Hrs.

UNIT VI Current Contours – GIS - Case studies (Not for Final Exam only for Discussion) - Recent scientific and technological development, advancement, Industrial application and Job opportunities.

REFERENCES

1. Burrough, P.A 1986: Principles of Geographical information Systems for Land Resources Assessment, Clarandone Press, Oxford.
2. Avery, T.V,Interpretation of Aerial Photography Burgass, Publishing Company.
3. Gautham, N.C 1970: Urban Landuse Study Through Aerial Photo binterpretations Techniques, Pink Publishing House, Mathura.
4. American Society of Photogrammetry, 1983: Manual of Remote Sensing (2nd Edition), ASP Falls Church, Virginia.
5. Campbell, J 1984: introductory Cartography, Printers Hall Englewood Cliffs, N.J
6. Dent B.D 1985: Principles of Thematic Map Design, Addition - Wesley, Reading, Mass.
7. Freeman, H and GG.Pieroni 1980: Map Data Processing, Academic Press, New York.
8. Monmonier, M.A 1982: Computer Assisted Cartography - Principles and Prospects, Prentice Hall, Englewood Cliffs, NJ
9. Tomlinson, RF Calkins, HS and D.F.Marble 1976: Computer Handling of Geographic Data, UNESCO, Geneva.
10. Graeme F. & Bonham – Carter; Geographic information Systems for Geoscientists; Modelling with GIS, Pergamon.
11. Lo, C.P. and Yeung, Albert K.W., “Concepts and Techniques of Geographic Information Systems”, Pearson, 2016.

OBJECTIVE

To understand the fundamentals, data handling, analysis techniques, and advanced applications of Geographic Information Systems (GIS).

COURSE OUTCOMES

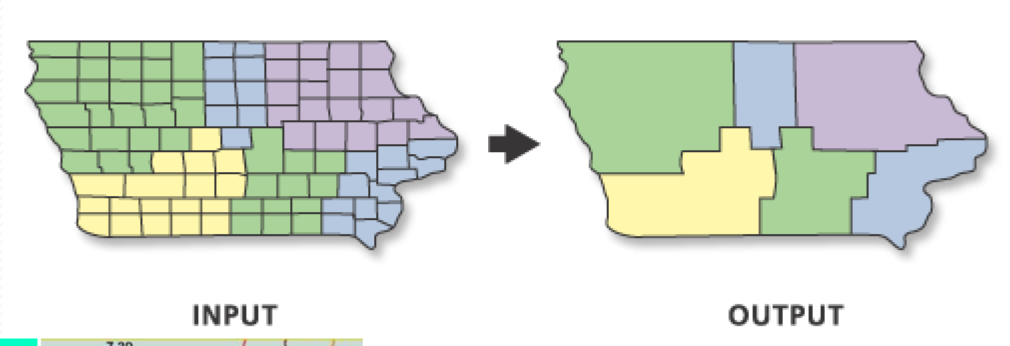
- The concepts of GIS, including data types, components, and data structures.
- Develop proficiency in GIS data input, verification, storage, and output processes.
- Gain knowledge in raster and vector data analysis, including interpolation methods and proximity analysis.
- Understand three-dimensional modeling, surface analysis, and network modeling techniques.
- Explore classification methods, AI integration, web-based GIS, and the use of open-source and proprietary GIS software.

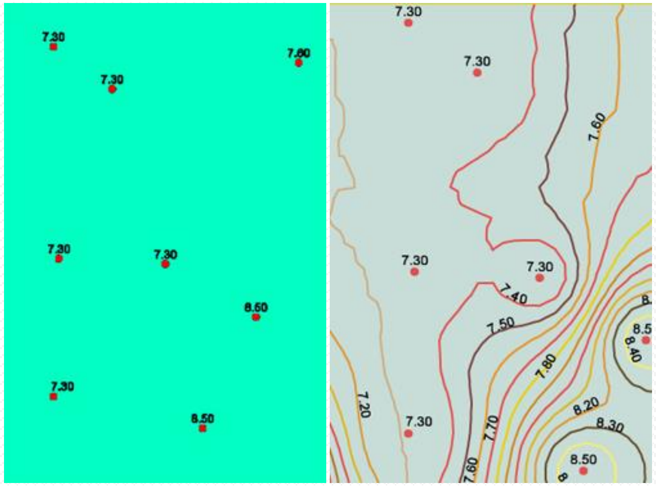
Unit-III

Raster and vector data Interpolation and Analyses - Raster data analysis: Local, neighborhood and regional operations - Map algebra - Vector data analysis: Topological analysis, point-in-polygon, line-in-polygon, polygon-in-polygon - Proximity analysis: Buffering, Thiessen polygon - Non topological analysis: Attribute data retrieval and analysis – Concepts of SQL – ODBC - Basic Principles of Interpolation – Methods of Interpolation - Global Methods of Interpolation, Local Interpolation (Trend Surface Analysis) – Local Interpolation (Splines) - Optimal Interpolation (Kriging). 12 Hrs.

GIS DATA PREPARATION / PRE-PROCESSING FOR SPATIAL ANALYSIS

Generalization, Buffering, Numerical spatial database generation, Spatial Interpolation, Data Resampling, Data Normalization, Data standardization, Data Classification/Grouping , Regrouping/Reclassification, Conversion of data formats, indexing, Assigning Ranks and Weightages …





DISSOLVE

MERGE

UPDATE

CALCULATE RASTER ……

Classification Methods

Classifying the ranges of values of a single continuous variables

The different types of making class intervals are:

- **Exogenous class intervals** – according to threshold values that are relevant to, but not derived from the data set under study
  - Fail to resolve the variations in any specific area
  - If exogenous intervals are derived from qualitative class definitions, they may not be unique and may in fact overlap

- **Arbitrary class intervals**
  - chosen without any clear aim in mind
  - very often the intervals are irregularly spaced and
  - Chosen without a proper examination of the data

- **Idiographic class intervals**
  - chosen with respect to specific aspects of the data set
  - They include methods that attempt to divide the data into multi-modal groupings
  - Using ‘Natural Breaks’, their location,
  - Objective comparison is difficult – not exactly followed from dataset to dataset

- **Serial class intervals**
  - have limits that are direct mathematical relation to one another.
  - intervals are chosen independent of the characteristics of the particular dataset,
  - then the results from different samples or datasets will be strictly comparable
    - **Normal percentiles** – class intervals subdivide a normal distribution with mean μ and standard deviation, with classes centred on the mean
    - **Equal arithmetic intervals** with no variation in class width, such as are normally used for elevation contours
    - **Equal intervals** on reciprocal, trignometric, geometric, or logarithmic scales – used when original data are not normally or rectangularly distributed

DATA SORTING and PERFORMING STATISTICAL OPERATIONS

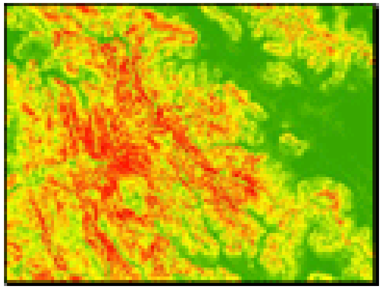
- Data Sorting
- Mean, Mode, Std. Deviation, Correlation, Factor analysis, Multivariate analysis, Multiple Regression analysis, etc.

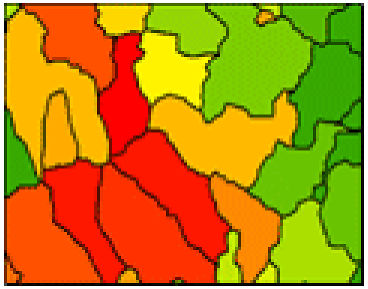
In order to perform statistical analysis in GIS, it is necessary to have samples of uniform length (equal number of samples) for all the thematic maps.

Hence, it would be better to convert the vector data to Raster format for statistical analysis.

DATA CONVERSION

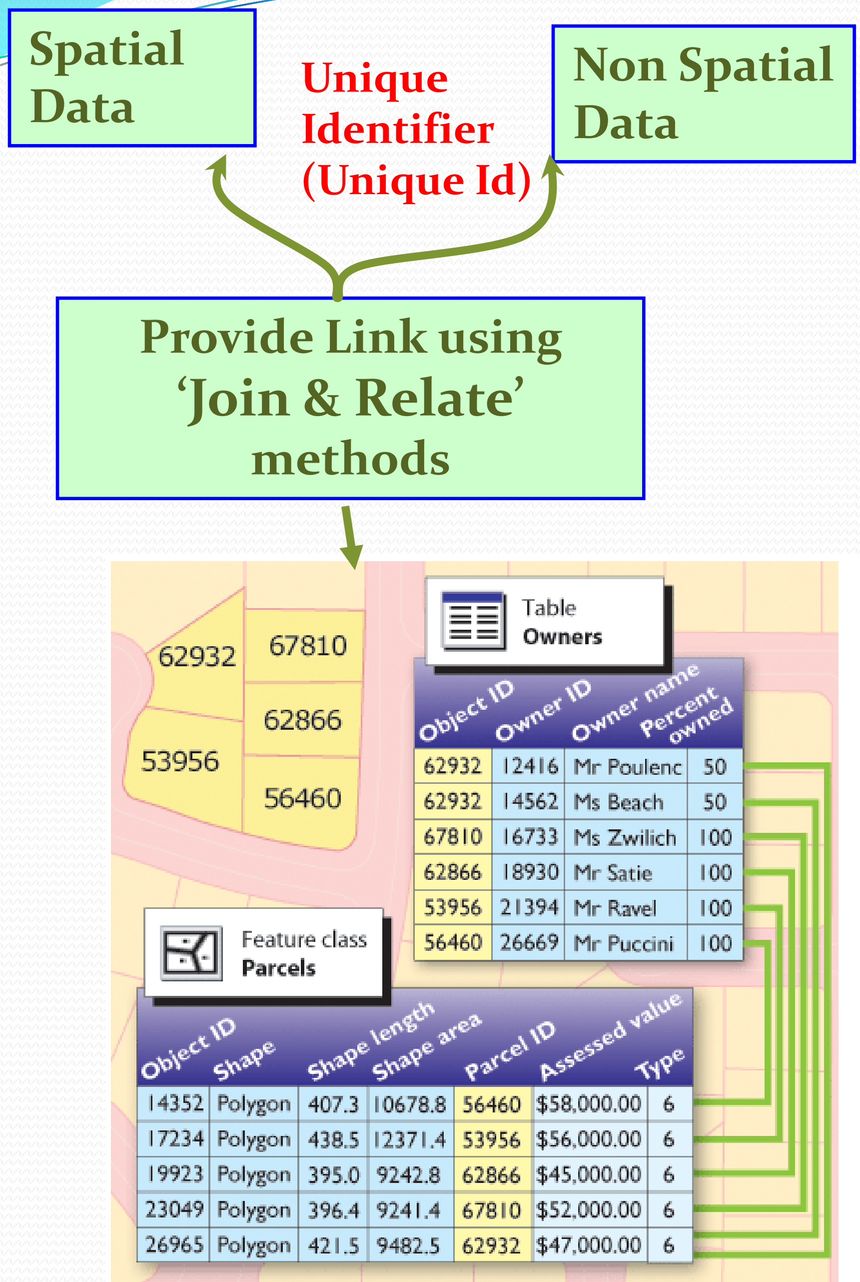
Raster ⇄ Vector





GIS ANALYSIS

- GIS has the analysis capabilities, that are not available with CAD
- Operates on Topology, spatial and/or non-spatial aspects combined
- The user can interactively work for getting the results
- Analysis capabilities range from
  - Retrieving subsets of information
  - up to univariate,
  - multivariate statistical and
  - spatial analysis using neighbourhood functions and using interpolate methods
- Spatial modelling
- Simulation programs



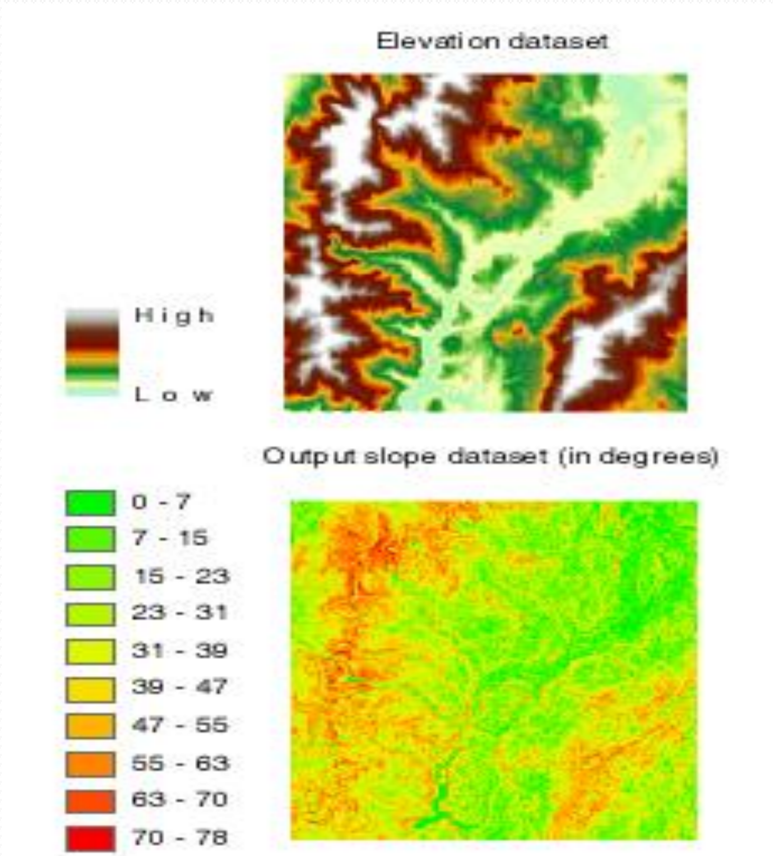
Object ID	Owner ID	Owner name	Percent owned
62932	12416	Mr Poulenc	50
62932	14562	Ms Beach	50
67810	16733	Ms Zwilich	100
62866	18930	Mr Satie	100
53956	21394	Mr Ravel	100
56460	26669	Mr Puccini	100

Object ID	Shape	Shape length	Shape area	Parcel ID	Assessed value	Type
14352	Polygon	407.3	10678.8	56460	$58,000.00	6
17234	Polygon	438.5	12371.4	53956	$56,000.00	6
19923	Polygon	395.0	9242.8	62866	$45,000.00	6
23049	Polygon	396.4	9241.4	67810	$52,000.00	6
26965	Polygon	421.5	9482.5	62932	$47,000.00	6

- It is necessary to establish a link between database and output that provides answer in different forms – maps, tables, figures, etc. using “**Unique_Identifier**”
- That link is any function used to convert data from one or more input maps into an output / **Action Plan** map
- Define the database – nature of geodatabase – **vector** / **raster**, and whether it contains **points**, **lines**, **polygons** & **associated attributes**, having **same projection**

GIS Data Analysis ...contd...

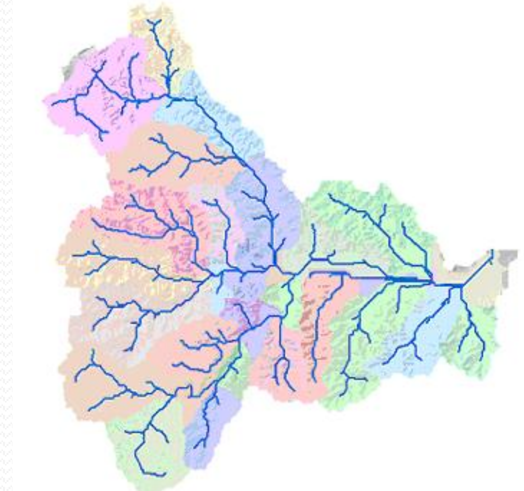
SLOPE ANALYSIS



Slope Categories – DEGREES , ASPECT, LENGTH

Intervisibility Analysis, etc.

DRAINAGE ANALYSIS



Delineation of Drainages,

Flow Direction, Flow Accumulation

Flow Length,

Demarcation of drainage basins,

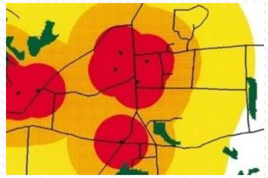
Runoff estimation,

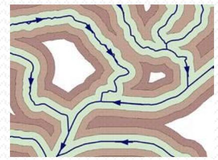
Quantum of soil erosion, etc.

Hydrological Modeling

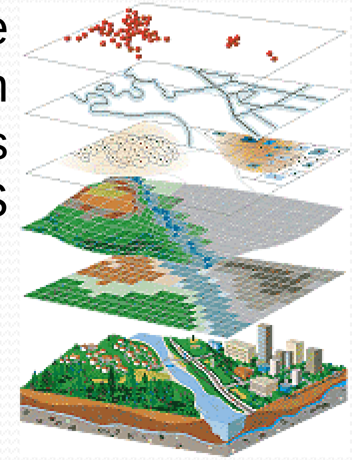
GIS Data Analysis & integration

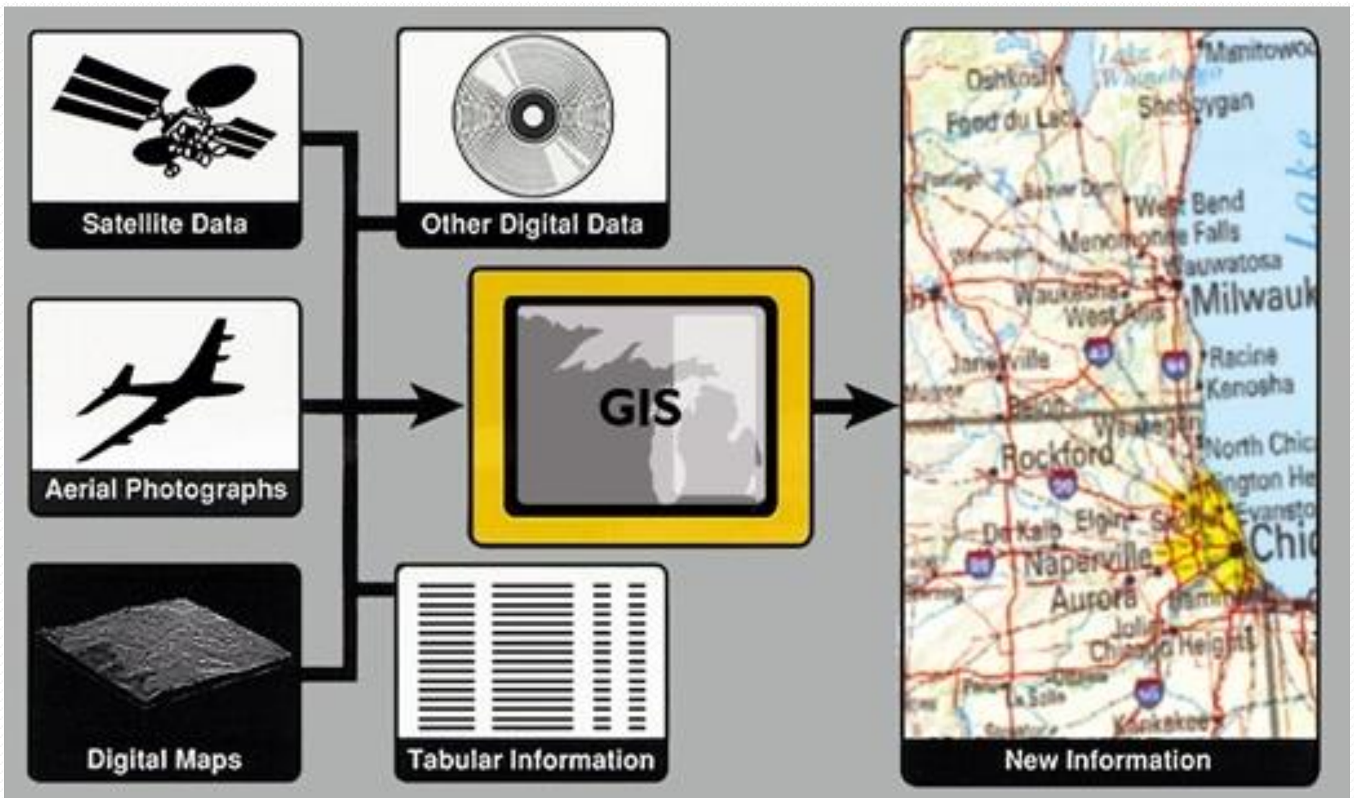
- Performing overlays
- Creating buffers
- Calculating statistics
- Merging datasets

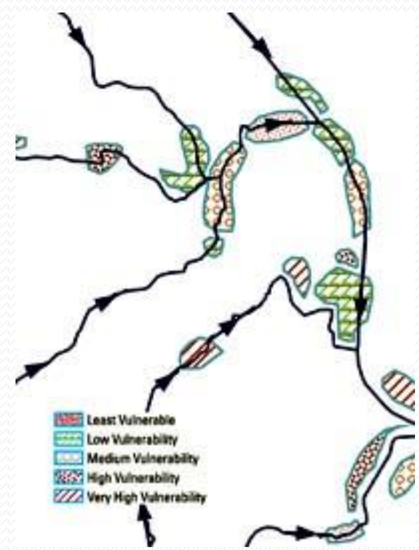




Data integration is the linking of information in different forms through a GIS







GIS OVERLAY – INTEGRATION OF SPATIAL LAYERS

- **Spatial integration (union / intersection) of multiple layers having :**
  - **Single feature attribute** –Resulted with unique polygons having positive / inducing / vulnerable polygons only - useful for further planning - prioritization (Union) and direct implementation (Intersection);
  - **Binary feature attributes** – Resulted with both +ve and –ve polygons – useful to know the parameter which are not potential / not loaded / not vulnerable or safe polygons / sites – so as to improve / develop their property to achieve the target / reduce vulnerability as minimum as possible – mitigation
  - **Multiple feature attributes** – complex but gives useful detailed information about the parameters involved.

GIS Can do many operations / Analysis like add, subtract, multiply, divide, square, buffer, corridor, integrate multiple layers, etc.

Raster Based Addition

GO BACK

InRaster1

1	1	2	2	2
3	1	3	4	4
3	3	2	4	4
3	2	2	1	4
3	2	2	1	1

\+

InRaster2

3	3	1	2	2
2	3	1	1	1
2	4	3	3	1
2	2	3	4	4
1	2	1	1	4

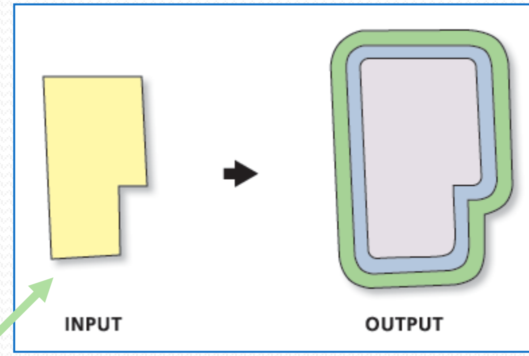
=

OutRaster

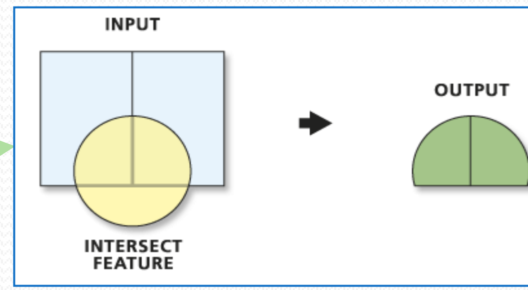
4	4	3	4	4
5	4	4	5	5
5	7	5	7	5
5	4	5	5	8
4	4	3	2	5

InRaster1 + InRaster2 = OutRaster

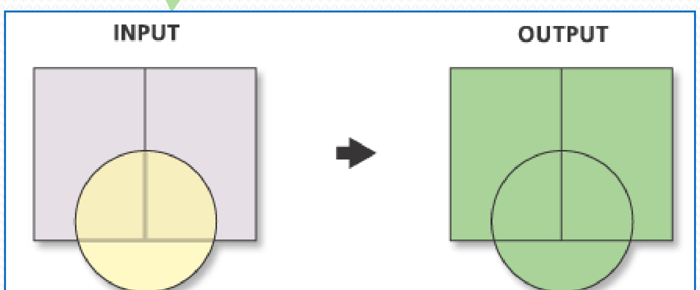
Vector Based Layer Integration



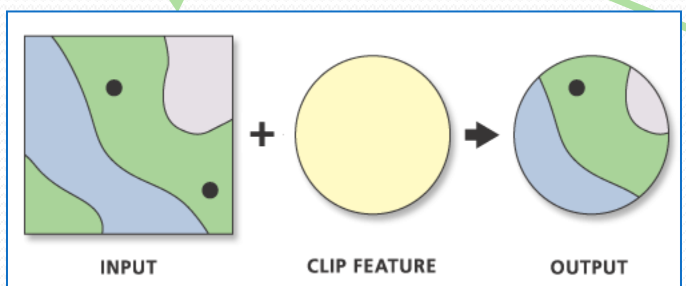
BUFFER



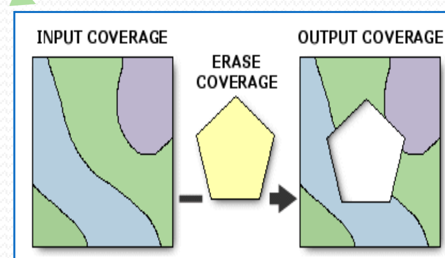
INTERSECT



UNION



CLIP



ERASE

Vector Topological Analyses: Point-in-polygon, Line-in-Polygon, Polygon-in-Polygon, Proximity: Buffering, Network Analysis

Non-topological Analysis: Attribute data retrieval and analysis

Simple Data Retrieval

- **Boolean logic**
- **Reclassification and display**
- **Boolean operations with two or more maps**

Boolean logic

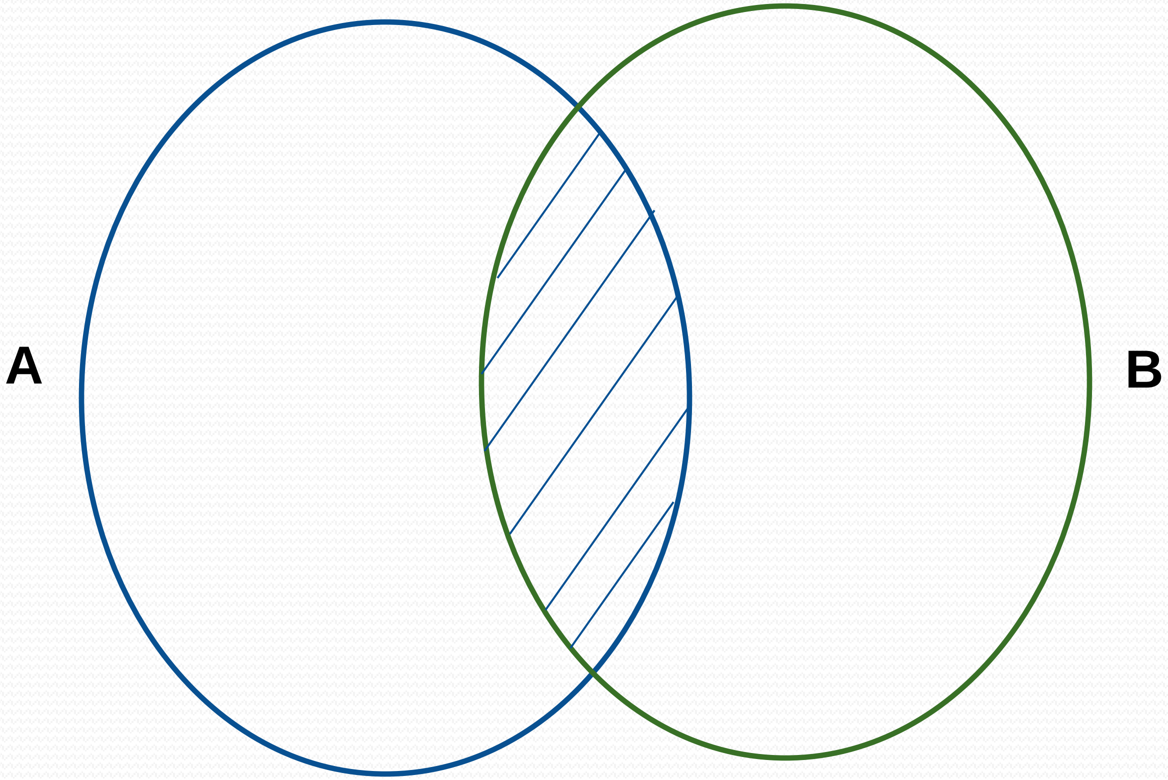
- Retrieve data by using the rules of Boolean logic to operate on the attributes and spatial properties
- Operators are : AND, OR, XOR, NOT
- The particular condition is ‘true’ or ‘false’
- Visual portray of Boolean logic is in the form of ‘Venn diagrams’
- A is the set of items having attribute ‘a’ and B is the set of items having attribute ‘b’, then the statements

A **AND** B,   A **OR** B,   A **XOR** B,   A **NOT** B,

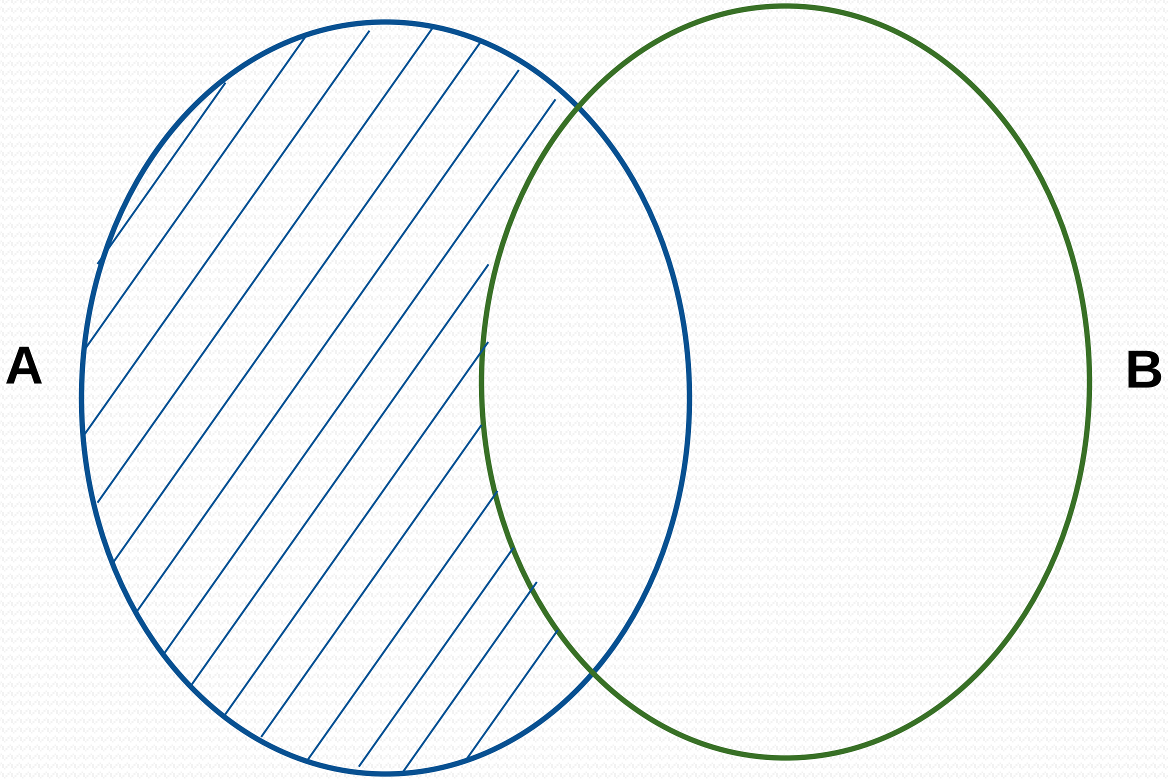
(A **AND** B) **OR** C,   A **AND** (B **OR** C)

will return items that have attribute combination covered by the **shaded portions** respectively.

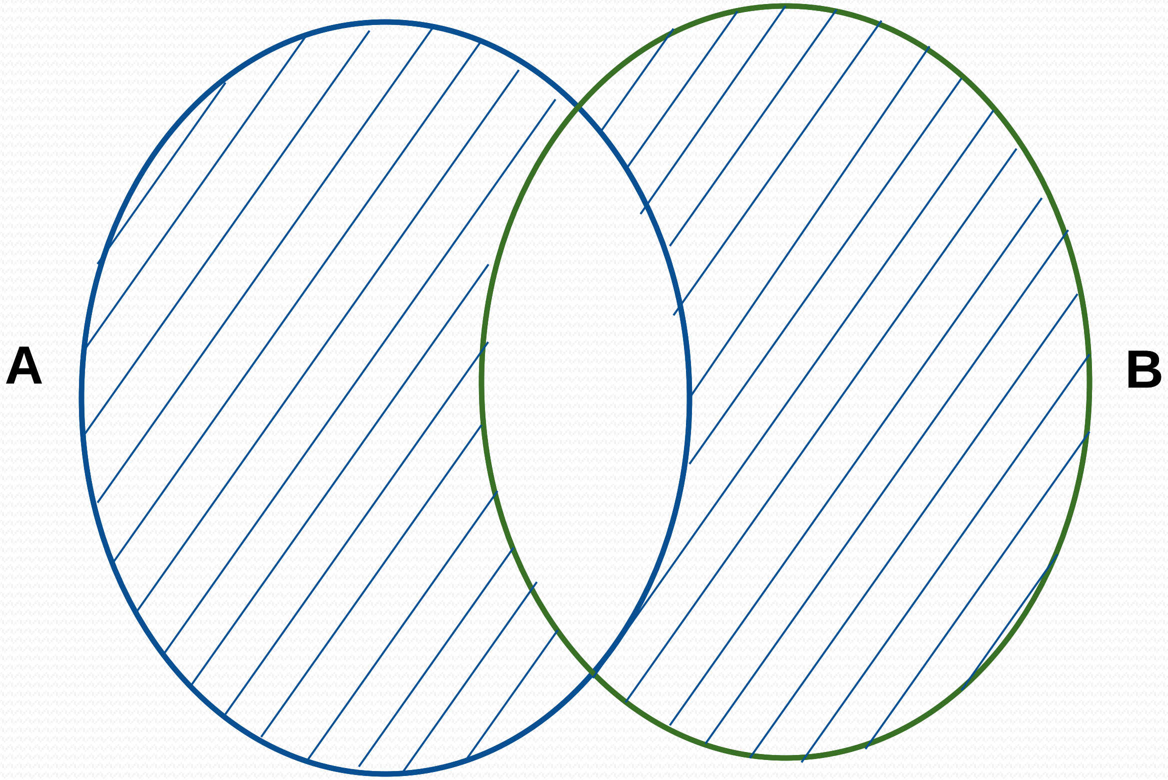
- On the basis of priority of AND with respect to OR
- Parentheses are usually used to represent the order of evaluation
- More complicated searches involve the shapes of areas, the properties of the boundaries of areas or the neighbouring areas. For e.g.,
  - All areas of woodland bordering urban areas
  - All open areas that have a particular shape and area bounded by hedgerow (a row of bushes)



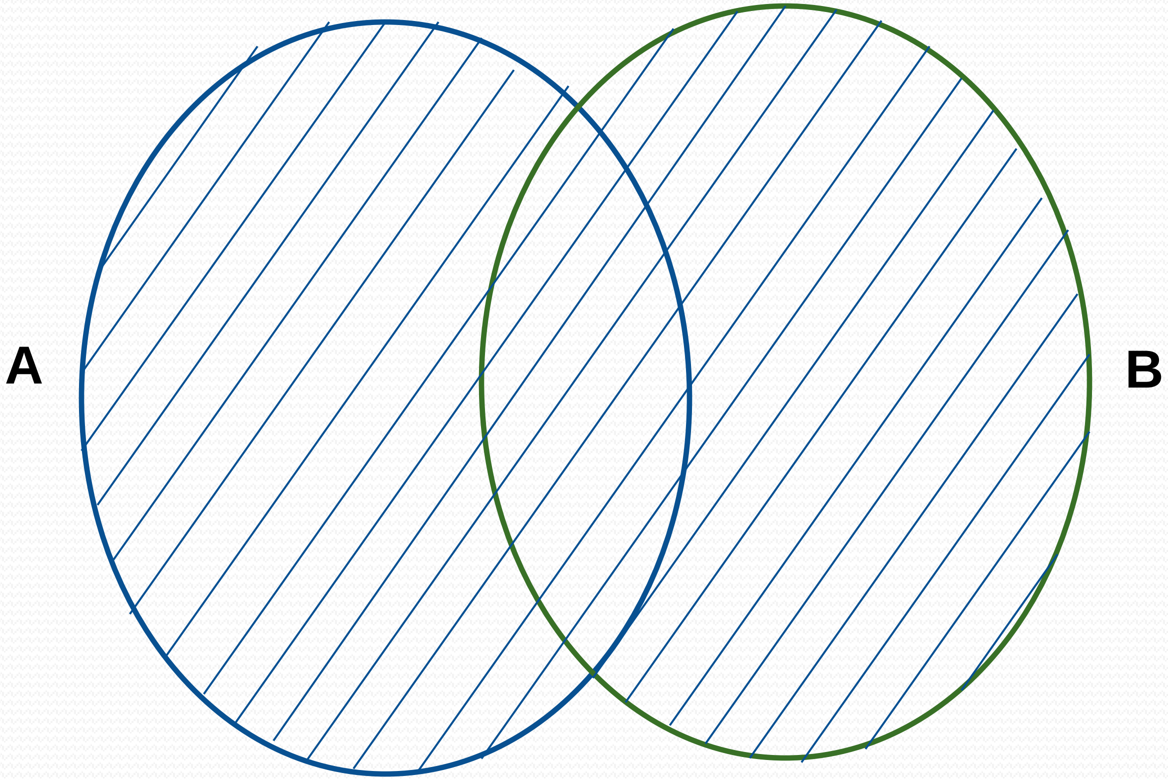
A AND B



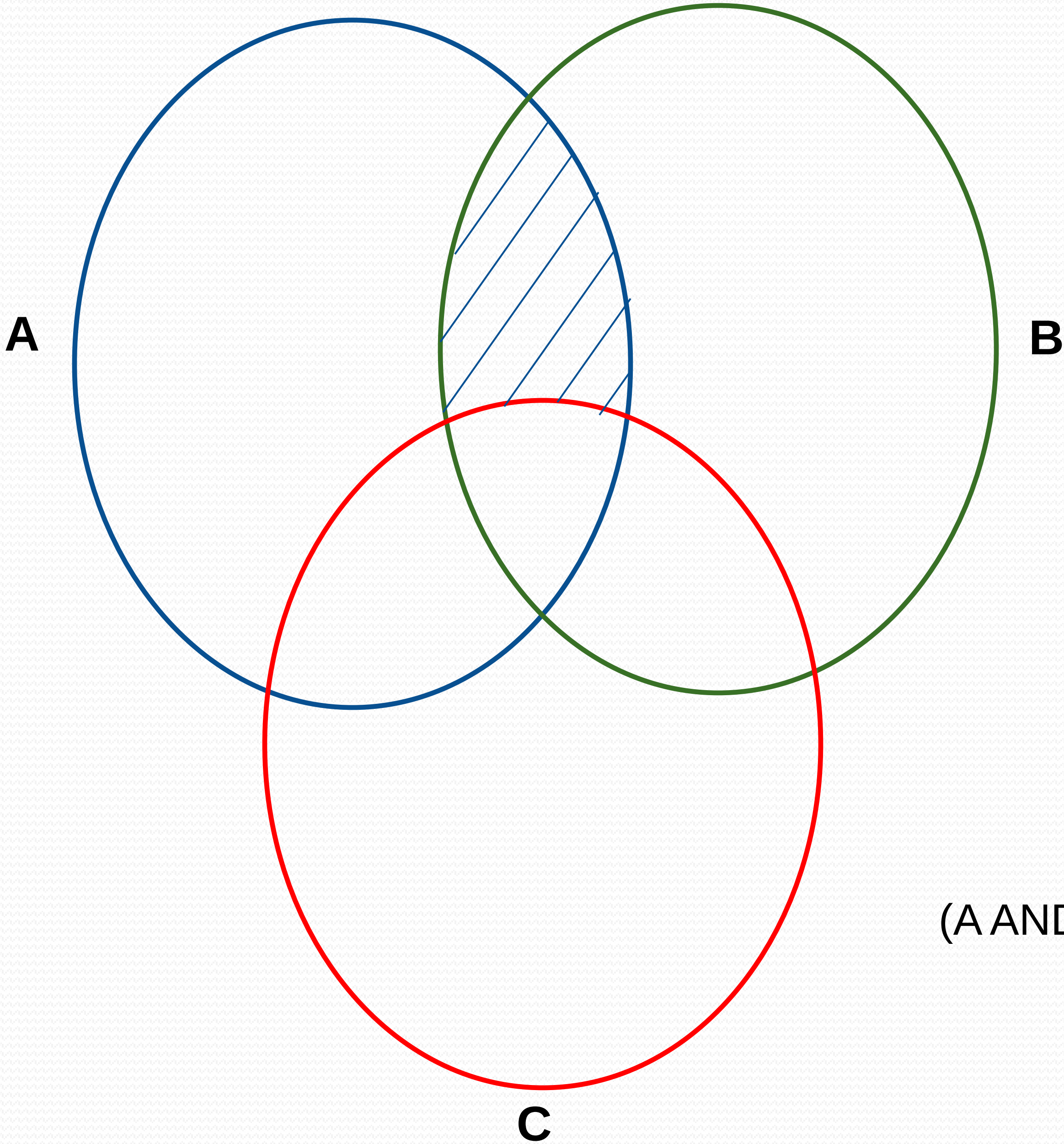
A NOT B



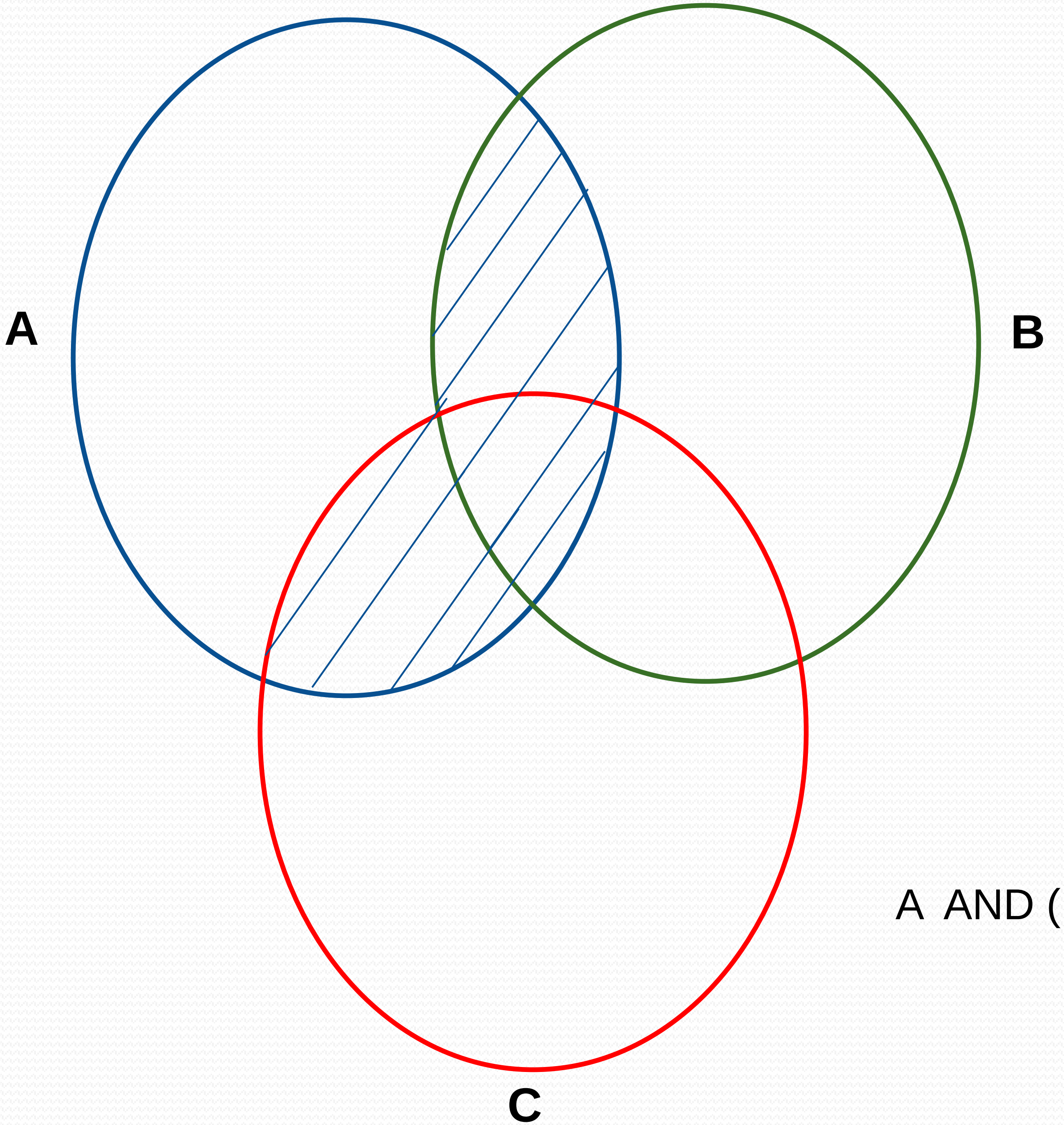
A XOR B



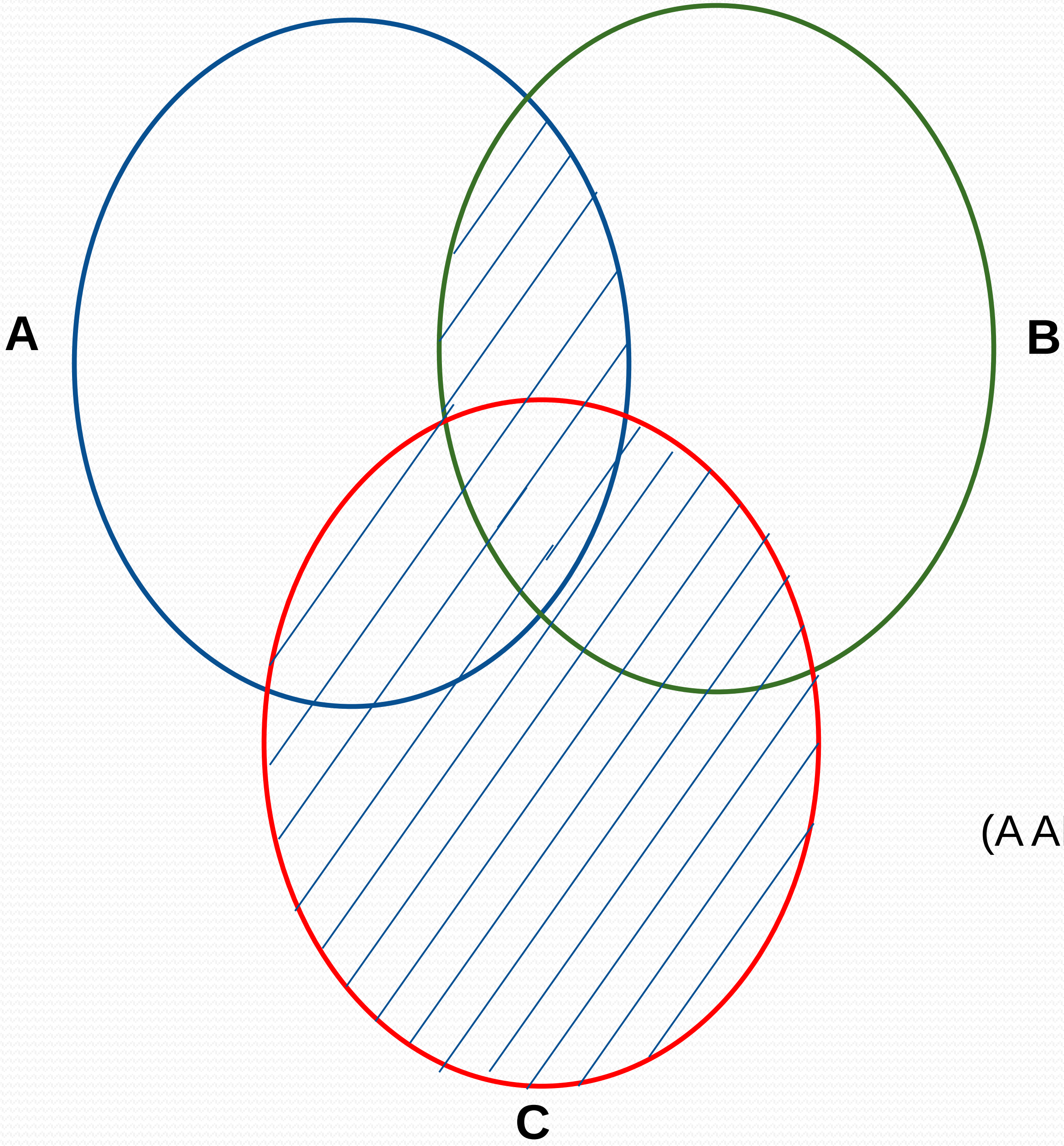
A OR B



(A AND B) NOT C



A AND (B OR C)



(A AND B) OR C

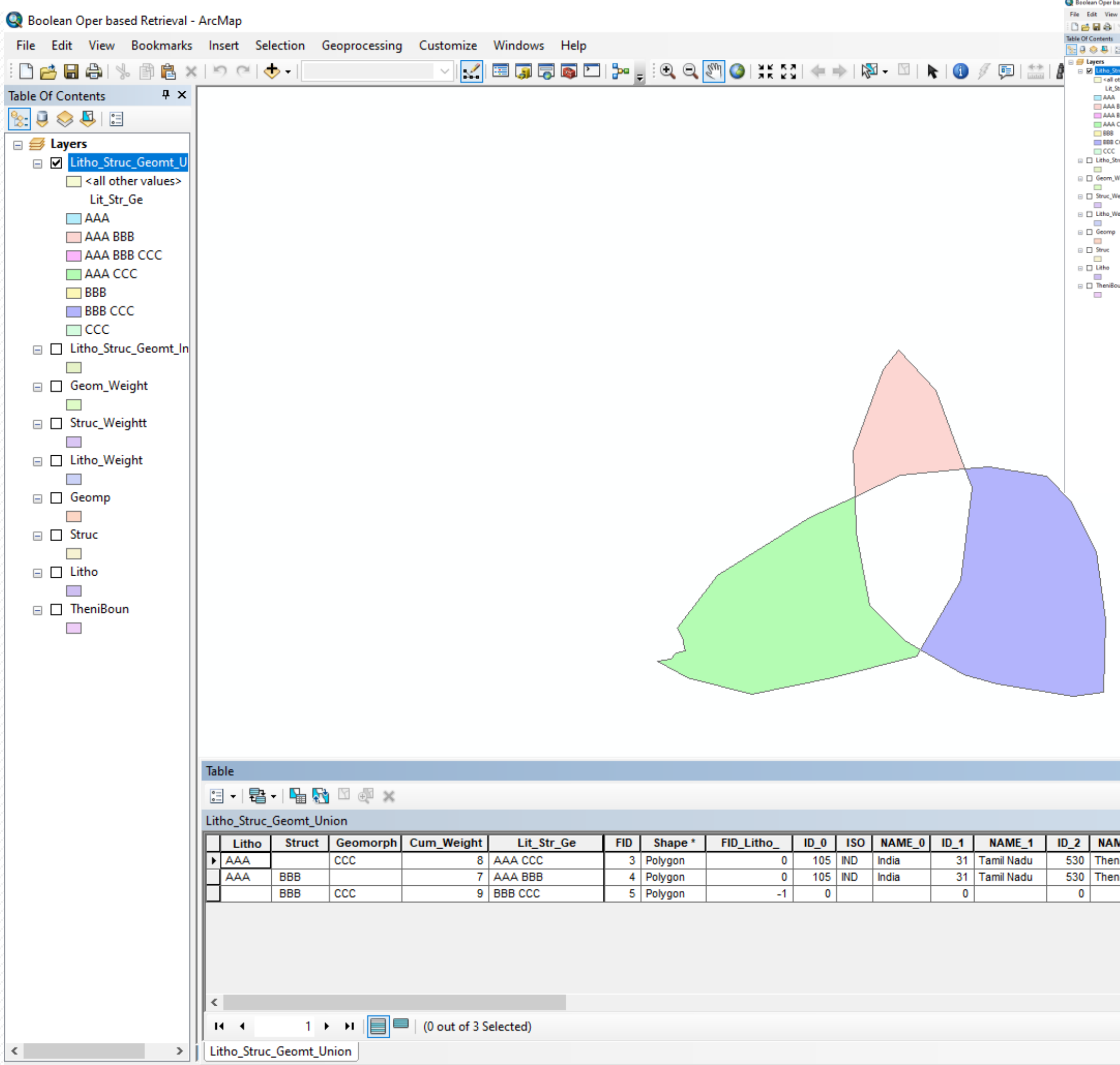
(i) AB OR AC OR BC

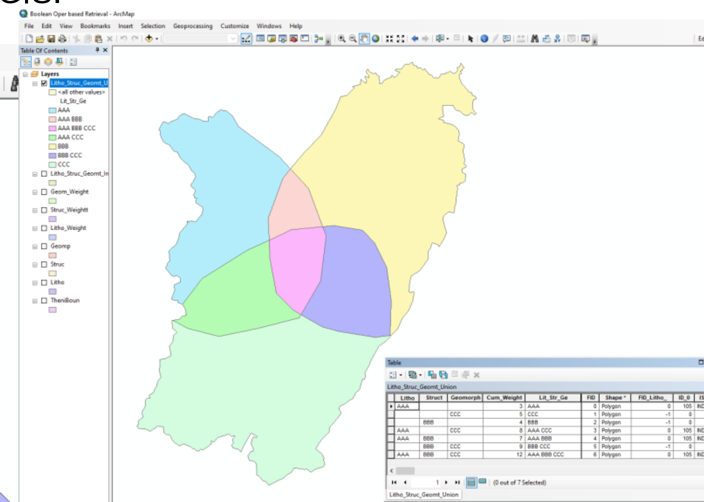
"Lit_Str_Ge" = 'AAA BBB' OR "Lit_Str_Ge" = 'AAA CCC' OR "Lit_Str_Ge" = 'BBB CCC'

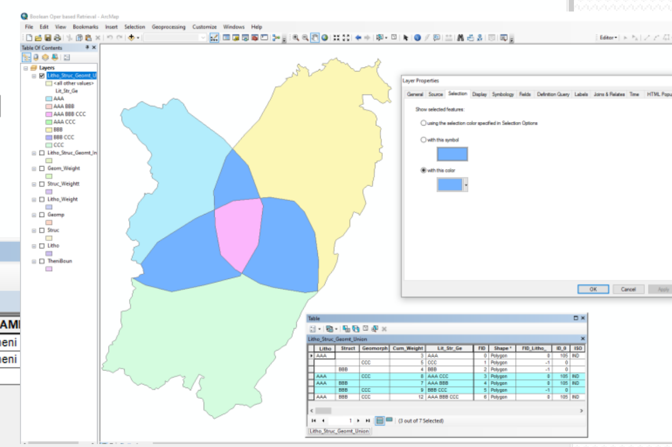
(ii) NOT ABC AND NOT (A OR B OR C)

NOT "Lit_Str_Ge" = 'AAA BBB CCC' AND NOT ("Lit_Str_Ge" = 'AAA' OR "Lit_Str_Ge" = 'BBB' OR "Lit_Str_Ge" = 'CCC')

(iii) "Switch selection" option – useful for selection of targeted spatial features in GIS.







SQL (Structured Query Language) is a programming language for storing and processing information in a relational database. SQL is used to create a database, define its structure, implement it, and perform various functions on the database. SQL is also used for accessing, maintaining, and manipulating already created databases. SQL is a well built language for entering data, modifying data, and extracting data in a database.

Open Database Connectivity (ODBC) is a standard that allows applications to access data from various database management systems (DBMSs).

Here are some concepts related to ODBC:

ODBC-enabled applications
These applications can use ODBC to communicate with databases and retrieve, analyze, and manipulate data. Examples include Microsoft Excel, Tableau, Crystal Reports, and Microsoft Power BI.

ODBC data sources
These are the files or databases that an application can access. Examples of data sources include SQL Server, Oracle RDBMS, a spreadsheet, and a text file.

ODBC drivers
These are software components that act as an interface between an application and a specific DBMS. Each DBMS requires a different ODBC driver.

ODBC API
This defines a standard version of the SQL database language and other functions needed for communication between an application and a database.

Connection information
This includes the server location, database name, logon ID, password, and various ODBC driver options.

DSN
This stands for Data Source Name. It can be defined using the ODBC Data Source Administrator or a DSN file.

To connect to a data source, an application creates ODBC API function calls, such as a SQL statement. The application sends SQL commands to the ODBC Driver Manager, which then routes those commands to the appropriate driver for the database in use.

Cartographic modeling

- The procedure of using map algebraic techniques such as mathematical operation –i.e., Layer Addition, to build models for spatial analysis and represent the real world earth system processes in GIS environments is known as Cartographic modeling.

- The method offers benefits that can improve modeling efficiency and demonstrate model reliability.

- Because GIS relies largely upon spatial data, original inputs as well as intermediate output layers can be shown in map form, which in turn permits visual determination that the modeling steps are moving toward a logical end point.

- Cartographic models can take three forms (**descriptive, prescriptive** and **normative**).

- **Descriptive models** are used to characterize position and form, or to synthesize spatial relationships and establish spatial association, suitability, and possibly determine if relationships are hierarchical (Tomlin, 2013).

- Descriptive models can be transformed into **prescriptive models** by focusing attention on one particular factor, examining the range of parameters that could be modified to meet a specific objective.

- **Normative models** are not covered here but could be achieved by transforming the prescriptive model so as to limit the choice of parameter values in accordance with existing regulations and policies in place in the county.

- In addition, it would be possible to add steps to the model to assign weights to each criterion and then compute a score for each biogas site reflecting how close is its location to optimality (i.e., the highest scored site would be established as most optimal for these criteria).

- Weights could be chosen in accordance with established regulations to bring the model closer to normative status.

Advantages of Cartographic modeling :

- **Easy for programming**
- **Simple operations are involved and hence it is a simple model**
- **Quick completion of the task in GIS**
- **Customization of several tedious steps – works are completed easily**
- **Can improve modeling efficiency and demonstrate model reliability.**

Disadvantages of Carto modeling:

- **Restricted to certain software modules**
- **Error handling is needed**
- **Limits with the size of the layers / coverages**

Cartographic Modeling using Natural Language Commands

- Useful for the programmers and those who willing to type commands
- Simple terms are used as "commands" to perform the GIS works. For e.g,.
  - Buffer Point <input layer> buffer distance
  - Intersect <Input layer> <intersect Layer>
  - Classify <input layer> No. of classes required; method (e.g.,Equal, natural breaks…)

- All the sequences of commands can be listed in a Run Model text box and the commands can be executed by providing the required inputs through interactive method.
- Typographical errors are not allowed – this includes unnecessary spaces, special codes / marks…
- SML – Simple Macro Language; AML – Arc Macro Language were the modules of DOS versions of ArcInfo s/w.

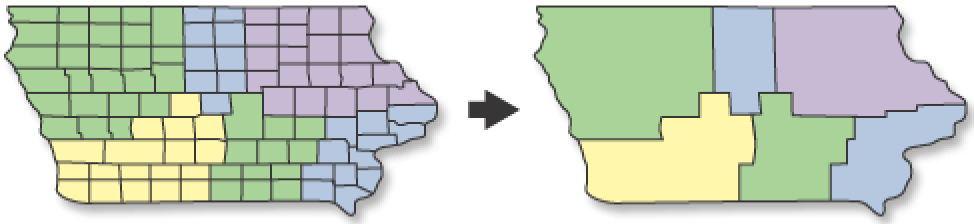
Reclassification / Post – Processing and display

- After the retrieval of required data, reclassification is necessary
- So that redisplay of the retrieved data can be done which will exactly reflect the user's perception
- The similar polygons in adjacent places / boundary shared polygons having same codes / values / attributes,
- Then Dissolving has to be done
- Then, recompute the polygon structure again

Due to the multiple fractionated classes of polygon groups with 1000’s of feature combinations in each, it is important to provide clarity to represent the output.

Similar to processing of data prior to analysis, post processing of out put is also necessary.

Data Classification/Grouping , Regrouping,/Reclassification



Finally using option like,

DISSOLVE

MERGE

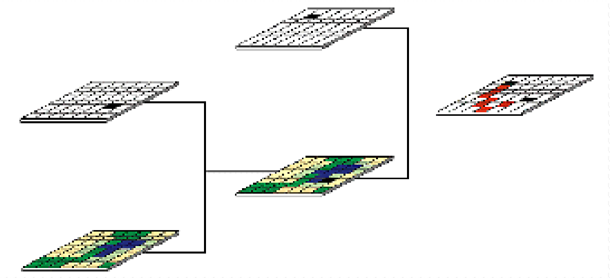
UPDATE

CALCULATE RASTER ……

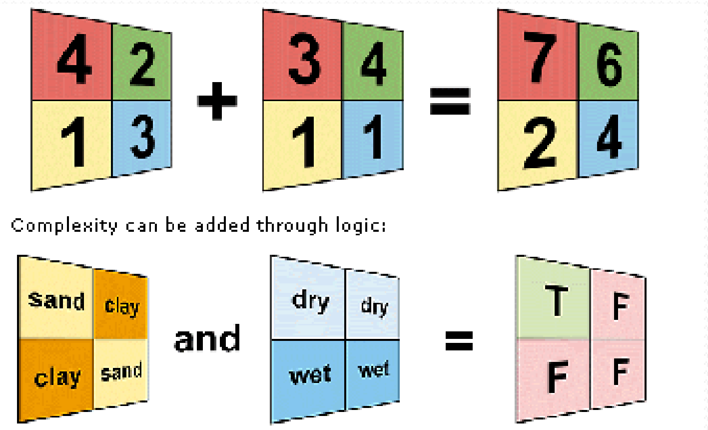
Simplified user friendly outputs can be presented.

GIS Modeling

- A model is a representation of reality.
- Due to the inherent complexity of the world and the interactions in it, models are created as a simplified, manageable view of reality.
- Models help to understand, describe, or predict how things work in the real world.

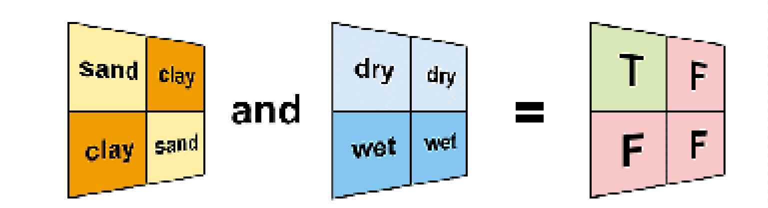


- **Binary model**
- **Process model**
- **Index model**
- **Regression model**
- …
- …



GIS MODELLING ...CONTD...

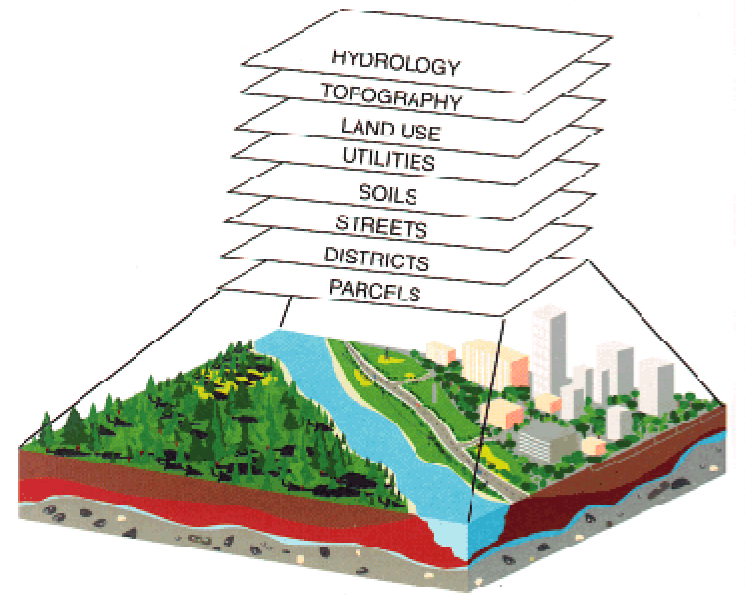
Binary Model



Yes / No, True / False, 1 / 0, etc.

[Back]

Process Model



Lookup Table (LUT)

Combined Attribute table containing the values / classes available from the both thematic layers involved in GIS integration.

1st **Thematic Map** / Layer having 3 classes – e.g. Thickness of Top soil (TTS) High, Moderate & Low

+

2nd **Thematic Map** / Layer having 3 classes – e.g. Water level (WL) - Deep, Moderate & Shallow

	TTS-H	TTS-M	TTS-L
WL-D	DH	DM	DL
WL-M	MH	MM	ML
WL-S	SH	SM	SL

In Binary model, Thematic layers can be classified with only two classes, Yes - No. In this case, once the GIS integration is performed then, the number of combined classes in the resultant layers can be easily determined by a simple formula,

$$N_{rc} = 2^n - 1 \,,$$

Where,

n = no. of layers having Y/N classes involved in GIS integration analysis.

N_{rc} = resultant number of classes

GIS Integration of **Two layers** in binary model

	TTS-1	TTS-0
WL-1	1 1	1 0
WL-0	0 1	0 0

	SLP-1	SLP-0
1 1	1 1 1	1 1 0
1 0	1 0 1	1 0 0
0 1	0 1 1	0 1 0
0 0	0 0 1	0 0 0

GIS Integration of **Three layers** in binary model

First step of GIS integration in Binary model will result 3 + 1 (1-unwanted) classes.

In the **Second step** of GIS integration, 7 +1(unwanted) classes are resulted.

$$N_{rc} = 2^3 - 1$$

	GEM-1	GEM-0
1 1 1	1 1 1 1	1 1 1 0
1 0 1	1 0 1 1	1 0 1 0
0 1 1	0 1 1 1	0 1 1 0
0 0 1	0 0 1 1	0 0 1 0
1 1 0	1 1 0 1	1 1 0 0
1 0 0	1 0 0 1	1 0 0 0
0 1 0	0 1 0 1	0 1 0 0
0 0 0	0 0 0 1	0 0 0 0

Similarly, for the GIS Integration involving **Four layers** in binary model will result into (**$2^4 - 1$**) = **15** + 1 (1-unwanted) classes.

Network Analysis / Model

- **Network** – Consists of connected linear features
  - E.g. Road, Railway, Bicycle paths, Streams & Shorelines.
- **Dynamic Segmentation**
  - A data model that is built upon lines of a network and allows the use of real-world co-ordinates with linear measures such as mileposts
  - To link linearly referenced data such as accidents & pavement conditions to geographically referenced road network
- For this, the networks need to be segmented and labelled by inserting nodes in proper locations
- For both **NW/DS** - Appropriate attributes are to be collected as per the real-world applications and entered in computer for 'Linking'.
  - For e.g., to use road network for Path Finding / Resource Distribution, etc., - we need the following attributes:

- Travel time
- Impedance at turn
- One-way streets
- Linear measures – to be added with arcs and nodes of the network – can be directly linked to road segments.

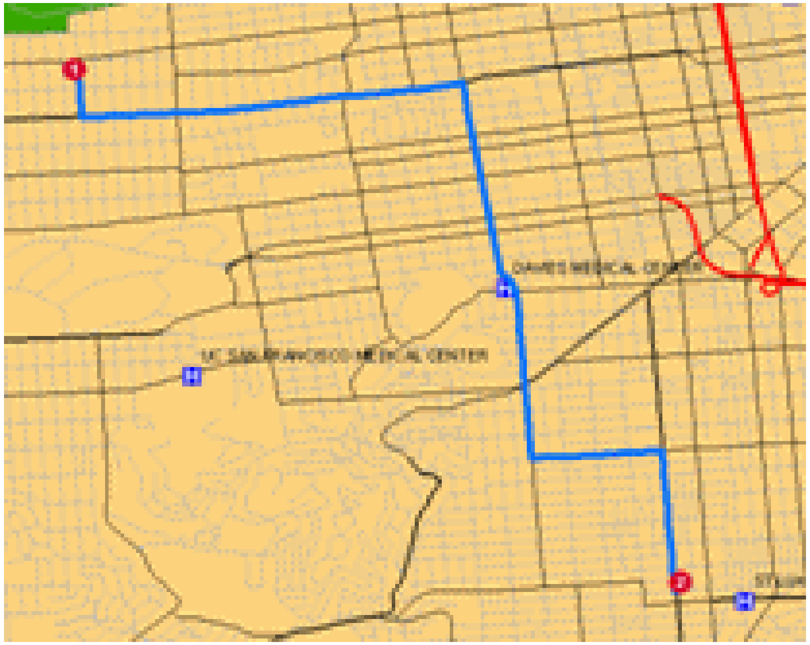
- 1. **Network** – is a line coverage, which is topology based and has the appropriate attributes for the flow of objects such as traffic.
  - Linear Feature + Attribute data, such as
    - Impedance values assigned to
      - network links – link impedance
      - Turns – turn impedance
    - One-way streets and
    - Overpasses and under passes.

Link impedance

- A link refers to a segment separated by two nodes in a road network

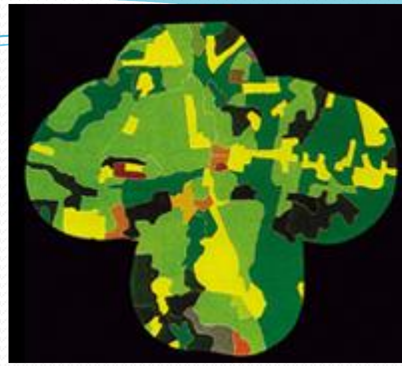
NETWORK ANALYSIS

Route Identification

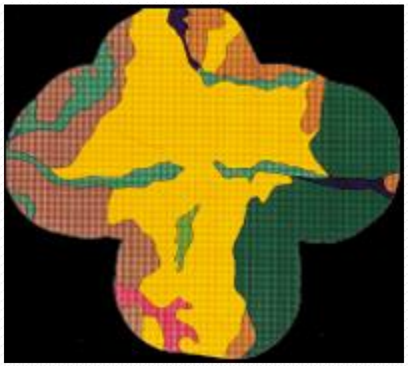


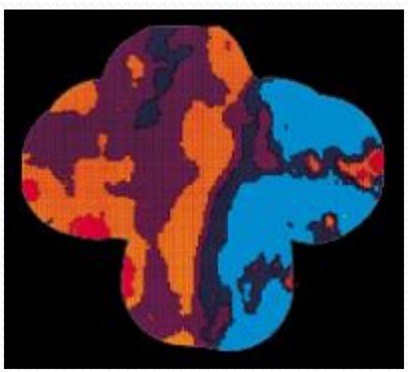
Identification of Potential groundwater well sites

Land use and land cover data for the area bounded by a half-mile buffer zone around the water company service area.



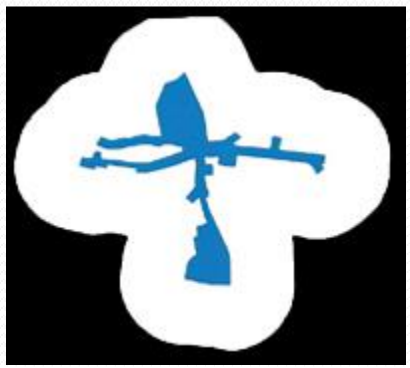
Map of surficial geology of the water service area





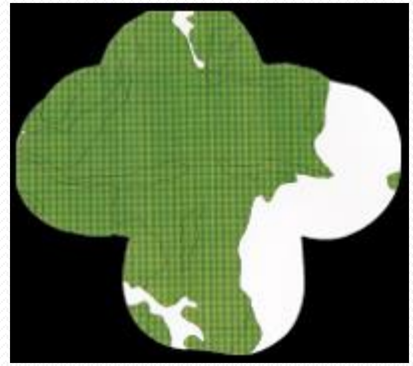
A bedrock elevation subtracted from water level elevation by a GIS to show the thickness of water-saturated sediment

A half-mile buffer zone drawn around the service area

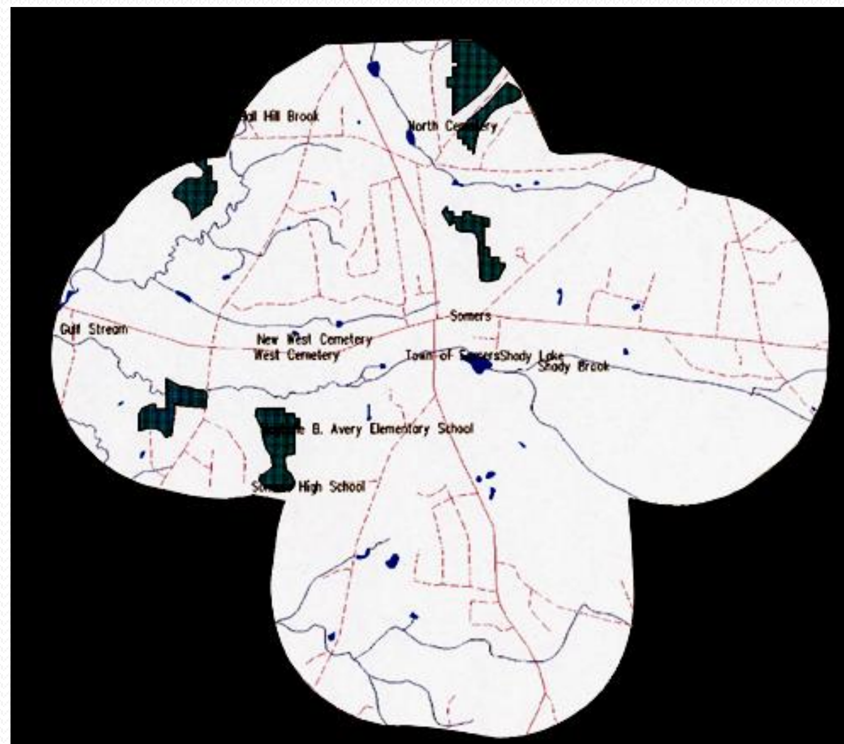


Potential sites with saturated thickness of sediments greater than 40 feet.

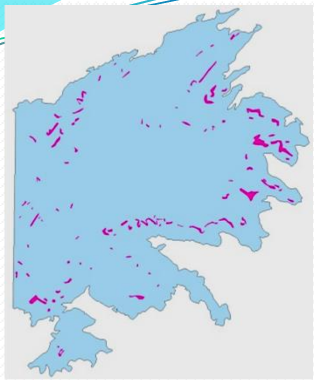
Selection areas of sand and gravel from the map of surficial geology



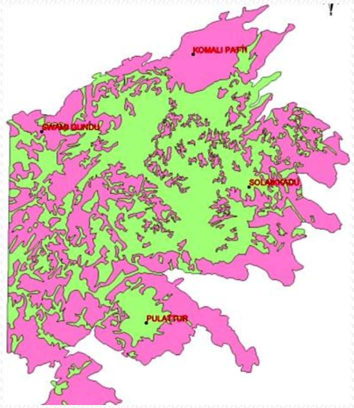
Buffer zones of 500 meters are drawn around the point sources of pollution.



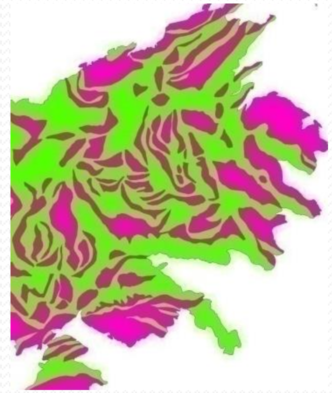
LANDSLIDE HAZARD ZONATION AND INDUCING PARAMETER IDENTIFICATION



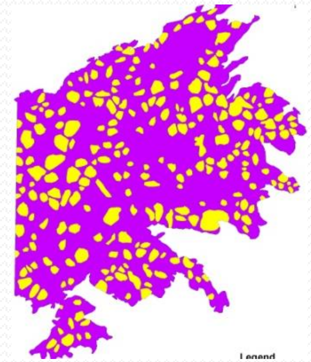
Mapping of Escarpments



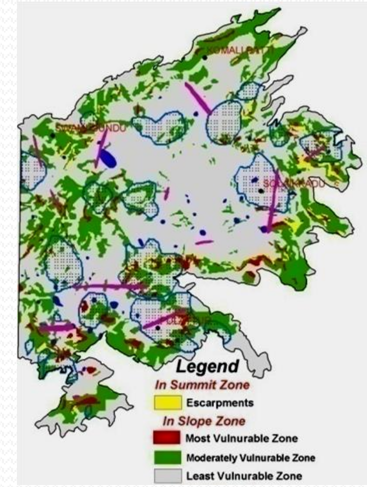
Identification of Active slope areas



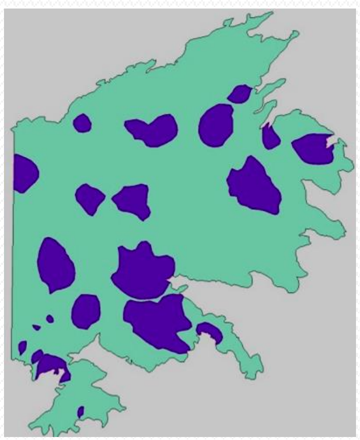
Interpretation of dip & obsequent slopes



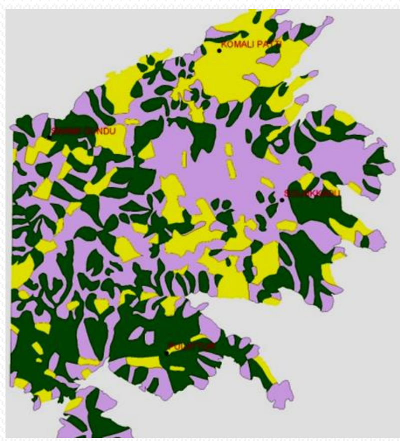
Map out dissected slopes



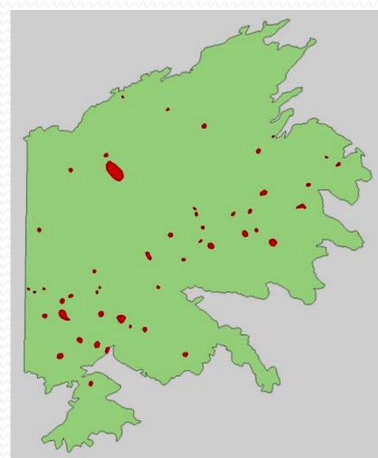
Landslide Hazard Zones



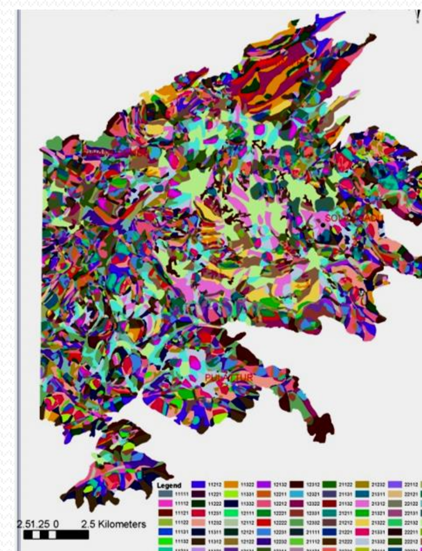
Zones of dendritic drainages



Filter out Convex & plain slopes

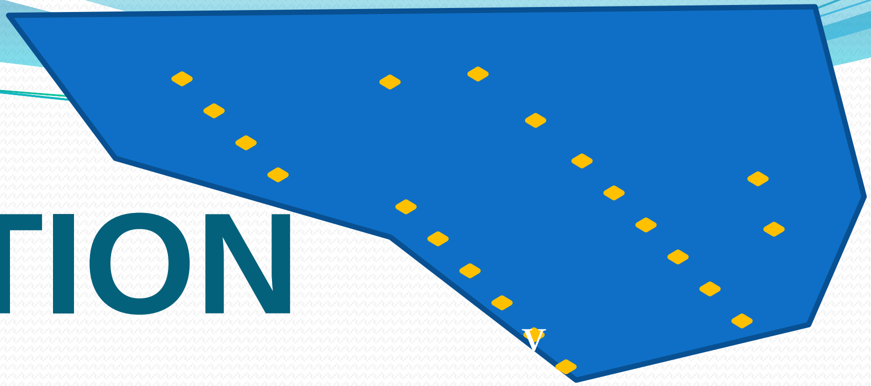


Buffer out sensitive toes



Landslide inducing parameters

SPATIAL INTERPOLATION



- A process of using points with known values to estimate values at other points or in data gap areas.
  - For e.g., the precipitation value at a location with no recorded data can be estimated through interpolation from known precipitation readings at nearby weather stations.

- In GIS, it is applied to a grid and estimates are made for all cells in the grid.

- A means of converting **point data to surface data / point-to-area conversions** – also known as **Spatial Manipulation**.

Necessity of spatial interpolation:

- To visualize an area by generating a map layer either in raster or vector mode.
- To analyse and make model as a semi-continuous or discontinuous piecewise surface, depending on the scale of the attribute.
- To generate a new set of data by re-sampling at a regular grid basis, and build point attribute table for multivariate analyses, and
- To improve the property of any unvisited sites.

Types of Spatial Manipulation:

- Methods of converting points to areas – i.e., Manipulation can be subdivided into two groups:
  - Non-interpolative method and
  - Interpolative method.

- Non-interpolative methods are particularly appropriate when the point attribute is measured on a categorical measurement scale, but can also be useful in some cases for an attribute measured on an ordinal, interval or ratio scale.

These are the four scales used mainly while collecting data:

- **Nominal/Categorical**: Used to categorize data into mutually exclusive categories or groups.
- **Ordinal**: Used to measure variables in a natural order, such as rating or ranking. They provide meaningful insights into attitudes, preferences, and behaviours by understanding the order of responses.
- **Interval**: Used to measure variables with equal intervals between values. This type of measurement is often used for temperature and time, allowing for precise comparisons and calculations.
- **Ratio**: Allows for comparisons and computations such as ratios, percentages, and averages. Great for research in fields like science, engineering, and finance, where ratios are used, percentages, and averages to understand the data.

Choropleth map:

- A map with line boundaries that are defining polygons enclosing areas that are assumed to be uniform or to which a single description can be applied.
- In choropleth map,
- Polygons cannot be finely divided into smaller entities
- we cannot say anything more precise about what is happening within their boundaries
- If we have point observations arranged either at random or on a regular lattice over the area of interest,
- Then it is possible to make more precise statements about the value of properties of interest at unvisited sites.

Choropleth 2D & 3D maps

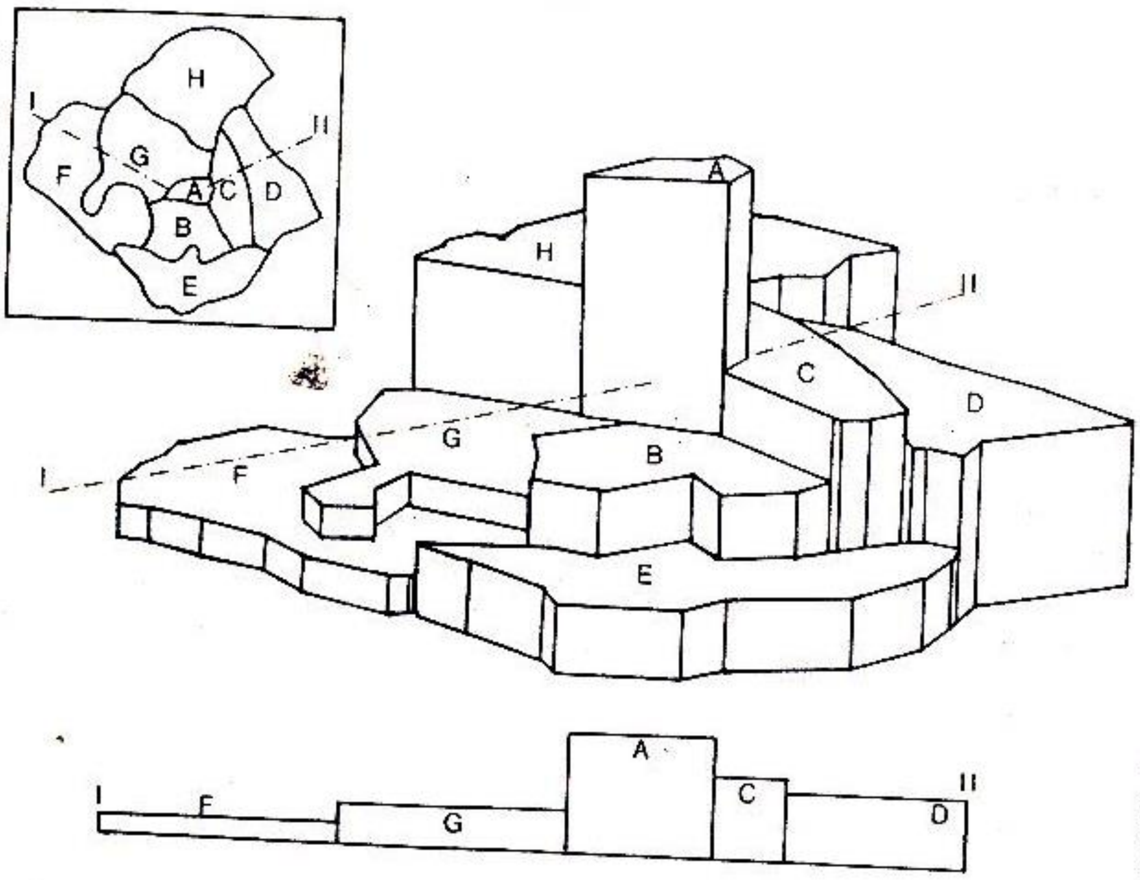
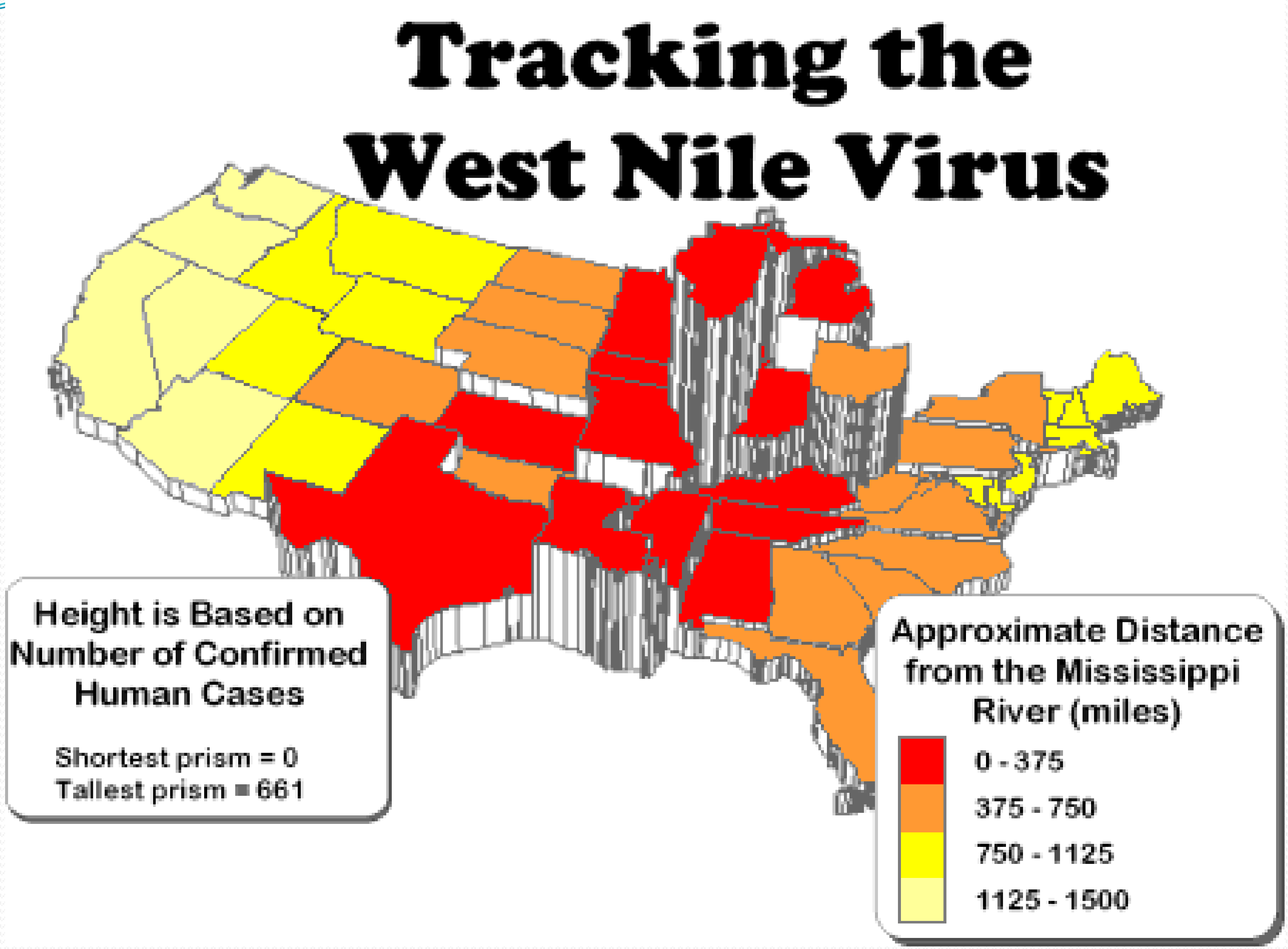
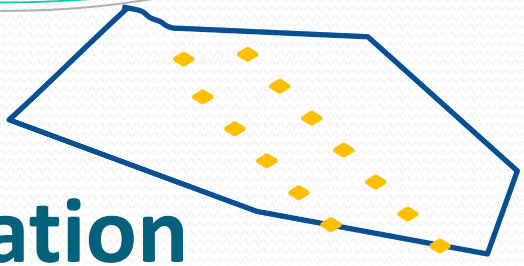


Fig. 8.1 The choropleth map is 'stepped' model of the landscape. (Based on a screen image, courtesy Computervision Corporation, The Netherlands.)


Tracking the
West Nile Virus
Height is Based on Number of Confirmed Human Cases
Shortest prism = 0
Tallest prism = 661
Approximate Distance from the Mississippi River (miles)
0 - 375
375 - 750
750 - 1125
1125 - 1500

Manipulation – Interpolative methods - 2 types:

1. Interpolation, & 2. Extrapolation



- This procedure of “Estimating the value of properties at unsampled sites **within** the area covered by existing point observations is called **interpolation**”.
- “Estimating the value of a property at sites **outside** the area covered by existing observations is called **extrapolation**.”
- The logic / basic idea behind Interpolation and Extrapolation is that,
- on average, points that are close together in space are more likely to have similar values than point further apart.

- For example,
- Two observation points a few metres apart on a single hill are more likely to have similar value for altitude,
- But, if the points on two hills some kilometers apart may not have similarity in height and we cannot say about in between point heights.
- Two different types of **spatial interpolation** methods are:
  - **Global** and
  - **Local** methods.
- The difference between these two groups lies in the use of control points, i.e., points with known values, in estimating unknown values.

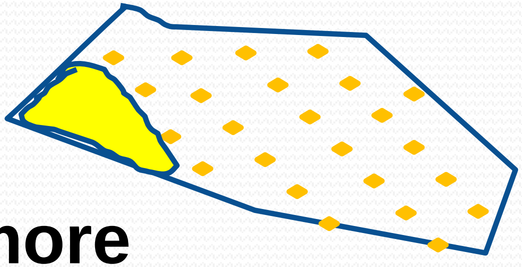
Methods of Spatial Interpolation / Manipulation

- **Global method** - uses every control point available in estimating an unknown value – to derive an equation or a model.

- **Local method -** uses a sample of control points in estimating an unknown value.

- Control points - are the points with known values.

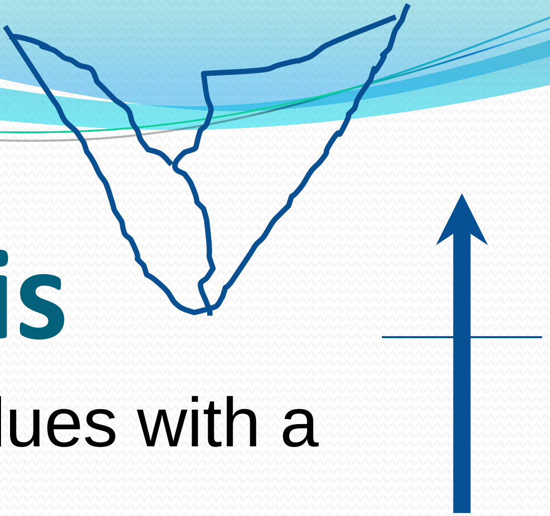
- Generally used terminology for manipulation as a whole, or representing either extrapolation or interpolation, it is holistically known as **Spatial Interpolation**. It becomes commonly used term now-a-days.



Basic assumption in spatial interpolation:

- The value to be estimated at a point is **more influenced** by **nearby control points** than those that are farther away.
- To be effective for estimation, **control points** should be **well distributed** within the study area - this ideal situation is rare in real-world applications.
- **Data-poor areas** – represent a major problem in estimation and cause problems in spatial interpolation.
- The <u>**number**</u> and <u>**distribution**</u> of control points can greatly influence the accuracy of spatial interpolation.

GLOBAL METHODS
Trend Surface Analysis



- Approximates points with known values with a polynomial equation.
- A linear or first-order trend surface uses the equation

$$Zxy = b_0 + b_1x+b_2y$$

Where,

z = attribute value, is a function of x and y

b = coefficients estimated from the control points

- It is computed by the "least squares method"
- A method of approximation based on the minimization of the squared distance between two sets of variables.

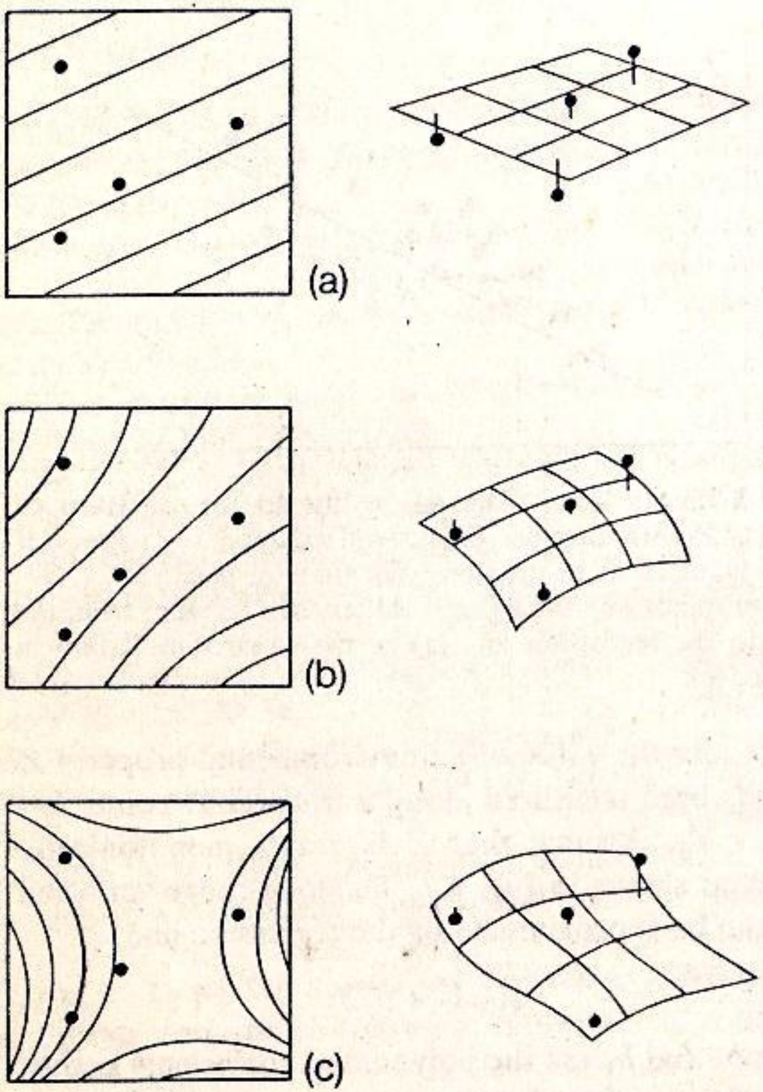
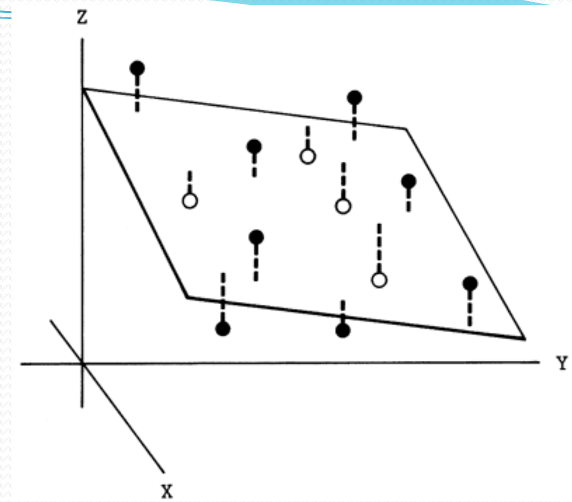
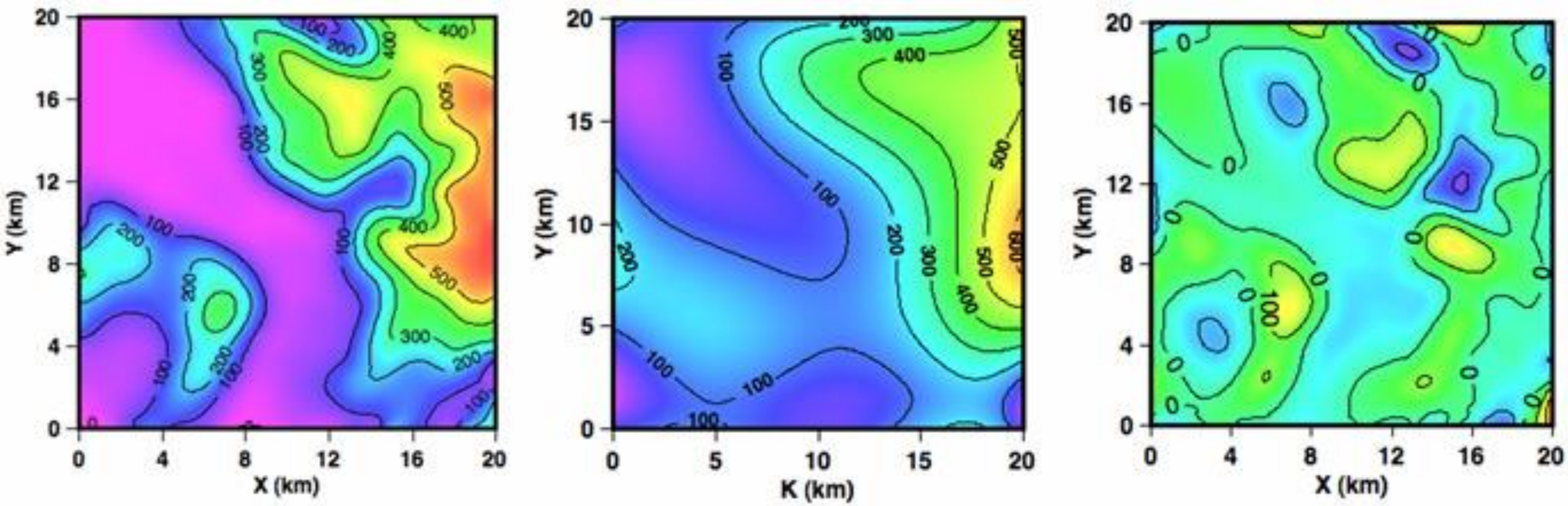
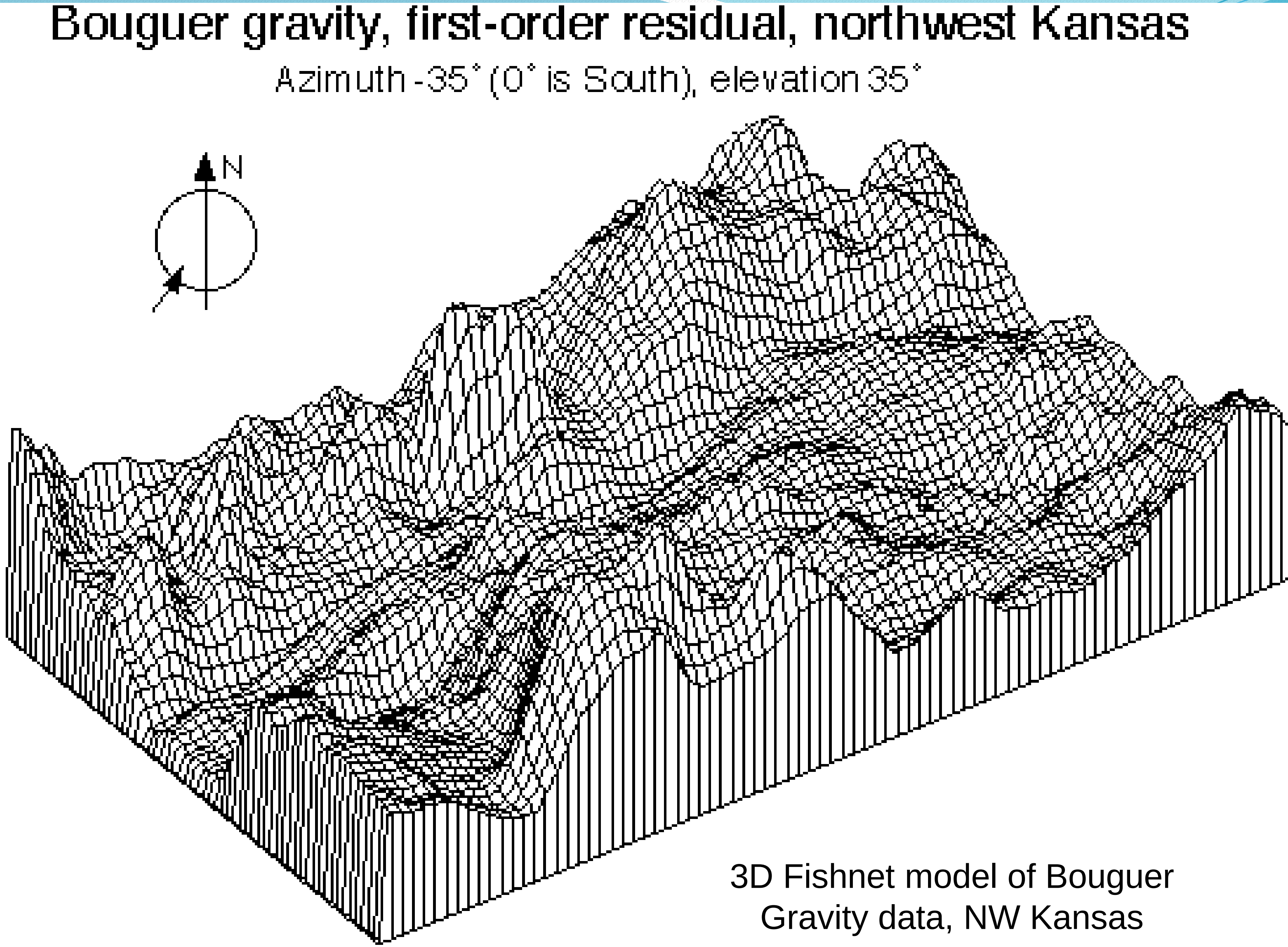


Fig. 8.5 Trend surfaces in two spatial dimensions. (a) Linear, (b) quadratic, (c) cubic.





A Graphical Example of Applying Trend Surface Analysis to Matrix Data: (a) Original Surface, (b) the 6th-orderTrend Generated Surface, and (c) the Residual Matrix



3D Fishnet model of Bouguer Gravity data, NW Kansas

Regression Models

- A regression model relates a **dependent variable** to a number of **independent variables** in an equation, which can then be used for prediction or estimation.

- Many regression models use non-spatial attributes such as income, education etc., and are not considered as methods for spatial interpolation.

- Some exceptional models are available which uses the location and topographic variables as the independent variables and the terrain parameter or thematic data as dependent variable.`

Regression Models...contd...

- Simple Linear Regression - By considering the values of the one Dependent and Independent parameters each in all the points of the study area, estimates the relationship of each location in numbers as well as sign (+ve / -ve).

- Multiple Linear Regression – Establishes the one-to-one relationships in numbers as well as their signs (+ve / -ve) by considering the values of the one Dependent and multiple Independent parameters in all the points / locations of the study area.

Global methods in ARC/INFO and ArcView

- TREND – command to run trend surface analysis from 1^{st} to 12^{th} order.

- REGRESSION – command for regression analysis, but without model-selection methods, maximum, R-square, etc.

- AVENUE Script – to run trend surface analysis.

LOCAL METHODS

1. Thiessen Polygons

- Thiessen polygons are constructed around a sample of known points so that any point within a Thiessen polygon is closer to the polygon's known point than any other known points.
- Originally proposed to estimate effective area coverage of precipitation.
- Thiessen polygons require initial triangulation among known points, that is connecting known points to form triangles / polygons.

- **Delaunay triangulation** – each known point is connected to its nearest neighbors, and that triangles are as equilateral as possible (shown as dashed lines in the following Figure-2 (b).
- After triangulation, Thiessan polygons can be easily constructed by connecting lines drawn perpendicular to the sides of each triangle at their midpoints – **Voronoi polygons / Dirichlet cells**.

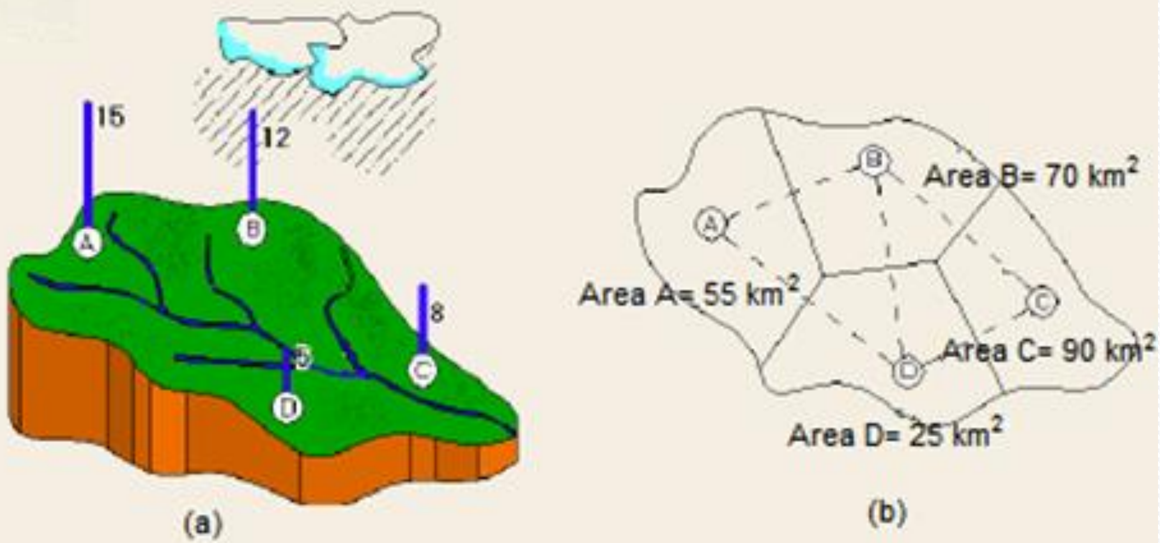


Fig-2: Rainfall measurement by Thiessen Polygon Method
(a) Rainfall Recorded, (b) Areas of Influence

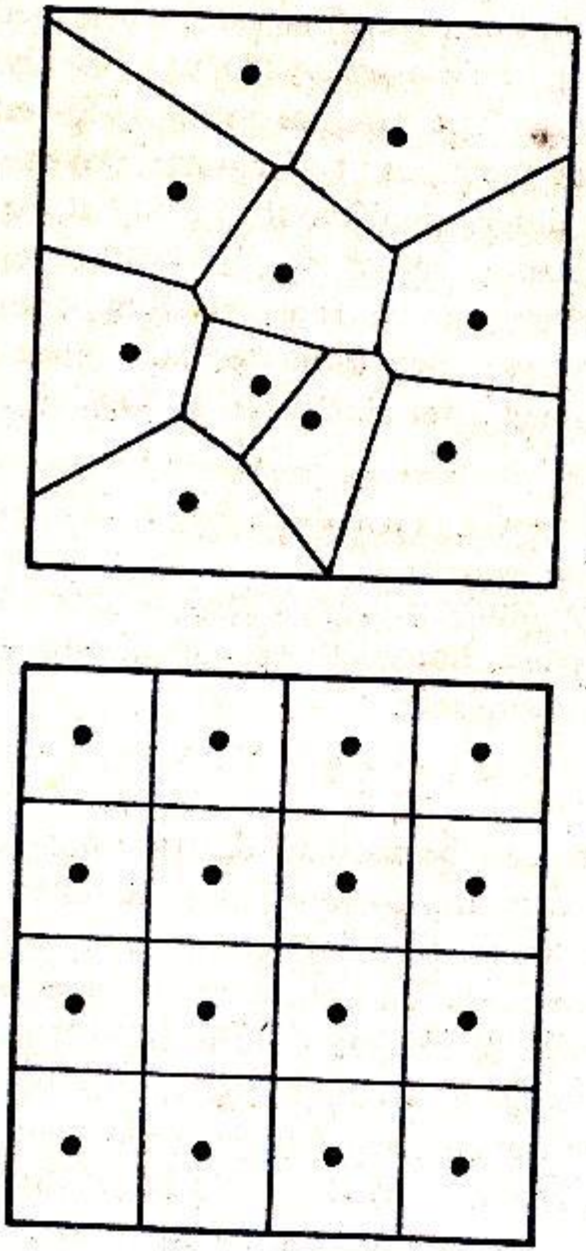


Fig. 8.2 Thiessen (or Dirichlet) polygons with irregular and regular sample point searching.

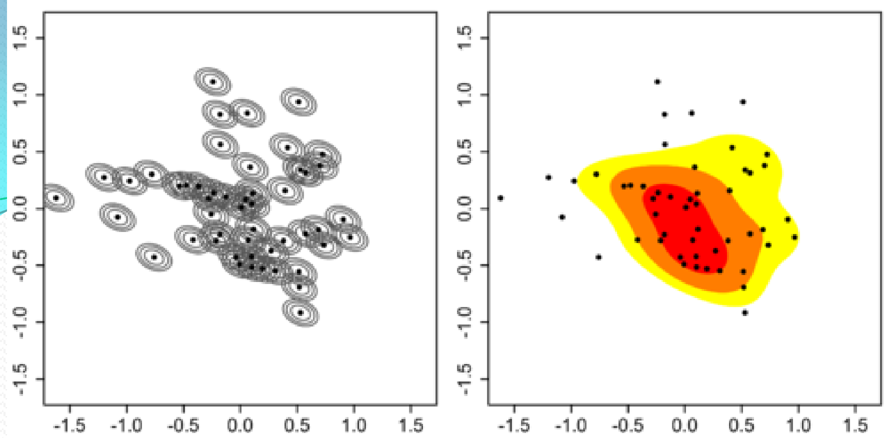
Unless there are many observations, this Thiessen polygon assumptions are not really appropriate for gradually varying phenomena like Rainfall, Air Temperature, Pressure ….
Because:

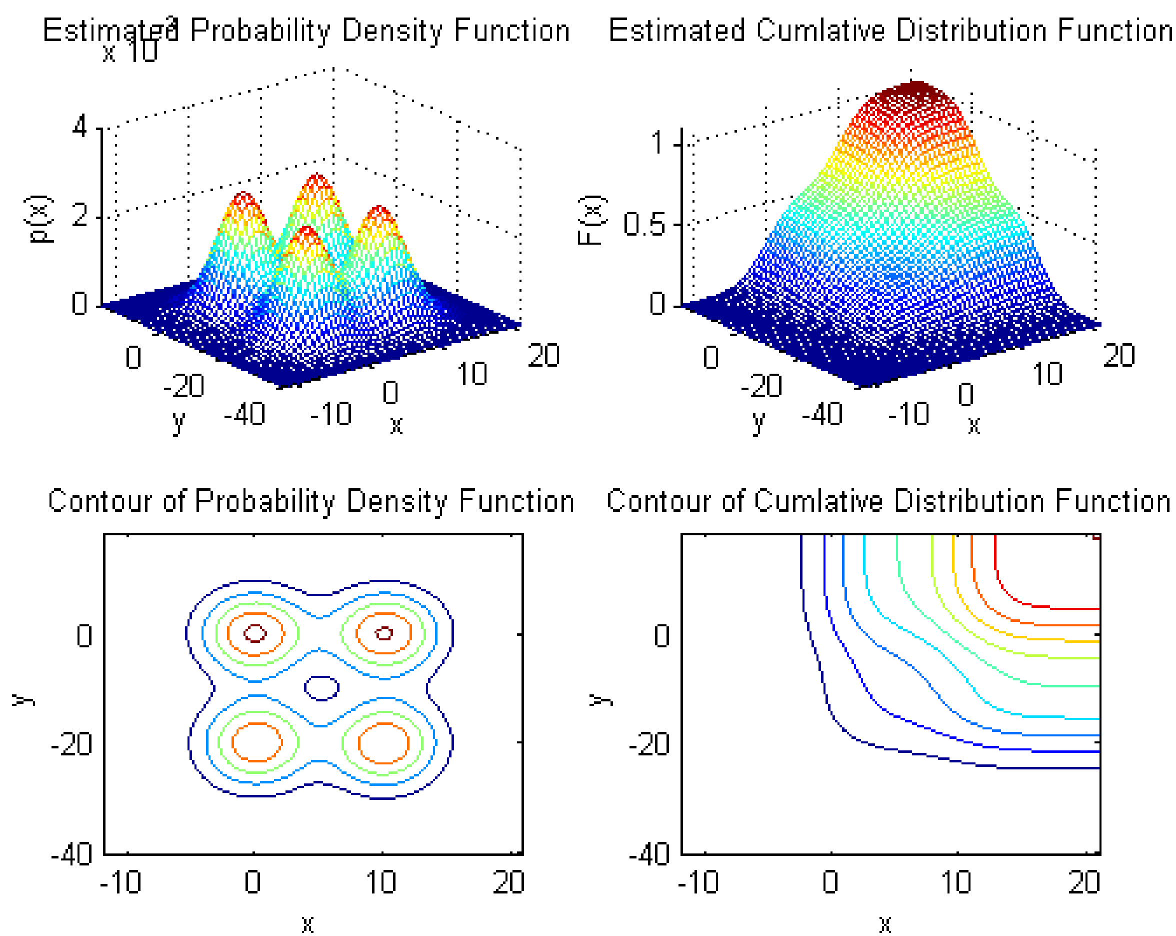
1) The form of the final map is determined by the distribution of the observations and
2) The method maintains the choropleth map fiction of homogeneity within borders and all changes only at borders.

As there is only one observation per tile available, there is no way to estimate **within-tile variability, short of taking replicate observations.**

2. Density estimation

- Density estimation measures densities in a grid based on a distribution of points and their known values.
- **Kernel estimation** is a different density estimation method, which associates each point or observation with a kernel function.
- Expressed as a bivariate probability density function, a kernel function looks like a “Bump”, centering at a point and tapering off to 0 over a defined bandwidth or window area.
- The kernel function and the bandwidth determine the shape of the bump, which in turn determines the amount of smoothing in estimation.



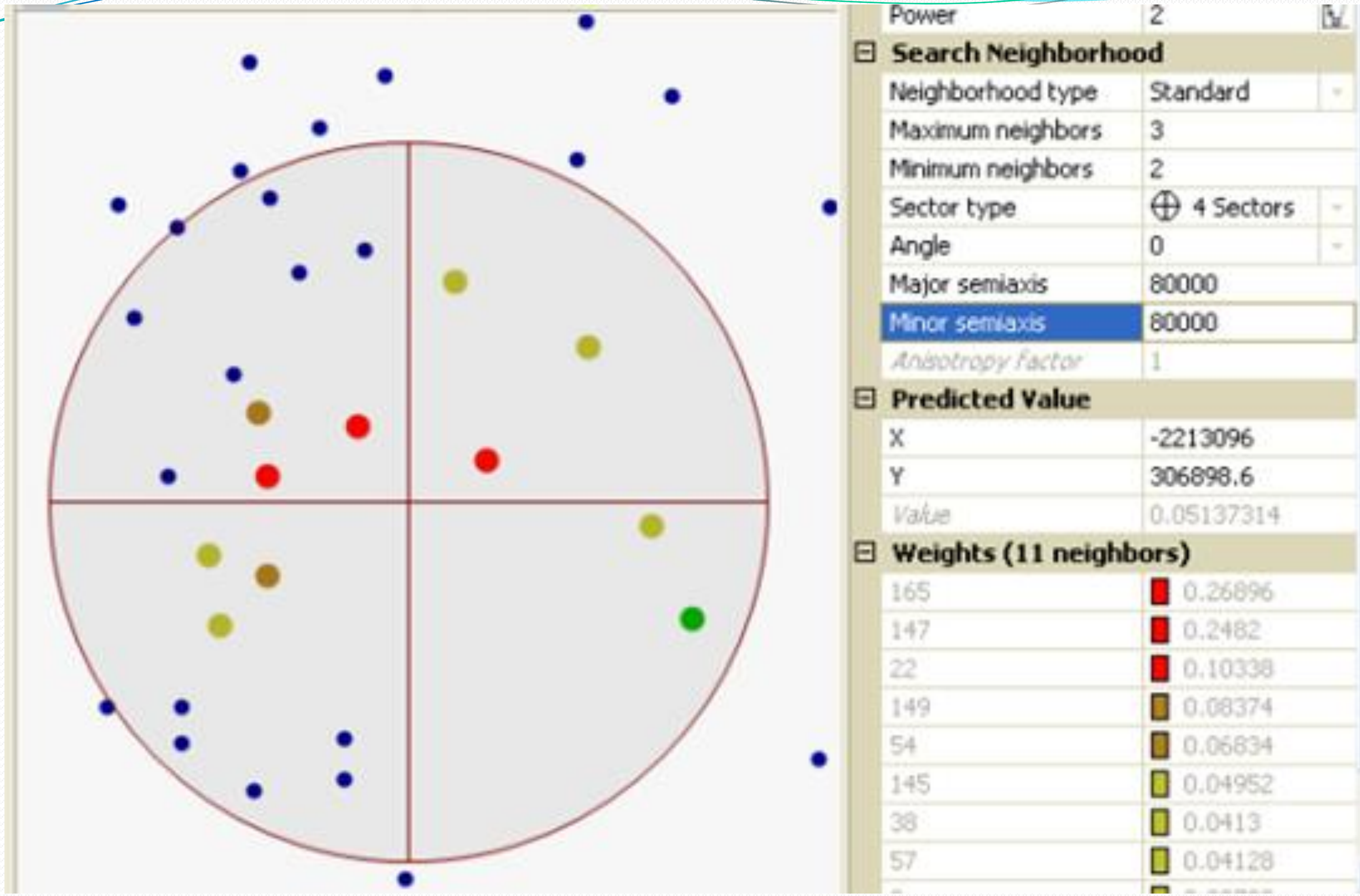


Estimated Probability Density Function
Estimated Cumlative Distribution Function
p(x)
F(x)
Contour of Probability Density Function
Contour of Cumlative Distribution Function

3. Inverse Distance Weighted Interpolation (IDW)

- It is a local method that assumes that the unknown value of a point is influenced more by nearby control points than those farther away.
- A type of deterministic method for multivariate interpolation with a known scattered set of points.
- The assigned values to unknown points are calculated with a weighted average of the values available at the surrounding known points.
- The weighted average applied since it resorts to the inverse of the distance to each known point ("amount of proximity") when assigning weights.

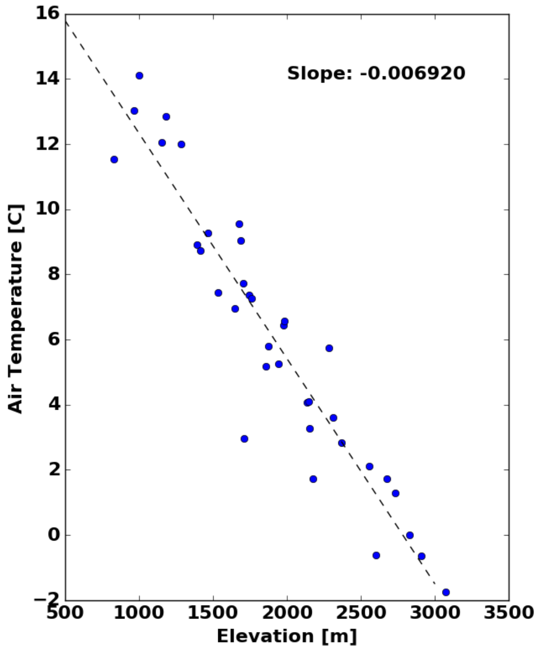
- To predict a value for any unmeasured location, IDW uses the measured values surrounding the prediction location.
- The **measured values closest to the prediction location** have **more influence** on the predicted value than those farther away.
- IDW assumes that each measured point has a local influence that diminishes with distance.
- It gives greater weights to points closest to the prediction location, and the weights diminish as a function of distance, hence the name **inverse distance weighted.**

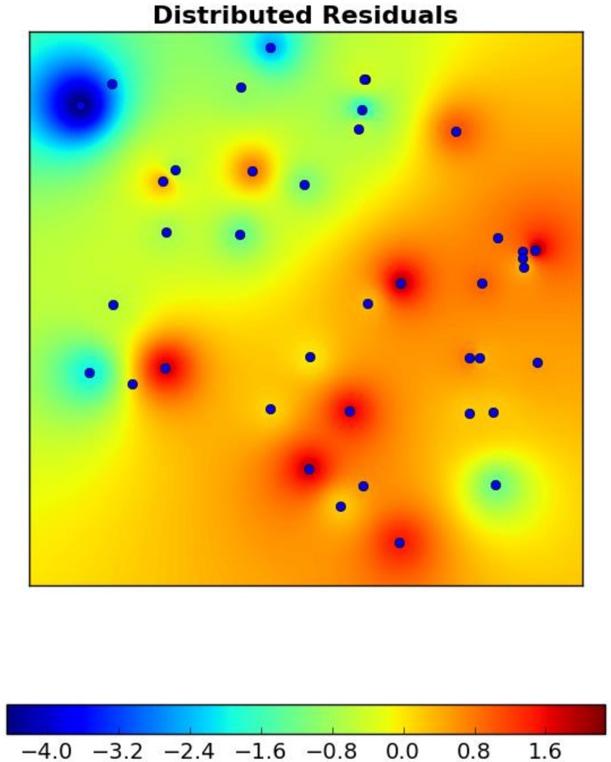


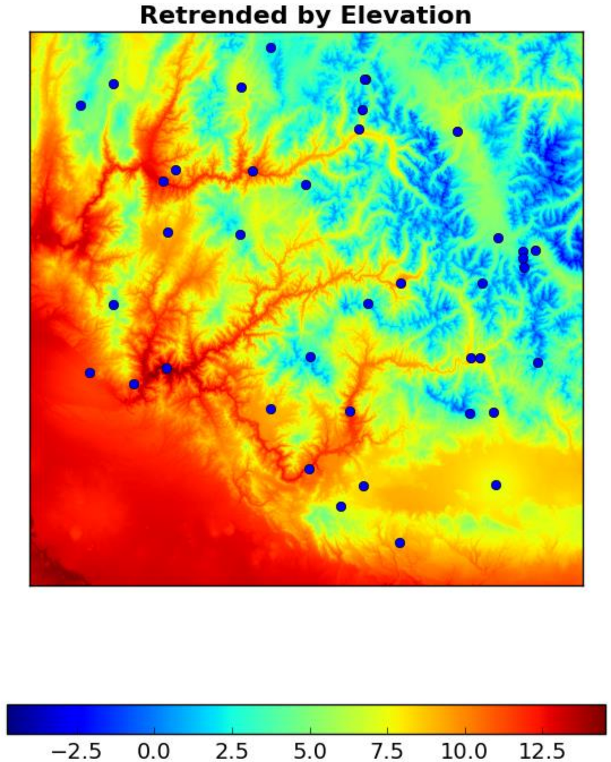
Power 2
Search Neighborhood
Neighborhood type Standard
Maximum neighbors 3
Minimum neighbors 2
Sector type 4 Sectors
Angle 0
Major semiaxis 80000
Minor semiaxis 80000
Anisotropy factor 1
Predicted Value
X -2213096
Y 306898.6
Value 0.05137314
Weights (11 neighbors)
165 0.26896
147 0.2482
22 0.10338
149 0.08374
54 0.06834
145 0.04952
38 0.0413
57 0.04128

In order to get a meaningful map derived through spatial local interpolation is:

1. Prepare IDW image based on the sample location and their values
2. Identify the controlling parameters
3. Derive the relation between them and
4. Use the relationship to retrend the IDW map to get a near accurate one relevant to the Earth System Processes.







4. Thin-plate Splines

- Splines for spatial interpolation are conceptually similar to splines for line generalization, except that in spatial interpolation they apply to surfaces rather than lines.
- The name *thin plate spline* refers to a physical analogy involving the bending of a thin sheet of metal or flexible curves (made of metal wire coated with rubber) used in cartography to make smooth curved line.
- Thin-plate splines creates a surface that passes through control points and has the least possible change in slope at all points.
- In other words, thin-plate splines fit the control points with a minimum-curvature surface.

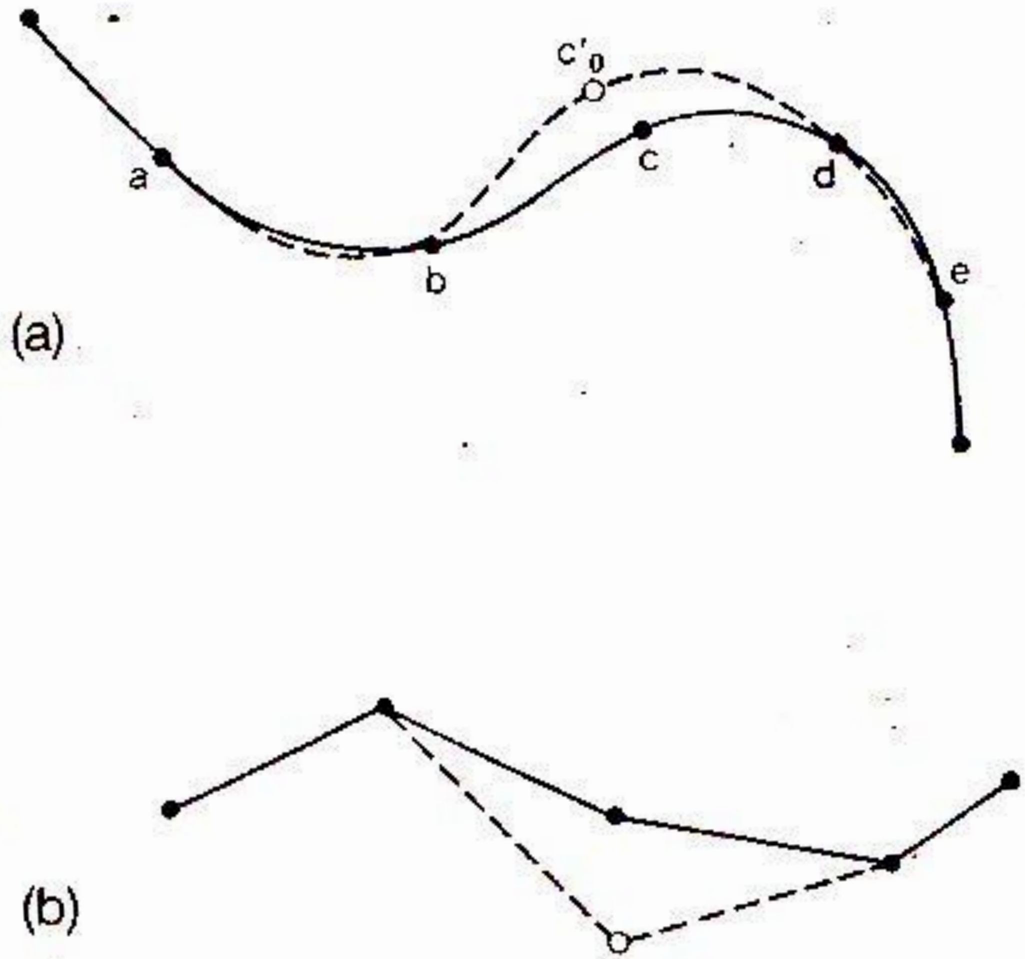


Fig. 8.6 The local nature of splines. When one point is displaced, four intervals must be recomputed for a quadratic spline (a) and two for a linear spline (b).

5. Kriging

- Kriging – the method named after the South African mining engineer D.G.Krige.
- Kriging is a geo-statistical method for spatial interpolation.
- The technique of Kriging assumes that the <u>spatial variation of an</u> <u>attribute</u>.
- For, e.g., changes in grade of ore within an ore body is neither totally random nor deterministic.

- Spatial variation – may consist of *three components*:
  - a spatially correlated component, representing the Variation of the regionalized variable
  - a drift or structure, representing a trend and
  - a random error term.
- The presence or absence of a drift and the interpretation of the regionalized variable have led to development of different Kriging methods for spatial interpolation.

They are:

- **Ordinary Kriging**
- **Universal Kriging**
- **Block Kriging and**
- **Co-kriging**.

Comparison of local methods

- <u>**Cross validation analysis**</u> **is possible in local methods**
- **For assessing the accuracy of estimates from different interpolation methods**

It involves,

- **Removal** of a **known point** from the data set and
- **estimating the value** at the point by **using** the **remaining known points** and **calculating the error** of the estimation by comparing the estimated with the known values.

Table 8.3 A comparison of methods of interpolation

Method	Deterministic/ stochastic	Local/ global	Transitions abrupt/ gradual	Exact interpolator	Limitations of the procedure	Best for	Output data structure	Computing load	Assumptions of interpolation model
'Eyeball'	Subjective/ deterministic	Global	Abrupt	No	Non-reproducible, subjective	Field data, aerial photo interpretation	Polygons	None	Intuitive understanding of spatial processes; homogeneity within boundaries
Edge-finding algorithm	Deterministic	Global	Abrupt	No	Often requires shapes to be defined and stored; better for man-made features than for natural landscapes	Raster images from remote sensors	Raster	Moderate	Homogeneity within boundaries
Proximal (Thiessen poly.)	Deterministic	Local	Abrupt	Yes	One data point per cell; no error estimates possible; tesselation pattern depends on data point distribution	Nominal data from point patterns	Polygons	Light/ moderate	'nearest neighbour' gives best information
Trend surface	Stochastic	Global	Gradual	No	Edge effects, outliers, complex polynomials do not necessarily have meaning; errors are rarely spatially independent.	Demonstrating broad features and removing them prior to other methods of interpolation	Points on a raster	Light/ moderate	Multiple regression—phenomenological explanation of trend surface; independent Gaussian errors
Fourier series	Stochastic	Global	Gradual	No	Not applicable to data; lacking periodicity	Periodic features such as sand dunes, ripple marks or gilgai, or man-made features	Points on a raster	Moderate	Strict periodicity in phenomenon of interest.
B-splines	Deterministic	Local	Gradual	Yes	No estimates of errors; masks all uncertainties in surface	Very smooth surfaces	Points on a raster	Light/ moderate	Absolute smoothness of variation
Moving average	Deterministic	Local	Gradual	No unless constrained	Results depend on configuration of data points and size of window; simple versions assume isotropy; no error estimates unless retrospectively calculated	Quick contour plots of moderately smooth data.	Points on a raster	Moderate	Continuous, differentiable surface is appropriate
Optimal interpolation (kriging)	Stochastic	Local	Gradual	Yes	Practical and theoretical problems of non-stationarity in data; large computing costs for mapping	Situations where the most detailed estimates and their errors are required	Points on a raster	Heavy (very heavy for universal kriging)	Intrinsic hypothesis (homogeneity of first differences); average local values can be represented by a continuous surface.


Thank you
For your kind cooperation
Patient listening & Learning GIS
Let us enjoy using GIS more efficiently & effectively…