# UNIT-III
# SPATIAL DATA BASE MANAGEMENT SYSTEM

# SPATIAL DATA INPUT AND EDITING

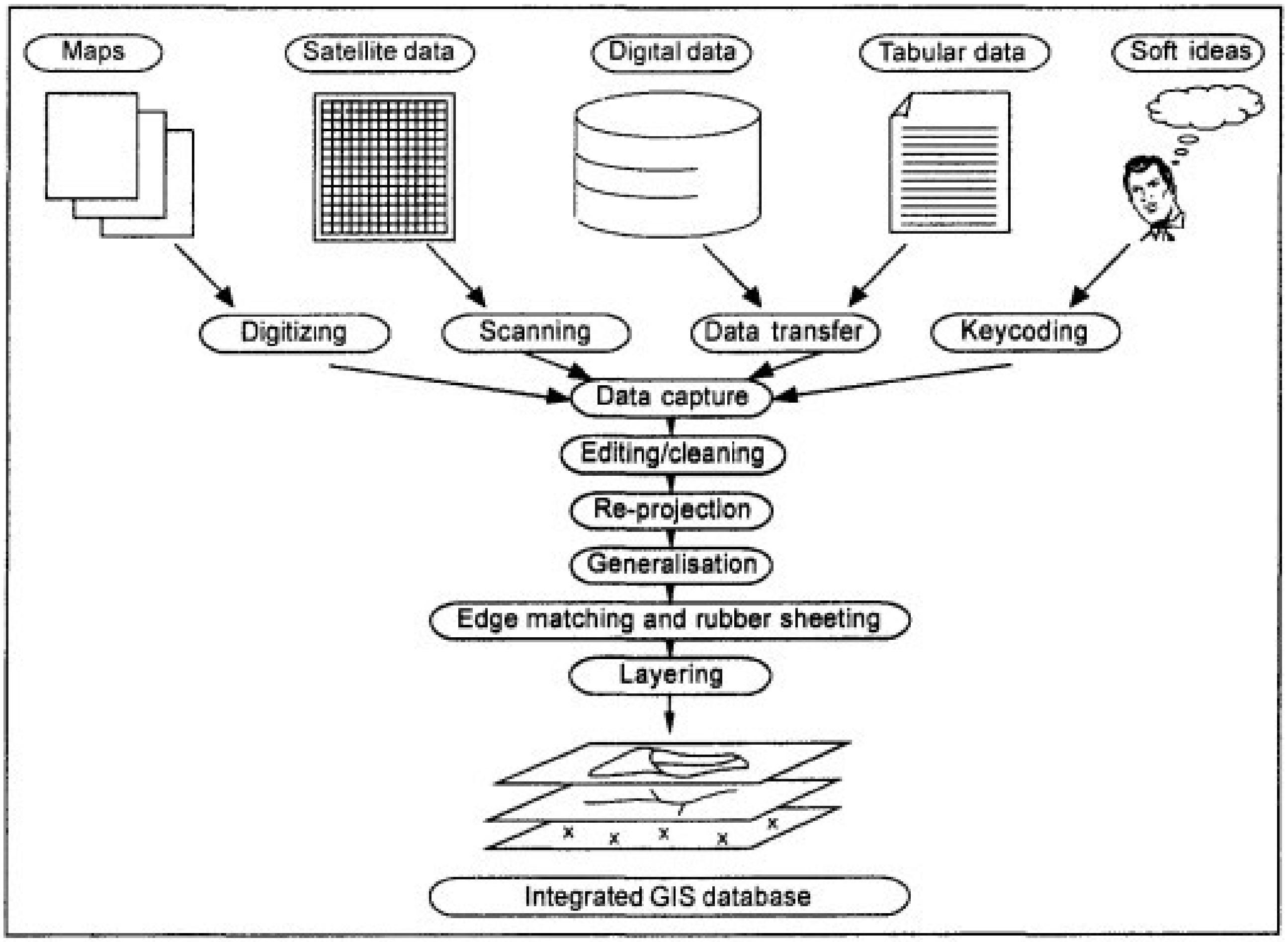





Fig. 10.1 The conceptual view of data stream in GIS.

**Table 10.1 Data types and their source in India**

| Data Type | Data Source |
|---|---|
| Topography | GSI, NATMO, SOI |
| Digital Elevation Model | |
| Digital Terrain Data | |
| Land Use and Land Cover | |
| Ownership and Political Boundaries | SOI, Land Records |
| Transportation | R & D, A.P.S.R.T.C. |
| Hydrography | Ministry of Water Resources<br>State & Central Ground Water Board |
| Socioeconomic and Demographic data | |
| Census Tract Boundaries | Department of Census |
| Demographic Data | Bureau of Economics & Statistics |
| Soils | NATMO |
| Wetlands | NATMO |
| Remotely Sensed Data | National Remote Sensing Agency |

GSI - Geological Survey of India

NATMO - National Atlas Thematic Mapping Organisation

SOI - Survey of India

APSRTC - Andhra Pradesh State Road Transportation Corporation

# DATA INPUTS METHODS

**There is distinction between analogue (non-digital) and digital sources of spatial data. Analogue data are normally in paper form and include paper maps, tables of statistics and hardcopy aerial photographs.**
**All these forms of data need to be converted to digital form before use in a GIS.**

**Digital data like remote sensing data are already in compute-readable formats and are supplied on diskette, magnetic tape or CD-ROM or across a computer network.**

**All data in analogue form need to be converted to digital form before they can be input into GIS.**

**There are four methods of data input which are widely used:**

**Keyboard entry, manual digitizing, automatic digitization, and scanning.**

Digital data must be downloaded from their source media and may require reformatting to convert them to an appropriate format for the GIS being used. **Reformatting or conversion may also be required after analogue data have been converted to digital form.**

**For example**, after scanning a paper map, the file produced by the scanning equipment may not be compatible with the GIS, so it needs reformatting. For both the analogue and digital data, keyboard entry method, manual digitizing and automatic digitizing and scanning methods are very important as detailed below.

## 1. KEYBOARD ENTRY

**Keyboard entry, often referred to as key coding**, is the entry data into a file at a computer terminal. **This technique is used for attribute data that are available only on paper**. This technique can be mixed with digitizing process for the creation of GIS database

The **coordinates of spatial entities like point, line and area features can be encoded by keyboard entry. This method is used when the coordinates of these spatial entities are known and there are not too many of them.** If the coordinates are more in number, this data can be encoded using digitizing. The **procedure of keyboard entry can be used to enter land record information.** This method leads to obtain very high level of precision data by entering the actual surveying measurements. This method is used for the development of cadastral information system.

## 2.Manual Digitizing

**Manual digitizing is the most common method of encoding spatial features from paper maps. It is a process of converting the spatial features on a map into a digital format. Point, line, and area features that form a map, are converted into (x, y) coordinates.**

A point is represented by a single coordinate, a line by a string of coordinates, and, when one or more lines are combined with a label point inside an outline, then an area (polygon) is identified. **Thus digitizing is the process of capturing a series of points and lines.** Points are used for two different purposes: to represent point features or to identify the presence of a polygon.

Manual digitizing requires a table digitizer that is linked to a computer work station. To achieve good results, the following steps are necessary.

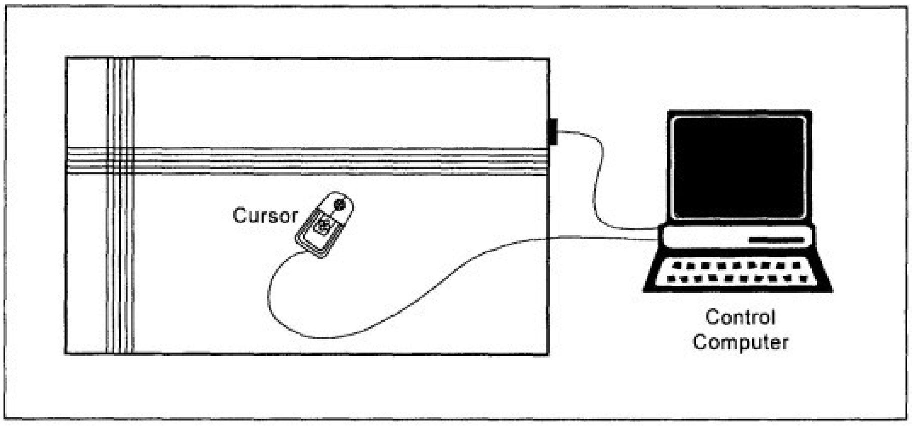



Fig. 10.2 Contemporary tablets using a grid of wires embedded in the tablet to generate a magnetic field which is detected by the cursor.

## The Digitizing Operation

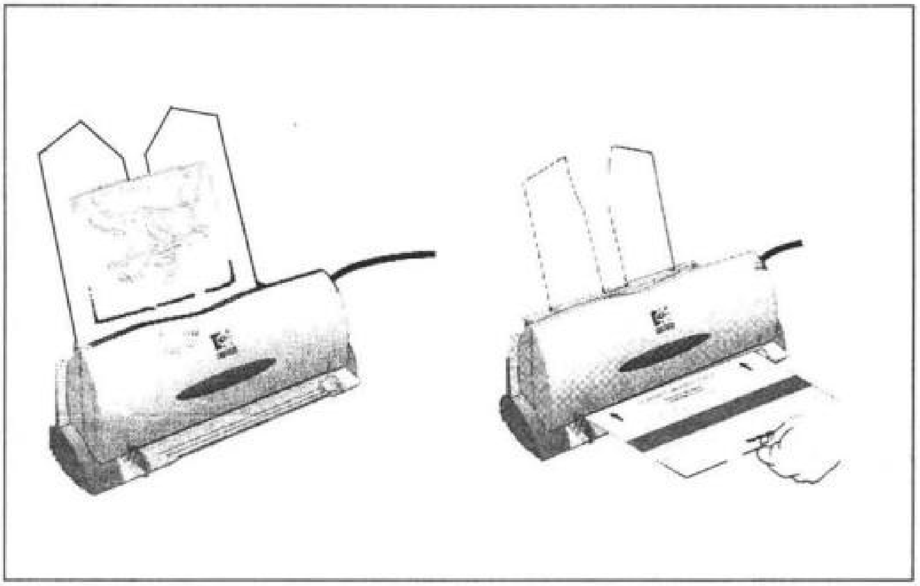
**The map is affixed to a digitizing table. Three or more control points are to be identified and digitized for each map sheet. These points should be those that can be easily identified like intersections of major streets and prominent landmarks.** The points are called reference points or tics or control points. The coordinates of these control points will be known in the coordinate system to be used in the final data, such as, latitude and longitude. The control points are used by the system to calculate the necessary mathematical transformations to convert all coordinates to the final system. The more the control points, the better the accuracy of digitization. Digitizing the map contents can be done in two different modes: point mode and stream mode. Point mode is the mode in which the operator identifies the points to be captured explicitly by pressing a button, and stream mode is the mode in which points are captured at set time intervals, typically 10 per second, or on movement of the cursor by filed distance. Most digitizing is currently done in point mode.

## 3. Scanning and Automatic Digitizing

Scanning is the most commonly used method of automatic digitizing. Scanning is an appropriate method of data encoding when raster data are required, since this is the automatic output format from most scanning software. Thus scanning may be used as a background raster dataset for the over-plotting of vector infrastructure data, such as, pipelines and cables. A scanner is a piece of hardware for converting an analogue source document to a digital raster format (Jackson Woodford, 1997).

There are two types of scanners,

(i) Flatbed scanner and
(ii) Rotating drum scanners. The cheapest scanners are small flatbed scanners, and high quality and large format scanners are rotating drum scanners in which the sensor moves along the axis of rotation.



A digital image of the map is produced by moving an electronic detector across the map surface. The size of the map area viewed by the detector and scanning should be processed or edited to improve the quality and convert the raster to vector after online digitization. The accuracy of the scanned output data depends on the quality of the scanner, the quality of the software used to process the scanned data, and the quality of the source document. A very important feature that a GIS user should observe after scanning the paper map is the occurrence of splines, which is black appearance on the scanned output. This can be removed by using a process called thinning. The resolution of the scanner used affects the quality and quantity of output data. The cheaper flat-bed scanners have resolutions of 200-500 mm whereas the more expensive drum scanners use resolutions of 10-50 mm. The higher the resolution, the larger the volume of the data produced. Fig. 10.7 shows scanned output of paper map and the output of automatic digitisation.

# DATA EDITING

**The storage and editing subsystem of GIS provides** a variety of tools for storing and maintaining the digital representation of study area. It also provides tools for examining each coverage for mistakes that may have crept into our preparations. **The input data that is encoding may consist of a number of errors derived from the original data source as well as errors that have been introduced during the encoding process**. There may be errors in coordinate data as well as in accuracies and uncertainty in attribute data. **Before successfully using the methods of data analysis for any specific application, it is better to intercept errors before they contaminate the GIS database. The process of detecting and removing errors through editing is called cleaning.**

**All the errors into 3 groups: (i) entity errors, (ii) attribute errors, and (iii) entity-attribute errors.** They occur during the execution of the project using vector based GIS and attribute and entity attribute agreement errors occur in the use of both raster and vector based GIS.

## Data editing and cleaning of GIS database are covered under three subheads:

*(a)Detecting and correcting errors*

*(b) Data reduction and generalization\and*

*(c)edge matching and rubber sheeting.*

## (A) Detecting and correcting errors

In any information system, **facilities must be provided to detect and correct errors in the database.** Different kinds of errors are common in different data sources. To illustrate some of the common varieties, will discuss **some of the errors that must be detected when generating a vector dataset, whether developed from a manual digitization or an automated digitization.**

**Polygonal areas, by definition, have closed boundaries**. **If a graphic object has been encoded as a polygon (rather than a vector or point), the boundary must be continuous.** **Software should be able to detect that polygon which is not closed.** The causes for this kind of error include encoding error (the object is a vector, rather than a polygon) and digitizing error (either points along the boundary of the polygon, or connections between the points, are missing).

**Table 10.2 Common errors in GIS database**

| Error | Description |
|---|---|
| Missing entities | missing points, lines or boundary segments |
| Duplicate entities | points, lines or boundary segments that have been digitised twice |
| Mislocated entities | points, lines or boundary segments digitised in wrong place. |
| Mislocated labels | unidentified polygons |
| Duplicate labels | two or more identification labels for the same polygon. |
| Artifacts of digitising | undershoots, overshoots, wrongly placed nodes, loops and spikes |
| Noise | Irrelevant data entered during digitising, scanning or data transfer. |

Pseudo nodes occur where a single line connects itself (an island) or where only two arcs intersect. Pseudo nodes do not indicate an error. **There can be pseudo nodes representing an island or an intermediate point having an attribute data attached to it.** A dangling node refers to the unconnected node of a dangling segment. **Every segment begins and ends at a node point. So if a segment does not close properly (undershoot), or was digitized past an intersection (overshoot), it will register a dangling node.** There can be instances where dangling nodes are representing some real-world feature. It is always better to have overshoots. It is much easier than editing an undershoot. Fig 10.8 shows some of the examples of spatial error in vector data.

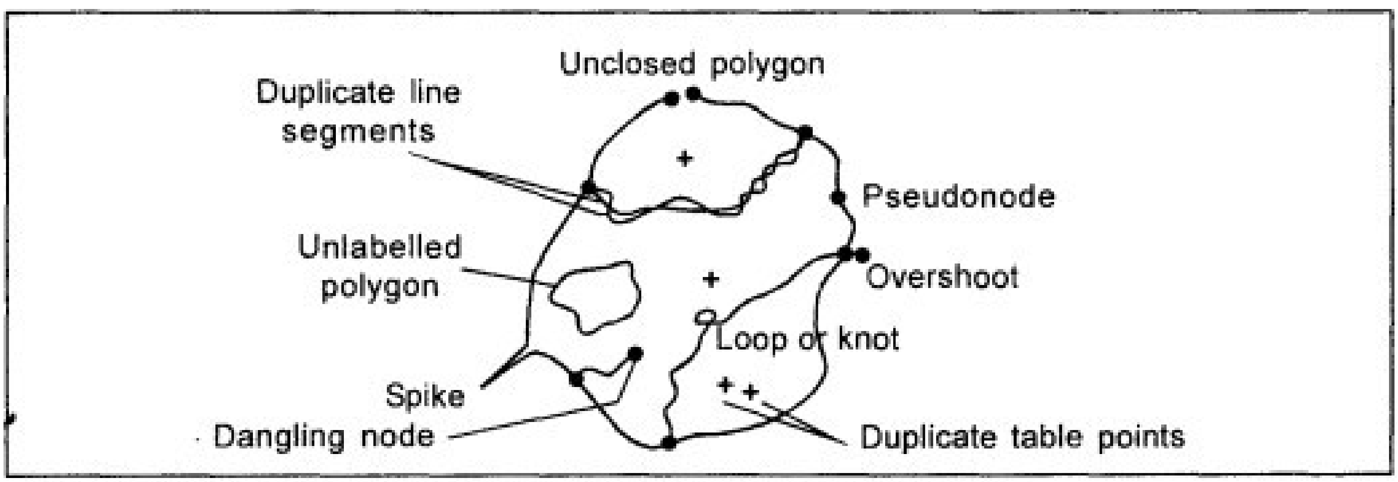
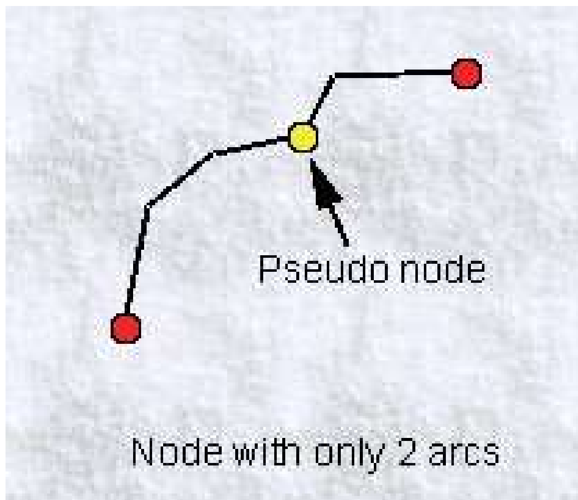



Fig. 10.8 Examples of commonly committed errors in vector data.

**Pseudo nodes occur:** Where a single line connects itself or when only two arcs intersects

**For example a long road segment may have been divided into two features**

**Dangling Node:** A dangling node refers to the unconnected node of dangling segment



## (B) Data Reduction and Generalisation

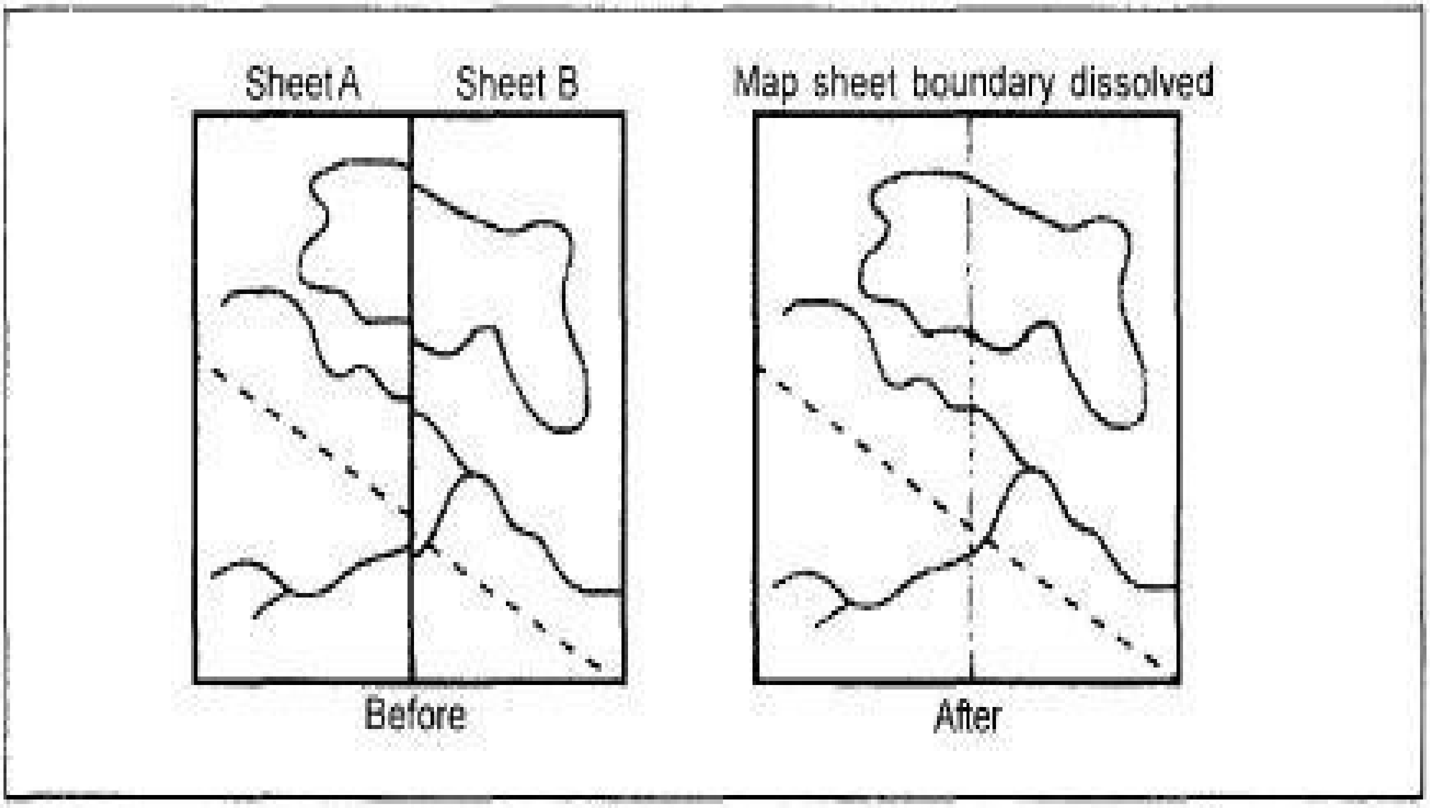
•In a given application, we may need to average the measurements of a single tree, and enter the average into our database. This is ***Data Reduction***

•We may need to assemble property ownership records in an area, and the **original surveyor's records may be more detailed than we require**. There are two obvious options: either accept the level of detail (and thus, incorporate a greater volume of data than necessary, with the attendant increased processing and storage costs), or develop a less precise representation from the original source data. The latter is called ***Generalisation***.

## ( C )Edge Matching and Rubber Sheeting

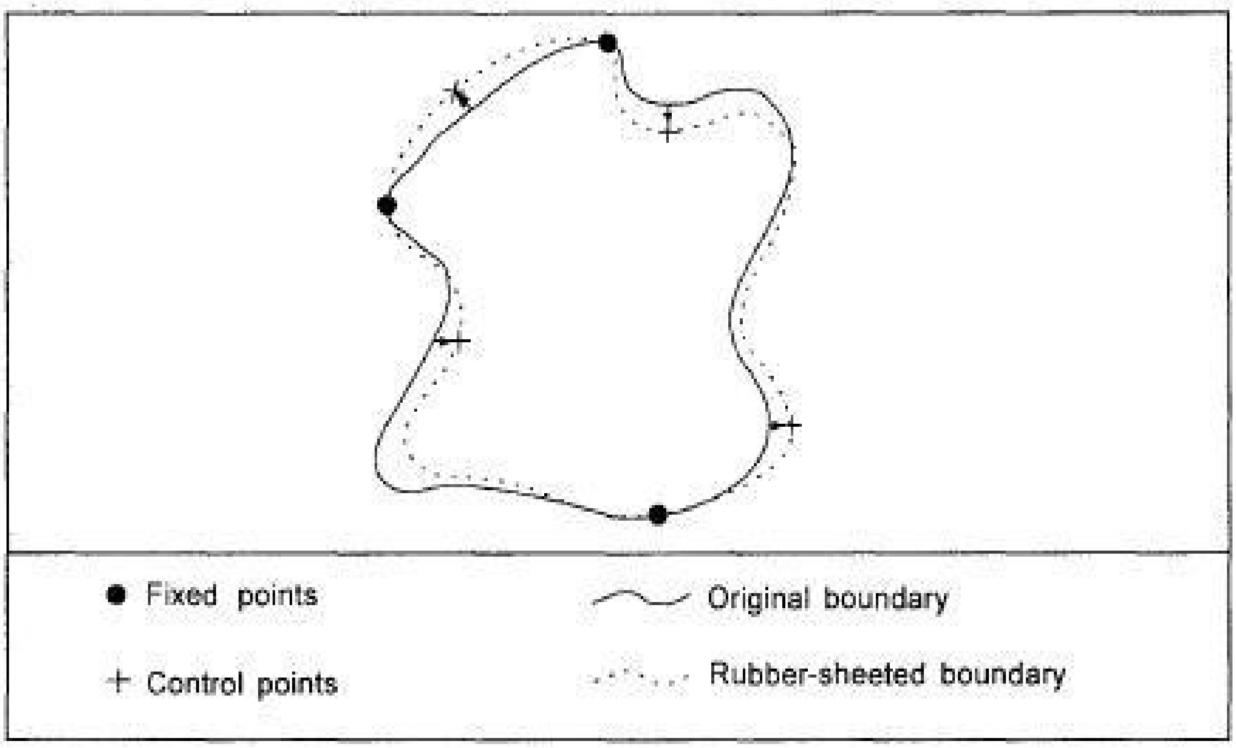
**Sometimes, any given project area under study, extends across two or more map sheets. small differences or mismatches between adjacent map sheets may need to be resolved.** Normally, each map sheet would be digitised separately and then adjacent **sheets joined after editing, projection, transformation, generalisation, and scaling. The process of joining is known as edge matching.**

Many vector GIS systems also allow you to separately stores potions of database as large, predefined subsections for archival purposes. This process is called tiling. This tiling process is commonly used to reduce the volume of the data needed for the analysis of extremely large database. Edge matching is a process of operating on more than one tile at a time to ensure that there is a correct match between the two tiles for entities for entries that extend across the tile boundaries.


Sheet A
Sheet B
Map sheet boundary dissolved
Before
After

## Rubber Sheeting

- **Certain data sources may give rise to internal distortion within individual map sheets. This is especially true of data derived from aerial photographs as the movement of the aircraft and the distortion caused by the camera lens can cause internal inaccuracies in the location of features within the image.**
- **These inaccuracies remain even after transformation and re projection.** These problems can be rectified through a process known as rubber sheeting or conflation. **Rubber sheeting involves stretching the map in various directions as if it were drawn on a rubber sheet.**
- Objects on the map that are accurately placed are tacked down and kept still while others that are in the wrong location or have wrong shape are stretched to fit with the control points, which are known points identified on the ground and on the image. The coordinates of these control points may be determined from field observation using GPS.


Fixed points
Control points
Original boundary
Rubber-sheeted boundary

# DATA ACCURACY

A useful starting point for discussing accuracy is the **0, which serves as the conceptual basis for most database implementations of real world phenomena**. **According to this model, 'entities' represent real-world phenomena,**

such as, streets, districts, and hazardous waste sites, and 'attributes' specify the relevant properties of these objects, such as, width and number of lanes, while 'values' give the specific qualitative and quantitative measurements pertaining to a particular attribute

**In this model error is defined as the discrepancy** between the encoded and actual value of a particular attribute for a given entity. **Accuracy is the inverse of error**. **This model can be used to define spatial, temporal, and thematic error for a particular entity as, the discrepancies in the encoded spatial, temporal, and thematic attribute values**

Accuracy is a **relative measure** rather than an absolute one, since it depends upon the intended form and content of the database. Different specifications can exist for the same general types of geospatial data. To judge the fitness-for-use of the data for some applications, one must not only judge the data relative to the specification, but also consider the limitations of the specification itself.

# Error and Corresponding action to remove that error

| Error | Action |
| --- | --- |
| Missing Segment | Draw it |
| A gap between two Segments | Indicate which arc to extend or which node to move |
| An overshoot | Delete node if needed |
| An Undershoot | Merge nodes or extend segment |

**Accuracy of the GIS data can be discussed in terms of**

1. **spatial accuracy,**
2. **temporal accuracy,**
3. **thematic attribute data accuracy, and**
4. **conceptual accuracy.**

## 1. Spatial Accuracy

**Spatial accuracy also called 'positional accuracy' refers to the accuracy of the spatial component of a database. Measurements of spatial accuracy depends upon the dimensionality.** This applies to both horizontal and vertical positions. Accuracy and precision are the functions of the scale. This means that when we see a point on a map we have its **"probable" location within a certain area**. This is a very great danger in computer systems that allow users to pan and zoom to an infinite number of scales. **Accuracy and precision are tied to the original map scale and do not change even if the user zooms in and out.**

**For points, error is usually defined as the discrepancy between the encoded location and the location as defined in specification.** Error can be measured in anyone of, or in combinations of, the three dimensions of space. The most common measures are horizontal error and vertical error

**Clearly this is not the case since a database can achieve a high level of temporal accuracy without being current and is dependant upon the availability of historical data.**

# 2. Temporal Accuracy

**Temporal accuracy and correctness are two distinct concepts . Temporal accuracy refers to the agreement between 'encoded' and 'actual' temporal coordinates.**

**Correctness is an application-specific measure of temporal accuracy.**

**A value is current if it is correct in spite of any possible time-related changes in value. Thus correctness refers to the degree to which a database is up-to-date** (Redman 1992). To equate temporal accuracy with correctness is to state, in effect, that to be temporally accurate a database must be up-to-date.

# 3. Attribute Accuracy

**Attribute accuracy indicates the attribute attached to the points, lines and polygons features** of the spatial database, which are how reliable and reasonably correct or free from bias.

- For categorical data, most of the research into quality has come from the field of classification accuracy assessment in remote sensing.

- Accuracy assessment is based on the selection of a sample of point locations and a comparison of the land cover classes assigned to these locations by the classification procedure with the classes observed at these locations on a reference source usually called '**ground truth**'

•The non-spatial data linked to location may also be inaccurate or imprecise.

• Non-spatial data can also vary greatly in precision. Precise attribute information describes phenomena in great detail.
For example, a precise description of a person living at a particular address might include gender, age, income, occupation, level of education, and many other characteristics.

• An imprecise description might include just income, or just gender

# 4.Conceptual Accuracy

•**GIS depends upon the abstraction and classification of real-world phenomena.**

•**The users determine what amount of information is used and how it is classified into appropriate categories. Sometimes users may use inappropriate categories or misclassify information.**

**For example, classifying cities by population size would probably be an ineffective way to study land patterns**

- Even if the correct categories are employed, data may be misclassified
- **A study of drainage systems may involve classifying streams and rivers by "order," that is where a particular drainage channel fits within the overall tributary network. Individual channels may be misclassified if tributaries are miscounted**

# SOURCES OF ERROR IN GIS

- Spatial and attribute errors can occur at any stage in a GIS project.
- **These errors may arise during the derivation of spatial entities.**
- ***Encoding these entities in the computer system, forms the use of data in analysis.***
- **Errors may also be present in source data, arising during data conversion arising** during data manipulations and can be produced

- There are many sources of error that may affect the quality of a GIS dataset.
- **Few of these will be automatically identified by the GIS itself.** It is the user's responsibility to prevent them.

Burrough divided source of error into three main categories:

(a) Obvious sources of error,

(b) Errors resulting from natural variations or from original measurements,

(c) errors arising through processing.

- during the presentation of results.

| | |
|---|---|
| (i) | Obvious sources of error<br>Age of data<br>Areal coverage<br>Map scale<br>Density of observations<br>Relevance<br>Format<br>Accessibility |
| (ii) | Error resulting from natural variations or from original measurements<br>Positional accuracy<br>Accuracy of content<br>Qualitative and Quantitative<br>Variation in data<br>Natural variation<br>Data entry/output faults |
| (iii) | Error arising through processing<br>Numerical error in computer<br>Faults due to topological analyses<br>Misuse of Logic<br>Problems associated with map overlay<br>Classification and generalisation problems<br>Interpolation |

# DATA ANALYSIS

Pre-processing procedures are used to convert a dataset into a form suitable for permanent storage within the GIS database for application development. Often, **a large proportion of the data entered into a GIS requires some kind of processing and manipulation** in order to make it conform to a data type, georeferencing system, and data structure that is compatible with the system

## Format Conversion

Format conversion covers a number of different problems, but can be discussed in terms of two families:

**conversion between different digital data structures problem of modifying one data structure into another**

2. **conversion between different data media**. **Involves converting source material, such as, paper maps, photographic prints, and printed tables into a computer based digital data either by means of manual digitisation or by means of scanning with automated digitisation**

# Data Medium Conversion

**Most of the spatial data available for development of various applications are not in computer-compatible formats.**

**These include maps on different scales, printed manuscripts, and imagery.**

Converting these materials into a format compatible with a digital geographical information system can be very expensive and time-consuming.

**The most common methods of converting maps and other graphic data to digital format is to use a technique called digitisation**

## 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. Calculating length, perimeters, and areas is a common application of GIS.**

**For example, measuring length of tank bund, perimeter of HussainSagar lake and its area are straightforward methods using in GIS.**

# SPATIAL ANALYSIS

Spatial analysis is **a process in which you model problems geographically, derive results by computer processing, and then explore and examine those results**. Several fundamental spatial analysis workflows form the heart of spatial analysis: spatial data exploration, modeling with GIS tools, and spatial problem solving.

# SPATIAL ANALYSIS

A distinction is made in this course between GIS and spatial analysis. In the context of mainstream GIS software, **the term *analysis* refers to data manipulation and data querying.** In the context of **spatial analysis**, the *analysis* focuses on the *statistical analysis* of patterns and underlying processes or more generally, spatial analysis addresses the question "what could have been the genesis of the observed spatial pattern?" It's an **exploratory process** whereby we attempt to quantify the observed pattern then explore the processes that may have generated the pattern.

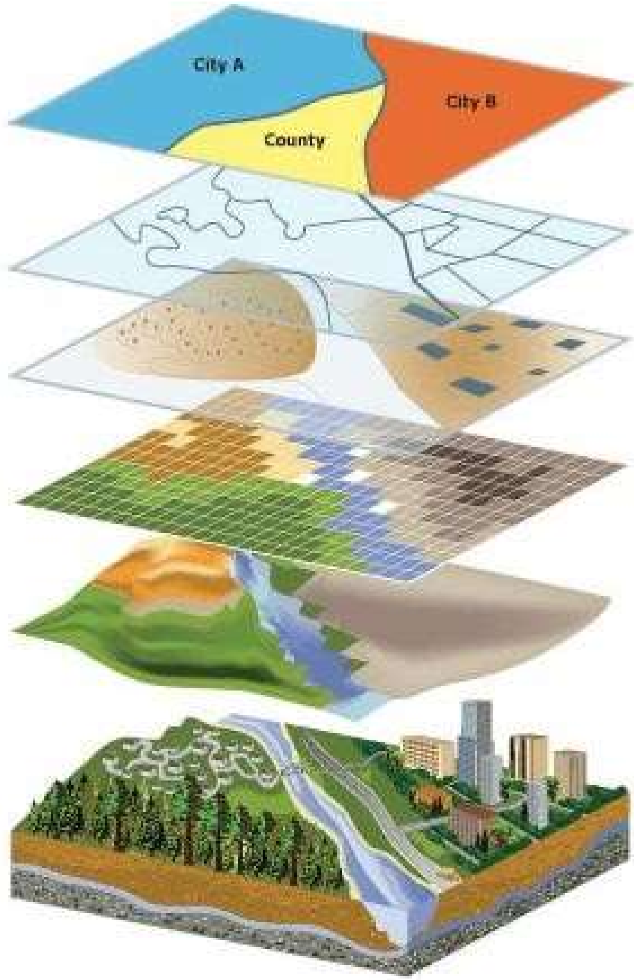
For example, you record the location of each tree in a well defined study area. You then map the location of each tree (a GIS task). At this point, you might be inclined to make inferences about the observed pattern. Are the trees clustered or dispersed? Is the tree density constant across the study area? Could soil type or slope have led to the observed pattern? Those are questions that are addressed in spatial analysis using quantitative and statistical techniques.

# What is Spatial Analysis?

The most important function of GIS is to enable the analysis of the spatial data and their attributes for decision support.

Spatial analysis is categorized as follows.

1. **Query:** retrieval of attribute data without altering the existing data by means of arithmetic and logical operations.
2. **Reclassification:** reclassification of attribute data by dissolving a part of the boundaries and merging into new reclassified polygons.
3. **Coverage Rebuilding:** rebuilding of the spatial data and the topology by "update", "erase", "clip", "split", "join" or "append".
4. **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.
5. **Connectivity Analysis:** analysis of connectivity between points, lines and polygon in terms of distance, area, travel time, optimum paths etc.

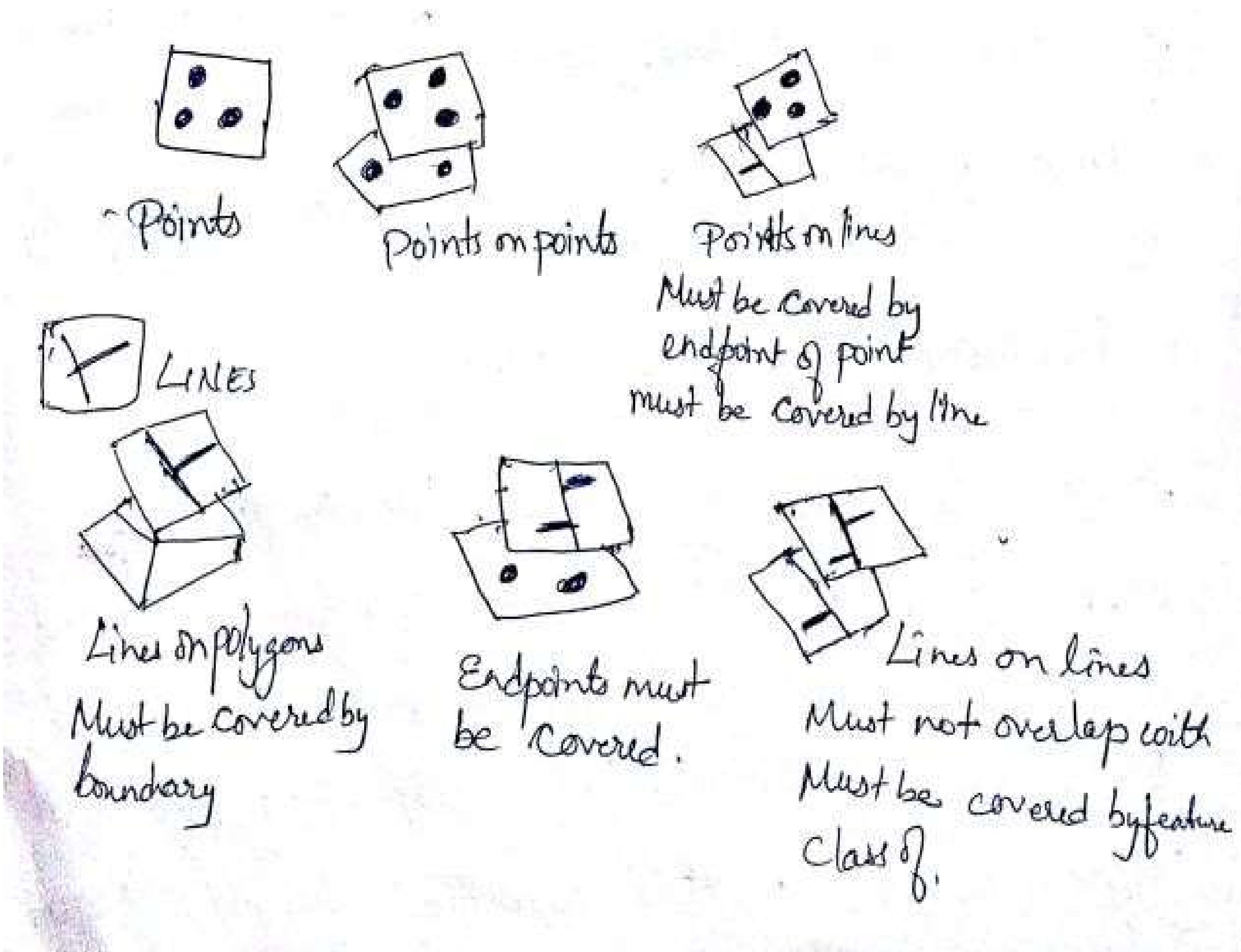

City A
City B
County

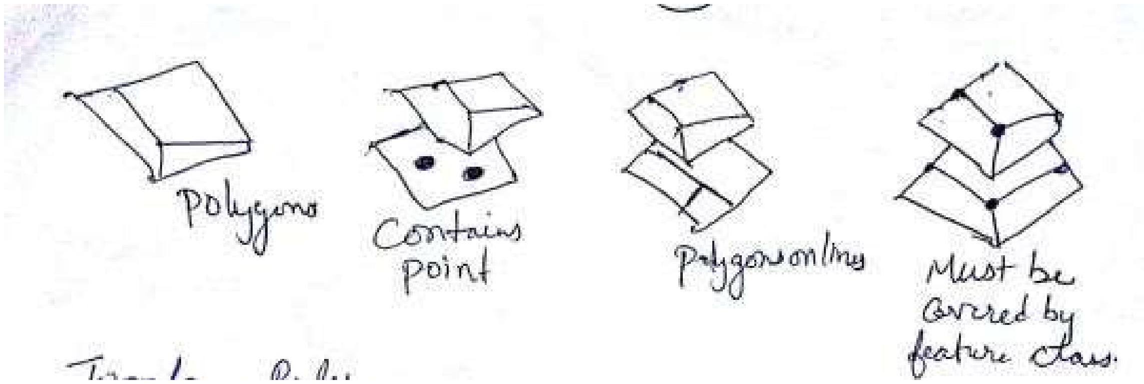
# TOPOLOGY

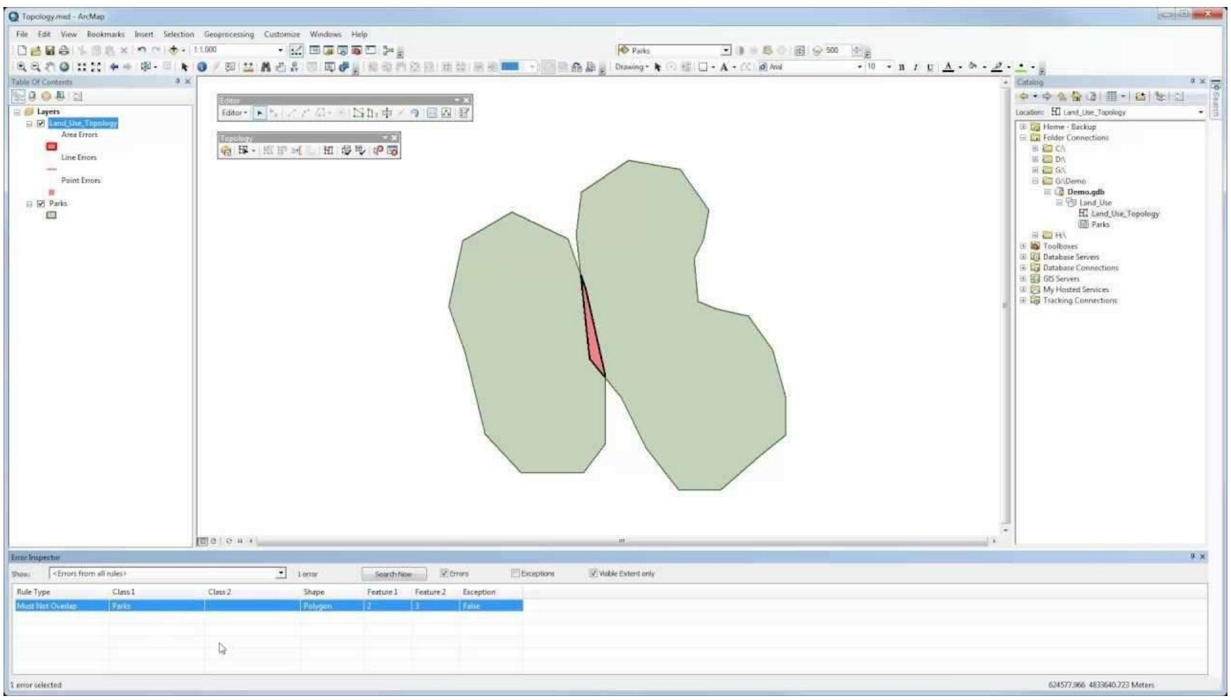
In **GIS Topology is a collection of rules that, coupled with a set of editing tools and techniques to enable the geospatial data** to more accurately mode geometric relationships topology is defined through a set of rules that **define how features may share geometry in an integrated fashion**. **A topology is commonly stored in geodatabase as one or more relationships that defined how the features in one or more feature classes share geometry.**

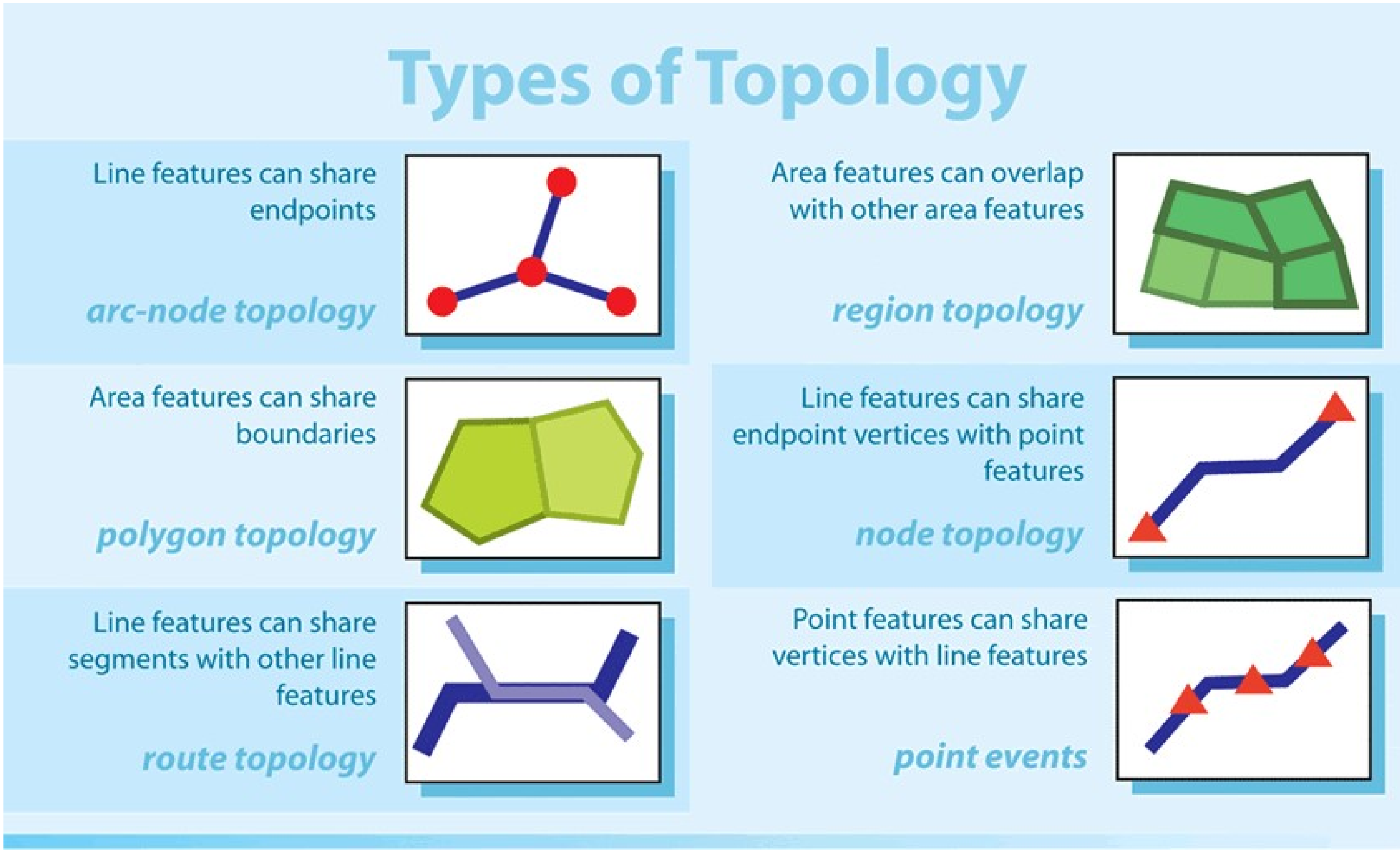
The feature participating in a topology are still simple feature class.

Rather than modifying the definition of the feature class, **a topology serves as a description of how the features can be spatially related**


Points
Points on points
Points on lines
Must be covered by
endpoint of point
must be covered by line
LINES
Lines on polygons
Must be covered by
boundary
Endpoints must
be covered.
Lines on lines
Must not overlap with
Must be covered by feature
Class of.


Polygons
Contains point
Must be Covered by feature class.


Topology.mxd - ArcMap
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
Table Of Contents
Layers
Land_Use_Topology
Area Errors
Line Errors
Point Errors
Parks
Catalog
Error Inspector
Rule Type
Class 1
Class 2
Shape
Feature 1
Feature 2
Exception
Must Not Overlap
Parks
Polygon
2
3
False


Types of Topology
Line features can share endpoints
arc-node topology
Area features can overlap with other area features
region topology
Area features can share boundaries
polygon topology
Line features can share endpoint vertices with point features
node topology
Line features can share segments with other line features
route topology
Point features can share vertices with line features
point events

# Topology Rules:

**Topology in "ArcGIS" allows you to model spatial relationship between feature classes in a feature dataset.**

**Topology rules allow you to define those relationships between two feature classes or Subtypes** . Topology rules allow you to define those relationships between features class or subtype or between two feature classes or subtypes. Topology rules allow you to define the relationship between two feature classes or subtypes.

**Topolology errors are violations of rules that you can easily find and manage using the editing tools found in Arc Map.**

# Vector Data Analysis

Vector data analysis uses the geometric objects of point, line and polygon.

The accuracy of analysis results depends on accuracy of these objects in terms of location and Shape.

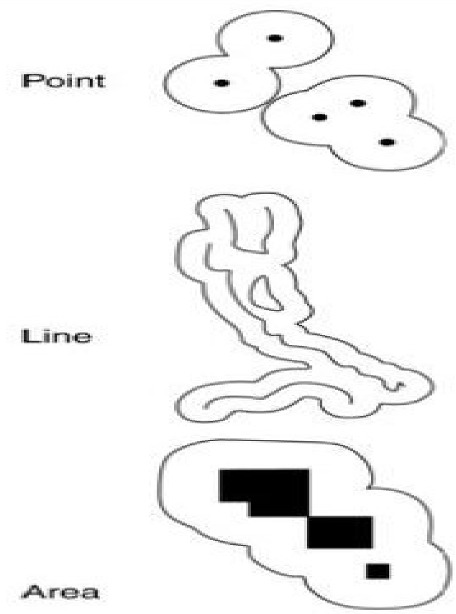
The accuracy of analysis results depends on the accuracy of these objects in terms of location and shape.
Topology can also be factor for some vector data analysis such as buffering and overlay

# Buffering

Based on the concept of proximity, buffering creates two areas: one area that is within a Specified distance of select features and the other area that is beyond.

The area that is within the specified distance is called the buffer zone.

The buffer distance can vary according to values of given field. Buffering around the line feature.



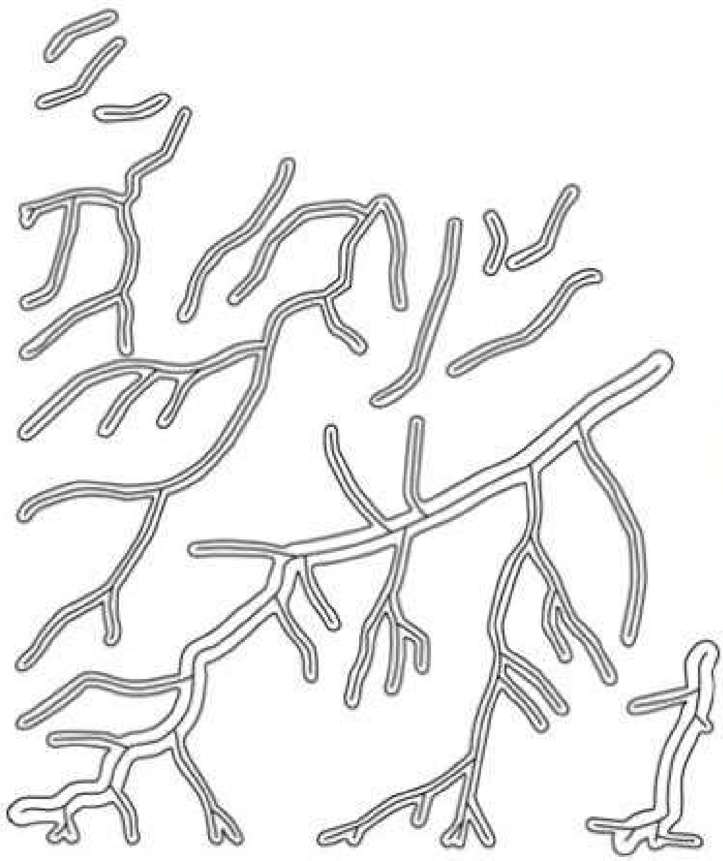


**Figure 11.2**
Buffering with different buffer distances.

# Overlay

- An overlay operation combines the geometries and attributes of two feature layers to create the output.
- The geometry of the output represents the geometric intersection of features from the input layers.
- Each feature on the output contains a combination of attributes from the input layers, and this combination differs from its neighbors.

# Feature Type and Overlay

Overlay operations can be classified by feature type into point-in-polygon, line-in-polygon, and polygon-on-polygon.

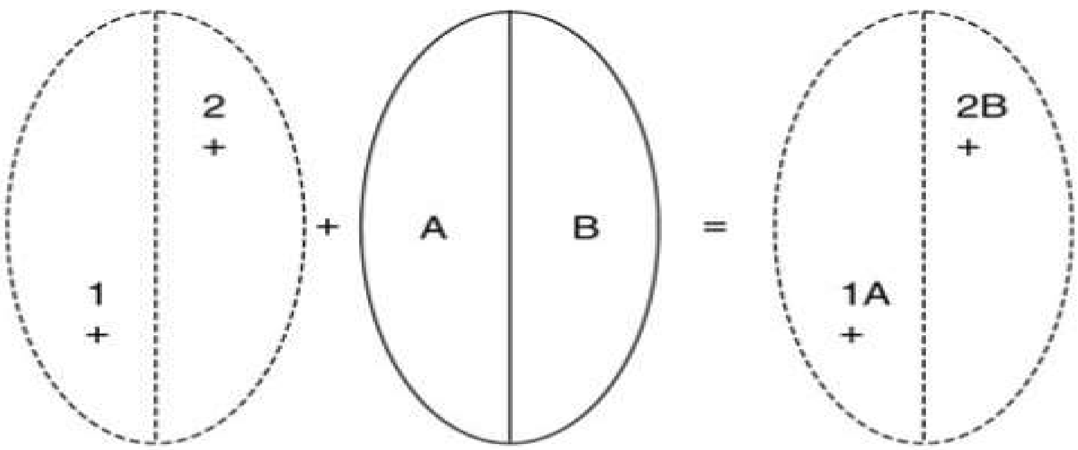



**Figure 11.6**
Point-in-polygon overlay. The input is a point layer (the dashed lines are for illustration only and are not part of the point layer). The output is also a point layer but has attribute data from the polygon layer.

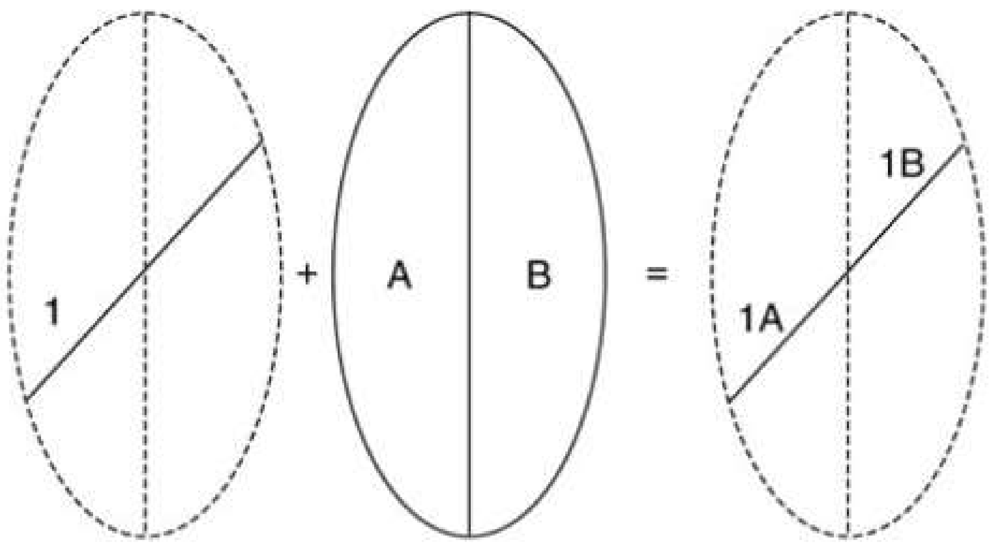



**Figure 11.7**
Line-in-polygon overlay. The input is a line layer (the dashed lines are for illustration only and are not part of the line layer). The output is also a line layer. But the output differs from the input in two aspects: the line is broken into two segments, and the line segments have attribute data from the polygon layer.

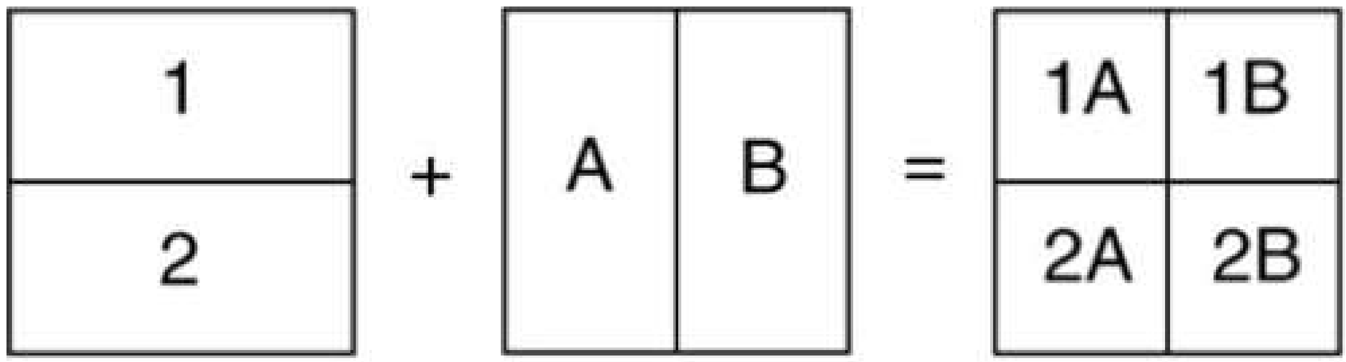



**Figure 11.8**
Polygon-on-polygon overlay. In the illustration, the two layers for overlay have the same area extent. The output combines the geometry and attribute data from the two layers into a single polygon layer.

# Slivers

- A common error from overlaying polygon layers is slivers, very small polygons along correlated or shared boundary lines of the input layers.
- To remove slivers, ArcGIS uses the cluster tolerance, which forces points and lines to be snapped together if they fall within the specified distance.

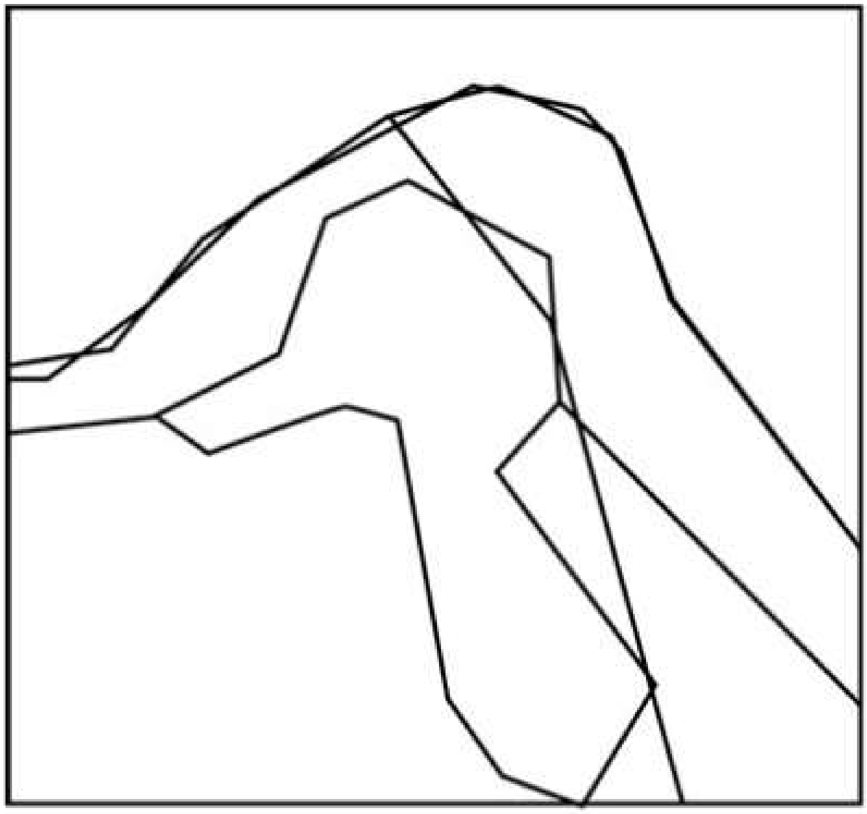


**Figure 11.13**
The top boundary has a series of slivers. These slivers are formed between the coastlines from the input layers in overlay.

# Feature Manipulation

- Tools are available in a GIS package for manipulating and managing maps in a database.
- These tools include Dissolve, Clip, Append, Select, Eliminate, Update, Erase, and Split.

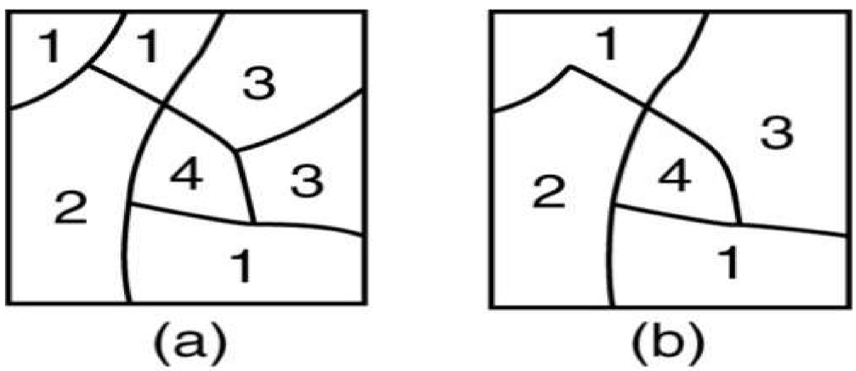



**Figure 11.22**
Dissolve removes boundaries of polygons that have the same attribute value in (a) and creates a simplified layer (b).

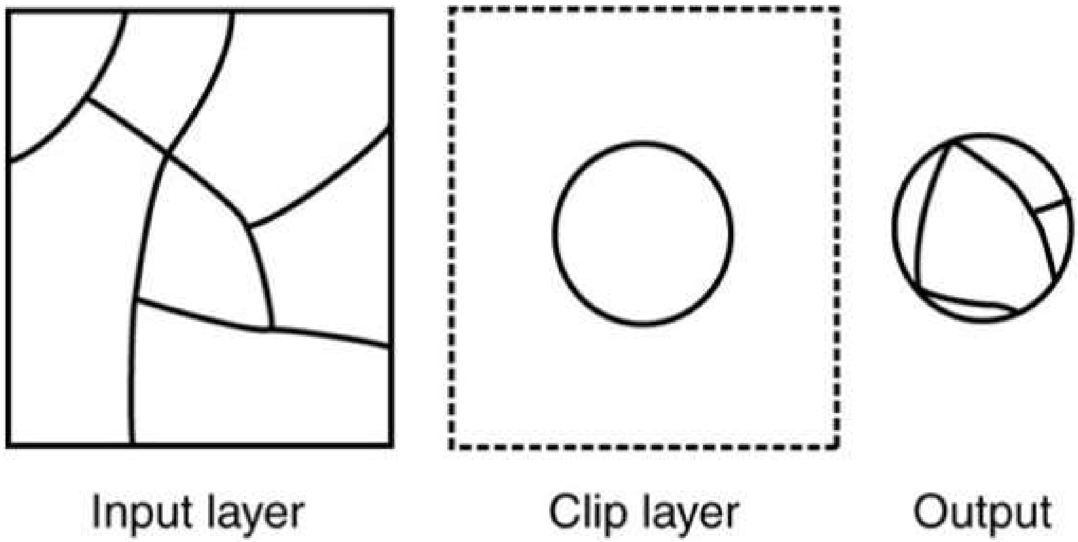



**Figure 11.23**
Clip creates an output that contains only those features of the input layer that fall within the area extent of the clip layer. (The dashed lines are for illustration only; they are not part of the clip layer.)

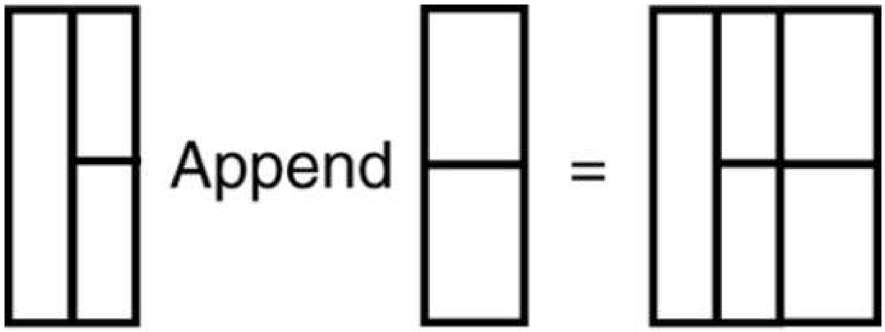



**Figure 11.24**
Append pieces together two adjacent layers into a single layer but does not remove the shared boundary between the layers.

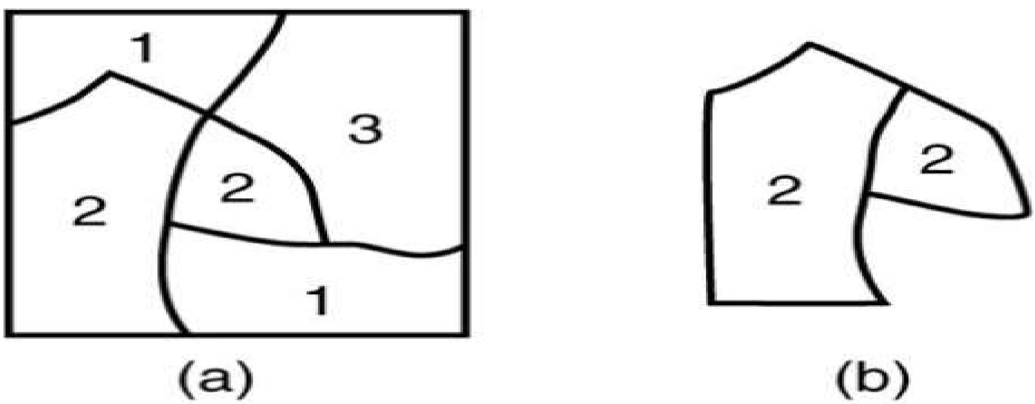



**Figure 11.25**
Select creates a new layer (b) with selected features from the input layer (a).

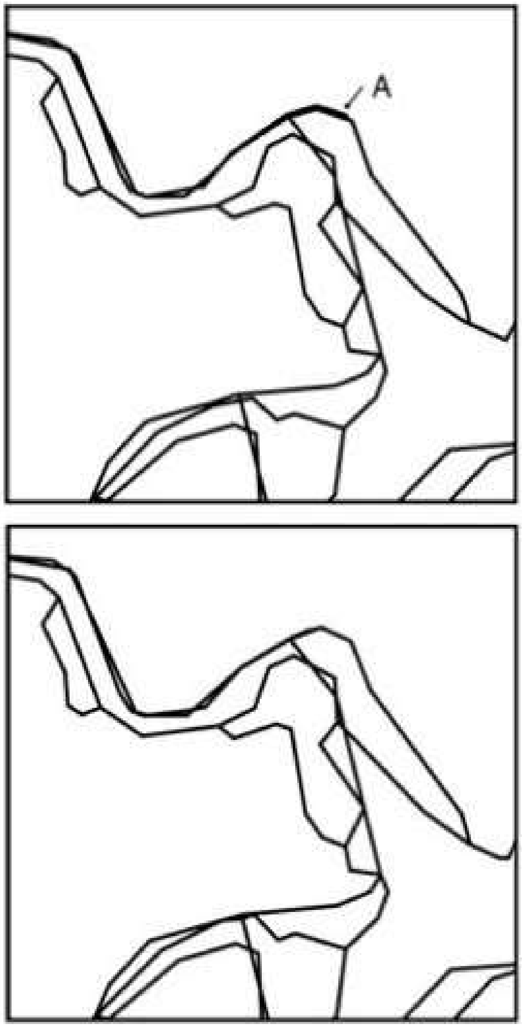



**Figure 11.26**
Eliminate removes some small slivers along the top boundary (A).

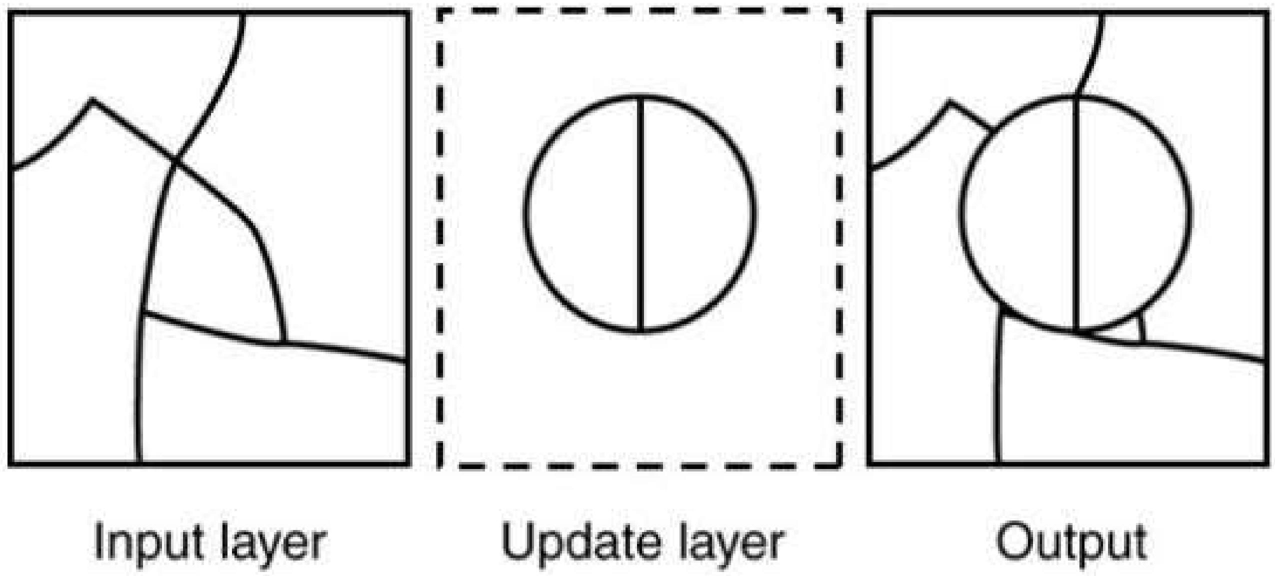



**Figure 11.27**
Update replaces the input layer with the update layer and its features. (The dashed lines are for illustration only; they are not part of the update layer.)

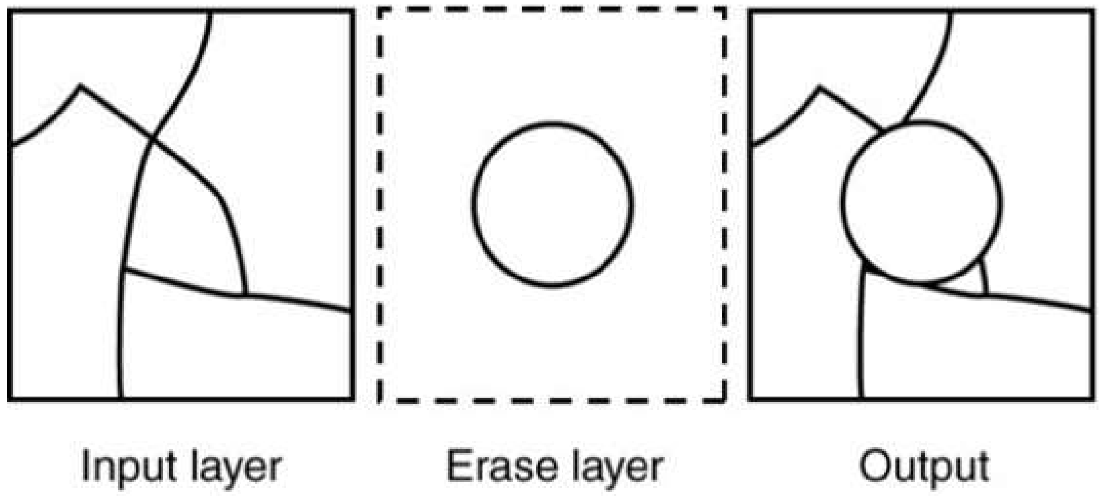



**Figure 11.28**
Erase removes features from the input layer that fall within the area extent of the erase layer. (The dashed lines are for illustration only; they are not part of the erase layer.)

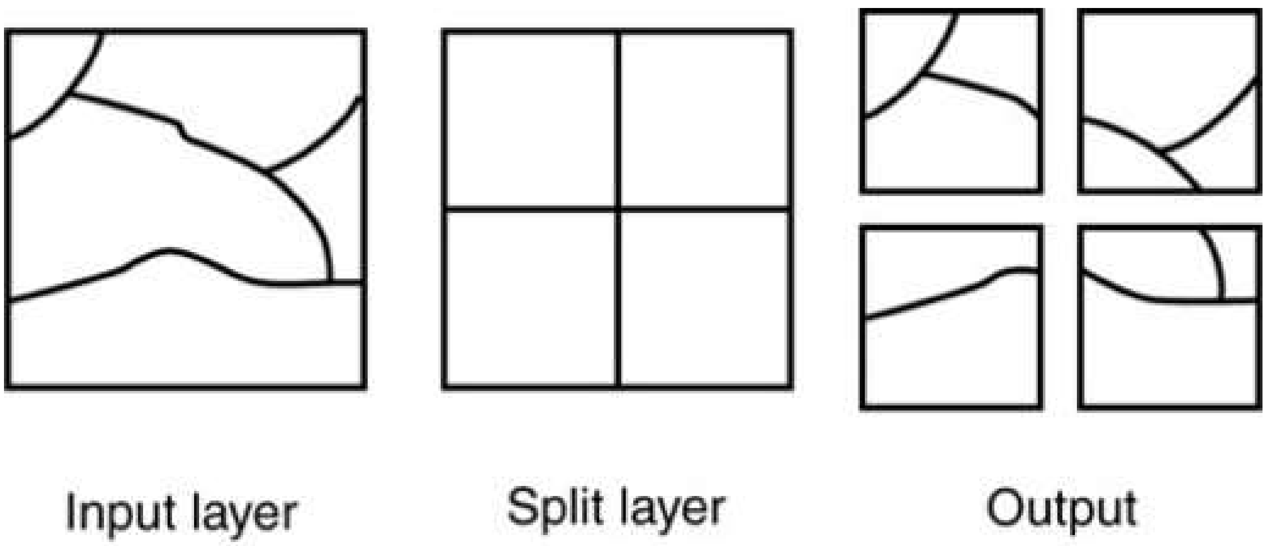



**Figure 11.29**
Split uses the geometry of the split layer to divide the input layer into four separate layers.

# What is Network Analysis?

Almost everyone has needed a **type of network analysis** in their life. **For example:**

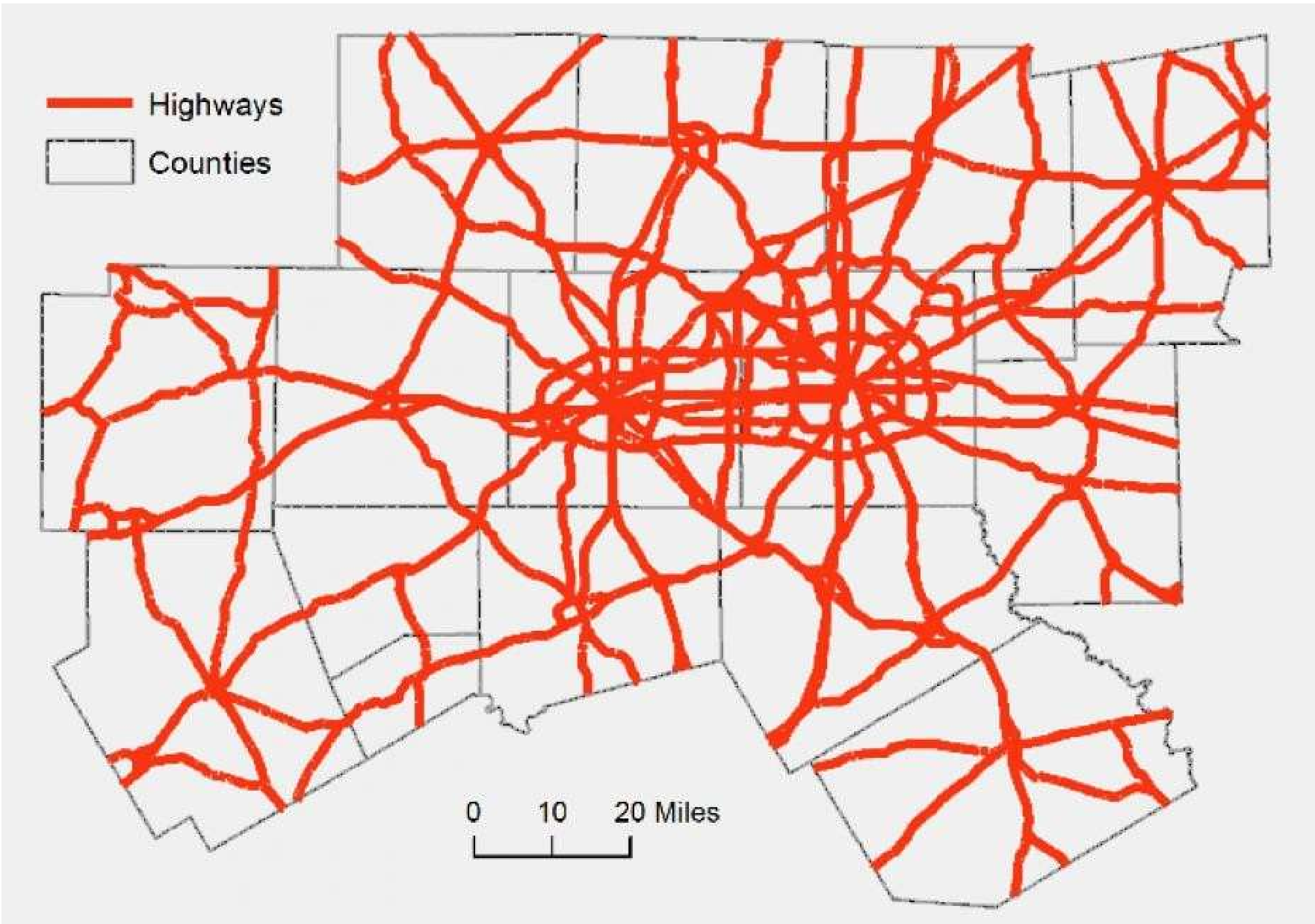
What's the shortest route to the beach?

Where should I build a hospital to best serve a community?

How can I optimize a vehicle delivery fleet?

Here are the 5 most common types of network analysis:

Network analysis is an operation in GIS which **analyses the datasets of geographic network or real world network**. Network analysis examine the properties of natural and man-made network in order to understand the behavior of flows within and around such networks and locational analysis.

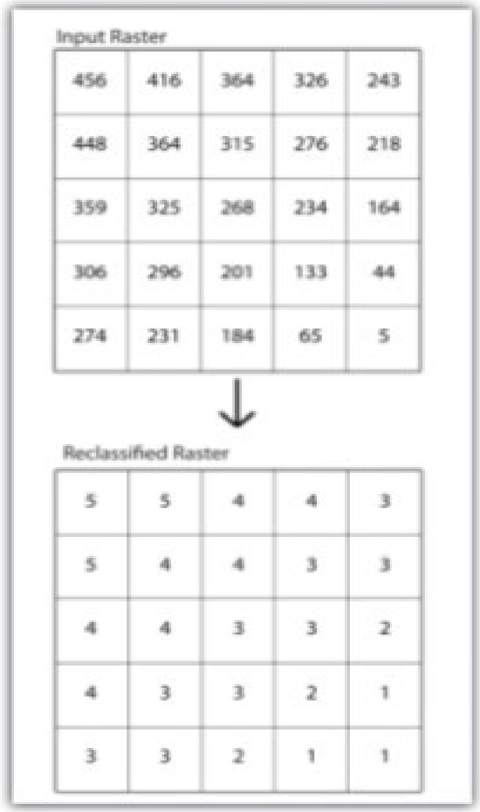

Highways
Counties
0
10
20 Miles

# RASTER DATA ANALYSIS

Raster data analysis is **based on cells and rasters**. Raster data analysis can be performed at the level of individual cells, or groups of cells, or cells within an entire raster. Some raster data operations use a single raster; others use two or more rasters.

# Single Layer Analysis

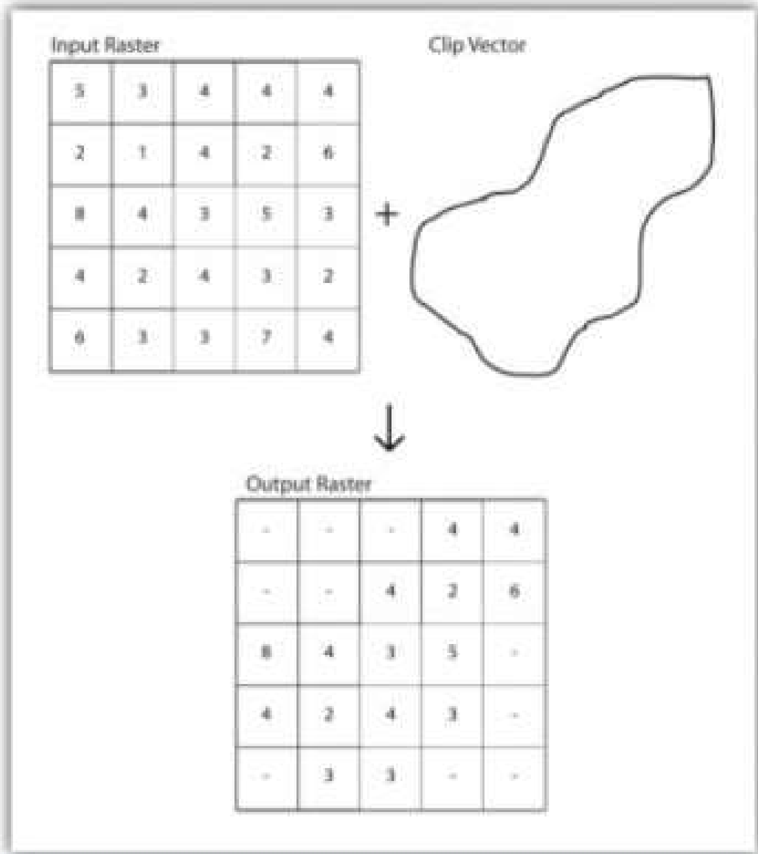
Reclassifying, or recoding, a dataset is commonly one of the first steps **undertaken during raster analysis.** Reclassification is basically the **single layer process of assigning a new class or range value to all pixels** in the dataset based on their original values



# Multiple Layer Analysis

A **raster dataset can also be clipped similar to a vector Here**, the input raster is overlain by a vector polygon clip layer. The raster clip process results in a single raster that is identical to the input raster but **shares the extent of the polygon clip layer.**

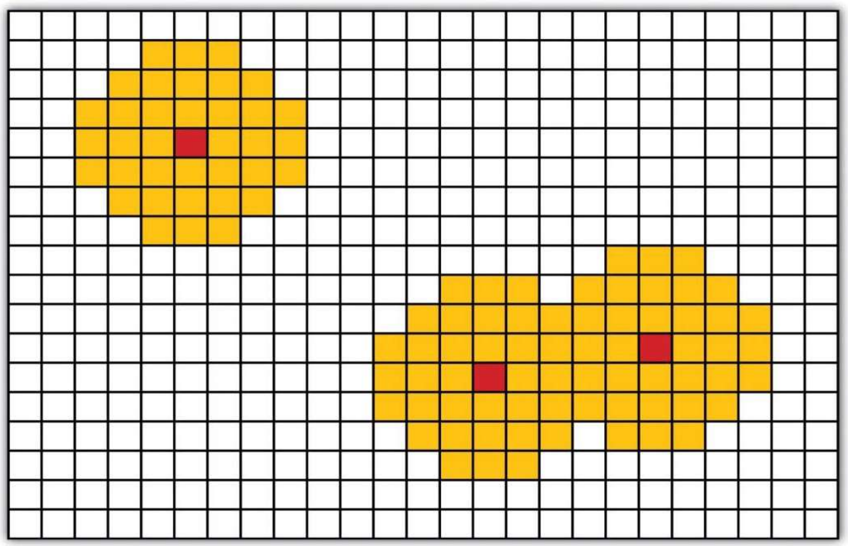
*Figure 8.3* Clipping a Raster to a Vector Polygon Layer




Raster overlays are relatively simple compared to their vector counterparts and require much less computational power

## Raster buffer

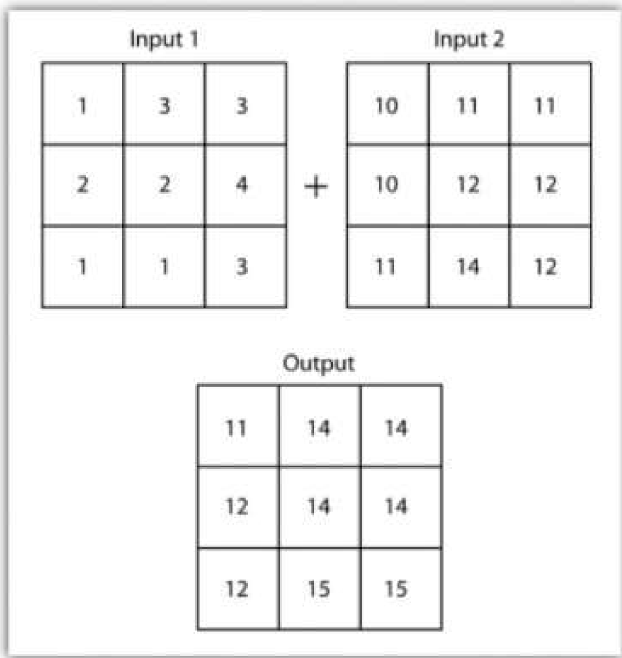
by creating a grid of **distance values from the center of the target cell(s)** to the center of the neighboring cells and then reclassifying those distances such that a **"1" represents those cells composing the original target, a "2" represents those cells within the user-defined buffer area, and a "0" represents those cells outside of the target and buffer areas.**



*Raster Buffer around a Target Cell(s)*

The **mathematical raster overlay** is the most common overlay method. The numbers within the aligned cells of the **input grids can undergo any user-specified mathematical transformation**. Following the calculation, an output raster is produced that contains a new value for each cell

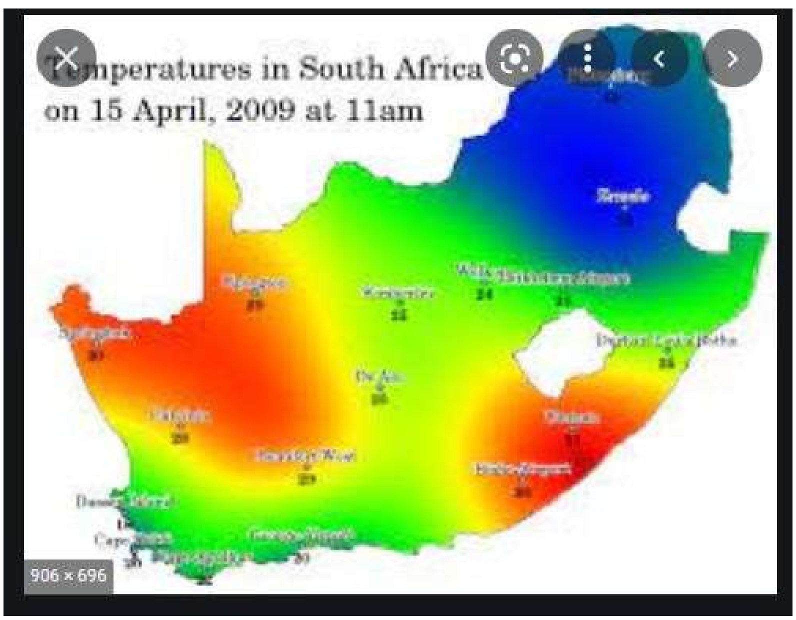
*Figure 8.4 Mathematical Raster Overlay*




*Two input raster layers are overlain to produce an output raster with summed cell values.*

# SPATIAL DATA INTERPOLATION TECHNIQUES

**Interpolation is the process of using points with known values or sample points to estimate values at other unknown points** be used to predict unknown values for any geographic point data, such as **elevation, rainfall, chemical concentrations, noise levels, and so on.**



# Spatial Interpolation

**Spatial interpolation is the process of using points with known values to estimate values at other unknown points. For example, to make a precipitation (rainfall) map for your country, you will not find enough evenly spread weather stations to cover the entire region.**

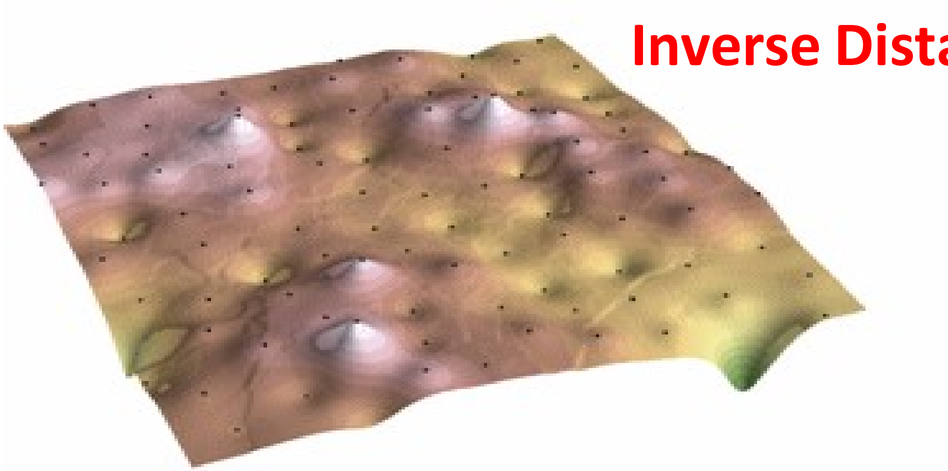
**Spatial interpolation can estimate the temperatures at locations without recorded data by using known temperature readings at nearby weather stations.**

This type of interpolated surface is often called a statistical surface. **Elevation data, precipitation, snow accumulation, water table and population density are other types of data that can be computed using interpolation.**

Interpolation uses vector points with known values to estimate values at unknown locations to create a raster surface covering an entire area. The interpolation result is typically a raster layer.

**The Inverse Distance Weighting interpolator assumes that each input point has a local influence that diminishes with distance.** It weights the points closer to the processing cell greater than those further away. **A specified number of points, or all points within a specified radius can be used to determine the output value of each location**. Use of this method assumes the variable being mapped **decreases in influence with distance from its sampled location.**

## Inverse Distance Weighted (IDW)



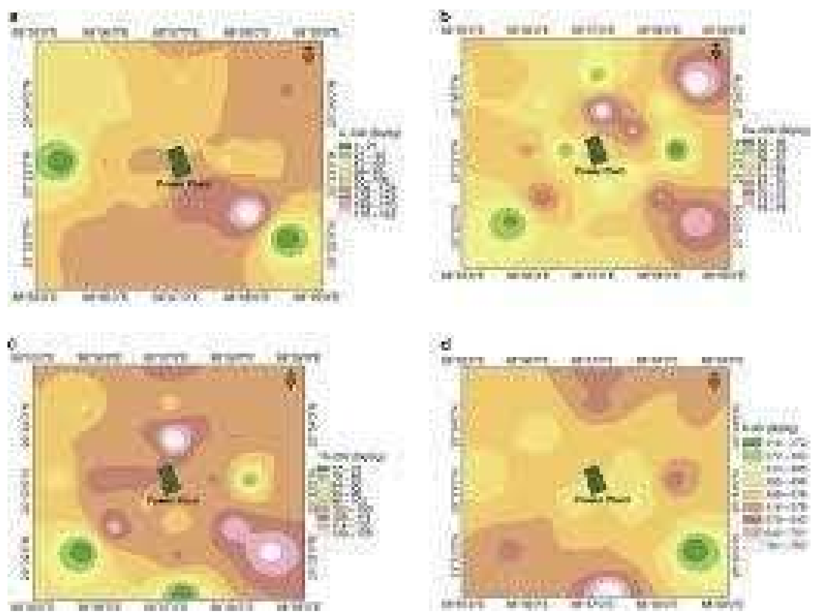
IDW Interpolated Surface; Courtesy:ESRI

IDW assumes that each measured point has a local influence that diminishes with distance. **The IDW function should be used when the set of points is dense enough to capture the extent of local surface variation needed for analysis.** **IDW determines cell values using a linear-weighted combination set of sample points. It weights the points closer to the prediction location greater than those farther away, hence the name inverse distance weighted.**

## Natural Neighbor Inverse Distance Weighted (NNIDW)

Natural neighbor interpolation has many positive features, can be used for both **interpolation and extrapolation**, and generally **works well with clustered scatter points.** This method can efficiently handle large input point datasets. When using the Natural Neighbor method, local coordinates define the amount of influence any scatter point will have on output cells.

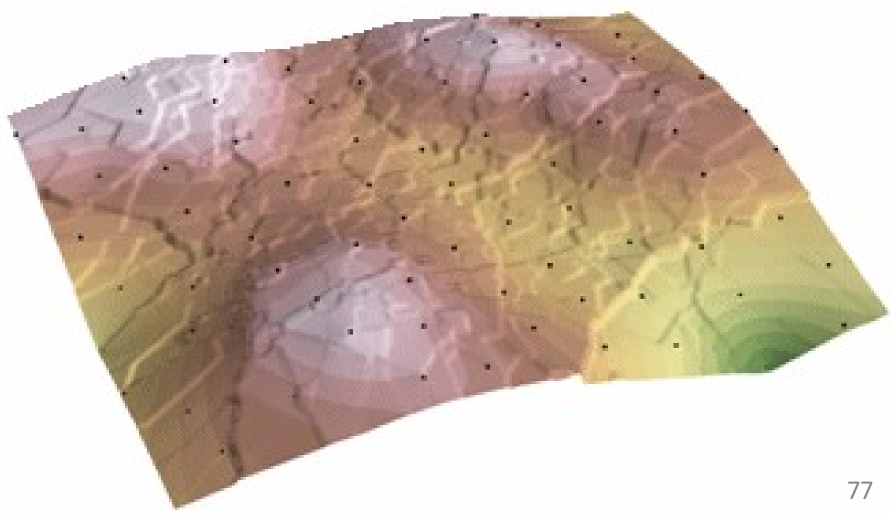
**The Natural Neighbor method is a geometric estimation technique that uses natural neighborhood regions generated around each point in the data set.**



# Kriging

Kriging is a geostatistical interpolation technique that considers **both the distance and the degree of variation between known data points** when estimating values in unknown areas. **A kriged estimate is a weighted linear combination of the known sample values around the point to be estimated.**

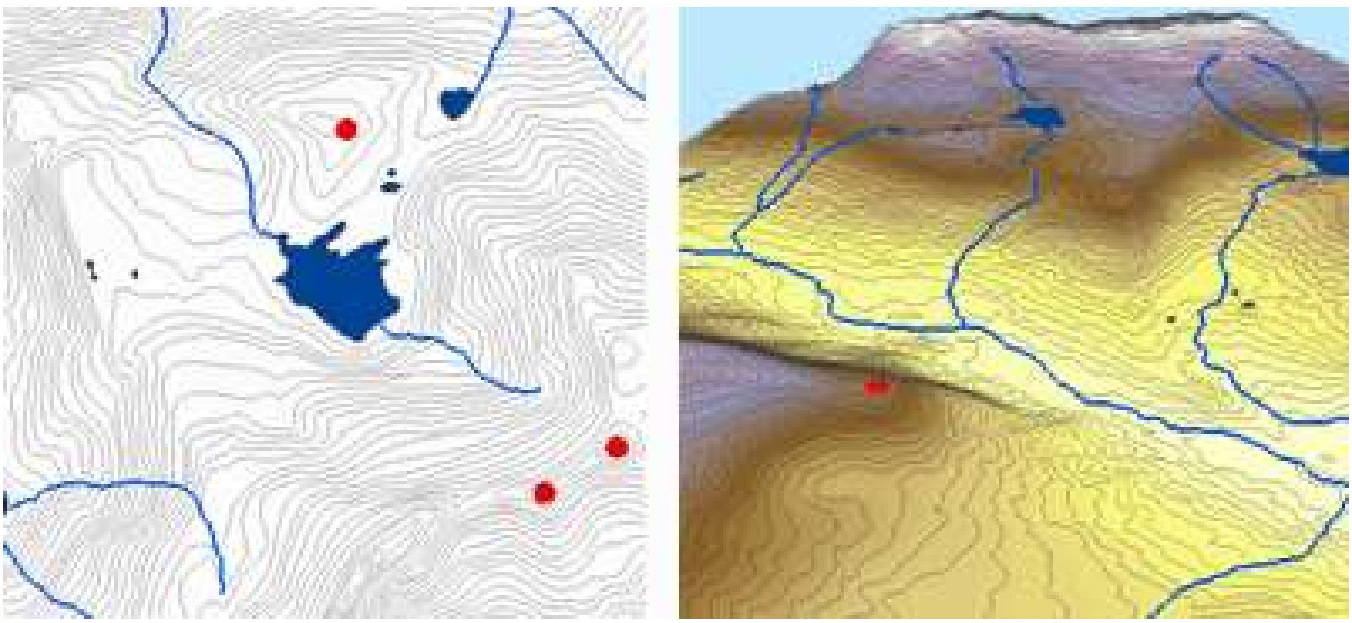
Kriging



# Topo to Raster

By **interpolating elevation values for a raster, the Topo to Raster method imposes constraints that ensure a hydrological correct digital elevation model**

**It was specifically designed to work intelligently with contour inputs.** Below is an example of a surface interpolated from elevation points, contour lines, stream lines, and lake polygons using Topo to Raster interpolation



Topo to raster