



**GIS 506/DGIS 506**

**M.A. /M.Sc. Geo-informatics/DGIS**

**ADVANCE GIS**

**DEPARTMENT OF REMOTE SENSING AND GIS  
SCHOOL OF EARTH AND ENVIRONMENT SCIENCE  
UTTARAKHAND OPEN UNIVERSITY  
HALDWANI (NAINITAL)**

**ADVANCE GIS**



**DEPARTMENT OF REMOTE SENSING AND GIS  
SCHOOL OF EARTH AND ENVIRONMENT SCIENCE  
UTTARAKHAND OPEN UNIVERSITY**

**Phone No. 05946-261122, 261123**

**Toll free No. 18001804025**

**Fax No. 05946-264232,**

**E. mail [info@uou.ac.in](mailto:info@uou.ac.in)**

**Website <http://uou.ac.in>**

---

**Board of Studies**

**Chairman**

Vice Chancellor  
Uttarakhand Open University, Haldwani

Professor R.K. Pande  
Dean Arts, DSB Campus  
Kumaun University  
Nainital

**Convener**

Professor P.D. Pant  
School of Earth and Environment Science  
Uttarakhand Open University, Haldwani

Professor D.D. Chauniyal  
Retd. Professor  
Garhwal University  
Srinagar

Professor Pradeep Goswami  
Department of Geology  
DSB Campus, Nainital

Dr. Suneet Naithani  
Associate Professor,  
Department of Environmental Science  
Doon ,University, Dehradun

Dr. Ranju Joshi Pandey  
Department of Geography & NRM  
Department of Remote Sensing and GIS  
School of Earth and Environmental Science  
Uttarakhand Open University, Haldwani

---

**Programme Coordinator**

Dr. Ranju J. Pandey  
Department of Geography & NRM  
Department of Remote Sensing and GIS  
School of Earth and Environment Science  
Uttarakhand Open University, Haldwani

## GIS-506/DGIS-506

S.No.	Units Written By	Unit No.
1.	Dr. Harish Karnatak, Scientist & Head Geo-web Services, IT & Distance Learning Department Indian Institute of Remote Sensing, ISRO, Dehradun	1
2.	Dr. Swati Thakur, Assistant Professor Department of Geography, Dayal Singh College University of Delhi, Lodi Road, Delhi 110003	2 & 11
3.	Dr. Manish Kumar, Assistant Professor Department of Geography School of Basic Sciences, Central University of Haryana & Sanjit Kumar	3,6 & 7
4.	Dr. Manish Kumar, Assistant Professor Department of Geography School of Basic Sciences, Central University of Haryana, & Sourav Bhadwal	4,8 & 12
5.	Dr. Manish Kumar, Assistant Professor Department of Geography School of Basic Sciences, Central University of Haryana & Rishi Kumar	5
6.	Dr. Sneh Gangwar, Assistant Professor Department of Geography Aditi Mahavidyalay College, University of Delhi Delhi Auchandi Road, Bawana, Delhi-110039	9,10 & 13

### Course Editor

Dr. Ranju J. Pandey  
Department of Geography & NRM  
Department of Remote Sensing and GIS  
School of Earth and Environment Science  
Uttarakhand Open University, Haldwani

Title	:	Advance GIS
ISBN No.	:	
Copyright	:	Uttarakhand Open University
Edition	:	First (2021) Second (2022)

Published By: Uttarakhand Open University, Haldwani, Nainital-263139

Printed By:

---

**CONTENTS**

---

**BLOCK 1: SPATIAL DATABASE**

<b>UNIT 1: GIS Database</b>	<b>01-23</b>
<b>UNIT 2: Characteristics of Spatial &amp; Non Spatial Data</b>	<b>24-37</b>
<b>UNIT 3: Topology Creation and Data Query</b>	<b>38-63</b>
<b>UNIT 4: Data Manipulation</b>	<b>64-73</b>

**BLOCK 2: SPATIAL DATABASE RASTER ANALYSIS**

<b>UNIT 5: Raster Data Manipulation and Reclassification</b>	<b>74-88</b>
<b>UNIT 6: Raster Data Analysis-Local, Focal, Zonal and Global</b>	<b>89-103</b>
<b>UNIT 7: Raster Data Analysis- Arithmetic Operations and Decision Rule Based</b>	<b>104-129</b>
<b>UNIT 8: Raster Data Formats</b>	<b>130-146</b>

**BLOCK 3: SPATIAL DATABASE VECTOR ANALYSIS**

<b>UNIT 9: Overlay Analysis- Union, Intersection</b>	<b>147-163</b>
<b>UNIT 10: Proximity Analysis- Buffering</b>	<b>164-176</b>
<b>UNIT 11: Networking Analysis: Optimal Path &amp; Neighborhood</b>	<b>177-187</b>
<b>UNIT 12: Map Manipulation</b>	<b>188-201</b>
<b>UNIT 13: Vector Data Formats</b>	<b>202-216</b>

## **BLOCK 1: SPATIAL DATABASE**

---

### **UNIT 1 - GIS DATABASE**

---

***1.1 OBJECTIVES***

***1.2 INTRODUCTION***

***1.3 GIS DATABASE***

***1.4 SUMMARY***

***1.5 GLOSSARY***

***1.6 ANSWER TO CHECK YOUR PROGRESS***

***1.7 REFERENCES***

***1.8 TERMINAL QUESTIONS***

---

## ***1.1 OBJECTIVES***

---

After reading this chapter, the student will understand:

- Various GIS or spatial data types such as raster, vector and attribute or non-spatial data;
- Understand the various topological relationship of vector data which are very important to minimize the errors in GIS database;
- Understand the concept of Layers and Coverage in GIS.

---

## ***1.2 INTRODUCTION***

---

Geographic information system (GIS) or Geospatial Information Systems or Geomatics is a integrations of tools that captures, stores, analyzes, manages, and presents data related to location(s). In the simplest terms, GIS is the merging of cartography, statistical analysis, and database system with information technology. GIS systems are used in cartography, remote sensing, land surveying, public utility management, natural resource management, precision agriculture, photogrammetry, geography, urban planning, emergency management, navigation, aerial video, and localized search engines and many more areas. Therefore, in a general sense, the term describes any information system that integrates stores, edits, analyzes, shares, and displays geographic information for informing decision making. GIS applications are tools that allow users to create interactive queries (user-created searches), analyze spatial information, edit data, maps, and present the results of all these operations. Geographic information science is the science underlying the geographic concepts, applications and systems.

Although the above definitions cover wide range of subjects and activities best refer to geographical information, sometimes it is also termed as Spatial Information Systems as it deals with located data, for objects positioned in any space, not just geographical, a term for world space. Similarly, the term 'a spatial data' is often used to refer location data in space and time. The discipline that deals with all aspects of spatial data handling is known as geospatial or geomatics or Geoinformatics.

During the initial phases of development, GIS has been extensively used for data conversion / digitization of paper maps, storing and generating map prints with little focus on spatial analysis. With the advent of time, the scenario has changed drastically wherein the spatial analysis took the pivotal role in many location specific planning and decision making. GIS also

facilitates modeling to arrive at locale specific solutions by integrating spatial and non-spatial data such as thematic layers and socio-economic data. With the simultaneous development of communication networks, the data storage boundaries have been erased and new areas like collaborative mapping and web map services have been developed. Now the present GIS technology enables map anywhere and serve anytime'. With recent developments, there is a leap in the development of spatial analysis tools and logical processing methods. This enabled the development of numerous spatial algorithms, spatial modeling techniques and better display and visualization of data.

Modern GIS technologies use digital information, for which various digitized data creation methods are used. The most common method of data creation is digitization, where a hard copy map or survey plan is transferred into a digital medium through the use of a computer-aided design (CAD) program, and geo-referencing capabilities. With the availability of ortho-rectified images both from satellite and aerial sources, onscreen digitization is becoming the main avenue through which geographic data is extracted. On-screen digitization involves the tracing of geographic data directly on top of the images instead of by the traditional method of tracing the geographic form on a separate digitizing tablet.

---

### ***1.3 GIS DATABASE***

---

The data is row facts and figures on a subject or theme with respect to qualitative or quantitative variables. The power of GIS is its 'data' which allows various analytics to present the information in meaningful ways. Without data, GIS will become a simple drawing tool which will be same computer aided design. The effort to bring stable, accurate data is an enormous task for any GIS. Data development and maintenance is the most costly, labor intensive component of GIS. There are several ways in which to bring spatial data into a GIS. This chapter provides a brief overview of some of the more common methods. GIS data represents real objects such as roads, land use, elevation, trees, waterways, etc. with digital data determining the mix. In general, the geography is represented either as a Field or an Object. A field is a phenomenon that has a value everywhere in the geographic space. It can be discrete or continuous. An object are usually well distinguishable, discrete bounded entities. The space between them is potentially empty. The real objects can be divided into two abstractions:

**Discrete Data:**

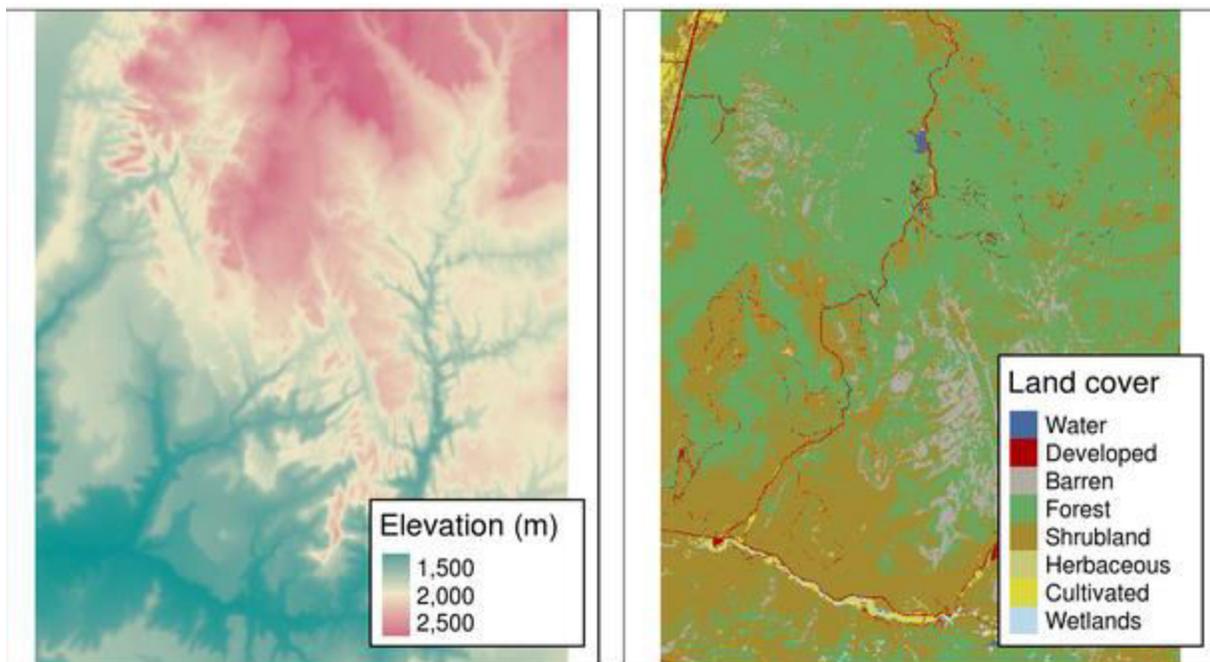
Discrete data, which is also known as thematic, categorical, or discontinuous data. The discrete data often represents objects in both the feature (vector) and raster data storage systems (please refer following section for raster and vector data models). A discrete object (s) will have known and definable boundaries in the geography. In discrete object it is easy to define precisely where the object begins and where it ends. Typically a house is a discrete object within the surrounding landscape. Its boundaries and corners can be easily marked. Other examples of discrete objects could be lake, roads, and ward boundaries in the city. In Figure 1.1 the Nainital Lake is shown as discrete object.



**Figure 1.1-** Nainital Lake as discrete object (Source- <https://theculturetrip.com>)

## Continuous Data:

A continuous surface represents a geographic phenomenon in which each location on the surface is have a value or its relationship from a fixed point in space or from an emitting source. Continuous data is also referred to as field, non-discrete, or surface data. One type of continuous surface is derived from those characteristics that define a surface, in which each location is measured from a fixed registration point. These include elevation (the fixed point being sea level) and aspect (the fixed point being direction: north, east, south, and west). Another type of continuous surface includes phenomena that progressively vary as they move across a surface. The most suitable example of progressively varying continuous data are fluid and air movement. These surfaces are characterized by the type or manner in which the phenomenon moves. The continuous and categorical data in a map representation are shown in Figure 1.2.



**Figure 1.2-** Continuous and Categorical Data

Traditionally, there are two broad methods used to store data in a GIS for both kinds of abstractions mapping references: raster images and vector. Points, lines, and polygons are the stuff of mapped location attribute references. A new hybrid method of storing data is that of identifying point clouds, which combine three-dimensional points with Red Green or Blue (RGB) information at each point, returning a "3D color image". GIS thematic maps then are

becoming more and more realistically visually descriptive of what they set out to show or determine.

### **GIS or Spatial Data Models:**

The real world observations (objects or events that can be recorded in 2D or 3D space) need to store in the computer system for effective understanding and analysis. Conversion of real world geographical variation into discrete objects or continuous field is done through data models. Typically, Model is an abstract representation of the reality. It represents the linkage between the real world domain of geographic data and computer representation of these features. Data models discussed here are for representing the spatial information. Primarily in GIS, data models are of two types: (a) raster and (b) vector.

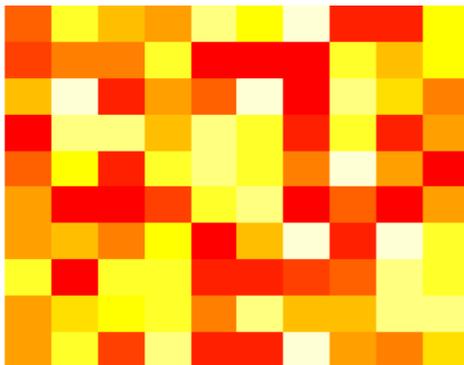
### **Raster Data Model:**

Most likely, you are already very familiar with this data model if you have taken or seen any digital photographs. Look into the photograph taken by you using your smart mobile phone. Check its file format in you mobile or computer. In general these photographs are stored in JPEG or JPG or PNG formats. Zoom these images till you see the small square size boxes. Yes, these are called Pixels. These images and file formats such as JPEG, PNG, BMP, and TIFF are based on the raster data model. The latest digital displays using Liquid Crystal Display (LCD) such as computer monitors are based on raster technology as they are composed of a set number of rows and columns of pixels. Notably, the foundation of this technology predates computers and digital cameras by nearly a century.

In raster type of representation of the geographical data, a set of cells located by coordinate is used; each cell is independently addressed with the value of an attribute. Each cell contains a single value and every location corresponds to a cell. A digital image is concerned with its output blending together its grid based details as an identifiable representation of reality, in a photograph or art image transferred into a computer, the raster data type reflects a digitized abstraction of reality dealt with by grid populating tones or objects, quantities, co-joined or open boundaries, and map relief schemas. Aerial photos and satellite images are commonly used form of raster data, with one primary purpose in mind: to display a detailed image on a map area, or for the purposes of rendering its identifiable objects by digitization. Additional raster data sets used by a GIS contains information regarding elevation, a digital elevation model, or reflectance

of a particular wavelength of light. In GIS, one set of cell and associated value is known as a LAYER. Raster models are simple with which spatial analysis is easier and faster.

Raster images consists of rows and columns of cells, with each cell storing a single value. Raster data can be images (raster images) with each pixel (or cell) containing a color value. Additional values recorded for each cell may be a discrete value, such as land use, a continuous value, such as temperature, or a null value if no data is available. While a raster cell stores a single value, it can be extended by using raster bands to represent RGB (red, green, blue) colors, color maps (a mapping between a thematic code and RGB value), or an extended attribute table with one row for each unique cell value. The resolution of the raster data set is its cell width in ground units. Raster data is stored in various formats; from a standard file-based structure of TIF, JPEG, etc. to binary large object (BLOB) data stored directly in a relational database management system (RDBMS) similar to other vector-based feature classes. Database storage, when properly indexed, typically allows for quicker retrieval of the raster data but can require storage of millions of significantly sized records. Sample raster representation is shown in Figure 1.3.



A simple raster image of 10 x10 array of cells or pixels



Sample satellite image (LISS IV) – Raster representation

**Figure 1.3-** Sample Raster representation in Image

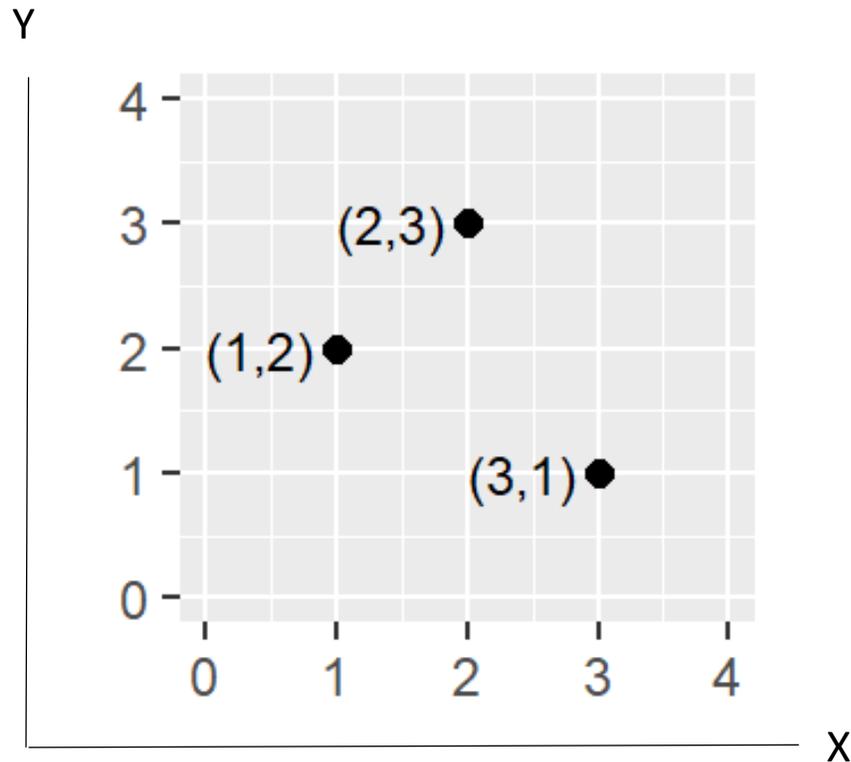
The raster data model consists of uniform series of square pixel and is referred to as a grid-based system. Typically, a single data value will be assigned to each grid location. Each cell in a raster carries a single value, which represents the characteristic of the spatial phenomenon at a location denoted by its row and column and is known as digital number or pixel value. The data type for that cell value can be either integer or floating-point. The advancements in database management system allows to link multiple attribute tables to link with raster graphics.

The raster data model averages all values within a given pixel to produce a single value for the region. Therefore, *the more area covered per pixel, the less accurate the associated data values.* The area covered by each pixel determines the spatial resolution of the raster model from which it is derived. Specifically, resolution is determined by measuring one side of the square pixel. A raster model with pixels representing 10 m by 10 m (or 100 square meters) in the real world would be said to have a spatial resolution of 10 m; a raster model with pixels measuring 1 km by 1 km (1 square kilometer) in the real world would be said to have a spatial resolution of 1 km; and so forth.

### **Vector Data Model:**

Vector data models use points and their associated X, Y coordinate pairs to represent the vertices of geographical feature (s). This representation looks as if they were being drawn on a map by hand. Vector data model uses line segments or points represented by their explicit x, y coordinates to identify locations. Discrete objects are formed by connecting line segments which area is defined by set of line segments. The data attributes of vector features are stored in a separate attribute table as a file or database management system. The spatial information and the attribute information for these models are linked via a simple identification number that is given to each feature in a map. Vector data models require less storage space, outputs are appreciable, estimation of area/perimeter is accurate and editing is faster and convenient. The vector model is extremely useful for describing discrete features, but less useful for describing continuously varying features such as soil type or accessibility costs for hospitals. In a GIS, geographical features are often expressed as vectors, by considering those features as geometrical shapes. Different geographical features are expressed by different types of geometry:

**Points:** is a simple vector map, using each of the vector elements: points for wells, lines for rivers, and a polygon for the lake. Points are zero-dimensional objects that contain only a single coordinate pair. Zero-dimensional points are used for geographical features that can best be expressed by a single point reference — in other words, by simple location. Examples include wells, peaks, features of interest, and trailheads. Points convey the least amount of information of these file types. Points can also be used to represent areas when displayed at a small scale. For example, cities on a map of the world might be represented by points rather than polygons. No measurements are possible with point features. The representation in X, Y coordinates are shown in Figure 1.4.

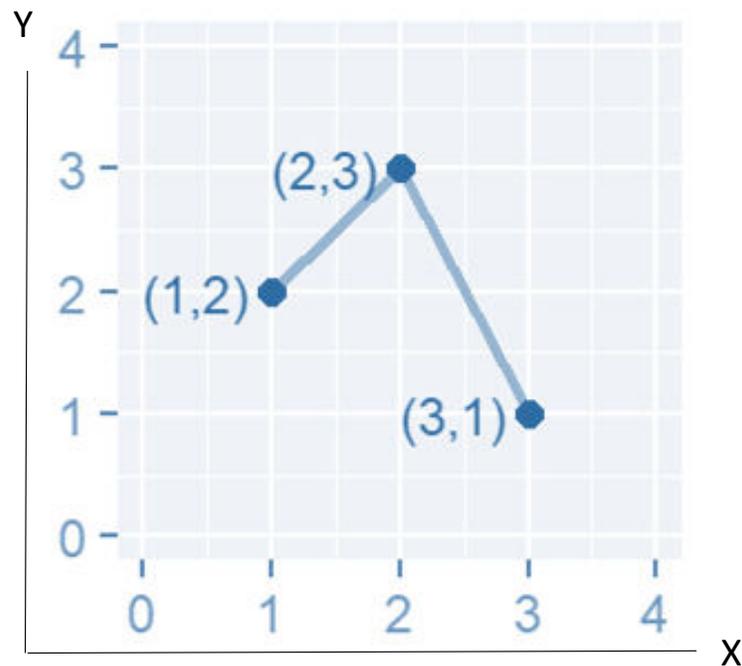


**Figure 1.4-** Vector points defined by X and Y coordinate values

Points are the most basic geometric type having no length or area. However the geographic feature represented points have both area and shape (e.g. circle, square, plus signs). We seem capable of interpreting such symbols as points, but there may be instances when such interpretation may be ambiguous (e.g. is a round symbol delineating the area of a round feature on the ground such as a large oil storage tank or is it representing the point location of that tank?).

***Lines or Polylines:*** One-dimensional lines or polylines are used for linear features such as rivers, roads, railroads, trails, and topographic lines. Again, as with point features, linear features displayed at a small scale will be represented as linear features rather than as a polygon. Lines are composed of multiple, explicitly connected points. Lines have the property of length. Lines that directly connect two nodes are sometimes referred to as chains, edges, segments, or arcs. In the line feature the measurement of distance is possible but area calculation is not. A polyline is composed of a sequence of two or more coordinate pairs called vertices. A vertex is defined by coordinate pairs just like a point, but what differentiates a vertex from a point is its explicitly

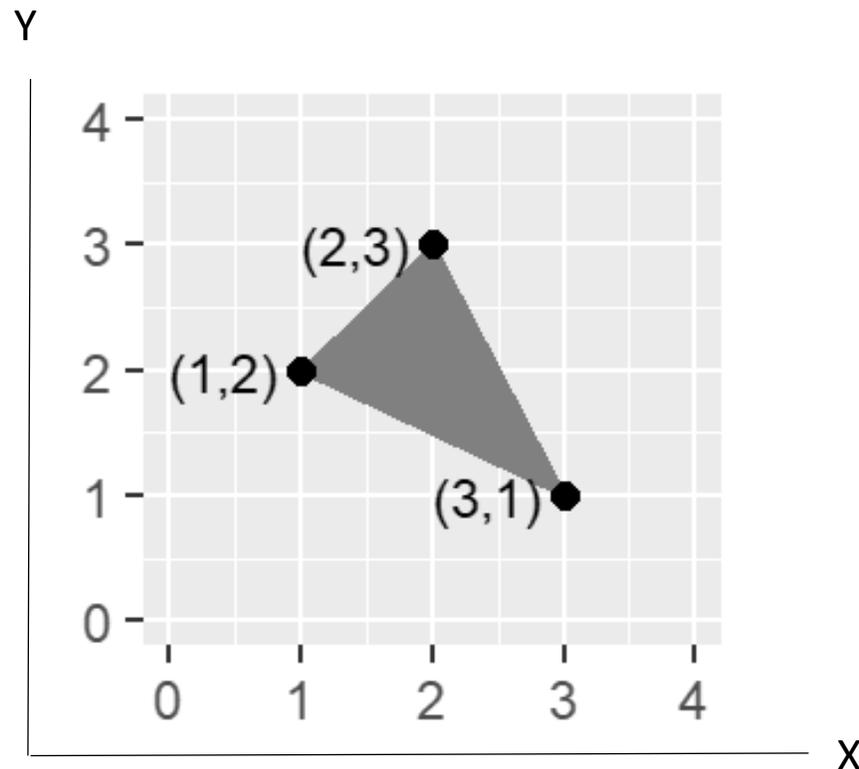
defined relationship with neighboring vertices. A vertex is connected to at least one other vertex. Like a point, a true line can't be seen since it has no area. And like a point, a line is symbolized using shapes that have a color, width and style (e.g. solid, dashed, dotted, etc...). A sample coordinate representation of line feature is shown in Figure 1.5.



**Figure 1.5-** A simple polyline object defined by three connected vertices

**Polygons:** Polygons are two-dimensional geometry used for geographical features that cover a particular area of the earth's surface. Such features may include lakes, park boundaries, buildings, city boundaries, or land uses.

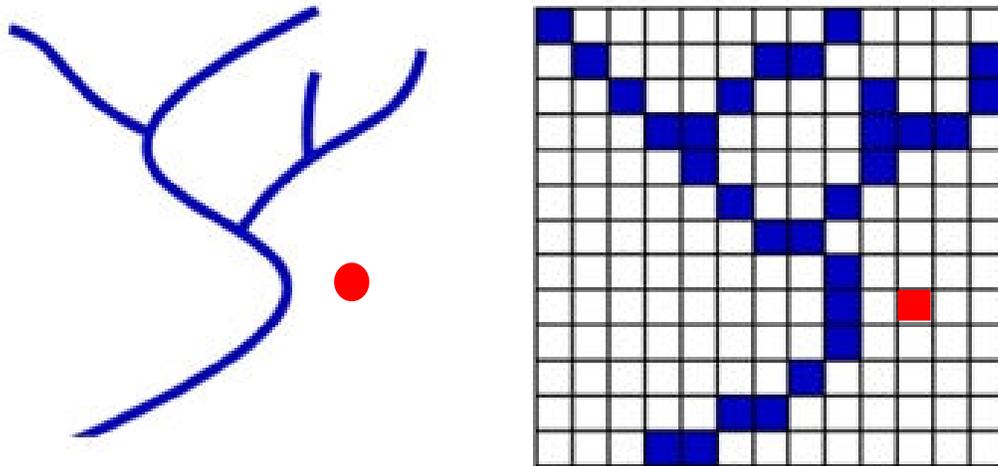
A polygon is composed of one or more lines whose starting and ending coordinate pairs are the same. Polygons have the topological relations such as inside and outside; in fact, the area that a polygon encloses is explicitly defined and calculated in GIS sometime automatically also. If you are working with a feature which looks to be closed area but area calculation is not possible then certainly it is a Polyline. If this does not seem intuitive, think of three connected lines defining a triangle: they can represent three connected road segments (thus polyline features), or the grassy strip enclosed by the connected roads (in which case an 'inside' is implied thus defining a polygon). A sample polygon representation in X, Y coordinates is shown in Figure 1.6.



**Figure 1.6-** A simple polygon object defined by an area enclosed by connected vertices

Polygons convey the most amount of information of the file types. Polygon features can measure perimeter and area. Each of these geometries is linked to a row in a database that describes their attributes. For example, a database that describes lakes may contain a lake's depth, water quality, pollution level. This information can be used to make a map to describe a particular attribute of the dataset.

The raster model has evolved to model such continuous features. A raster image comprises a collection of grid cells rather like a scanned map or picture. Both the vector and raster models for storing geographic data have unique advantages and disadvantages. Modern GIS packages are able to handle both models. Representation of same feature in vector geometry and raster cells is shown in Figure 1.7.



**Figure 1.7-** Vector and Raster representation of Point and line features

Vector features can be made to respect spatial integrity through the application of topology rules such as 'polygons must not overlap'. Vector data can also be used to represent continuously varying phenomena. Contour lines and triangulated irregular networks (TIN) are used to represent elevation or other continuously changing values. TINs record values at point locations, which are connected by lines to form an irregular mesh of triangles.

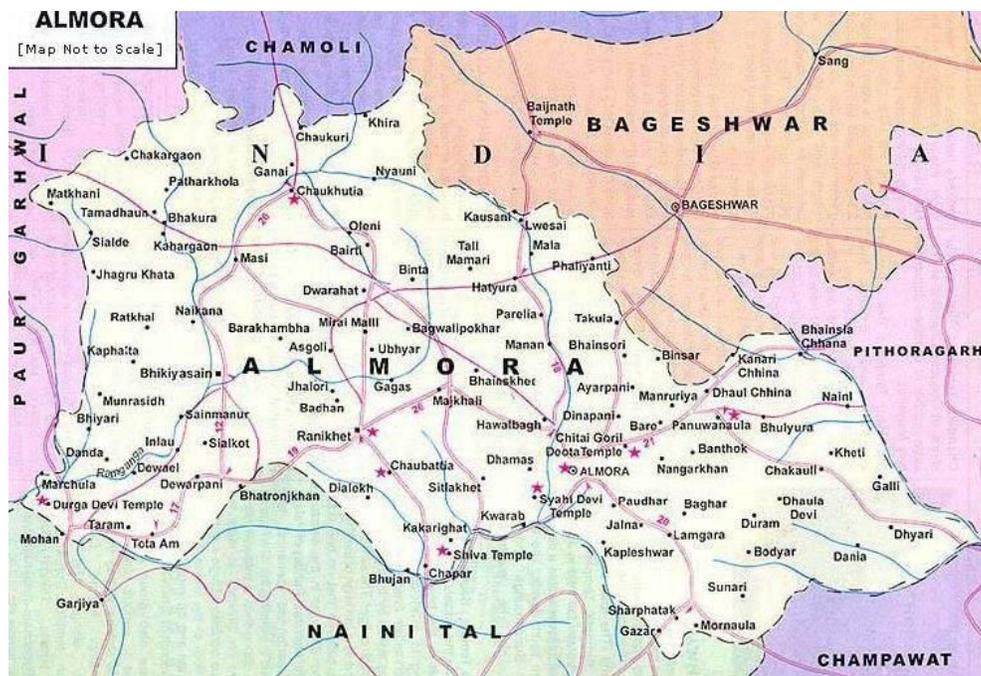
### **Topology:**

Topology is a mathematical relationship between earth objects. It is the method to structure the data based on the principles of feature adjacency and feature connectivity. It is in fact the mathematical method used to define spatial relationships. Without a topologic data structure in a vector based GIS most data manipulation and analysis functions would not be practical or feasible.

Topology is an important aspect of vector-based models is that they enable individual components to be isolated for the purpose of carrying out measurements of, for example, area and length, and for determining the spatial relationships between the components. Spatial relationships of connectivity and adjacency are examples of topological relationships and a GIS spatial model in which these relationships are explicitly recorded is described as topologically structured. In a fully topologically structured data set, wherever lines or areas cross each other, nodes will be created at the intersections and new areal subdivisions defined. In two dimensions,

this may be regarded as part of the process of planar enforcement referred to previously. *In GIS, topology is implemented through data structure.*

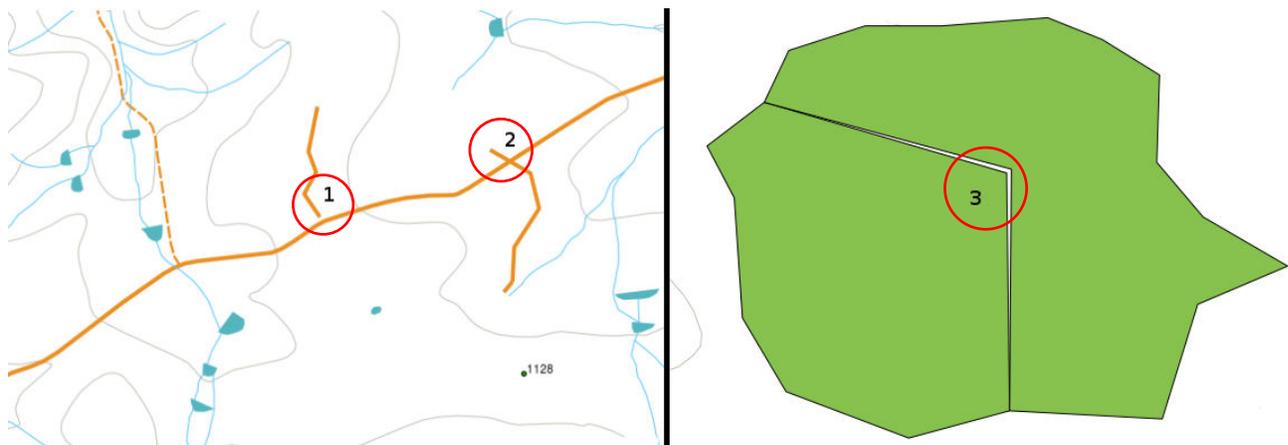
Topological structure is important in keeping track of the components of complex objects and in determining the spatial relationships of connectivity and adjacency between recorded phenomena. Thus if two lines cross each other they will share a common node. If two areas are adjacent to each other, such as two neighboring counties, they will share a common boundary arc. If the boundary of a county coincides with the path of a river they might also share the same arc. The inclusion of one area in another, such as a specific type of forest within a county, will result in their sharing common polygons. The presence of these various spatial relationships can be determined by relatively simple comparisons of the identifiers of their topological components, rather than requiring possibly computationally demanding geometric calculations based on coordinates. It may also be noted that because shared spatial objects are only stored once, though perhaps referenced many times, storage space is saved by avoiding duplication of the same geometric data. This in turn assists in the maintenance of the integrity of the database by avoiding the possibility of two different versions of the same geometric components. The topology of tourist destination and road network of tourist map of Almora is shown in Figure 1.8.



**Figure 1.8-** Topology of tourist destinations and Road network of Almora (Map source- <http://www.uttarakhand-tourism.com/>)

## Topology errors:

There are different types of topological errors and they can be grouped according to whether the vector feature types are polygons or polylines. Topological errors with **polygon** features can include unclosed polygons, gaps between polygon borders or overlapping polygon borders. A common topological error with polyline features is that they do not meet perfectly at a point (node). This type of error is known as **undershoot** error if a gap exists between the lines and an **overshoot** if a line ends beyond the line it should connect. The **slivers** created when digitizing the polygons. The example of three topological errors is shown in figure 1.9.



**Figure 1.9-** Undershoots (1) occur when digitized vector lines that should connect to each other don't quite touch. Overshoots (2) happen if a line ends beyond the line it should connect to.

Slivers (3) occur when the vertices of two polygons do not match up on their borders.

The result of overshoot and undershoot errors are so-called '**dangling nodes**' at the end of the lines. Dangling nodes are acceptable in special cases, for example if they are attached to dead-end streets.

Topological errors break the relationship between features. These errors need to be fixed in order to be able to analyze vector data with procedures like network analysis (e.g. finding the best route across a road network) or measurement (e.g. finding out the length of a river). In addition to topology being useful for network analysis and measurement, there are other reasons why it is important and useful to create or have vector data with correct topology. Just imagine you digitize a municipal boundaries map for your province and the polygons overlap or show

slivers. If such errors were present, you would be able to use the measurement tools, but the results you get will be incorrect. You will not know the correct area for any municipality and you will not be able to define exactly, where the borders between the municipalities are.

It is not only important for your own analysis to create and have topologically correct data, but also for people who you pass data on to. They will be expecting your data and analysis results to be correct!

### **Topology rules:**

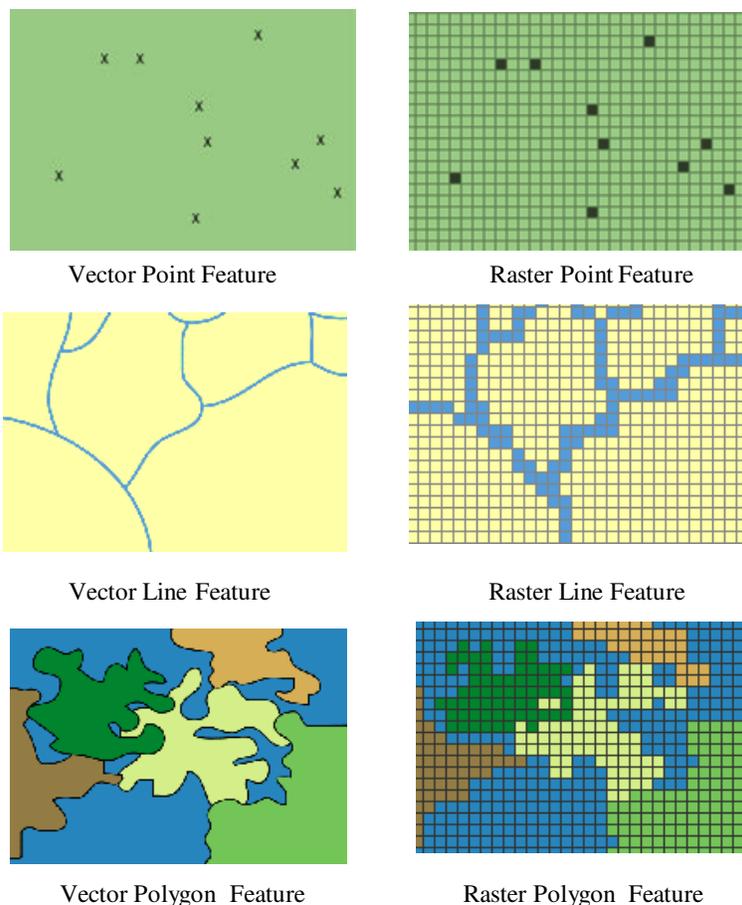
Fortunately, many common errors that can occur when digitizing vector features can be prevented by topology rules that are implemented in many GIS applications. Except for some special GIS data formats, topology is usually not enforced by default. Many common GIS, like QGIS, define topology as relationship rules and let the user choose the rules, if any, to be implemented in a vector layer. The following list shows some examples of where topology rules can be defined for real world features in a vector map:

- Area edges of a municipality map must not overlap.
- Area edges of a municipality map must not have gaps (slivers).
- Polygons showing property boundaries must be closed. Undershoots or overshoots of the border lines are not allowed.
- Contour lines in a vector line layer must not intersect (cross each other).

### **Raster versus Vector Data Models:**

- Raster datasets record a value for all points in the area covered which may require more storage space than representing data in a vector format that can store data only where needed.
- Raster data allows easy implementation of overlay operations, which are more difficult with vector data.
- Vector data can be displayed as vector graphics used on traditional maps, whereas raster data will appear as an image that may have a blocky appearance for object boundaries. (depending on the resolution of the raster file)
- Vector data can be easier to register, scale, and re-project, which can simplify combining vector layers from different sources.

- Vector data is more compatible with relational database environments, where they can be part of a relational table as a normal column and processed using a multitude of operators.
- Vector file sizes are usually smaller than raster data, which can be 10 to 100 times larger than vector data (depending on resolution).
- Vector data allows much more analysis capability, especially for "networks" such as roads, power, rail, telecommunications, etc. Examples: Best route, largest port, airfields connected to two-lane highways. Raster data will not have all the characteristics of the features it displays (Figure 1.10).



**Figure 1.10-** Vector and Raster representation of geographic features

In GIS, additional non-spatial data (sometimes referred to as attribute data) can also be stored along with the spatial data represented by the coordinates of vector geometry or the position of a raster cell. In vector data, the additional data contains attributes of the feature. For example, a forest

inventory polygon may also have an identifier value and information about tree species. In raster data the cell value can store attribute information, but it can also be used as an identifier that can relate to records in another table.

### **Non-spatial or Attribute Data:**

Non-spatial data is also known attribute or characteristic data. It consists of the characteristics of spatial features which are independent of all geometric considerations. Let us illustrate this with the help of an example. The non-spatial data of town comprise of name of the town, its population, settlement type, means of transportation and communication, administration set-up, education institutions, occupations and facilities. It is important to note that all the above mentioned data of town are not dependent on their location identities. Hence, non-spatial data is independent from location information.

Non-spatial data are stored in GIS as tables. Such tables are known as non-spatial (attribute) tables. A non-spatial table is represented by rows and columns in which each row shows a spatial feature and each column represents a characteristic. The intersection of a row and a column gives the value of a specific characteristic for a particular feature as shown in Table 1.1. A row is also known as a record or a tuple and a column is known as a field or item

**Table- 1.1** Arrangement of rows and columns of a non-spatial data

River_ID	Rivers_Name	Total_Length	Number_of_Dam
1	Ganga	3500	5
2	Kosi	600	2
3	Song	300	3
4	Harsu	100	0

To understand it better let's say we have a spatial data model that stores the location of Community Service Center (CSC) in your locality. For each CSC, to represent the object, we would store the location/positional of CSC. In addition to the positional information, we will also store attributes that will describe the various services available in the CSC. In this example, we are storing Net banking service, revenue service such checking of land record and generation of certificates as three attributes that describe with this particular CSC at this particular position on your locality. The location, net banking service, revenue service and generation of certificates

will be stored as one row in an attribute table that will contain four columns because there are four descriptors for this CSC.

Attributes can store all kinds of different descriptive statistical information, which can be broken down into four different categories: **nominal**, **ordinal**, **interval**, and **ratio**. A **nominal attribute data** provides descriptive information about the object such as the name of an object so for instance a city name, or the type of an object. What's important here is that this descriptive information does not imply any order, size, or any other quantitative information. That means that you cannot state that one attribute is greater than or less than another attribute or you cannot multiply attributes together, so for instance, it does not make sense to multiply the city name by the district. The only comparisons you can do with nominal attributes are to check whether attributes are equal or not equal. In addition to text descriptions, the nominal attribute category includes descriptive information such as images, movies, and sounds. What could be the example of it?

The next attribute category is **ordinal attribute data**, which imply a ranking or order based on their values. These values can be descriptive text, or numerical. For example, I can describe an object as having a high/medium/low ranking, or a ranking of 100/50/1. In either case, these ordinal attributes allow us to specify rank only, and not scale. So for instance, we can state that high is ordered higher than low, and high is ordered higher than medium, and low is ordered lower than high, but we cannot say that high is twice as high as medium, and medium as twice as high as low. Additionally, if the numerical attributes are of the ordinal attribute category, again we can say that 50 is ordered higher than 20 and 20 is ordered higher than 10 but we cannot say that 50 is twice as high as 25 and 25 is twice as high as 12 ½. Even though we are using numbers to describe a rank, do not let that confuse you into thinking that a scale is implied. The third entry category is **interval attribute data**. Interval attributes imply a rank order and magnitude or scale. Interval attributes use numbers, however, those numbers do not have a natural zero, and use an arbitrary zero point instead. For instance if we look at temperature on the Fahrenheit scale, 0°F is not a natural zero point for temperature, it is a human defined zero point. Therefore, while we can say that 50°F is 10°F more than 40°F, we cannot say that 50°F is twice as hot as 25°F, again, because 0°F is a human created zero, and not a natural phenomenon. With an interval attribute, addition and subtraction to make sense but not multiplication since values are relative from that arbitrary zero.

The fourth and final category is the **ratio attribute data**. A ratio attribute implies both rank order and magnitude about a natural zero. Ratio data, unlike interval attribute data, use numerical attributes of addition, subtraction, multiplication, and division where there is an absolute natural zero. So for example, if we are measuring speed in Kilometer per hour, then a car not moving at all is moving at zero Kilometers per hour. In terms of temperature, the only measurement that uses a natural zero is Kelvin, which has absolute zero.

Now you know the four different attribute categories, let's take a look at an example data set and its related attribute table, and try to identify each column as holding nominal, ordinal, interval, or ratio data.

Let's finish talking about attribute *data types*. Computers fundamentally “think” differently than humans. While humans see numbers, letters, pictures, and sounds, a computer only sees zeros and ones, or **binary data**. Therefore, we need a way to translate the numbers, sounds, and videos, as humans know it, to a form in which a computer can understand, and store the information. Computer scientists have created data structures that can be used by us to translate information into a format which the computer can store in its memory, called a **datatype**. There are four typical data types that we use in GIS: *integer*, *float/real*, *text/string*, and *date*. It is important that we specify which data type we are going to use to store information in the computer's memory so that we may use the memory in the most efficient manner and let the computer know which operations are allowed for each data point stored in that memory location using that the data type.

The first data type is the **integer**, which is a whole number, such as the number one, the number 2458, and the number -54. Integers can be used for mathematical calculations; however, any resulting fraction of a whole number will be rounded, or truncated.

The **float, or real**, data type holds a decimal number such as the number 1.452, the number 254,783.1, or -845.157. Like the integer data type, the float or real data type can be used for mathematical calculations. No rounding or truncation will take place when using float or real numbers, depending on the number of significant digits you have specified.

The **text, or string**, data type contains characters such as character “A”, the characters “GIS”, the characters “House No. 61 Kalidas Road.”, or the number “61”. Even though the text may contain numbers, it is important to note that they cannot be used for mathematical calculations. However, strings can be manipulated to find substrings, or to cut strings and locations.

The last common data type is *date*. The date data type holds time and date information such as 12/10/2018, or 10/12/18, or December 10, 2018. The date data type cannot be used for mathematical calculations however, it can be used to determine and calculate lengths of time between two different dates or times. Additionally, the computer stores the date information in its own internal data structure, but can be formatted to output the date in many different ways, as shown in these examples.

### **Layers and Coverages in GIS:**

The common requirement to access data on the basis of one or more classes has resulted in several GIS employing organizational schemes in which all data of a particular level of classification, such as roads, rivers or vegetation types are grouped into so called layers or coverages. The concept of layers is to be found in both vector and raster models. The layers can be combined with each other in various ways to create new layers that are a function of the individual ones. The characteristic of each layer within a layer-based GIS is that all locations with each layer may be said to belong to a single Aerial region or cell, whether it be a polygon bounded by lines in vector system, or a grid cell in a raster system. But it is possible for each region to have multiple attributes.

---

## ***1.4 SUMMARY***

---

Geographic information system (GIS) or Geomatics is a integrations of three major disciplines viz. geography, information technology and mathematics. Technically it has emerged as tools that captures, stores, analyzes, manages, and presents data related to location(s). In the simplest terms, GIS is the merging of cartography, statistical analysis, and database system with information technology. The GIS also provides an abstract representation of geographical features in the computer system for its better understanding and analysis. The geographical features are represented by using GIS data models. The two basic types of data exists in GIS i.e. discrete and continuous data. There are two basic data models viz. raster and vector data models which are used to represent geographical features in GIS. In raster representation of geographical features, the raster cells or pixels are used as a unit to store the information. The vector data model uses there major geometries i.e. Point, line and polygon to represent the GIS data. The topology plays a critical role to establish mathematical relationship between earth object in

vector data models. While creating the GIS data in vector data models the topological errors must be handled very carefully. The characteristics of geographical feature are stored as attribute data which is also known as non-spatial data. The non-spatial data is organized and managed in a database management system using standard data types.

---

## ***1.5 GLOSSARY***

---

- Raster Data- A raster consists of a matrix of cells or pixels organized into rows and columns or a grid where each cell contains a value representing information.
- Vector Data- Data in a format consists of points, lines or polygons.
- Spatial Data- Comprise the relative geographic information about the earth and its features.
- Non-spatial data- It is an independent of Geographic location.

---

## ***1.6 ANSWER TO CHECK YOUR PROGRESS***

---

- 1 What is Data? Explain the difference between data and Information.
- 2 What is Data Model in GIS?
- 3 What is Raster and vector representation in GIS?
- 4 What is discrete and continuous data? Explain with example.
- 5 What is Topology? Explain Topological Errors.

---

## ***1.7 REFERENCES***

---

1. Aronoff, S. 1989. Geographic Information Systems: A Management Perspective. Ottawa, Canada: WDL Publications
2. Burrough, P. A., and Mc. Donnel, R. A. Principles of Geographical Information System,- Oxford: Oxford University Press, London. 306 pp.
3. Campbell, James B. Introduction to Remote sensing.-2nd ed. – London: Taylor and Francis, 1996. 622 pp.
4. Gupta, R P, Remote Sensing geology.-2<sup>nd</sup> ed. Springer-verlag.

5. Harish Chandra Karnatak, R Shukla, VK Sharma, YVS Murthy, V Bhanumurthy, 2012 “Spatial mashup technology and real time data integration in geo-web application using open source GIS—a case study for disaster management”, *Geocarto International* 27 (6), 499-514.
6. Harish Chandra Karnatak, S Saran, K Bhatia, PS Roy, 2007 “Multicriteria spatial decision analysis in web GIS environment”, *Geoinformatica* 11 (4), 407-429.
7. Karnatak Harish, Karamjit Bhatia and Sameer Saran, (2008) “Multi-criteria decision analysis using Spatial Compromise Programming”, *Proceedings of the 2nd National Conference-INDIACom-2008, on Computing for Nation Development*, ISBN No-ISSN 0973-7529, ISBN
8. Jensen, John R. *Remote Sensing of the Environment: An Earth Resource Perspective*. – New Delhi: Pearson Education, 2006. 560 pp.
9. Jensen, John R. *Introductory Digital Image Processing: A Remote Sensing Approach*. – Raglewood cliffs: Prentice Hall, 1986. 379 pp.
10. Joseph, George. *Fundamentals of Remote Sensing*.- Hyderabad: University Press, 2003. 433 pp.
11. Lillesand, Thomas M and Kiefer, Ralph W. *Remote Sensing and Image Interpretation*.- New York: John Willey and Sons. 1987. 721 pp.
12. Maguire, David J, (Ed.), Goodchild, Michael F, (Ed.) and Rhind, David W, (Ed.). *Geographical Information System: Vol. 1: Principles*.-Essex:Longman Scientific and Technical, 1992. 649 pp.
13. Reddy Anji, M. *Textbook of remote sensing and geographical information system – 2nd ed.*-Hyderabad: B S Publications, 2001. 418 pp.
14. Sabins, Floyd F. *Remote Sensing : Principles and interpretation*. – San Fransisco. W.H.Freeman, 1978. 426 pp.
15. Schowengerdt, Robert A. *Remote Sensing: models and methods for image precessing*.-2nd ed.-San Diego: Academic Press, 1997. 522 pp.
16. Swain, Philip H and davis, Shirley M. *Remote sensing: The quantitative approach*.: Newyork, Mc Grow Hill, 1978. 396 pp.
17. Williams, Jonathan. *Geographical Information from Space: Processing and application of geocoded satellite images*. – Chichester: John Willey and Sons, 1995. 210 pp.

**Web URLs:**

1. <http://en.wikipedia.org/wiki/>
2. <http://geog.hkbu.edu.hk/geog3600/> (Hongkong Baptist University)
3. <http://geosun.sjsu.edu/paula/137/ppt/lecture13/sld008.htm>
4. <http://rst.gsfc.nasa.gov/>

5. <http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/tutore.html>
6. <http://www.cla.sc.edu/gis/avshtcrs/handouts.html>
7. <http://www.ed.ac.uk/>
8. <http://www.geoplace.com>
9. <http://www.gisdevelopment.net>
10. <http://www.gislinx.com/Software/Programs/MicroStation/index.shtml>
11. <http://www.gisqatar.org.qa/conf97/links/b4.htm>
12. <http://www.innovativegis.com/basis/>
13. <http://www.isro.gov.in>
14. <http://www.nasa.org>
15. <http://www.ncgia.ucsb.edu/~spalladi/thesis/Chapter3.html>
16. <http://www.planweb.co.uk/>
17. <http://www.sbg.ac.at/geo/idrisi/wwwtutor/tuthome.htm>
18. <http://www.sli.unimelb.edu.au/gisweb/menu.html>
19. [https://saylordotorg.github.io/text\\_essentials-of-geographic-information-systems/s08-data-models-for-gis.html](https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s08-data-models-for-gis.html)
20. [https://docs.qgis.org/testing/en/docs/gentle\\_gis\\_introduction/topology.html](https://docs.qgis.org/testing/en/docs/gentle_gis_introduction/topology.html)
21. <https://mgimond.github.io/Spatial/feature-representation.html>
22. <https://opengeospatial.weebly.com/22-non-spatial-data.html>

---

## ***1.8 TERMINAL QUESTIONS***

---

1. What is overshoot, undershoot and slivers in vector data? Explain with example.
2. How we store the characteristic of geographic features in GIS? Explain Non-spatial data.
3. What is Data types? Explain different data types used in GIS to store non-spatial data.
4. What is binary data?
5. What do you understand by Layers and Coverage in GIS?

---

## **UNIT 2 - CHARACTERISTICS OF SPATIAL & NON SPATIAL DATA**

---

**2.1 OBJECTIVES**

**2.2 INTRODUCTION**

**2.3 CHARACTERISTICS OF SPATIAL & NON SPATIAL DATA**

**2.4 SUMMARY**

**2.5 GLOSSARY**

**2.6 ANSWER TO CHECK YOUR PROGRESS**

**2.7 REFERENCES**

**2.8 TERMINAL QUESTIONS**

---

## 2.1 OBJECTIVES

---

After studying this unit you will be able to:

- Define data and differentiate between data and information.
- Define spatial and non spatial data and distinguish their characteristics.
- Enumerate conceptual model of spatial and non spatial information.
- Know the use of spatial and non spatial data in geographic analysis.

---

## 2.2 INTRODUCTION

---

Information system plays important role in any decision making. Starting from a common user who wishes to find path and direction to reach a desired destination to a politician who is concerned about prioritizing developmental activity in an area or business community interested to find optimum location of market or city planner and wants to know the areas of population concentrations, all rely on set of information. Every set of information has a concern with geographic locations, pattern of change and processes on surface of earth. Information which pertains to space other than human body representing all that surrounds is the geographic information. We are more interested in this unit to know how this geographic information can be described, measured and stored in different forms to be able to facilitate decision making tools.

---

## 2.3 CHARACTERISTICS OF SPATIAL AND NON SPATIAL DATA

---

The term Data and Information are generally used interchangeably but there is vital difference between them. *Data* are facts, or numbers representing facts. It is the one describing status in a raw state until interpreted to offer meaning. There are different terms commonly used to describe data. The most commonly heard is *Geographic Data*, which refers to any data on Earth surface or near Earth surface e.g. terrain height, drainage density or land use, *Geospatial Data* has much precision in terms of reference to location on Earth surface and/or near Earth surface, e.g. latitude/longitude, GPS location.

The word *Data* also find its close association with the Latin word '*datum*' meaning 'having being or 'given'. Technically it is referred to 'datum is' and 'data are' to highlight the plurality of its forms. Hence, *spatial data* can relate to things on which conceptualization, analysis and inference are based to understand the real world phenomena.

*Information* entails significance to collective data. It is a description of the meaning of data. It could be noted that information for one researcher can be data for another. The informational

aspect hence becomes an important aspect to understand the nature of data. The illustration given in the figure 2.1 can help understand this.

Sample Site	Measurement Quantity
A	4
B	8
C	6
D	2
E	3
F	5
Total	28

Datum

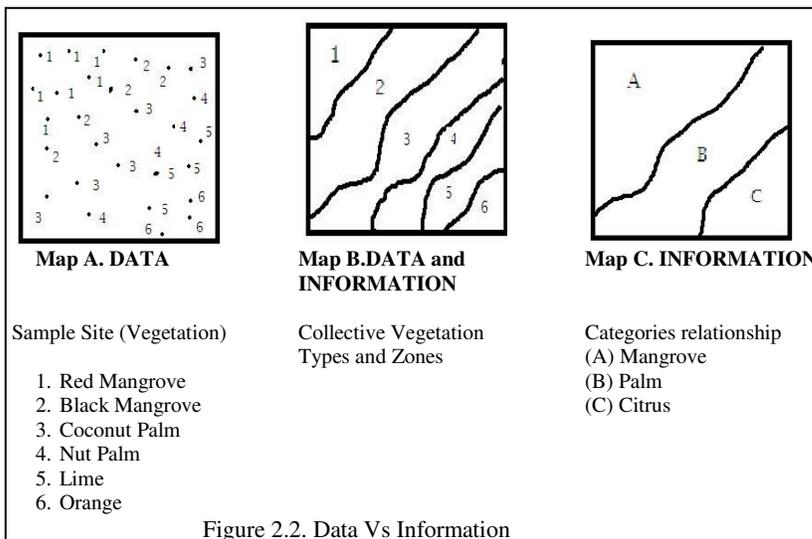
Data

Information

Figure 2.1: Data, Datum and Information

The characteristics of information which can be easily distinguished describing where things are using location or reference system; relationship between those locations which represent spatial interactions; qualitative and quantitative description of associations forming pattern or form of phenomena can be grouped as *Spatial Information*.

The difference in the term would be much easily understood with the help of illustration in figure 2.2. Map shows the study area representing samples with single observation/measurement. These points refer to one type of details in this case vegetation type found in the area. These are referred to as ‘data’. Now, look Map B, now the data has been combined to provide detail about the area building vegetation zones, this is ‘information’. The zones were identified plotting line separating one vegetation type from the other on the basis of corresponding data. Further in Map C, data and information can be combined to illustrate zone of vegetation as Mangrove (Red and Black Mangrove), Palm (Coconut and Nut) and Citrus (Lime and Orange). This can be both data and information depending upon end user.



**Check Your Progress I**

Q1. Define Data

.....  
 .....  
 .....  
 .....

Q2. Differentiate between data and information.

.....  
 .....  
 .....  
 .....

**Types of Data:**

All facts and figures collected for specific purposes can be grouped to several categories. On the basis of method of collection of data they can be grouped as *Primary data* (data collected directly from the source), *Secondary data* (data which has already been collected and currently made available). Data can also be grouped on the basis of their characteristics like *Categorical data* (representing character, like name, gender, age, language etc.) and *Numerical data* (represented by numeric values, like age, population numbers, income etc.). Data and Information related to specific location on Earth surface is generally identified as *Geospatial data* and can be grouped into two broad categories for its storage, analysis and manipulation. They are: spatial and non-spatial data (Fig 2.3). Primarily these data sets are used to catalogue and create database for computer based application to be used further for data processing and analysis.

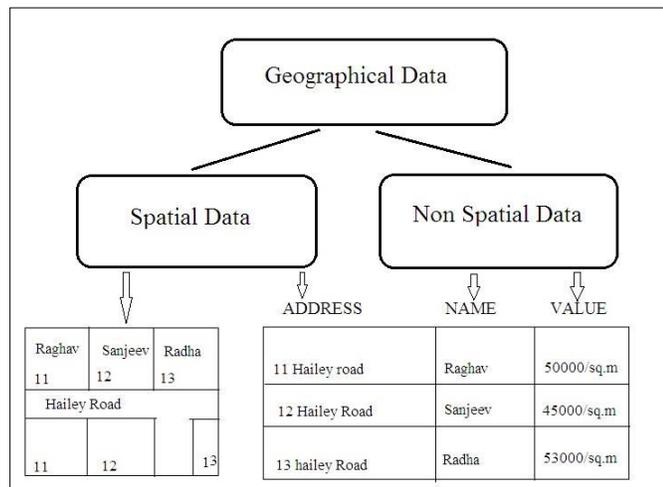


Figure 2.3. Types of Geographical Data

Geo spatial data can be identified with basic geographical structures and can be represented in the form of precise location, connectedness between locations and enclosed section of locations

in the space in reference to any theme of information and adding a label to it stating what they are and about its character. A detailed note on spatial and non spatial data follows in the preceding section.

### **Spatial Data:**

Spatial Data are data that seek connectedness to a place in the Earth. Dictionary defines spatial data as data that occupies cartographic (map able) space that usually has specific location according to some geographic referencing system (latitude/longitude) which enable them to be located in two-dimensional or three-dimensional space.

Spatial data defined by physical characteristics usually include location and position representing known location on earth.

Spatial data can simply give an address (precise location) and can give magnitude.

Spatial data are data /information about the location and shape of, relationship among geographic features which is generally stored as co-ordinate and topology (spatial proximity of object).

### **Characteristics of Spatial Data:**

The following can be enlisted as characteristics of spatial data:

- i) Position (location) is the starting point of measurement. The location is identified by the precision and accuracy of position normally using geographic co-ordinate system. Every spatial data must have reference to location on Earth surface.
- ii) Time is an important part of spatial data. The date of data becomes meaningful when temporal change is determined.
- iii) Spatial data also depict spatial characteristics in form of shape of features where dimensions like area and perimeter becomes significant.
- iv) Spatial relationship between and among features also becomes important where distance becomes a characteristic. Distance from one feature to the other through simple measurements describes proximity, nearness or connectedness in spatial relations.

### **Using Spatial Data for Geographic Analysis:**

Geographic analysis allows us to study and understand the real world processes. The method of representation of spatial data is central to its analysis as it enables the user adopt models to analyze, describe and map the real world phenomena. Computer based operations tools like GIS (Geographic Information System) enhances the process of spatial analysis combining meaningful sequences to reveal new or unidentified relationship between datasets which help better understand the real world phenomena. The scope of spatial analysis ranges from simple query about spatial phenomena to complex combinations of original and derived data sets.

**Data Forms** - The computer based aided tools recognizes three data forms to represent the spatial data as represented in Figure 2.4A and 2.4B: **A Point** indicating specific location of feature. It can also represent non physical entities like address of a location or point location of an accident. It is shown by convenient visual symbol on map viz. a dot or X mark. To be precise in dimension it should be noted that it do not have real length or width depicted for the feature. **A Line** depicting linear feature. It is one dimensional meaning ‘length only’. It has a beginning and an end. It can also be seen as line joining two point locations, e.g. roads, canal, river etc. or administrative boundaries. **A polygon** which is a two dimensional feature and gives spatial information its magnitude. It is an enclosed area comprising at least three sides. Concepts like area; perimeter becomes functional and adds detail to its analysis.

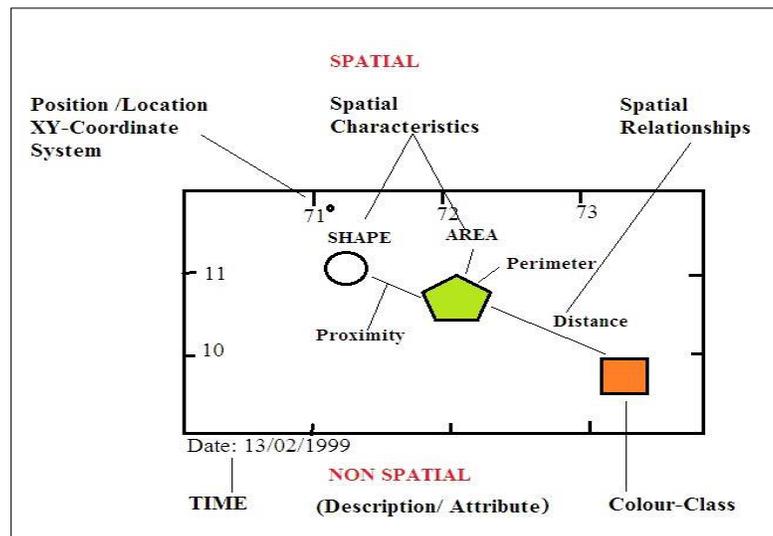


Figure 2.4 A. Data Forms

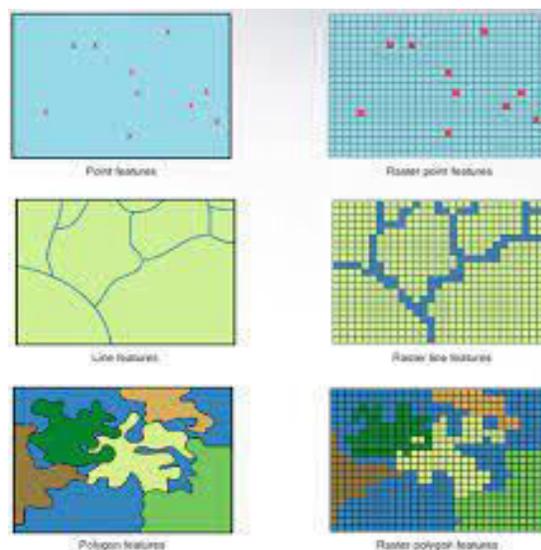


Figure 2.4B. Data Forms in Computer based Geographical Analysis

**Conceptual Model** - The spatial data is being organized and processed within GIS as objects, network or field respectively.

*Object based model* are preferred for the entities having well defined boundary. It can be studied as individual phenomena provided it can be separated conceptually from the neighboring phenomena as discrete entity e.g. river, building, forest, utility centre, roads etc. It can also be evaluated having specific relation with other objects.

*Network based model* also subsets of object based models but the emphasis is on the specific characteristics of interaction within and across multiple objects. The discrete flow of connectivity is important rather than shape of the phenomena e.g. flow of gas pipeline, air traffic route or sea navigation routes.

*Field models emphasize* phenomena that have continuous variable across some region of space. This may represent two either three dimensional extent e.g. air pollution extent, direction of wind flow etc.

**Data Structure** - In order to store and display data in computer the data structure are framed and data models are created. There are two models or data structures adopted for representing the spatial data in GIS: raster data and vector data. Raster and Vector data structures are way of defining spatial data in the computer.

- i) *Raster Spatial Data Model* –The most commonly adopted structure of data is the grid cell tessellations which regard space as unit of tessellation in a grid. Raster data structure represents the real world phenomena as a matrix of grid cells. Each cell in the grid has unique identity usually a code number which refers to a specific attribute measure, e.g. specific vegetation type in a forest land use, amount of rainfall at a station or its elevation. It should be noted that the single value in the given space would represent specific criteria and the overall representation of the landscape would include several of such codes to represent its varied characteristics.

Raster model also uses layered approach. Each layer indicates a specific theme and value of individual cell in a layer represents categories of classes within that theme. Each cell is also known as pixel (picture element). The size of the pixel identified as number of matrix division of that particular layer relative to the depicted feature of interest measures effectiveness of its representation.

- ii) *Vector Spatial Data Model* – When the object is likely to be represented as accurately as possible occurring in the real world, vector data models are used. Vector features are defined by ‘co-ordinate’ points. The term co-ordinate means the X-Y plane of reference where the position can be defined precisely. These plane surfaces refer to latitude and longitude in the spherical co-ordinate system.

Vector data model treat phenomena as sets of composed spatial entities each defined precisely by a set of coordinates. A vector point is expressed as single X-Y coordinate position and is represented by a dot or any other symbol for visual convenience. A vector line has two nodes with a specific beginning and ending vector points. A straight line would have no vertex whereas a complex line would have vertices with X-Y coordinate pairs. When an arrangement is as such that there are set of pairs of X-Y coordinates on the boundary and there is same point as the beginning and ending node it makes it a self enclosed line, this represents a polygon (Figure 2.5).

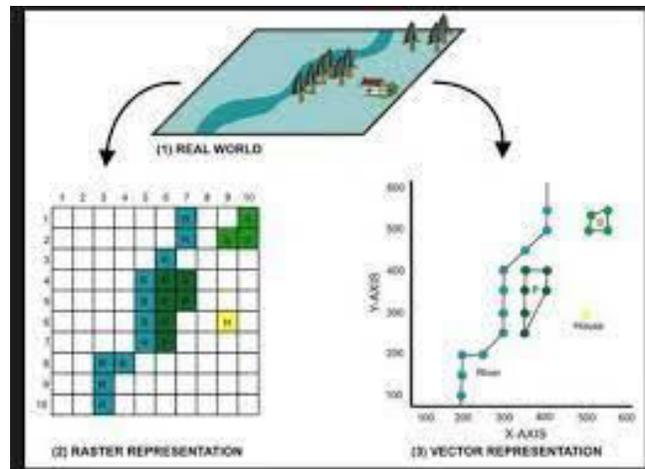


Figure 2.5. Raster and Vector representation

### Choice between Raster and Vector:

The raster and vector methods to represent spatial data structures are mutually exclusive. As seen in figure the storing and display of the spatial data have different mode in both the representations but the choice of the method would certainly depend upon the identified real world problem and spatial analysis.

#### Raster methods

##### *Advantages*

- Simple data structure
- Computation and spatial analysis is easy  
reduce data volume reduces spatial
- Technology is cheap

##### *Disadvantages*

- Volumes of graphic data
- Use of Large cell to  
accuracy of representation
- Network linkages difficult to establish

Vector methods

*Advantages*

- Good representation of data structure
- Accurate graphics with network linkages
- Updating and generalization of information is possible

*Disadvantages*

- Complex data structures
- Simulation is difficult as each unit has different topological form
- Expensive Technology

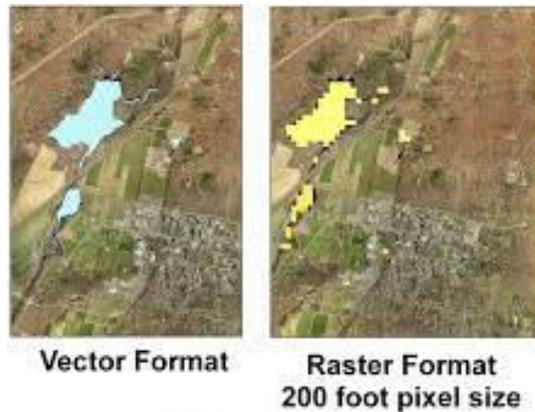


Figure 2.6. Vector and Raster Data Images

The problem of raster or vector data structure choice disappears once it is realized that both are valid method of spatial data representation and both structures are inter-convertible. But some of the uses in terms of best representation of spatial data can be enlisted as:

- Vector data are best suitable for soil type, land use and digital terrain mapping.
- Network Analysis such a communication and transport network is best represented by vector spatial data model
- Raster data structure is chosen for quick map overlay, map combinations and spatial analysis.

**Check your Progress II**

Q. What is Spatial Data?

.....

.....

.....

Q. Discuss different characteristics of Spatial Data.

.....  
.....  
.....

Q. What are different data structures used to represent spatial data in a geographical analysis?

.....  
.....  
.....

**Non- Spatial Data:**

Non spatial data is also known as attribute data. An attribute is the description of a feature that defines the spatial data. It does not account the geometric considerations. There are many forms of non spatial data including text description, numbers indicating quantities of some sort, codes or short description of character etc.

An illustration would help in understand more. A non spatial data is generated by asking common generic questions which are exclusive of spatial information. E.g. In a city where every area is coded with a ward number, a simple form of non-spatial data would be the query about which specific land use type it belongs, population and what the land value within that land use type is. Non spatial data is independent of the location based identity of features. As in the example cited above the description of land use, land values, population are not dependent on their location identities.

**Types of Non-Spatial Data:**

Attribute data/ non spatial data can be explained in terms of their qualitative and quantitative characteristics.

*Qualitative Non Spatial data-* The data obtained in this category do not have any numeric description. They are devoid of any measurement and magnitude. A name, explanation, labels serves as description and letter or number codes are proxy to word description and do not poses any mathematical meaning. These have no role in any statistical analysis and averages of numeric scores are meaningless.

As the illustration shows the map a shows a classification of high, medium and low with numerals 1, 2 and 3. But these numbers are mere cumbersome codes and have been used to represent the characters in the legend which otherwise would have taken much space and would look visually messed up. Similarly map B with codes 1, 2 and 3 represents city names which otherwise was difficult to write on map.

*Quantitative Non Spatial Data* – With reference to the nature of non spatial data having no mathematical meaning, the numbers refer mere to measurements of the magnitude of the feature which they represent. As given in the illustration map C the area depicts land value and map D refers city population for the point location.

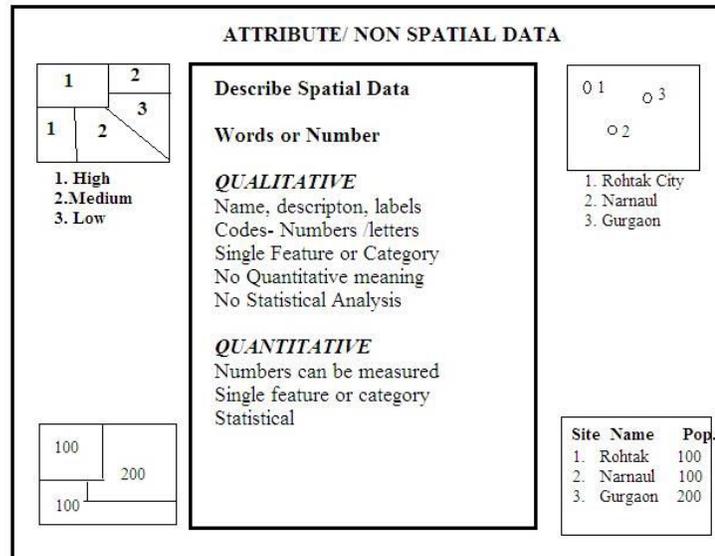


Figure 2.7. Non Spatial Data Characteristics

Hence data serves as raw material from which the information base is built up. They are collected and amassed into records and files. A database is of vital importance as it is collection of data which can be further used by different users. They are structurally organized and the categories include quantitative and qualitative data sets.

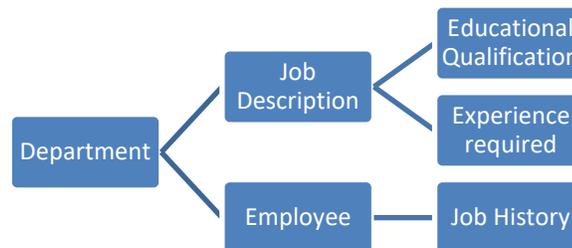
### Using Non-Spatial data for Geographical Analysis:

The data are vital tool of any analysis. The size of data becomes an asset and utmost care has to be put in to ensure its non redundancy, loss or damage. Data Base Management System (DBMS) can be referred to as a tool for representing real world oriented model of data on computers. The entire process of data entry, its classification, abstraction and representation are associated with it.

**Data Storage** - Non spatial data stored in GIS are known as attribute tables. The row in the table represents a special feature and broad characteristics are represented by column. Technically the row is called a record or tuple where as column is depicted as field or item. Queries to the database (finding desired dataset in the computer) require database management software to find the named data or classes of data items. Hence it is necessary to arrange the data so as the entities and attributes are based on some conceptual models of arrangement of data in a set format of structure so that the retrieval becomes an easy task. This theoretical foundation helps in storage, organization and manipulation of the datasets. The following data models are generally used for non spatial information based on the analysis required. :

## Data Models -

*Hierarchical Data Models* – Based on tree structure relationship where one too many concepts are implied. Common generic GIS questions of what, who, where and how can be asked and the data retrieved can be evaluated to show the connection. It composes of hierarchy of nodes (entities of data) where each lower node is connected to the primary node called root/parent.



*Network Data Model*– Relies on the principle that an item in the data set can be linked to any other item. Each entry data set is classified as node and relationship sets are seen as linkages by using pointers and the relationships can be one too many and many too many. The generic question pertains to analysis of patterns and relationships.

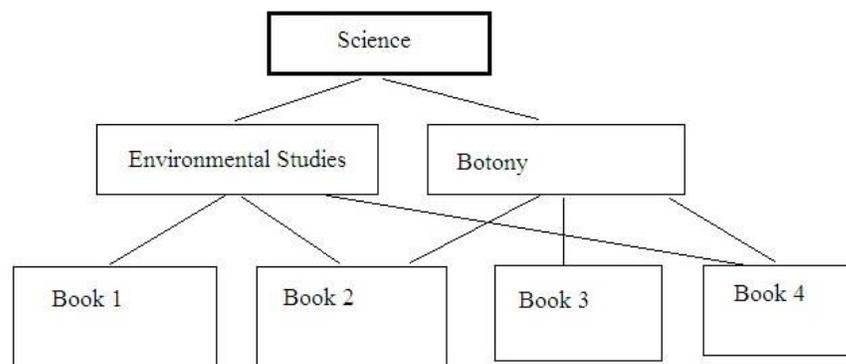


Figure 2.9 Network Data Model.

*Relational Model* – This model is based on the design to relate one set of data with another. Dataset are chosen from one field which meets the condition and it is then moved to the next field. In this type, data are organized in two dimensional tables which are easy for users to both develop and understand. The relation can further be described mathematically.

*Objects Oriented Model* – The data recognizes object as classes of real world object and uses the additional information to describe the object through attribute, procedures or method which operate on them. It uses messages to send to the object to identify them depending on the property e.g. object identifier would send message to inquire for co-ordinates, area, perimeter

for a polygon which further can be grouped in classes or new ones may be created which has combinations of sub classes.

**Check Your Progress III**

1. Explain types of non spatial data and give suitable example.

.....  
 .....  
 .....

2. Identify spatial and non spatial data items

- (i) Phone book .....
- (ii) Recipe book.....
- (iii) Crime hot spot for police patrol routes.....
- (iv) Unusual warming of Pacific Ocean.....
- (v) Best schools in locality.....

---

**2.4 SUMMARY**

---

In this unit you have learnt the following:

- *Geographic Data* refers to any data on Earth surface or near Earth surface and *Geospatial Data* has much precision in terms of reference to location on Earth surface and/or near Earth surface, e.g. latitude/longitude, GPS location.
- *Data* and can be grouped into two broad categories: Spatial and Non-spatial data. Spatial Data are data that seek connectedness to a place on the Earth and that occupies cartographic (map able) space. It has specific location according to some geographic referencing system (latitude/longitude) which enables them to be located in two-dimensional or three-dimensional space. Non spatial data is also known as attribute data. An attribute is the description of a feature that defines the spatial data. It does not account the geometric considerations.
- Both spatial and non spatial data are required to understand phenomena on space and are used in geographical analysis.

---

**2.5 GLOSSARY**

---

- Spatial data- is used to describe any data related to or containing information about a specific location on the Earth’s surface.
- Non spatial data- is data that is independent of geographic location.

---

## **2.6 ANSWER TO CHECK YOUR PROGRESS**

---

Q1. Explain different types of Geographic Data and give a detail account of its characteristics.

Q2. Differentiate between data structures of spatial data and discuss its advantages.

---

## **2.7 REFERENCES**

---

- Fischer, M. and Wang, J (2011) Spatial Data Analysis: Models, Methods and Techniques, Springer Publication, USA.
- Chou, H.Y (1997) Exploring Spatial Analysis in Geographical Information, Onward press, USA
- Davis, E. B (1996) GIS: A visual Approach, Onward press, USA.
- Dalamagas T., Sellis T., Sinos L. (1998) A Visual Database System for Spatial and Non-spatial Data Management. In: Ioannidis Y., Klas W. (eds) Visual Database Systems 4 (VDB4). VDB 1998. IFIP — The International Federation for Information Processing. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-35372-2\\_6](https://doi.org/10.1007/978-0-387-35372-2_6)

---

## **2.8 TERMINAL QUESTION**

---

Q1. What is data and how is it different from information?

Q2. Write a detail note on non spatial database for geographical analysis.

---

## **UNIT 3 - TOPOLOGY CREATION AND DATA QUERY**

---

**3.1 OBJECTIVES**

**3.2 INTRODUCTION**

**3.3 TOPOLOGY CREATION AND DATA QUERY**

**3.4 SUMMARY**

**3.5 GLOSSARY**

**3.6 ANSWER TO CHECK YOUR PROGRESS**

**3.7 REFERENCES**

**3.8 TERMINAL QUESTIONS**

---

### **3.1 OBJECTIVES**

---

After reading this unit you will be able to know about:

- i. Spatial relationship between different entities.
- ii. Errors associate with point, line & polygon.
- iii. Importance of topological file format.

---

### **3.2 INTRODUCTION**

---

The study of geometric properties that do not change when forms are bent, stretched, or undergo similar transformations is known as topology. Because the list of neighbors to any given polygon does not change during geometric stretching or bending, polygon adjacency is an example of a topologically invariant property. Topology deals with spatial properties that do not change under specific transformations: a) The relationships between neighborhoods continue and the boundary lines have both the beginning and end nodes & b) The areas still are bounded by the same borders, only their perimeter shapes and lengths have changed.

In the topological data files, topological relationships such as adjacency and connectivity are explicitly recorded. These relationships can be recorded independently of the coordinate data and thus do not change when the data is stretched or bent, as when converting between coordinate systems. Topology is the mathematics branch used to determine spatial connections between entities (ESRI, 1999). Topology is a specific component of the vector representation model. Topology is said to be present in a vector layer if it contains the spatial relations between its features. Topology is required for certain analyses and alters how some GIS operations, such as geometry editing, operate. GIS transmits information through graphic symbolization (points, lines and polygons), and mathematically retains relationships through topology concept. For example, you can easily identify crossing streets and adjacent properties when you stand on a hill and look into the countryside. To identify these links, the mathematical logic used by a computer is topology. Topology can be stored as a topological data model (geometric data correction), but topology can also be used for non-topological data analyses. Creating and storing topological relationships have variety of benefits: a) Data is efficiently stored to allow fast processing of large datasets, b) Enables the computer to quickly determine the spatial relations of all characteristics and analyze them, c) Ensure that data is geometrically correct, d) Improved data quality - detects and corrects digitizing errors and validates data to ensure accuracy & f) Carrying out some types of spatial analysis (selections, network analysis)

### 3.3 TOPOLOGY CREATION AND DATA QUERY

In geographic data, topology assists us with the close spatial relationship between different entities or nearby features. Geometric relationships between spatial entities and their attributes are critical for spatial analysis and integration in GIS. (M. Anji Reddy, 2008). Because the structure is included in the data model, a single line can represent the common boundaries to indicate which side of the line belongs to which polygon. Lists are used to express spatial relationships (for example, a polygon is defined by a list of arcs that make up its boundaries). Although most vector layer operations can be performed without topology, some, such as network analysis, cannot. If we consider a roads layer, there is no way to build a network from it if it only contains lines representing roads but no information about how they are connected. The points at which lines intersect may be crossing surround about (allowing movement from one road to another), but they may also be points with no connection between the roads (one passing above the other). We are missing information if we do not know this, and the network analysis cannot be performed. Topological relationships are constructed by combining simple elements with more complex elements:

- Nodes define line segments
- Line segments connect to define lines
- Lines define polygons

Figure 3.1 depicts the following spatial relationships: disjoint, meets, equals, inside, covered by, contains, covers, and overlaps. What are the applications of spatial relationships? These relationships can be used in queries on a spatial database. Topological relationships can also be used to ensure to pological consistency of space.

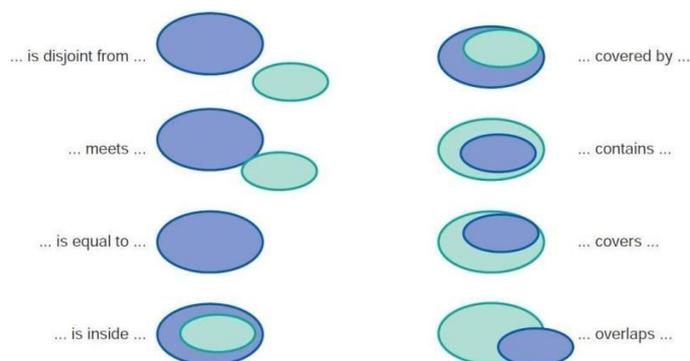


Figure: 3.1. Spatial relationships between two regions derived from the topological invariants of intersections of boundary and interior.

Topology in the context of spatial data is made up of three components: adjacency,

containment, and connectivity. The geometric relationships that exist between area features are described by adjacency and containment. Containment is a subset of adjacency that describes area features that are entirely contained within another area feature. These three topological relationships will make certain of the following:

- There is no duplication of nodes or line segments.
- Line segments and nodes can be linked to multiple polygons.;
- Each polygon has a unique identifier.;
- and
- Polygons representing island sand holes can be adequately represented.

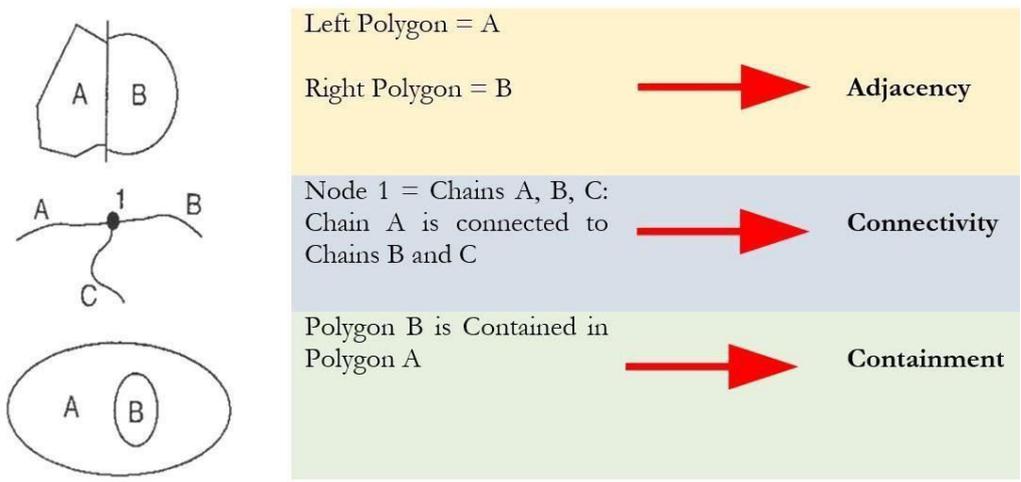


Figure:3.2.Topologicalspatial relationships

**Connectivity:**

Connection is a geometrical property used to describe the connections between line functions, such as the road network. You can connect to the airport; connect the rivers to the streams, or take the water treatment plant to a house. You can find a routing route. This is the basis for many operations for network tracing and tracking. The arc-node data structure has arcs connecting to each other at nodes. The two arcs have a Node from which the arc begins and a Node to which the node ends. It is called the topology of the arc-node. By searching for common node numbers, connected arcs are determined by the list. Arcs 1, 2 and 3 all intersect in Figure 3.3, since node 11 is shared. Arc 1 can be passed to arc 3 due to its common node meeting at 11. On the other hand nodes starting from arc 1 to arc 5 are not in a state to turn either of its direction due to absence of common node. For network analysis, connectivity answers which line segments are connected?"

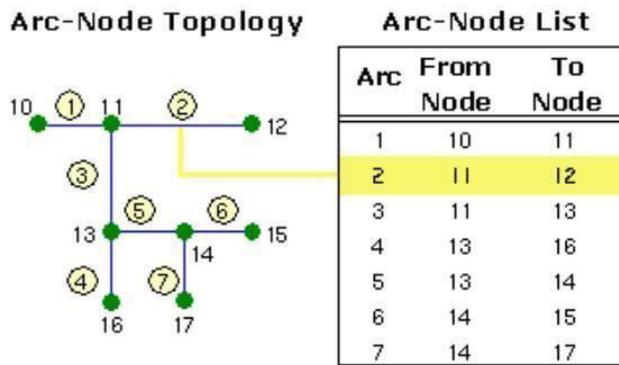


Figure3.3: Arc-Node Topology example

**Are a definition/containment:**

Containment is an adjacent extension that describes area characteristics that can be wholly included in another area feature. For example, the inner limits (or hole) of the polygon are defined on an island. An island describes about inner portion of a polygon of the vector model. The arc node structure reflects polygon basically through arc list instead of closed loop of set of X & Y values. Polygons are an ordered arc list instead of a closed loop, known as polygon-arc topology, of (x,y)co-ordinates.

Polygon F is comprised of 8, 9, 10 and 7 arcs in Figure 3.4. (The 0 arc creates an island in the polygon before seven indicates). Each arc is represented by two polygons (in the illustration below, arc 6 appears in the list for polygons B and C).The arc co-ordinates are only stored once because the polygon is simply a list of border arcs, reducing the amount of data and preventing border overlap between immediate polygons. Containment responses to "Which spatial characteristics are included in which?"

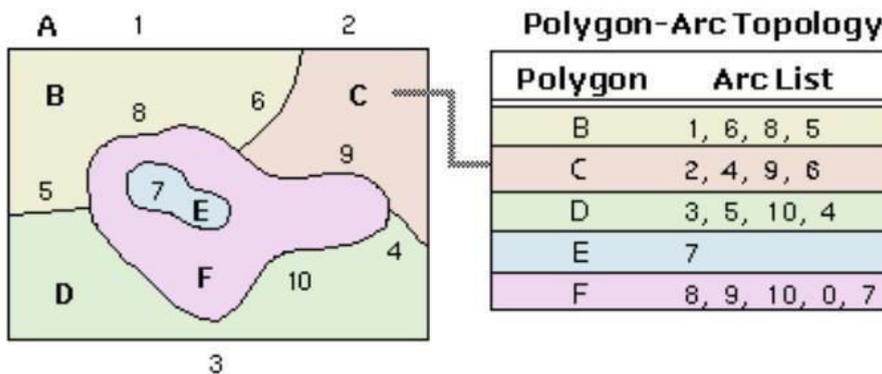


Figure: 3.4.Polygon-ArcTopologyexample

**Contiguity or adjacency:**

Contiguity is the topological concept that allows the vectors to determine the neighborhood of the characteristics that share a border. This is the basis for many surfacing operations in neighboring countries. When they share a common frontier, areas can be described as adjacent. The arc is defined by the from node and the to-node. The arcs have a right and left sides, so that the polygons can be determined at both sides. In Figure 3.5, polygon B is on the left side of Arc 6, and polygon C is on the right. Hence both the polygons located side by side. The polygon of the universe ensures every arc has the right and the left. "what polygons are contiguous to which polygon on ground?" and used for spatial analysis of the areal data.

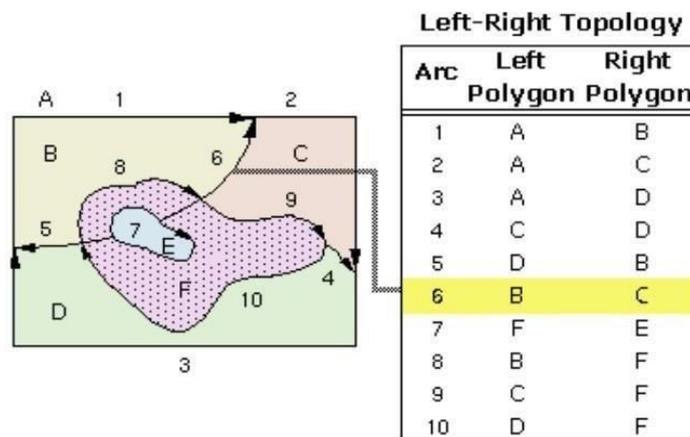


Figure:3.5.Topologycontiguityexample

**Rules of topological consistency:**

Topology is a set of rules that allows the geo-database to model geometric relations more accurately together with arrange of editing instruments and technique. There are certain dedicated rules designed to extract how spatial features share a space and also provides a great range of editing tools which work in a précised and assimilated way with geometry-specific features are applied to topology. One or more relationships in geometry are saved as topology that defines how the features share geometry in any individual or group of vector file. The topology features are still relatively simple classes – A topology describes how features can be spatially linked instead of changing the feature class definition. And what is done, especially to help provide information on data collection. In many cases it is also used to analyze the spatial relations of the Journal, so that the smaller of the decomposition of the topological graph may draw out in the mouth of the feature (polygons) with especially when neither the features a network of identical values. Journal of to be and by what means, in each portion it can be composed of different functional classes.

It is generally the inter relationship of different features among them which is shown to us through the proper implications of topological rules. The main function of all these topographical rules is to define the relationship between different objects drawn in the form of vector data. These rules are highly manageable through the geo database format and this format helps us to fix diverse errors associated with vector data. "Must not overlap" is a rule used to maintain the integrity of features in the same feature class, As an example, When two features' geometries overlap, they are highlighted in red (as the red overlap in nearby polygons and a linear segment of the following two lines show). For example, suppose you have two types of road properties: normal roads (which are connected to the other roads at both nodes) and hillside roads (which are connected to other roads at both nodes) (those which are at a dead end node). A topology rule may require that road features at both ends be connected to other road features, with the exception of roads that are of the Dead End subtype. Topology rules can be divided into three categories based on the type of geometry. These categories are as follows:-

- Point based topology
- Line based topology
- Area/Polygon based topology

**Point Based Topology:**

- **Must be coincident with (Point)-:** This rule is useful when point of one file compulsorily coincident with point of another class. This type of error are emerged when point of one file or its lower order file are not purely covered by points available in other files.

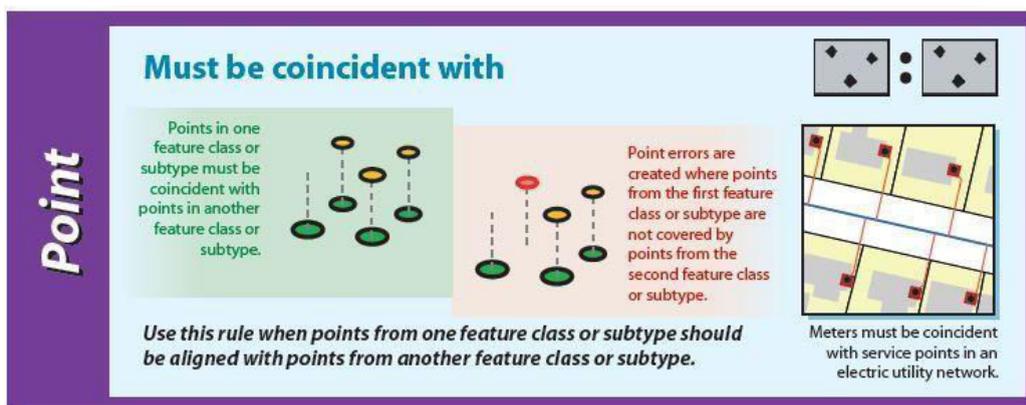


Figure3.6:Must be coincident with(Source-ESRI)

- **Must be disjoint (Point)-:** It is very fundamental with vector data that points of same feature class or its subtype need not to be overlap in any condition. Violating this rule will always create error in the database. It is very necessary to eliminate these errors by

making use of this rule.

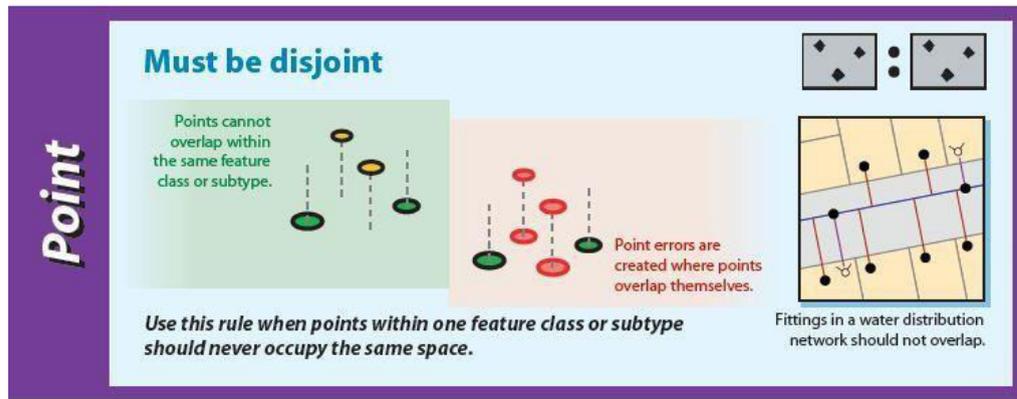


Figure3.6:Must be disjoint(Source-ESRI)

- **Must be covered by endpoint of (Point)-:** These errors are mainly created when point of feature class is not properly covered by the last point of same feature class or its subtype. This topological rule is useful to mark errors and to correct those errors which are not on the straight line.

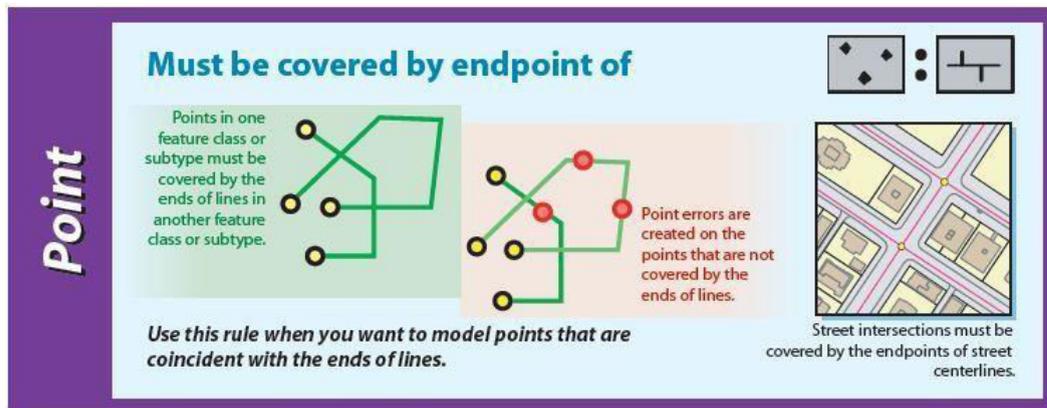


Figure3.7: Must be covered by end point of(Source-ESRI)

- **Point must be covered by line (Point)-:** This kind of rule is appropriate to eliminate errors when a point which lies outside any line feature or not covered by the dimensions of a line feature.

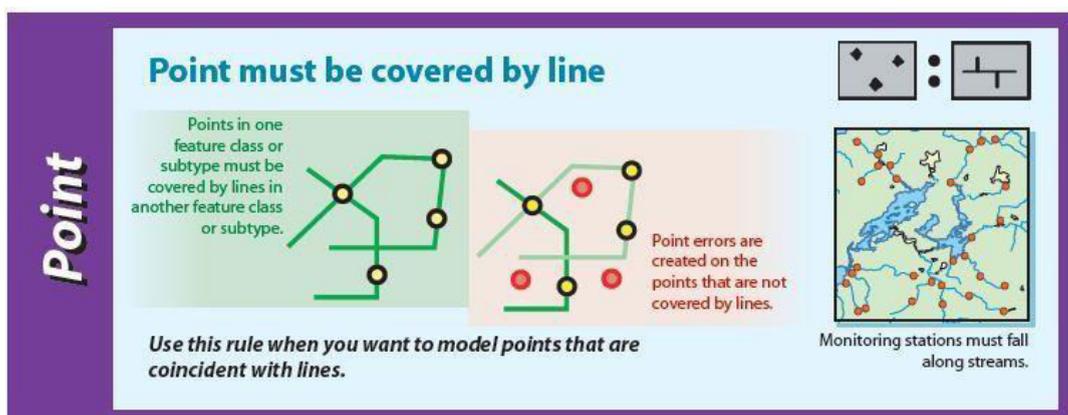


Figure3.8: Point must be covered by line (Source-ESRI)

- **Must be properly inside polygons (Points)-:** Each point must be in a polygon, if points are located outside any polygon then the particular case is considered as an error in vector data.

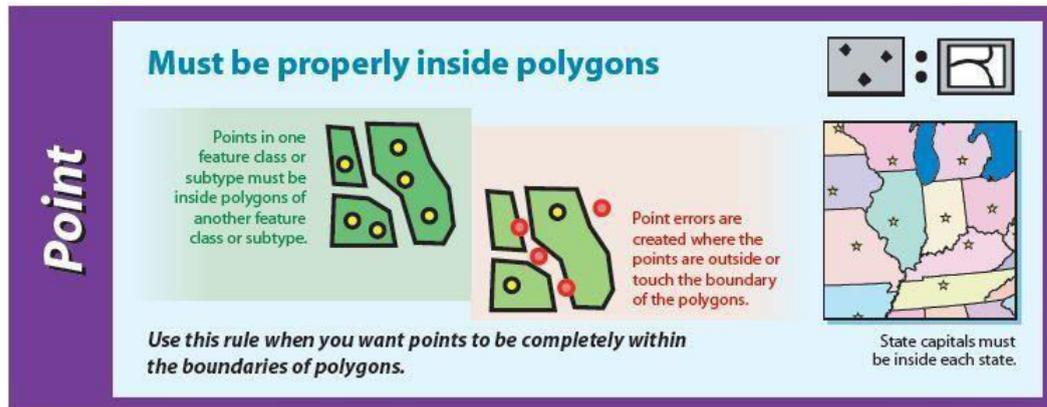


Figure3.9: Must be properly inside polygon (Source-ESRI)

- **Must be covered by boundary of (Point)-:** A feature class or sub-polygons types must intersect with points from another feature class or sub-type. When points must be on or within the boundaries of polygons this rule is quite useful.

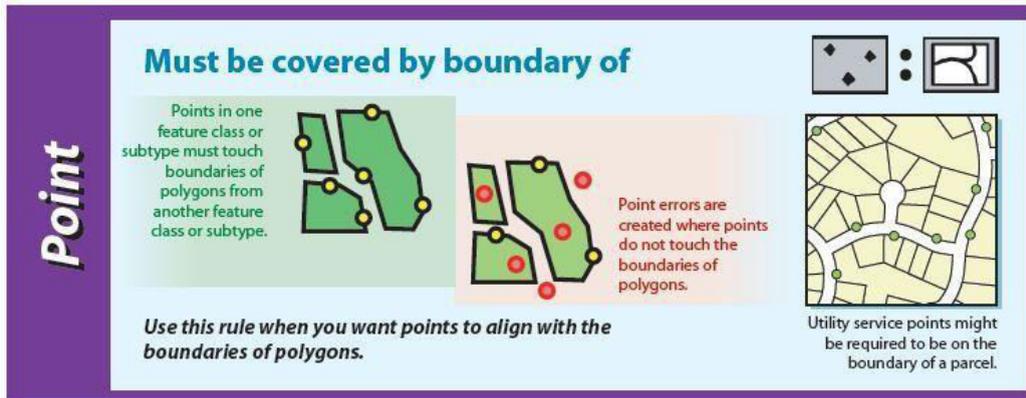


Figure3.10: Must be covered by boundary of (Source-ESRI)

**Line Based Topology:**

- **Must Be Larger than Cluster tolerance (Line) -:** The minimum distance between the vertices that make up a function is called the cluster tolerance. Cluster tolerances are used to determine which vertices match. This rule applies to all poly line feature groups and is required for the topology. When checking the topology, the reduced poly line function is an error. Properties that violate the law do not affect the law.

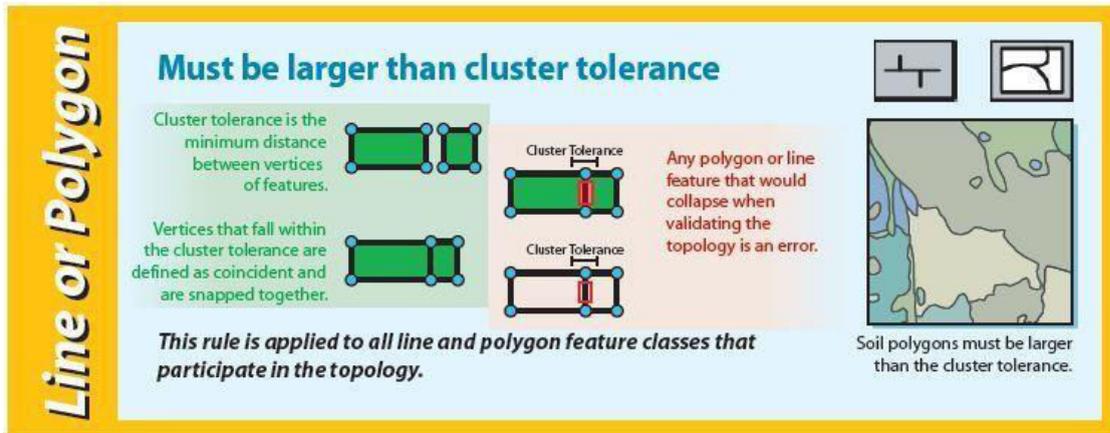


Figure3.11: Must be larger than cluster tolerance (Source-ESR)

- Must Not Overlap (Line)-:** This rule is very useful to identify the errors in which a line shares some of its length with other line or its sub type. Lines touch, intersect, and overlap with each other. This rule will be implemented to the scenarios where two of the lines will share the same space on ground in real. When two of the lines will overlap in any condition then this error will occur.

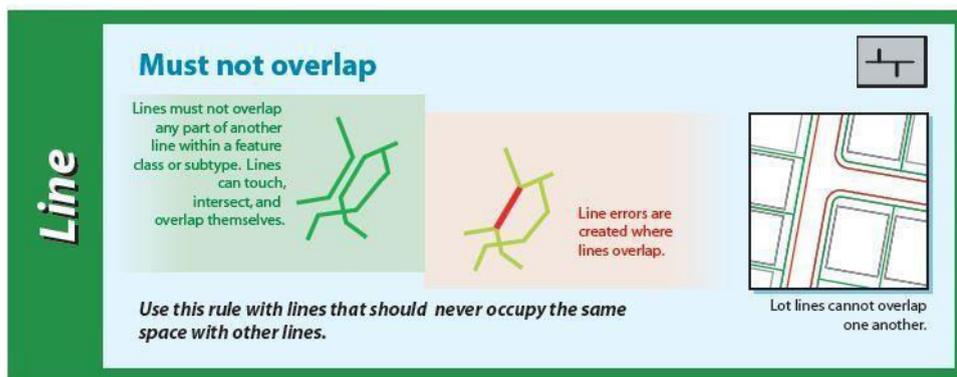


Figure3.12: Must not overlap(Source-ESRI)

- Must Not Intersect (Line)-:** Crossing and overlapping of one line with other line or any arc of same lines from same feature class or its subtype is generates this sort of error when working with line. This rule will be of highly use when any line or segments of any line drawn over a same space as other lines. Line errors occur when lines overlap and dot errors occur when lines cross.

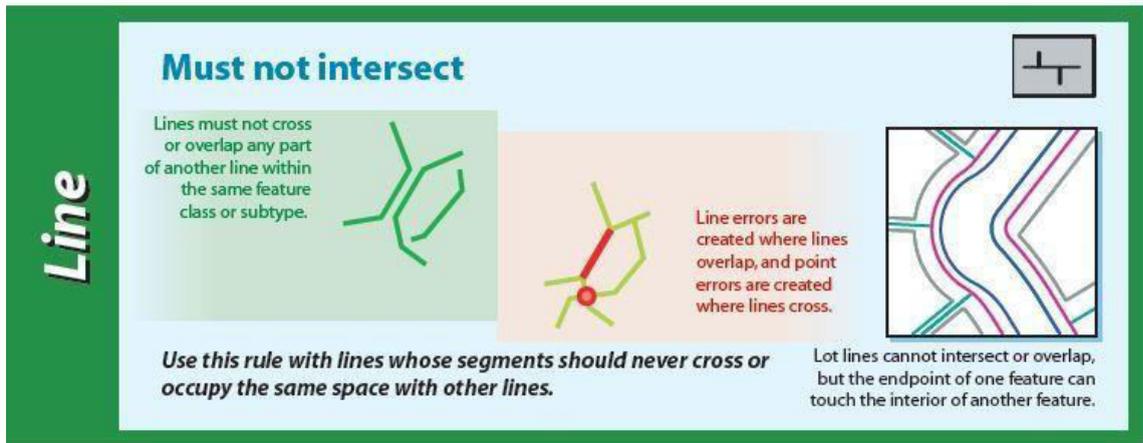


Figure3.13: Must not intersect (Source-ESRI)

- Must Not Intersect With (Line)-:** The lines of the same class or sub-structure shall not cross or surpass part of another line. For instance, if many lines cannot cross or cross, but one function can affect the internal function of another, use this rule with lines that never cross segment and occupy the same space as other lines. Line errors occur when lines overlap and dot errors cross lines.

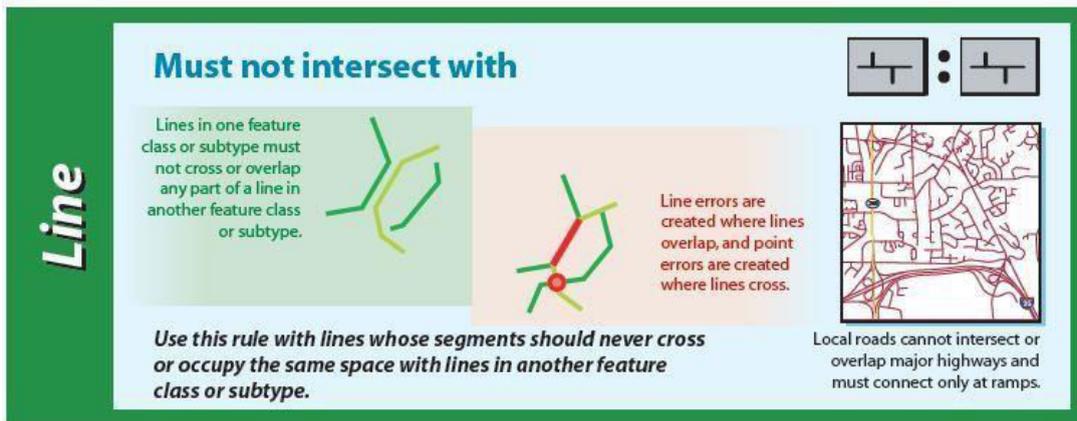


Figure3.14: Must not intersect with (Source-ESRI)

- Must Not Have Dangles (Line)-:** Any part or portion of a line within a single feature or subtype shall affect the end of a line. Use this rule for example, if lines are connected in a class of features or subtypes. In this example, for stretches ending in cul-de-sac or close dead ends, you can set exceptions to this rule. At the end of a line no other line or line will be affected by point mistakes.

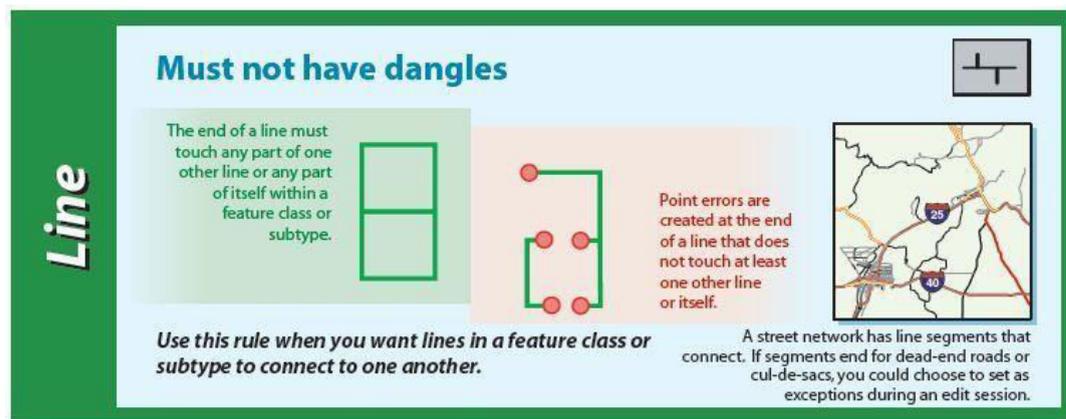


Figure3.14: Must not have dangles (Source-ESRI)

- **Must Not Have Pseudo Nodes (Line)-:** The end of a line cannot touch a different line in a class or subtype. Any part of itself may influence the end of a line. Use this rule for cleaning data on subdivided lines. Segment of a river system, for example, can be limited to ends or junctions of hydrological analysis. Point errors occur when the end of a line only affects the end of a line.

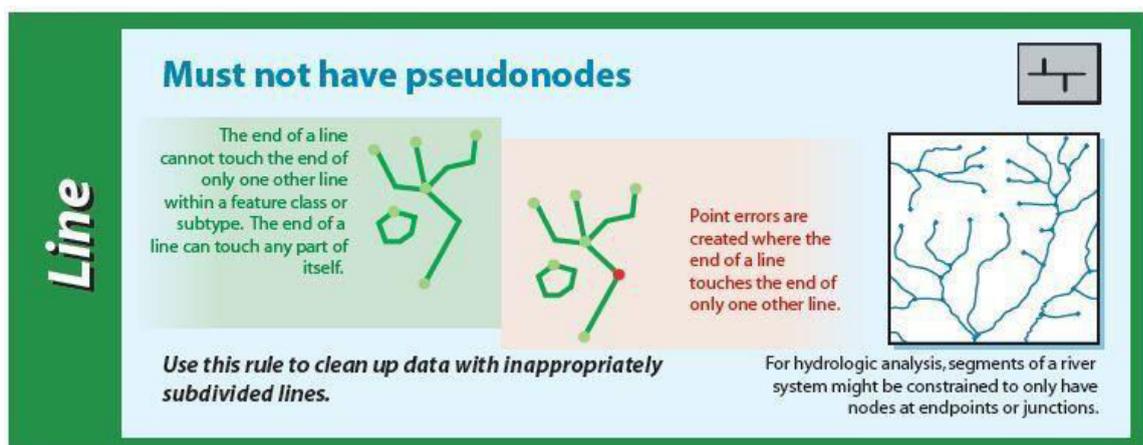


Figure3.15: Must not have Pseudo nodes (Source-ESRI)

- └ **Must not intersect or Touch Interior (Line)-:** Lines can touch their ends only and cannot overlap in a class or subtype of features. Use this rule only when touching the line at its ends instead of crossing or overlapping the line, for example if lots cannot cross or overlap the line and only connect at the end of each line function. Line mistakes occur when two or more lines overlap and when two or more lines cross or touch, dot errors arise.

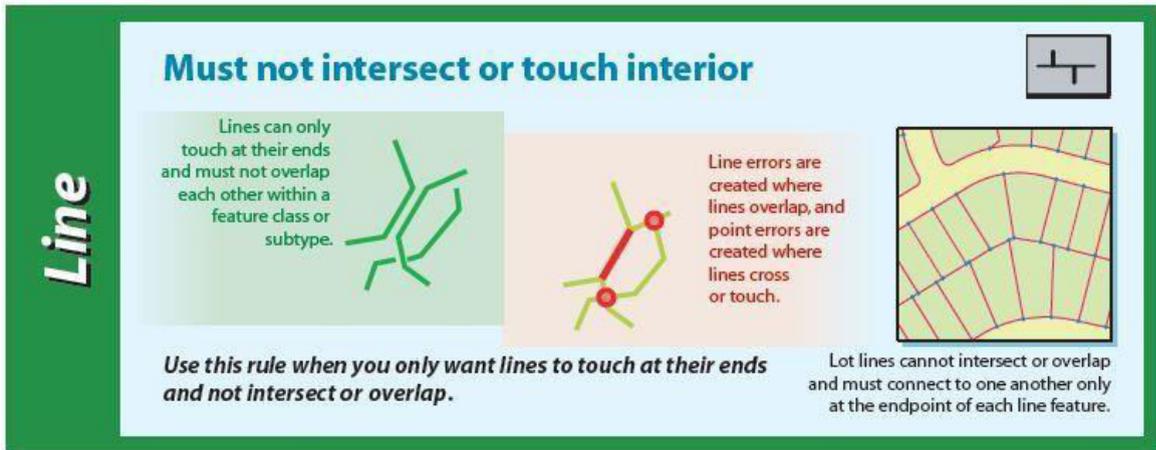


Figure3.16: Must not intersect or touch interior with (Source-ESRI)

- Must not intersect or Touch interior with (Line)-:** Lines in one class or subtype of functions may only affect ends and cannot overlap lines in another class or subtype of features. This rule should only be used when you want to touch a line at its end but not intersect or overlap it with any other feature class or subtype, such as when plot lines cannot intersect or overlap block lines. When two or more lines overlap, a line error occurs, and a dot error occurs when two or more lines cross or touch.

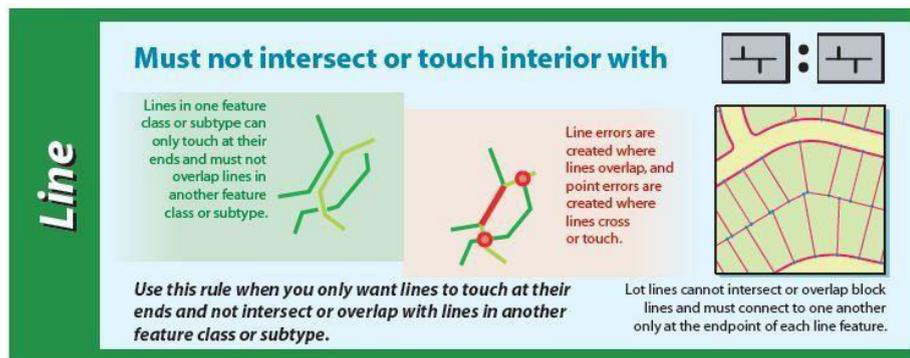


Figure3.16:Must not intersect or touch interior with (Source-ESRI)

- Must not overlap with (Line)-:** Lines of the same class or subtype of function shall not cover the function class or subtype of any other line. For instance, road segments cannot overlap on flat segments when roads cross and get near rivers. Line errors occur when the overlap between two feature classes and lines of subtypes. This rule is applicable to lines that should never share space with lines from another class or subtype of function.

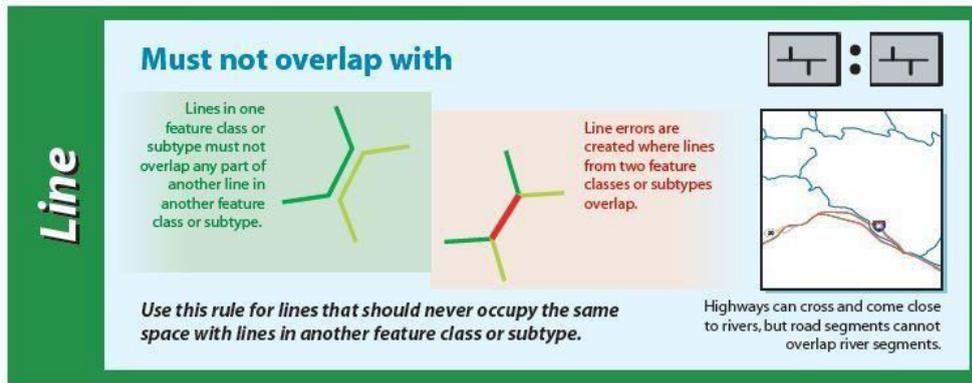


Figure3.17: Must not overlap with (Source-ESRI)

- Must be covered by Feature Class of (Line)-:** Lines in one class or subtype of feature should be lines in another class or subtype of feature. Use this rule when you have multiple lines describing the same geographic location, like when bus lines must be above road lines. Line errors must be generated by the first-class lines not covered by second-class lines.

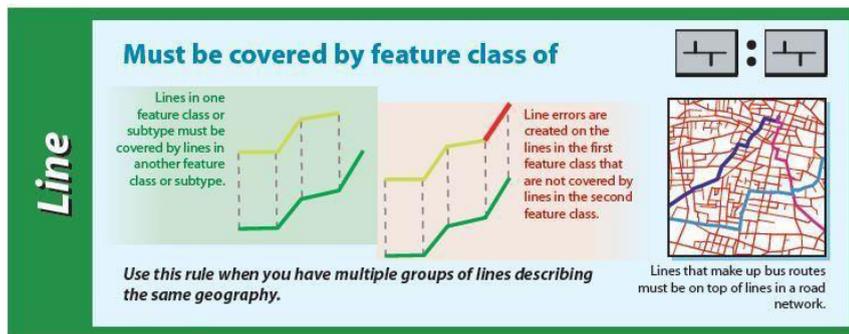


Figure3.18: Must be covered by feature class of (Source-ESRI)

- Must be covered by boundary of (Line)-:** The boundaries of lines in one feature class or sub type must match the boundaries of polygons in another feature class or subtype. Use this rule to model lines that match polygon borders, such as poly line characteristics used to display block and lot limits, to cover parcel limits. Please follow this instruction. Line errors occur when lines are not covered by polygon boundaries.

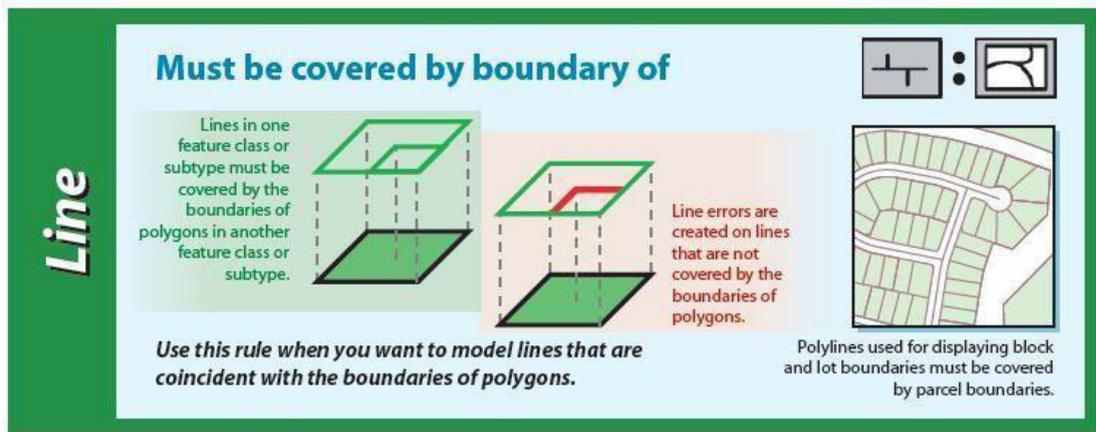


Figure3.19: Must be covered by boundary of (Source-ESRI)

- Must be inside (Line)-:** Lines of one feature class or subtype must be contained within polygons of another feature class or subtype. If lines are included within polygons, use this rule, for example, when streams are within watersheds .Lines that are not in polygons are created by mistake.

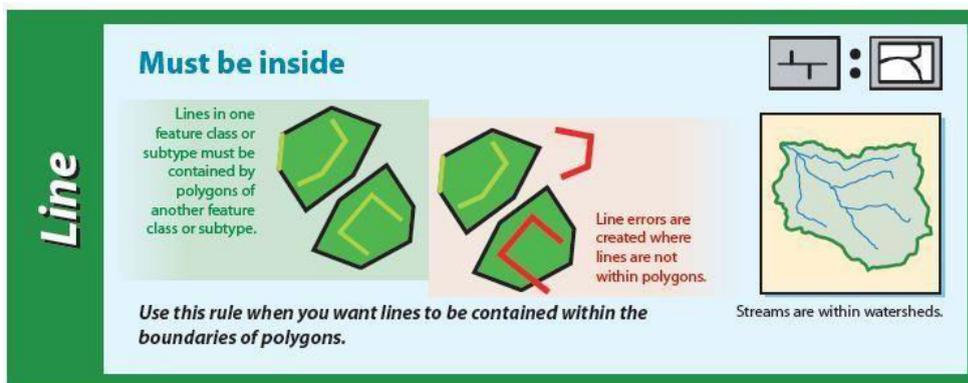


Figure3.20: Must be inside (Source-ESRI)

- Endpoint must be covered by (Line)-:** The ends of lines in one feature class or subtype must be covered by points from another feature class or subtype. When the endpoints of secondary electric lines must be capped by a transformer or metre, use this rule to model the ends of lines in one feature class or subtype that are coincident with point features in another feature class. Point errors happen when a line's end isn't covered by a point.

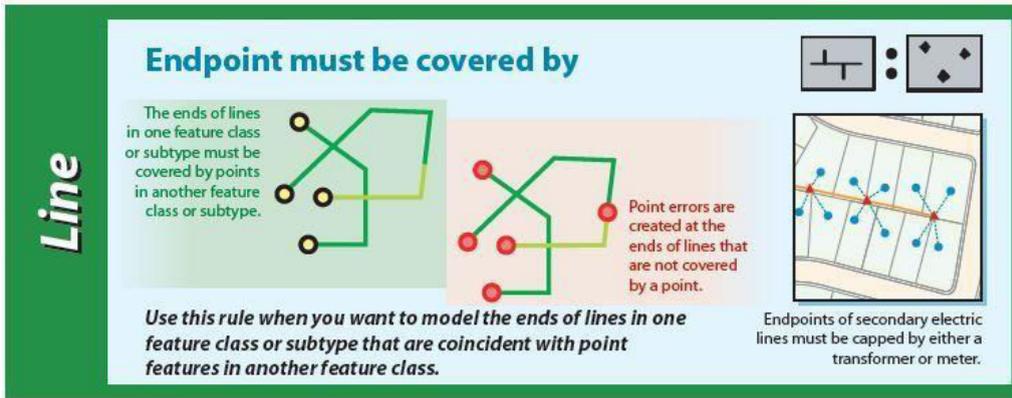


Figure 3.21: End point must be covered by (Source-ESRI)

- Must not self- overlap (Line)-:** Lines must not cross within a feature class or sub type? Lines from one feature class or subtype can intersect, touch, and overlap with lines from another feature class or subtype. This rule applies to lines whose segments should never be placed next to one another on the same line. For example, in transportation analysis, street and highway segments of the same feature should not be overlapping. When two or more lines overlap, a line error occurs.

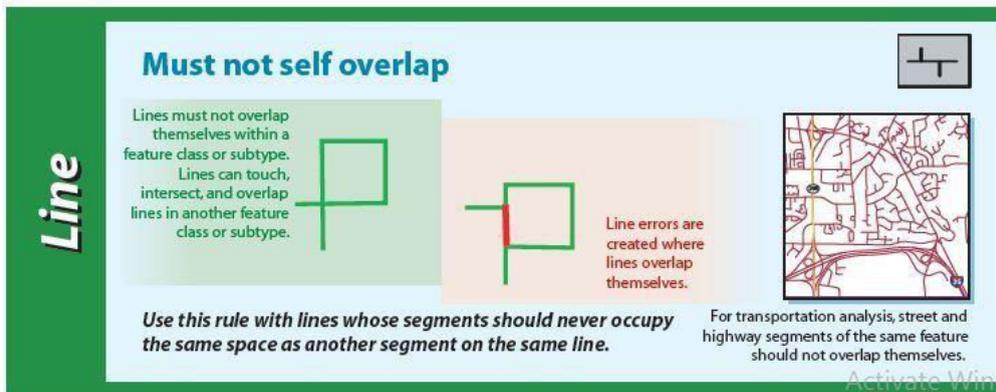


Figure 3.22: Must not self overlap (Source-ESRI)

- Must not self- intersect (Line)-:** Within the class or subtype of the feature, lines must not cross or overlap? Lines can intersect, cross, and overlap each other. This rule is used, for example, when contour lines cannot intersect or overlap on their ends without intersecting or overlapping. Where lines overlap or cross, line mistakes occur.

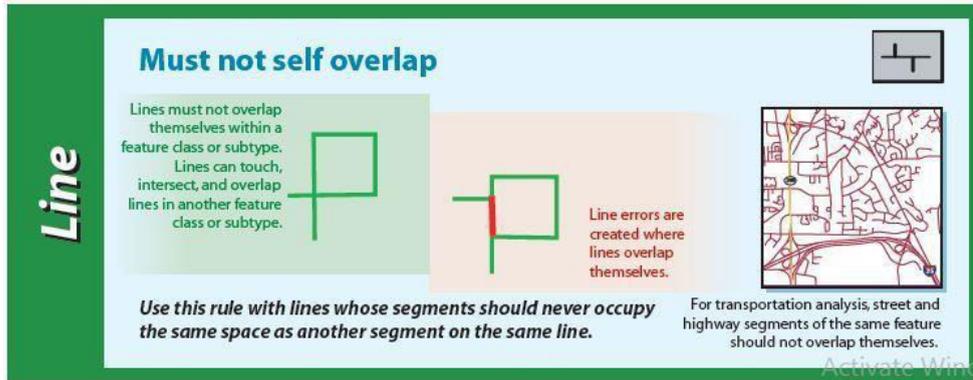


Figure 3.23: Must not self overlap (Source-ESRI)

- **Must be single part (Line)-:** Within the feature class or subtype, lines must only have one part. Use this rule if, for example, a highway system is made up of individual features, each of which is made up of only one part.

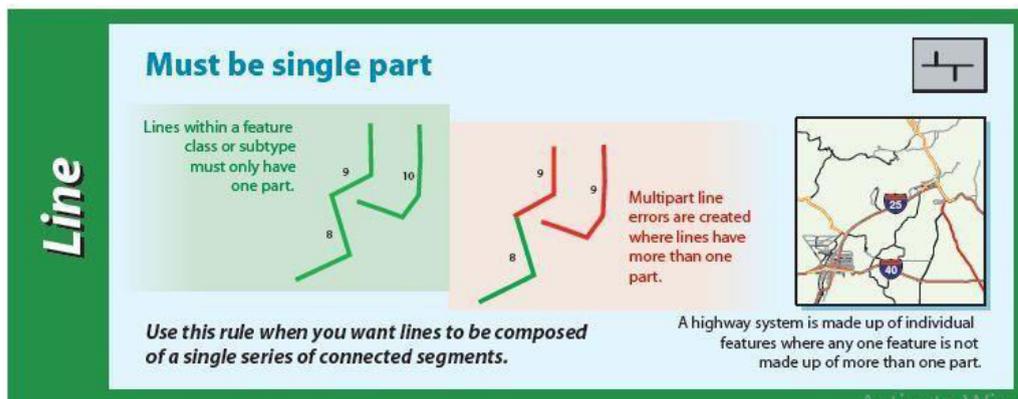


Figure3.24: Must be single part (Source-ESRI)

**Polygon/Area Based Topology:**

- **Must be larger than cluster tolerance (Polygon)-:** The tolerance of a cluster is the minimal distance from the vertices of a feature. The coincidences of vertices falling within the cluster tolerance. This rule applies to all classes of polygons. This rule is required for a topology.

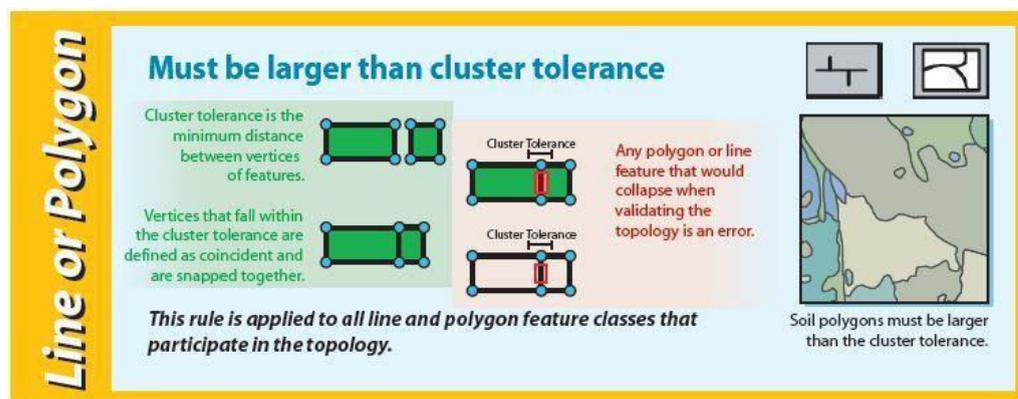


Figure3.25: Must be larger than cluster tolerance-Polygon (Source-ESRI)

- Must not overlap (Polygon)-:** Within a feature class or subtype, polygons must not overlap. Polygons can be linked together, either at a point or along an edge. This rule ensures that no polygon feature in the same feature class or subtype overlaps another polygon feature in the same feature class or subtype, such as when administrative boundaries such as ZIP Codes or voting districts, or mutually exclusive area classifications such as landform types, cannot overlap. When polygons overlap, polygon errors occur.

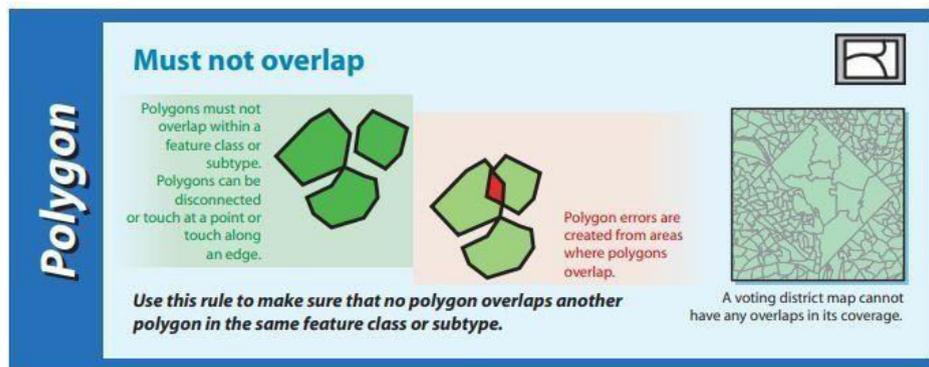


Figure3.26: Must not overlap (Source-ESRI)

- Must not have gaps (Polygon)-:** There must be no gaps between polygons within a feature class or subtype. Use this rule when all of your polygons must form a continuous surface with no voids or gaps, such as when soil polygons must form a continuous fabric with no gaps or voids. Line errors are created by the outlines of void areas within a single polygon or between polygon boundaries that are not coincident with other polygon boundaries.

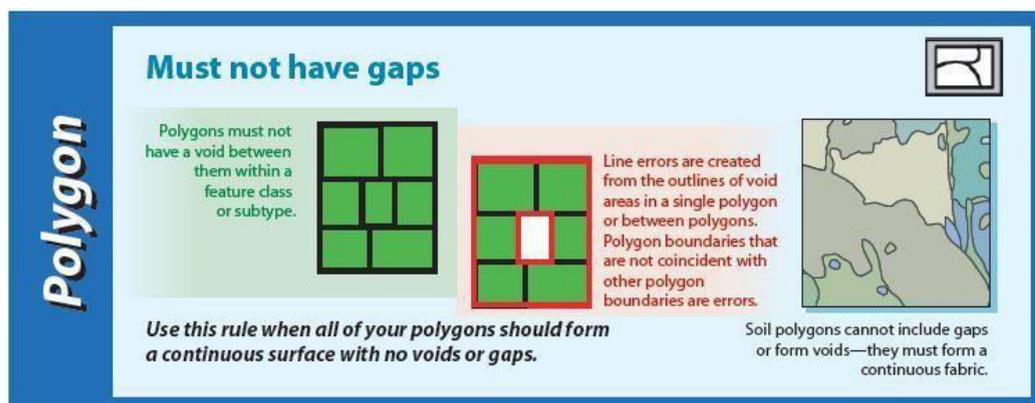


Figure3.27: Must not have gaps (Source-ESRI)

- Must not overlap with (Polygon)-:** The polygons of the first feature class or sub type must not overlap the polygons of the second feature class or subtype. Apply this rule when polygons from one feature class or subtype must not overlap polygons from another feature class or sub type, such as lakes and land parcels from two different feature classes. Polygon errors occur when polygons from two feature classes or sub types overlap.

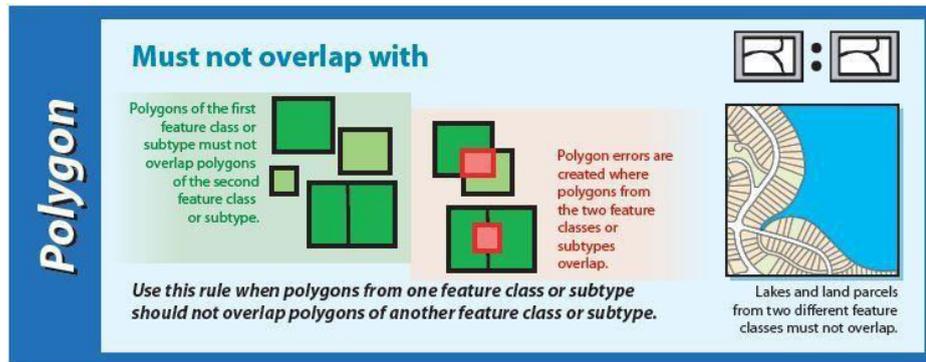


Figure3.28: Must not overlap with (Source-ESRI)

- Must be covered by feature class of (Polygon)-:** The second class or sub type of feature polygons should be covered by polygons in the first class or subtype. Use this rule if each feature class or subtype includes all polygons from another feature class or sub type, such as if states are covered by counties. The uncovered areas of the polygons cause polygon errors in the first feature class or subtype.

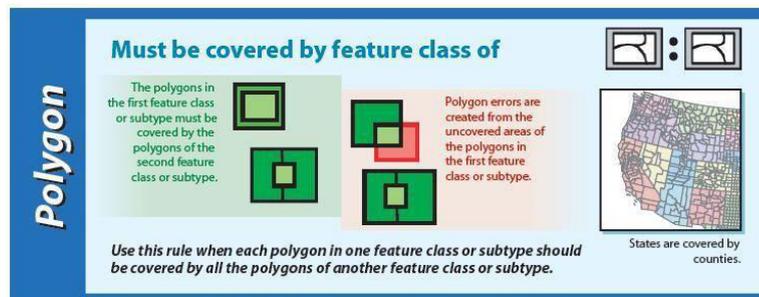


Figure3.29: Must be covered by feature class of (Source-ESRI)

- Must cover each other (Polygon)-:** All first-class polygons and all second-class polygons must be mutually exclusive. This implies that class 1 (1) must be class 1 (2) and class 2 (2) must be class 1 of the class 1 feature (1). Use this rule if you want to cover the same area with polygons from two feature classes or subtypes, such as when plants and soil need to be covered. If a part of a polygon does not cover one or more polygons in another feature class or subtype, a polygon error occurs.

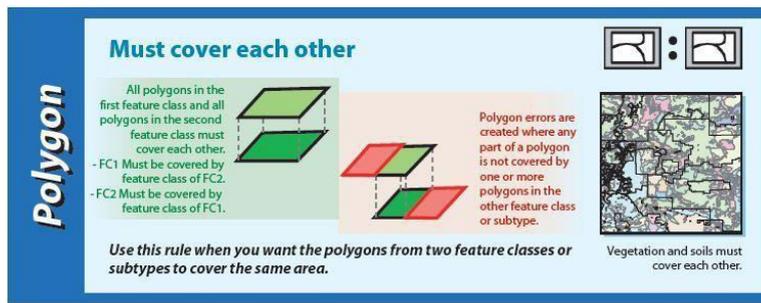


Figure 3.30: Must cover each other (Source-ESRI)

- Must be covered by (Polygon)-:** A single polygon in one class or subtype of feature must be covered by a polygon in another class or subtype. Use these rules if you want to cover some polygons in another functional class with some parts of another polygon, such as when counties must be covered by states. Polygon errors are caused by features that have polygons in the first class or subtype that are not covered by the second class or subtype of a single polygon.

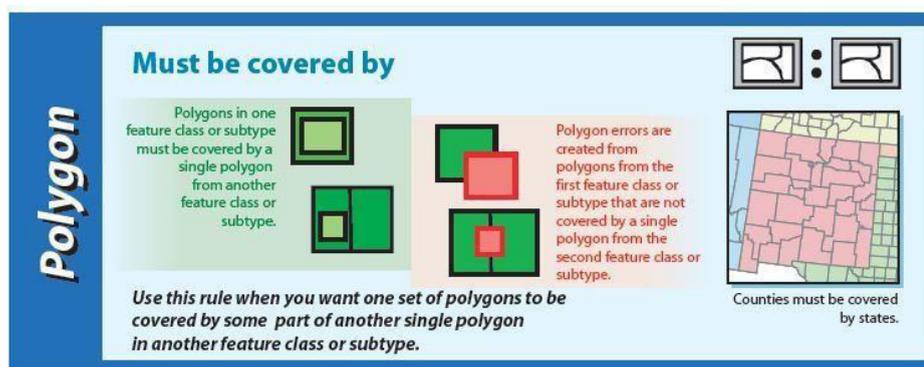


Figure 3.31: Must be covered by (Source-ESRI)

- Boundary must be covered by (Polygon)-:** Polygon boundaries in one feature class or subtype must be covered by lines from another feature class or subtype. When polygon boundaries must coincide with another line feature class or subtype, such as major road lines that are part of census block outlines, use this rule. When the boundaries of a polygon are not covered by a line of another feature class or subtype, line errors occur.

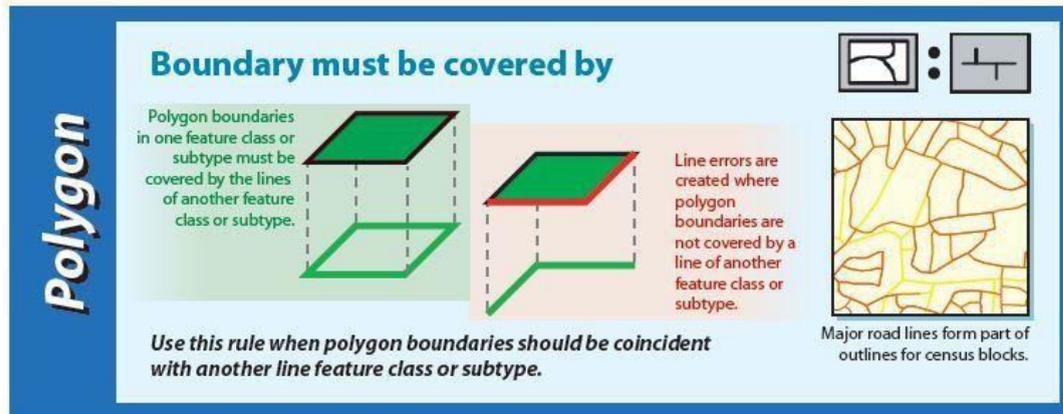


Figure 3.32: Boundary must be covered by (Source-ESRI)

- Area Boundary must be covered by boundary of (Polygon)-:** Polygon boundaries from one feature class or subtype must be surrounded by polygon boundaries from another feature class or subtype. Apply this rule whenever the boundaries of polygons in one feature class or subtype must align with the boundaries of polygons in another feature class or subtype, such as when residential area boundaries coincide with parcel boundaries but do not cover all parcels. Line errors occur when the boundaries of polygons in one feature class or sub type are not covered by the boundaries of polygons in another feature class or subtype.

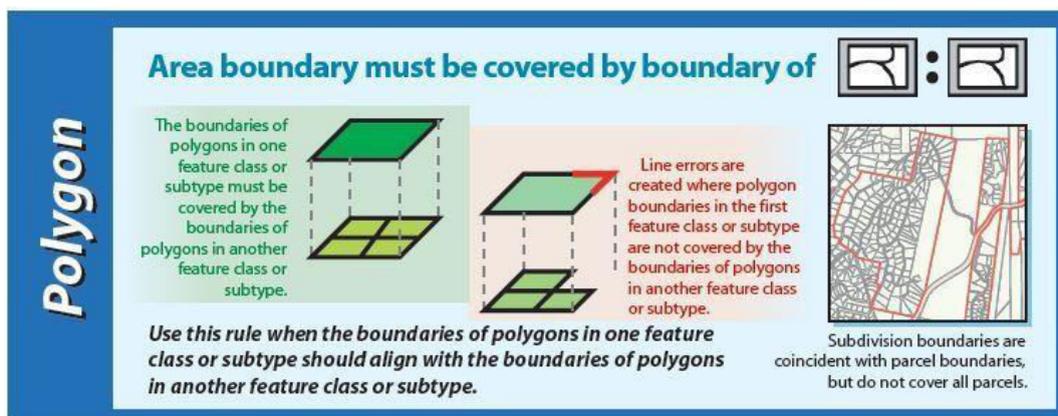


Figure 3.33: Area boundary must be covered by boundary of (Source-ESRI)

- Contain one point (Polygon)-:** It should be exactly one point for each polygon. It must be a polygon for each point. To ensure, for example, that the features of a polygon class and a point feature class are a one-to-one correlation, if parcels must have exact 1 point of address. The polygons that do not contain exactly one point create polygon errors.

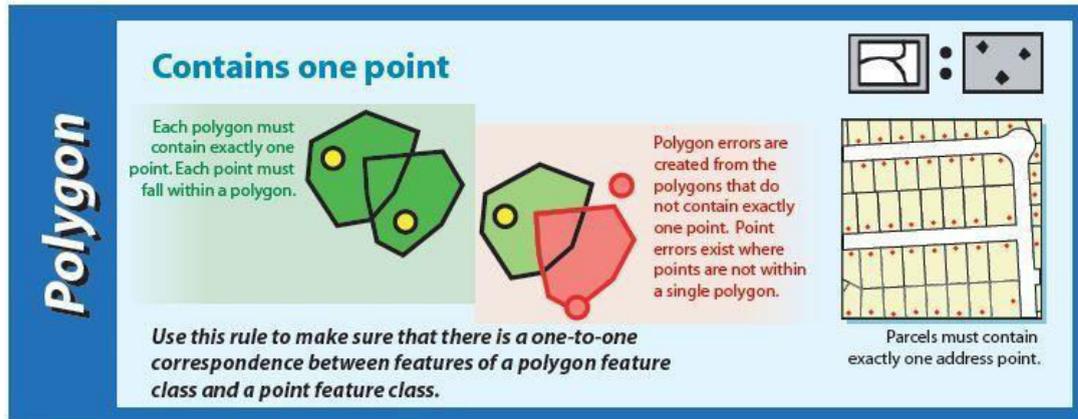


Figure 3.34: Contains one point (Source-ESRI)

These above explained rules of topology are used to extract out the errors from vector data and topology also allow us to remove all these unintentional error and make data error free. Error free data in GIS is always beneficial for all sort of analysis done with these datasets. Initially when most of the vector data are geo-relational in nature then finding error was very difficult. At that time only coverage file allows us to find out errors based on topology. But with the passage of time and also with the emergence of object based data model (Geo-database) topological editing become very common with vector database. Hence topological file format always plays an important role in any sort of analysis in GIS.

---

### 3.4 SUMMARY

---

Study of geometric properties that do not change when forms are bent or stretched is known as topology. Polygon adjacency is an example of a topologically. Topology deals with spatial properties that don't change under specific transformations. Topology is the mathematics branch used to determine spatial connections between entities. GIS transmits information through graphic symbolization (points, line sand polygons) and mathematically retains relationships through a topology concept. For example, you can easily identify crossing streets and adjacent properties when you stand on a hill and look into the countryside. Topology can be stored as a topological data model (geometric data correction) and can also be used for non-topological data analyses. The data is efficiently stored to allow fast processing of large datasets. Enables the computer to quickly determine the spatial relations of all characteristics and analyze them. Ensures that data is geometrically correct. Geometric relationships between spatial entities and their attributes are critical for spatial analysis and integration in GIS. Because topology is included in the data model, a single line can represent the shared boundary to denote which side of the line belongs to which polygon. Although most vector layer operations can be performed

without topology, some, such as network analysis, cannot. If we consider a roads layer, there is no way to build a network from it if it only contains lines representing roads but no information about how they are connected.

---

### 3.5 GLOSSARY

---

- **Topology-:** It is the spatial relationship between different entities which remain intact in any condition.
- **Spatial Relationship-:** A spatial relation describes how an object is located in space in relation to another object.
- **Node-:** A starting or ending point of a rim, linked topologically to all the rims of that rim
- **Contiguity-:** Contiguity is the topological concept for determining adjacency in vector data models.
- **Containment -:** A spatial relationship in which a point, line, or polygon feature or set of features is completely enclosed within a polygon.

---

### 3.6 ANSWER TO CHECK YOUR PROGRESS

---

#### Q-1-: What is Topological Editing?

Ans-: Topological editing is a type of editing that limits coincident geometry to a topologically connected graph of edges and nodes.

#### Q-2-: What is Connectivity?

Ans-: Connection is a geometrical property that describes how line functions, such as the road network, is connected.

#### Q-3-: What is Rules of Topology?

Ans-: Topology rules can also be defined between feature subtypes within one or more feature classes.

#### Q-4-: Write the major significance of Topology?

Ans-: topology is used to extract out the errors from vector data and topology also allow us to remove all these unintentional error and make data error free.

---

### 3.7 REFERENCES

---

- Batty, M and Xie, Y., Model structures, exploratory spatial data analysis, and aggregation, *International Journal of Geographical Information Systems*, 1994, 8:291-307.
- Bhalla, N., Object-oriented data models: a perspective and comparative review, *Journal of Information Science*, 1991, 17:145-160.
- Bregt, A. K., Denneboom, J, Gesink, H. J., and van Randen, Y., Determination of rasterizing error: a case study with the soil map of The Netherlands, *International Journal of Geographical Information Systems*, 1991, 5:361-367.
- Carrara, A., Bitelli, G., and Carla, R., Comparison of techniques for generating digital terrain models from contour lines, *International Journal of Geographical Information Systems*, 1997, 11:451-473.
- Congalton, R.G., Exploring and evaluating the consequences of vector-to-raster and raster-to-vector conversion, *Photogrammetric Engineering and Remote Sensing*, 63:425-434.
- Holroyd, F. and Bell, S. B. M., *Raster GIS: Models of raster encoding*, *Computers and Geosciences*, 1992, 18:419-426.
- Joao, E. M., *Causes and Consequences of Map Generalization*, Taylor and Francis, London, 1998.
- Kumler, M.P., An intensive comparison of triangulated irregular networks (TINs) and digital elevation models, *Cartographica*, 1994, 31:1-99.
- Langram, G., *Time in Geographical Information Systems*, Taylor and Francis, London, 1992.
- Laurini, R. and Thompson, D., *Fundamentals of Spatial Information Systems*, Academic Press, London, 1992.
- Lee, J., Comparison of existing methods for building triangular irregular network models of terrain from grid digital elevation models, *International Journal of Geographical Information Systems*, 5:267-285.
- Maquire, D. J., Goodchild, M. F., and Rhind, D. eds., *Geographical Information Systems: Principles and Applications*, Longman Scientific, Harlow, 1991.

- Nagy, G. and Wagle, S. G., Approximation of polygonal maps by cellular maps, *Communications of the Association of Computational Machinery*, 1979, 22:518-525.
- Peuquet, D. J., A conceptual framework and comparison of spatial data models, *Cartographica*, 1984, 21:66-113.
- Peuquet, D. J., An examination of techniques for reformatting digital cartographic data. Part II: the raster to vector process, *Cartographica*, 1981, 18:375-394.
- Piwowar, J. M., LeDrew, E. F., and Dudycha, D. J., Integration of spatial data in vector and raster formats in geographical information systems, *International Journal of Geographical Information Systems*, 1990, 4:429-444.
- Peucker, T. K. and Chrisman, N., *Cartographic Data Structures*, *The American Cartographer*, 1975, 2:55- 69.
- Rossiter, D. G., A theoretical framework for land evaluation, *Geoderma*, 1996, 72:165-190.
- Shaffer, C.A., Samet, H., and Nelson R. C., QUILT: a geographic information system based on quadtrees, *International Journal of Geographical Information Systems*, 1990, 4:103-132.
- Sklar, F. and Costanza, R. Quantitative methods in landscape ecology: the analysis and interpretation of landscape heterogeneity. in: Turner, M. and Gardner, R., editors. *The development of dynamic spatial models for landscape ecology: A review and prognosis*. New York: Springer-Verlag; 90:239-288.
- Tomlinson, R. F., The impact of the transition from analogue to digital cartographic representation, *The American Cartographer*, 1988, 15:249-262.
- Wedhe, M., Grid cell size in relation to errors in maps and inventories produced by computerized map processes, *Photogrammetric Engineering and Remote Sensing*, 48:1289-1298.
- Worboys, M. F., *GIS: A Computing Perspective*, Taylor and Francis, London, 1995.
- Zeiler, M., *Modeling Our World: The ESRI Guide to Geodatabase Design*, ESRI Press, Redlands, 1999.
- Chang, Kangtsung *Introduction to Geographic Information Systems 5th edition 2009* McgrawHill.
- *Textbook of Remote Sensing and Geographical Information System*, M.Anji Reddy, Second

Edition, Pp 1-23.

- C. P. Lo and Albert K. W. Yeung (2002) Concepts and Techniques of Geographic Information Systems, Upper Saddle River, New Jersey: Prentice Hall, 2002)
- Bhatta, (2008) Remote Sensing and GIS Oxford University Press

---

### ***3.8 TERMINAL QUESTIONS***

---

Q-1 What is Vector Data Format?

Q-2 Explain role of topology in GIS data creation.

Q-3 What do you mean line based topology?

Q-4 Explain must not have dangles rule of topology.

---

## **UNIT 4 - DATA MANIPULATION**

---

**4.1 OBJECTIVES**

**4.2 INTRODUCTION**

**4.3 DATA MANIPULATION**

**4.4 SUMMARY**

**4.5 GLOSSARY**

**4.6 ANSWER TO CHECK YOUR PROGRESS**

**4.7 REFERENCES**

**4.8 TERMINAL QUESTIONS**

---

## **4.1 OBJECTIVES**

---

After going through this unit the learner will able to learn:

1. Understanding the meaning of Data manipulation.
2. Basic knowledge of data manipulation language (DML).
3. Explore the necessity of data manipulation tools.

---

## **4.2 INTRODUCTION**

Data manipulation is the process used to modify or alter the information in a more orderly and readable manner. We use DML to do this. Now question is what does DML mean? Well, the Data Manipulation language Terminology means that we can add, exclude, and switch records, i.e. change the records so that we can interpret. In other words, manipulation of data is the modification of information to make comprehension easier or more formal, for example, a data log in alphabetical order should be sorted, which makes it easier to find individual entries. Data manipulation on Web server logs also enables the administrator of the website to track its most popular sites and traffic sources.

---

## **4.3 DATA MANIPULATION**

---

### **What is Data?**

Data are the facts, quantities or statistics which are collected and stored together for analysing which helps in providing an information. With evolution in time data can be used in scientific research, financial or business matters and e-governance. Many a times, use of data can be differentiated in 4 categories -

- Descriptive- it explains the 'what' of any phenomena or condition. It generates the accurate and fast analytic which reduces time and enables in better decision making.
- Diagnostics - its deals with the 'why' of the condition. Diagnostic analysis of any data helps in optimising the future activities.
- Predictive- it explains the 'what will happen' circumstances of the data retrieved or analysed as it improves the decision making process. For example in business this process will help in determining where and how business should invest in the market.

- Prescriptive- it illustrates the 'what should happen' condition of the data as it will help in processing and prescribing to what extent or measures to improve the outcomes or correct the problem.

Today's data transform corporate operations. All depends on data, from corporate decision-making to daily operations. None of this can be done without transforming raw data into accessible information, particularly when there are several data and various sources. This is where the processing or manipulation of data is concerned. Now, question arises what is data manipulation?

Data manipulation is the process used to modify or alter the information in a more orderly and readable manner. We use DML to do this. Now question is what does DML mean? Well, the Data Manipulation language Terminology means that we can add, exclude, and switch records, i.e. change the records so that we can interpret. In other words, manipulation of data is the modification of information to make comprehension easier or more formal, for example, a data log in alphabetical order should be sorted, which makes it easier to find individual entries. Data manipulation on Web server logs also enables the administrator of the website to track its most popular sites and traffic sources.

### **Purpose of Data Manipulation:**

For enterprise processes and optimisation, data manipulation is essential. In order to better use and transform it into usable knowledge such as analysis of financial data, customer behaviour and trend analysis, you must be able to work with the data in the required format. Data manipulation thus offers many advantages to a business, including:-

- **Consistent data:** In a standardised format, data can be organised to make it readable and understandable. One may not have the united perspective while taking data from multiple sources, but one will ensure that the data is consistently organised and processed through data manipulation and commands.
- **Project data:** In particular when it comes to finance, it is critical for businesses to use historical information to predict their future and to do more thorough analysis. The data manipulation makes this function possible.

- **Create more value from the data:** In addition, one can do something about data by converting, modifying, removing and adding data into a database. If information continues to be stagnant, it is worthless. However, if you know how to use your results, you can get a strong insight into better business choices.

- **Remove or ignore unwanted data:** Data that cannot be used will still interfere with what is essential. Inaccurate or unnecessary data must be removed and cleaned. Through data manipulation you can quickly clear records so, that we can deal with the documents that we need.

### **Data Manipulation Language:**

DML or data manipulation language is used to make data more organized or readable. DML is a language for the computer programming used to insert, omit and modify data into a database. It makes cleansing and mapping data simple for further analysis. Structured query language is a widely used language for data manipulation. We need SQL (structured query language) to communicate with the database, and there can be four functions during this communication:

- Select
- Update
- Insert
- Delete

Through these commands we will command a certain machine what to do with data or at least a chunk of chosen data.

- **SELECT:** The selected declaration helps users to draw a data base selection into the work. You say what to choose to the computer and where to choose.

- **UPDATE:** One uses the UPDATE statement to modify data which already exist. One will instruct the archive to upgrade certain data sets and new information to be entered either with one or multiple records at a time.

- **INSERT:** By using the INSERT statement, one can transfer data from one position to the next.

- **DELETE:** By using the DELETE expression to get rid of current records in a table. One instructs the machine when to remove and what files you want to remove.

Since SQL cannot import or export data from external sources, certain vendors can store data and provide you with the necessary resources to manipulate data for your business needs.

### **Types of Data Manipulation Language:**

Data manipulation language (DML) is a database language that helps users to view or modify information on a structured data model. Basically, there are two categories of data manipulation language:

- Procedural data manipulation language.
- Declarative data manipulation language.
- **Procedural DML:** It needs a user to decide what data is necessary and how the data can be collected.
- **Declarative DML:** This demands a user to specify the data without deciding how the data can be obtained, often, referred to as Non-procedural DML.

**Note: DML component of SQL language is Non-Procedural**

### **Need of Data Manipulation Tools:**

Manipulation of data is a major problem for the process of optimization as it makes available data to produce insights including financial information analysis, consumer behaviour analysis and pattern analysis. During integration, the technique is commonly used for data compatibility with the target device. For example, accountants handle raw data collected by retailers and marketing in order to understand the prices of products, pricing rates or future tax needs. Likewise, stock market analysts leverage data so that they can plan their investment portfolio accordingly in order to predict market patterns.

There are a number of uses of the data manipulation. Some more ways in which manipulation can be useful for organisations:

- **Data Consistency:** It's easier to organise, read and interpret data in a consistent data format. When data are derived from various sources, a single format needs to be

transformed and manipulated. It is easier to type or use data for reporting after standardising the format in the enterprise system.

- **Data Projection:** Data manipulation permits the use of historical data for future projections and a systematic study especially in the field of urban planning.
- **Value Generation:** One can update, alter, suppress, and input data in a database by using data manipulation. This ensures that one can use data for in-depth insights and smarter business choices.
- **Redundant Data Removal:** Data from sources also contain redundant, incorrect or unnecessary data. In order to use these data, it needs to be checked by accuracy and filters applied to extract information that is important to your business. You can quickly clean the data using data manipulation so that the data matter can be filtered out.
- **Data Interpretation:** It is almost difficult to make sense out of it without distortion when working with dynamic data involving different formats and market conditions. One must be able to view data and transform it into meaningful and understandable material. An instrument for handling of data will solve the issue by converting and integrating data into the right format with different techniques to enhance the visual experience. This makes data easier for users to comprehend and use.

### **Steps involved in Data Manipulate:**

The best way to manipulate the data is by the use of software that has integrated, automatic data management features including data cleaning, mapping, aggregating and storage. These tools spare one from the difficulty of manually inserting the data and repeating low-value tasks. In addition, these tools support the workflow functions for producing and delivering reports without human interaction.

The main five measures that are used for the efficient handling of data are given below:

- Developing a database from data sources would be the first step.
- Next, before reorganising and reconstruction, clear the data collected from the source scheme.

- Import and create the database that one need.
- Combine or filter out redundant content.
- Analyse data and generate valuable observations that inform the process of decision making.

### **Data Manipulation tips:**

Python data manipulation and R data manipulation are key elements of data manipulation. Now let's consider how to handle data before going through the more detailed concepts of Python and R data handling.

One knows how to use MS Excel most certainly. Some tips for manipulating Excel information are provided here.

1. Formulas and functions – One of the nice things about excel is that one can rely on important mathematical features to make the data more worthwhile.
2. Autofill in Excel- This function is helpful if one choose to use the same equation in several cells. One way to do this is to re-type it. Another method is to drag the mouse down to the bottom right corner of the cell. It helps in using the same formula in several rows at the same time.
3. Sort and Filter- When reviewing results, users can save a lot of time by sorting and filtering Excel options.
4. Removing duplicates- In the course of data collection and assimilation there are always chances of duplication of data. The Delete Duplicate function in Excel will help to clear duplicate table entries.
5. Dividing, combining, and joining columns or rows may be inserted or omitted in Excel frequently. Data organisation also needs several data sheets to be integrated, separated or combined.

### **Data Manipulation vs. Data Modification:**

One should also learn about data modification now as data manipulation has been discussed. Although these two words sound identical, they cannot be interchanged. Data manipulation typically results in the analysis of new and more sophisticated data through logic or raw data calculation. Modification of data on the other hand means that the same data values or the data themselves are changed. It can sound very complicated, for example, assume we have an  $X=5$  value. We can present the value as  $X=2+3$  or  $X=1+4$ ,  $X=6-1$  etc. this represents data manipulation, which is an examples of how the given value can be read by logic. Data modification implies to change the reference to  $X=7$  itself.

Now, how can we use data modification to support market choices through data manipulation? Well, data modification can be used to calculate financial objectives if several data sources are processed through manipulation.

---

## **4.4 SUMMARY**

---

In this chapter we have learned about data manipulation techniques that are a process to organise and make data easy to understand and analyse. This is accomplished by data manipulation language and is divided in declarative DML which tells what problem to be solved without specifying the exact steps. In fact, procedural DML explains how to solve the problem. Structured query language is an example of declarative DML which occurs through four function- select, update, insert and delete. On the other hand, examples of procedural DML are FORTRAN, COBOL, and ALGOL etc. Data manipulation is a useful and important function for analysing financial data, for business purposes, performing research analysis. The chapter also explains the basic steps of data manipulation which includes - defining why one needs data analysis, collection of data from sources, cleaning of unnecessary data, analysis of Data and finally interpretation of results and their applications.

---

## **4.5 GLOSSARY**

---

- **DATABASE** -As simply as possible, this is a storage space for data. We mostly use databases with a Database Management System (DBMS), like PostgreSQL or MySQL. These are computer applications that allow us to interact with a database to collect and analyze the information inside.



Q.3 What is Procedural DML and Declarative DML?

Q.4 What is database?

---

## 4.7 REFERENCES

---

- <https://www.jigsawacademy.com/blogs/data-science/data-manipulation/amp/#>
- <https://www.computerhope.com/jargon/d/datamani.htm>
- <https://www.astera.com/type/blog/data-manipulation-tools/>
- <https://whatagraph.com/blog/articles/data-manipulation>
- <https://www.digitalvidya.com/blog/data-manipulation/amp/>
- Chang, Kang-tsung Introduction to Geographic Information Systems 5 th edition 2009 Mcgraw-Hill.

---

## 4.8 TERMINAL QUESTIONS

---

Q.1 Write a note on Data manipulation language and their types.

Q.2 Write a short note on why we use Data manipulation tools.

Q.3 What is the difference between Data manipulation and Data Modification?

Q.4 Write a steps that involved in Data manipulation.

## **BLOCK 2: SPATIAL DATABASE RASTER ANALYSIS**

---

### **UNIT 5 - RASTER DATA MANIPULATION AND RECLASSIFICATION**

---

**5.1 OBJECTIVES**

**5.2 INTRODUCTION**

**5.3 RASTER DATA MANIPULATION AND RECLASSIFICATION**

**5.4 SUMMARY**

**5.5 GLOSSARY**

**5.6 ANSWER TO CHECK YOUR PROGRESS**

**5.7 REFERENCES**

**5.8 TERMINAL QUESTIONS**

---

## 5.1 OBJECTIVES

---

After going through this unit, the learner will be able to learn

1. Understand and meaning of Data and their types.
2. Learn about Raster data manipulation tools and techniques.
3. Learn about Reclassification of Raster data.

---

## 5.2 INTRODUCTION

---

Data is a Latin word that refers to information that is expressed in the form of a digit/number, symbol, or letter and is used to reflect the status of any geographical object, as well as its behaviour or outcome. The position and attributes of spatial features on surface of the earth are defined by data that is geographically referenced to the earth's surface. Location refers to the location on the earth's surface, while characteristics refer to the name of the location, the number of people going to or visiting that location, the form of settlement, transportation and communication options, and so on. Geoinformatics considers two types of data. First is **spatial data** and second is **non-spatial data**. Spatial data provides the information about the location, shape and size of objects, and non-spatial data, also called attribute data provides the information about spatial characteristics. Non-spatial data is independent from geometrical information of objects.

---

## 5.3 RASTER DATA MANIPULATION AND RECLASSIFICATION

---

### TYPES OF SPATIAL DATA:

There are two types of spatial data: vector data and raster data. Separate or discrete characteristics, such as locations, land use, streets, data summarized by area, parcels, and so on, are represented by vector data. The spatial geometry of features is expressed in vector data using coordinate pairs. In vector data format, real world characteristics/features are defined in the form of lines, points and polygons.

Raster data depicts continuous numeric values like elevation and continuous categories like vegetation types, water etc. In Raster data, real world characteristics/features are described as grids. Raster data consists a fixed grid dimension and register data/information about each grid. One or more than one characteristics/features are connected with each grid cell. One set of cell and connected value is called a layer. Digital satellite images or remote sensing images are most common examples of raster data.

### RASTER DATA MANIPULATION:

You have understood the definition of data and different types of data in earlier sections. Now, in this section, we are going to learn about the different operations and processes used

in Raster data manipulation. The raster data model covers the space with a regular grid, and each grid cell's value represents the characteristics of a spatial phenomenon at that cell's position. This simple raster with fixed cell positions data structure is not only computationally effective, but it also facilitates a large number of data analysis operations.

Raster data analysis is focused on cells and rasters, as opposed to vector data analysis, which is based on geometric objects such as point, line, and polygon. Individual cells, groups of cells, and cells within an entire raster can all be analysed using raster data. Some raster data operations only use one raster, while others use two or more. The cell type value is an important factor in analysis of raster data. Mean and standard deviation are optimised for numeric values, while majority (the most frequently occurring cell value) is designed for both numeric and categorical values.

Raster format is used to store a variety of data types. Raster data analysis, on the other hand, is limited to software-specific raster data, such as ArcGIS ESRI grids. As a result, to use DEMs and other raster data in data processing, they first be processed and transformed to software-specific raster data.

The general tools for raster data manipulation are covered in this chapter. The analysis of raster data environment, including the area for analysis and the output cell size, is described in the following section.

### **DATA ANALYSIS ENVIRONMENT:**

The data analysis environment refers to the size of the output cell and the area for analysis. The analysis area can be defined by a single raster, or area defined by the raster's min and max x-y coordinates, or a composite of rasters. The extended area for study can be depending on the union or intersection of raster data with distinct area extents. The union relates an extended area to all input rasters, while the intersect relates an extended area to all input rasters.

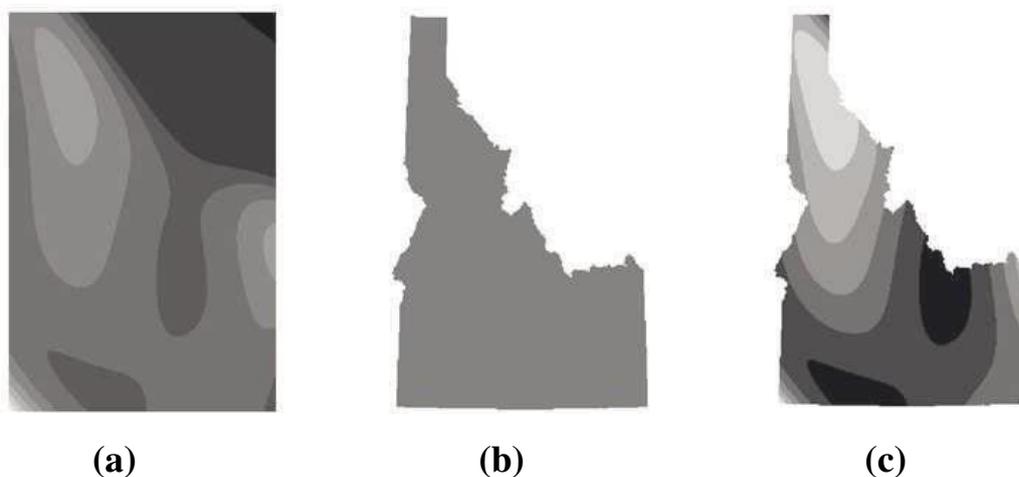
An **analysis mask** can also be used to determine the size of the area to be analysed. An analysis mask restricts analysis to cells which does not contain the cell value "no data." These no data points vary from zero. No data is the absence of data, while zero is a true cell value. For example, an elevation raster created from a DEM often includes no-data cells along its boundary. In several other cases, however, the user enters no-data cells on purpose to restrict the region that can be analysed. For example, one option for limiting soil erosion analysis to only private lands is to encode public lands with no data. A feature layer or a raster may be used to create an analysis mask.

We can describe the output cell size at any scale that we think is appropriate. The output cell size is usually set to be the same as or greater than the largest cell size in the input rasters. This is based on the idea that the output should have the same accuracy as the least accurate input raster. If the input cell size is between 10 and 30 metres, the cell size of output should be at least 30 metres. A GIS package resamples all input rasters, using a resampling technique to manipulate all input rasters to the required output cell size before data analysis, based on the output cell size. Nearest neighbour, bilinear interpolation, and cubic convolution are common resampling methods.

## RASTER DATA OPERATIONS:

The majority of raster data operations are local, neighbourhood, zonal, and distance measure operations. However, some raster data operations do not perfectly fit into the preceding classification scheme.

**Management of Raster Data** - We often need to clip or combine raster data found online to match the study area in a GIS project. To clip a raster, we can use the larger raster as the input and assign an analysis mask or the min and max x-y coordinates of a rectangular region for the analysis environment. **Mosaic** is a technique for merging collective input raster data into a discrete raster data. If the input rasters overlay, maximum GIS packages supply options for editing the cell values in the overlaid areas. For overlapping regions, ArcGIS, intends the user to choose the data from the foremost input raster or the mixing of data from the input data of rasters. If there are little gaps between the input data of rasters, one way is to put data in unknown values using neighbourhood mean operations.



**Figure 5.1:** An analysis mask (b) is used to clip an input raster data (a). The output raster is (c), which has the similar extended area as the analysis mask.

Source: Chang, K. T. (2019)

**Raster Data Extraction** - By extracting data from an existing raster, raster data extraction generates a new raster. The procedure is similar to that of a raster data query. To describe the area to be extracted, you can use a data set, a graphic object, or a query language. If there is a point layer in data set, the extraction technique retrieves the values at the feature position of points (for example, using bilinear interpolation) and stores them in a new field in the point feature table. If there is a raster or polygon feature in data set, the extraction technique extracts values from cells within the raster or polygon's specified area and assigns no values to other cells.

A set of points, a rectangle, a polygon, or a circle, or May all is used to extract raster data. The x-y coordinates of the object are entered. A circle, for example, can be represented by a pair of coordinates in x-y, for its centre and radius length. We may retrieve data of elevation in a range of 50-mile radius of an earthquake epicentre or elevation data from a collection of one or several nearby weather monitoring stations working on graphic materials.

Extract by attribute function generates a new raster with cell values that match the query expression. We may, for example, generate a new raster file inside a specific elevation region. On the output, those cells outside the elevation zone have no data.



**Figure 5.2:** A circle, in white, is used to retrieve values of cell from raster that is inputted. The output raster covers the same area as the input raster, but there lies no data, which is outside of that particular circular area.

Source: Chang, K. T. (2019).

**Raster Data Generalization** - A number of operations may be used to generalise or simplify raster data. Resampling is one such operation that can be used to create distinct levels of pyramid for a vast raster data set. **Aggregate** is similar to resampling in that it provides an output in raster format with cell size which is large than the input. Here, each output cell value is calculated, can be as either mean, median, the sum, or the low or the high of the cells inputted that fall inside the output cell, rather than using nearest neighbour, bilinear interpolation, or cubic convolution.

Zones, or groups of cells with the same value, are used in certain data generalisation operations. For example, ArcGIS has a tool called RegionGroup that defines the zone to which each cell in the output raster is related. RegionGroup can be considered of as a classification system that uses both cell values and cell spatial connectivity as classification criteria.

For certain applications, generalising or simplifying the values of cell of a raster file can be advantageous. A raster gathered from an image of satellite or LiDAR data, for example, typically has a lot of local variations. These small variations can add up to unwanted noise. We can use Aggregate or a resampling technique to get rid of them.

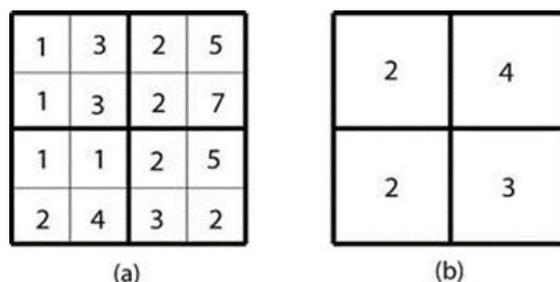


Figure 5.3: An Aggregate technique generates a raster with lower-resolution from the input (a). The procedure makes use of the mean statistics as well as a factor of two. To perform this operation, the cell value of 4 in (b) is the mean of values in second box in right in (a).

Source: Chang, K. T. (2019).

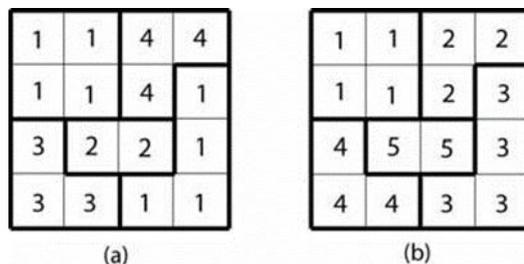


Figure 5.4: In the output, each cell (b) has a discrete number that relates the connected region to which it lies in the input (a). For example, the connected region which has the similar cell value of 3 in (a) has a discrete number of 4 in (b).

Source: Chang, K. T. (2019).

**PHYSICAL DISTANCE MEASURE OPERATIONS:**

Distances can be represented as physical or cost distances in a GIS project. The cost distance calculates the cost of traversing the physical distance, while the physical distance evaluates the straight-line or Euclidean distance. In real-world applications, the difference between the two forms of distance measurements is essential. For example, a truck driver, is more concerned with the route's time or fuel cost than with its physical distance. In this case, the cost distance is determined not only by the physical distance, but also by the speed limit and road conditions.

**Physical distance measurement technique** determines distances of straight-line far from the source of cells. To find the distance between cells (1,1) and (3,3), the formula used is as follows:

$$\text{Cell size} \times \sqrt{(3-1)^2 + (3-1)^2}$$

Or cell size X 2.828. If size of the cell would 30 meters, the distance would be 84.84 meters.

A physical distance measurement process effectively buffers the source cells with wavelike running distances across the complete raster or up to a given largest distance. This is why, this measurement techniques are also known as extended neighbourhood operations or universal (i.e., the entire raster) operations.

(0, 0)

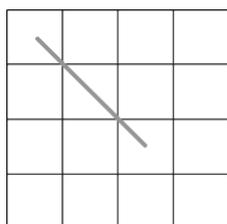


Figure 5.5: A straight-line distance is calculated from centre of a cell to another centre of cell. This figure depicts the straight-line distance between cell (1,1) and (3,3).

Source: Chang, K. T. (2019).

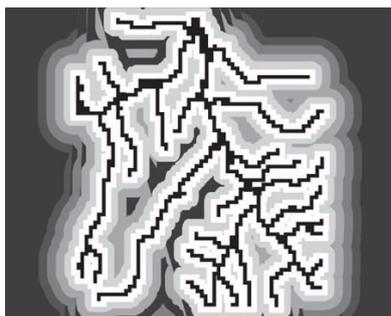


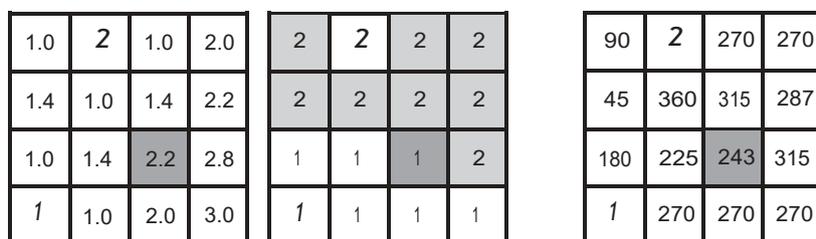
Figure 5.6: Continuous distance measures from a stream network.

Source: Chang, K. T. (2019).

A physical distance measure operation in a GIS could use a feature layer (e.g., a stream shape file) in the form of a source. Since the layer is transformed from vector to raster data before the process begins, this choice is based on convenience.

Following a physical distance measure method, the continuous distance raster can be used right away in other methods or operations. However, it can be processed to generate a particular distance zone or a sequence of distance zones from source cells. A continuous distance raster may be reclassified to a raster with one or more distinct distance zones. **Slice**, a variation of Reclassification, which divides a continuous distance raster into equal-area or equal-interval distance zones.

**Allocation and Direction** - Physical distance measurement techniques can generate allocation and direction rasters in addition to calculating straight-line distances. An allocation raster's cell value coincides to the cell's nearest source cell. The value of cell in a direction raster coincides to the cell's direction from the nearest source cell, in degrees. The values of direction are depended on the direction of compass like 90 degree to the east, 180 degree to the south, 270 degree to the west, and 360 degree to the north. And 0 degree is assigned to the source cell.



(a)

(b)

(c)

Figure 5.7: On the basis of source cells described as 1 and 2, (a) displays the physical distance measures in cell units from each cell to the nearest source cell; (b) displays the allocation of every cell to the nearest source cell; and (c) displays the direction in degrees from every cell to the nearest source cell. The dark shade cell (row 3, column 3) has the similar distance to both source cells. Thus, the cell can be allocated to any source cell. The direction of 243 degree is to the source cell 1. Source: Chang, K. T. (2019).

**Applications of Physical Distance Measure Operations** - You have understood the Physical distance measure operation in earlier section. Now, in this section, we are going to learn about the applications of physical distance measure operations. Physical distance

measure operations, including buffering around vector-based functions, have numerous applications. For example, a stream network or regional fault lines can be used to establish equal-interval distance zones. One more example is the use of distance measurement processes as methods for putting a model into action, such as Herr and Queen's potential nesting habitat model of greater sandhill cranes in north-western Minnesota (1993). The model categorises potentially suitable nesting vegetation as optimal, suboptimal, marginal, and unsuitable based on zones of continuous distance calculated from uninterrupted vegetation, agricultural land, roads, houses and other buildings. Physical distance measurements are useful in the cases above, but they are unrealistic in other cases.

### **Comparison of Vector and Raster-Based Data Analysis:**

You have understood the types of data, that is, vector data and raster data. Now, we are going to learn the comparison of vector and raster-based data analysis by explaining the different operations. The two most common types of GIS analyses are vector data analysis and raster data analysis. Since a GIS programme can't run them both at the same time, they're handled separately. Although some GIS packages permits vector data to be used in raster data processes (such as extraction), the data must first be converted to raster data before the operation proceeds.

In terms of data sources and targets, each GIS project is unique. Furthermore, vector data can be transformed to raster data and vice versa with ease. As a result, we must choose the most effective and suitable data analysis method. Overlay and buffering, the two most occurring GIS operations, are used as illustration to compare operations based on vector and raster in the following sections.

**Overlay** - A vector-based overlay operation is often compared to a local operation with multiple rasters. Both operations use multiple data sets as inputs, which makes them identical. However, there are significant variations between them.

An operation of vector-based overlay must first calculate intersections or junction between attributes and points are inserted at the junction in order to integrate the geometries and features from the input layers. Since the input rasters are of the same cell size and area extent, this form of calculation is not required for a local operation based on raster. Even though the input rasters must be sampled again to the similar size of cell, the calculation is still simpler than computing intersections of line. Second, a local operation based on raster can generate the output using a variety of tools and operators, while a vector-based overlay operation can only combine attributes from the input layers. All attribute computations must be done after the overlay process. For the reasons mentioned above, overlay based on raster is used often for projects including a number of layers and a significant amount of calculation.

Despite the fact that a local operation based on raster is more computationally effective than an overlay operation based on vector, the vector-based operation has its own set of advantages. Multiple attributes from each input layer can be combined by an overlay operation. All attributes can be queried and measured separately or together, once they've been combined into a layer. For example, a stand layer of vegetation, might have features like stratum, height, crown closure and crown diameter, whereas a layer of soil might have features like texture, organic matter, depth and pH value. The features of both layers are merged into a single layer, which is allowing all attributes to be queried and analysed. A local

operation, on the other hand, associates each input raster with a single collection of cell values. In other words, a local operation based on raster will require a raster for each characteristic to query or analyse the same stand and soil features as above. If the data sets, which is going to be examined have a large number of features or attributes which share the same geometry, an overlay operation based on vector is more effective than a raster-based local operation.

**Buffering** - Both buffering operation based on vector and a physical distance measure operation based on raster computes distances from specific features. However, in at least two ways, they vary. To begin, a buffering operation measures distances using x and y coordinates, while a raster-based operation measures physical distances using cells. As a result, a buffering operation will construct more precise buffer zones than a raster-based operation. This disparity in precision can be critical, for example, when executing functions of riparian zone management. Second, a buffering process is more adaptable and versatile. A buffering process, for example, can produce several rings (buffer zones), while a process based on raster produces continual distance measurements. To identify buffer zones from continuous distance measures, additional data processing is needed. A buffering operation may create individual buffer zones for each selected attribute or a buffer zone in dissolved state for all selected attributes. Using a raster-based process, it would be difficult to construct and control separate distance measurements.

## MOSAIC:

A mosaic is comprised of two or more images that have been combined or merged. By mosaicking multiple raster datasets together in ArcGIS, there can be generated a single dataset of raster file. There can be also made a mosaic dataset and a virtual mosaic from a series of datasets of raster.



Figure 5.8: The figure illustrates how six neighbouring datasets of raster are mosaicked into a single dataset of raster.

Source: <https://desktop.arcgis.com>

In certain instances, the edges of the raster datasets that are being mosaicked together can overlap.

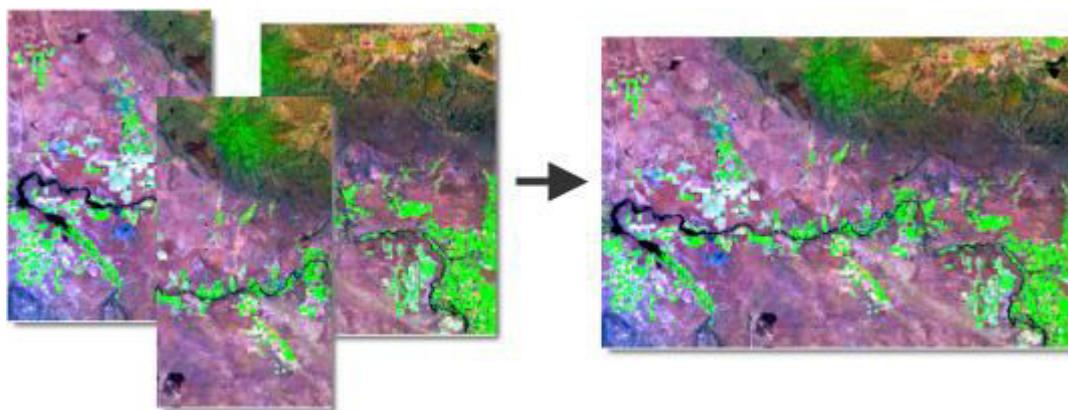


Figure 5.9: The figure illustrates the edges of the raster datasets that are being mosaicked together can overlap. Source: <https://desktop.arcgis.com>

These overlapping areas can be done in a variety of ways, including keeping only raster data from the initial or end dataset, using a weight-based algorithm to blend the overlapping cell values, taking the mean of the overlaying values of cell, or taking the least or highest value. The First, Least, or Highest options produce the maximum significant results when mosaicking discrete data. Continuous data is better served by the Blend and Mean options. The output is floating point if all of the input rasters are floating point. The output is integer if all of the inputs are integer and First, Least, or Highest are used.

You may use the mosaic dataset to apply a number of other methods of mosaicking to a dynamic mosaic or an exported mosaic, mosaicked raster dataset. Sorting by attributes, using a seamline, and other techniques are among them. If the raster dataset contains a colour map, there are many choices for handling it. You may use the colour map from the first or last dataset of raster in the mosaic, or confirm all of the colours in the final colour map are similar. You may also select whether or not to mosaic any rasters that contain a colour map.

You may also perform colour corrections on the raster datasets that are being mosaicked by choosing to colour balance or colour match them. The colour correction is done with a dodging technique in colour balancing. For each band, a global gamma value and contrast adjustment are calculated, and these values are then used to calculate the final value for each pixel in output. When displaying a catalogue of raster in ArcMap or on a dataset of mosaic, this option is available, and it can be added perpetually during the use of Raster Catalogue to Raster Dataset tool. The pixel values of the overlapped regions between the reference and source rasters are synchronised by colour matching. The matching algorithm is applied to the source rasters after it has been calculated in the overlap regions. To interpolate the correct matching of colour from the reference raster to the source rasters, colour matching can use one of three methods:

- ❖ Statistics Matching - The colour transformation is applied to the source datasets after the statistical variations between the reference overlap region and the source overlap region are matched.
- ❖ Histogram Matching - The colour transformation is added to the source datasets after the histogram from the reference overlap region is compared with the histogram from the source overlap region.

- ❖ Linear Correlation - A weighted average can be used to align overlapped pixels and interpolate to the remainder of the source pixels that do not have a one-to-one relationship.

When viewing a raster catalogue in ArcMap, when applying mosaicking methods, or when showing a mosaic dataset, colour matching can be conducted.

There are some other key points of mosaicking raster data:

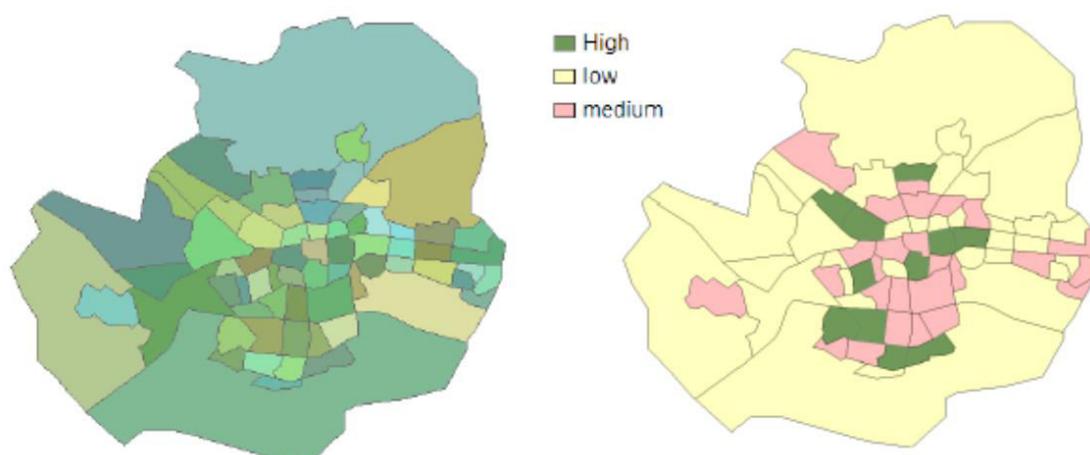
- The schema of a mosaicked raster dataset is the same as that of every other raster dataset.
- The number of bands in all of the datasets of raster and the produced mosaic of raster should be the similar; else, the mosaic can't be generated.
- Mosaicking two or more rasters of the similar spatial reference and pixel size into a single raster is feasible. If the spatial reference of second raster dataset dissent from the dataset of first raster, the spatial reference of second raster dataset will be ignored, and its data will be converted into the first raster dataset's spatial reference. In this situation, the Project Raster function is suggested to ensure that the data is not affected.

### Reclassification:

In classification, basically we group or assigns certain attributes to a class on basis of attribute value, that is, vector data or we group pixels to a class on basis of pixel values, that is, raster data. This process is called classification.

When we group or classify the already grouped data or classified data, that process is called the reclassification.

Classification is a technique of purposefully removing the details form an input data to reveal or to get the pattern out from the data.



**Figure 5.10:** Reclassification of data

**Source:** Study material IIRS Outreach Programme

Reclassification removes the details from an input dataset in order to get or to reveal the important spatial patterns. Reclassification reduces the number of classes and eliminates the details. If the input data set itself is the resultant of a classification, then it is called as reclassification. Reclassification of data can be in different systems for different-different purposes. Also, based on specific attribute values, some codes are assigned.

**Example:** A soil map.

Soil map is already a classified data, classified in few classes of different types of soil. And this soil map or soil types is reclassified into soil suitability analysis for a particular crop. Thus soil type is reclassified in two classes.

1. Soil suitability class – to that particular crop.
2. Soil unsuitability class – to that particular crop.

Classification – reclassification is based on Automatic classification and Manual or User controlled classification.

**User Controlled Classification** - In user controlled classification we indicates the classification based on attribute values and based on the classification method. This process is normally done via a classification table.

Table 5.1: Two examples of classification tables are given below:

Old Value	New Value
391 – 2474	1
2475 – 6030	2
6031 – 8164	3

Code	Old Value	New Value
10	Planned Residential	Residential
20	Industrial	Commercial
30	Commercial	Commercial

**Source:** Study material IIRS Outreach Programme

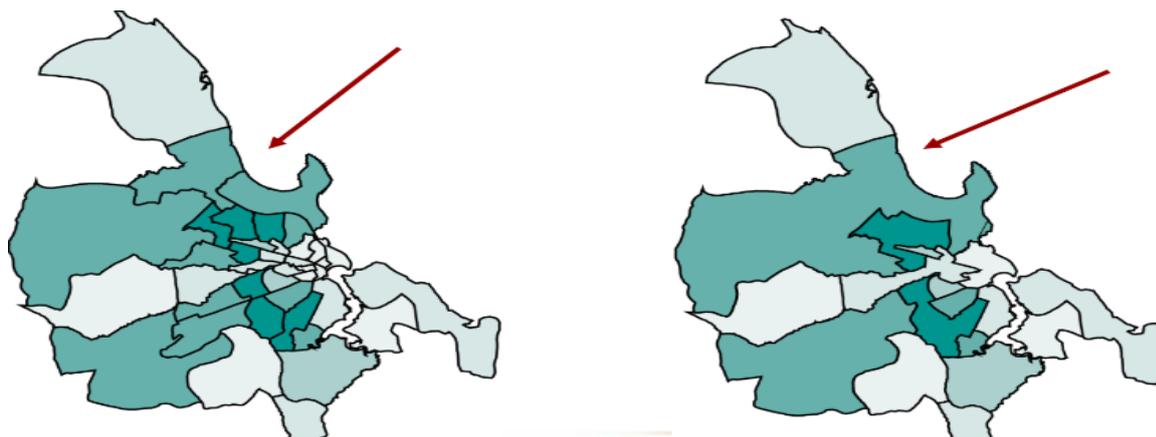
In left table, the original values are ranges. While in right table, the old values already were a classification.

**Automatic Classification** - In Automatic classification, the number of output classes are mentioned by the user. Computer selects the class break points on the basis of Equal frequency and Equal interval.

In Equal interval classification, class interval is fixed. And in Equal frequency classification, number of elements in a class are fixed.

**Reclassification – Merge** - After post processing of classification, the polygons can be merged, means polygons belonging to one class can be merged to make a single feature. Once the polygons are merged that means there is a change in geometry, spatial relationship.

Five classes of household income with original polygons intact



**Figure 5.11:** Polygons belonging to one class can be merged to make a single feature

**Source:** Study material IIRS Outreach Programme

---

## 5.4 SUMMARY

---

In this unit, we have discussed about data and their types, types of spatial data. Then we have discussed about raster data manipulation tools and techniques like data analysis environment in which an analysis mask is defined which restricts analysis to cells which does not contain the cell value "no data." Then we have discussed and learned about different raster data operations such as raster data management, raster data extraction and raster data generalization. After this, we have learned about physical distance measure operations, in which, allocation and direction and applications of physical distance measure operations are discussed.

Further, we have seen comparison of vector and raster based data analysis, which is compared on the basis of Overlay and Buffering operations. After this, we have learned about Mosaic technique, which is comprised of two or more images that have been combined or merged. By mosaicking multiple raster datasets together in ArcGIS, you can generate a single raster dataset. After this, we have learned about characteristics of mosaicked raster data.

in the end, we have discussed about Reclassification and their types. Then we have learned about Merge process in reclassification.

---

## 5.5 GLOSSARY

---

- **Analysis mask:** A mask that restricts analysis of raster data to cells which don't acquire the cell value of no data.
- **Mosaic:** An operation of raster data that can compile multiple input rasters into a single raster.
- **Physical distance:** Physical distance is a straight-line distance, which is in between the cells.

- **Physical distance measure operation:** An operation of raster data that measures distances of straight-line far from the source cells.
- **Raster data extraction:** An operation that extracts data from an existing raster using a collection of data, a pictorial entity, or a query expression.
- **Reclassification:** A local operation which classify the already grouped data or classified data, that process is called the reclassification.
- **Slice:** An operation of raster data which divides an uninterrupted raster into classes of equal-interval or equal-area.
- **Aggregate:** Aggregate is similar to resampling in that it provides an output raster with a large cell size than the input.

---

## 5.6 ANSWER TO CHECK YOUR PROGRESS

---

Q.1 What is data and describe the data types?

Q.2 Define vector and raster data?

Q.3 What is data analysis environment?

Q.4 Describe the raster data extraction process in detail?

Q.5 What is physical distance measure operation?

Q.6 Describe the applications of physical distance measure operations?

Q.7 What is Raster data generalization?

Q.8 Define Overlay operation in detail?

Q.9 What is User controlled classification?

Q.10 What is Automatic classification?

---

## 5.7 REFERENCES

---

1. Chang, K. T. (2009). Introduction to Geographic Information Systems, 5e.
2. Chang, K. T. (2019). Introduction to Geographic Information Systems, 9e.
3. Tomlin, C. D. (1990). *Geographic information systems and cartographic modelling* (No. 910.011 T659g). New Jersey, US: Prentice-Hall.
4. Beguería, S., & Vicente-Serrano, S. M. (2006). Mapping the hazard of extreme rainfall by peaks over threshold extreme value analysis and spatial regression techniques. *Journal of applied meteorology and climatology*, 45(1), 108-124.
5. Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation*. John Wiley & Sons.

6. [https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-a-mosaic.htm#:~:text=A % 20mosaic % 20is % 20a % 20combination,a % 20collection % 20of % 20raster % 20datasets.](https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-a-mosaic.htm#:~:text=A%20mosaic%20is%20a%20combination,a%20collection%20of%20raster%20datasets.)
7. [https://eclassems.iirs.gov.in/cms\\_admin/projectFile/23 % 20April % 202020 Spatial % 20Analysis % 20 % E2 % 80 % 93 % 20Introductory % 20Concept % 20and % 20Overview % 20by % 20Shri. % 20Prabhakar % 20Alok % 20Verma.pdf](https://eclassems.iirs.gov.in/cms_admin/projectFile/23%20April%202020_Spatial%20Analysis%20%E2%80%93%20Introductory%20Concept%20and%20Overview%20by%20Shri.%20Prabhakar%20Alok%20Verma.pdf)
8. [https://8945e053-a-62cb3a1a-sites.googlegroups.com/site/ignouhelpbooks302/Block-2 % 20Concept % 20of % 20Geospatial % 20Data.zip?attachauth=ANoY7copxVti-tN0OhwyM6gyE6j153NLKVnDQimQLI1KjBxW8\\_I97XfZ0aaLOGlvgXaHaAnxNEI8SgRCw15APiYqbBhzpA69jAd7YN3WHSabckIgjZCbM1f2F4CYFpDNEp yCpdLsadTFcXGZ69Y01WibVb7Dp5Ru6z2ePJv9JzgBzDDNFuqLXhsb067fFRI4PgQX-urtjdMjdcIbvfcC0imzXcAusOcRev-jb73kq6BjQJVfop-4kF37-iDqv09aUsxjPoMGdFs-&attredirects=0&d=1](https://8945e053-a-62cb3a1a-sites.googlegroups.com/site/ignouhelpbooks302/Block-2%20Concept%20of%20Geospatial%20Data.zip?attachauth=ANoY7copxVti-tN0OhwyM6gyE6j153NLKVnDQimQLI1KjBxW8_I97XfZ0aaLOGlvgXaHaAnxNEI8SgRCw15APiYqbBhzpA69jAd7YN3WHSabckIgjZCbM1f2F4CYFpDNEp yCpdLsadTFcXGZ69Y01WibVb7Dp5Ru6z2ePJv9JzgBzDDNFuqLXhsb067fFRI4PgQX-urtjdMjdcIbvfcC0imzXcAusOcRev-jb73kq6BjQJVfop-4kF37-iDqv09aUsxjPoMGdFs-&attredirects=0&d=1)
9. Chang, Kang-tsung Introduction to Geographic Information Systems 9th edition 2016 Mcgraw-Hill.
10. Lillesand, Thomas M., Ralph W. Kiefer, and Jonatham W. Chipman, 2004

---

## 5.8 TERMINAL QUESTIONS

---

- Q.1 Define Raster data operations in detail?
- Q.2 Define Physical distance measure operation in detail?
- Q.3 Compare the vector and raster based data analysis on the basis of different operations?
- Q.4 Give a detail description of mosaic operation?
- Q.5 Define the reclassification process?

---

## **UNIT 6 - RASTER DATA ANALYSIS-LOCAL, FOCAL, ZONAL AND GLOBAL**

---

**6.1 OBJECTIVES**

**6.2 INTRODUCTION**

**6.3 RASTER DATA ANALYSIS-LOCAL, FOCAL, ZONAL AND GLOBAL**

**6.4 SUMMARY**

**6.5 GLOSSARY**

**6.6 ANSWER TO CHECK YOUR PROGRESS**

**6.7 REFERENCES**

**6.8 TERMINAL QUESTIONS**

---

## **6.1 OBJECTIVES**

---

After reading this unit you will be able to understand:

- i. Different analysis based on raster data.
- ii. Find and execute different forms of raster data analysis based on requirements.
- iii. Raster data processing based on local, focal, zonal and global operations.

---

## **6.2 INTRODUCTION**

---

Raster analysis is similar to vector analysis in many ways. There are, however, some significant differences. The primary distinctions between raster and vector modelling are determined by the nature of the data models. Because datasets are stored in a common coordinate framework, all operations are possible in both raster and vector analysis. Every coordinate in the planar section is contained or falls within.

---

## **6.3 RASTER DATA ANALYSIS-LOCAL, FOCAL, ZONAL AND GLOBAL**

---

Raster analysis is similar to vector analysis in many ways. There are, however, some significant differences. The primary distinctions between raster and vector modelling are determined by the nature of the data models. Because datasets are stored in a common coordinate framework, all operations are possible in both raster and vector analysis. Every coordinate in the planar section is contained or falls within. In vector analysis, it is possible to carry out all operations because the features in one layer are explicitly located in relation to other layer existing features. The chirality or left and right hand of the arcs are part of the arc node vector data model (as shown in the polygon data model image from Spatial Data Model). Containment and superposition are inherent relations between layers as a corollary to this. For instance, in one layer, a point is located in another layer on one side of an arc or in another layer on or off the polygon.

On the other hand, raster analysis only strengthens its spatial relations at the location of the cell. Raster operations are generally carried out with multiple raster data sets, which are the

result of cell-by-cell computations. The output value for a single cell is typically independent of the input or output cell's value or location. Output cell values, for example, are in some cases influenced by the surrounding cells or cell groups. Raster data are particularly suitable for ongoing data. Data change across a landscape or surface smoothly. The raster data structures are much more concerned with phenomena such as chemical concentration, slope, elevation and look than with vector data structures. This makes many analyses more appropriate or possible only with raster data. This section and the following section will explain the basics and some of the most common analytical instruments for raster data processing. Spatial operations may occur with raster data. Although the actual calculations are substantially different from their vectors, their conceptual support is similar. The analysis of raster data covers geo-processing techniques with single as well as multiple layer operations.

**Single Layer Analysis-:** One of the first phases in a dataset is often reclassified or coded by raster analyses. Reclassification is basically a layer process in which all data pixels are assigned a new class or range value based on their original values. For example, for almost every cell within its reach, a different value is usually stored in an elevation grid. If each pixel value is added in several discrete classes (e.g., 0–100 = "1," 101–200 = "2," 201–300 = "3," and so on), these can be simplified. This simplification allows for fewer unique values and less expensive storage.

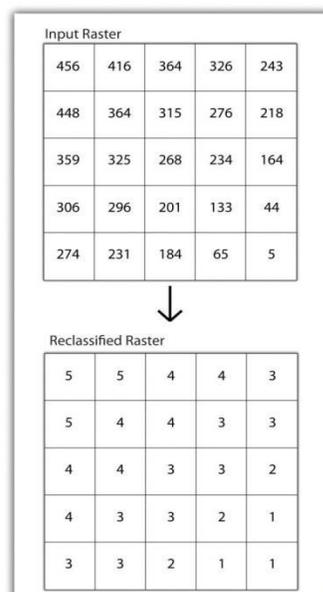


Figure 6.1 Raster Reclassification

Vector data commonly use different operations which always play an important role while analyzing different dataset. Buffering and overlay analysis are two of them which always enable researchers or professional to acquire their result in precise manner. Buffering is the process of producing an output dataset with a specified width zone (or zones) surrounding an input feature. In raster datasets, these input features are represented as a grid cell or a group of grid cells with a uniform value (e.g., buffer all cells with a value of 1). Buffers come in handy when determining the influence area around specific features. Vector data buffers produce a precise area of influence at a specified distance from the target feature, whereas raster buffers are approximations representing those cells within the target feature's specified distance range (Figure 6.2 "Raster Buffer around a Target Cell(s)"). The majority of geographic information systems (GIS) calculate raster buffers by establishing a grid for distance from the centre of the target cell(s) to the centre of the nearby cells and reclassifying those distances so that "1" represents the cells that comprise the original goal. These cells could be further classified by including values of "3," "4," "5," and so on, as the representative of multiple ring buffers (s).

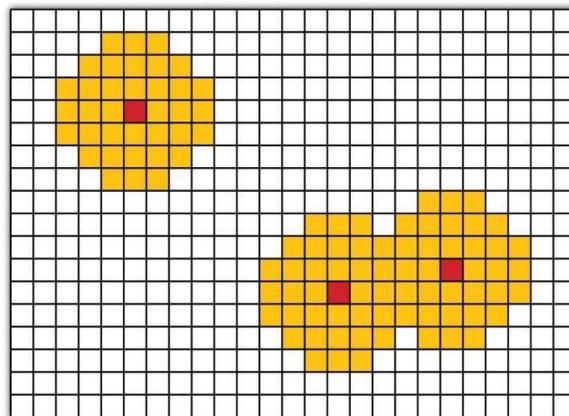


Figure 6.2 Raster Buffer around a Target Cell(s)

**Multiple Layer Analysis:** - A vector data set, like, raster can also be cut (Figure 6.3 "Clipping a Raster to a Vector Polygon Layer"). The input raster is overlaid with a vector polygon clip layer. The raster clip process leads to a raster identical to the raster, but which shares the raster clip layer.

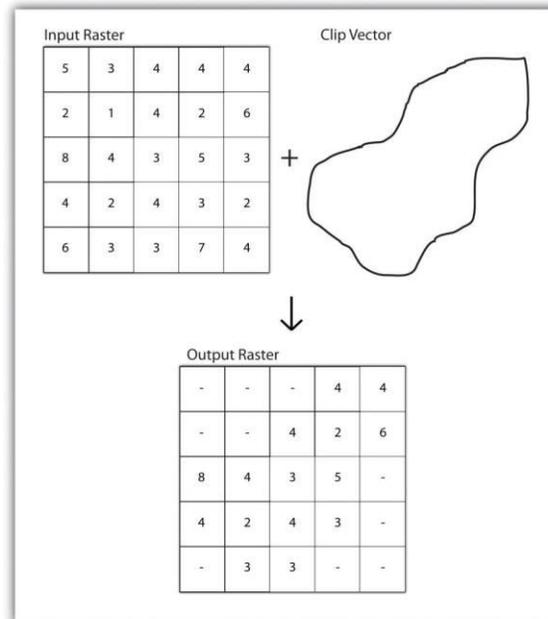


Figure 6.3 Clipping a Raster to a Vector Polygon Layer

### Scale of Analysis:

Raster analyses are basically of four types which differ from each other on the basis of its basic operations. All these operations are providing a unique set of functions which offers distinct functions to GIS analyst.

### Local Operations:

The output pixel is a weighted combination of the grey values of the pixels in the vicinity of the input pixel, hence the term local neighbourhood operation is found in many common image processing operations. The quarter size and the pixel weights determine the operator's action. This concept was introduced already when the discussion of realistic picture sampling considered pre-filtering. The discussion of transformations in the image will now be formalized and serves as a basis. On a single or multiple rasters, local operations can be performed. A single raster's local operation usually entails applying a mathematical transformation to each individual cell in the grid. For example, a researcher could obtain a digital elevation (DEM) model in which each cell value represents a feet elevation. If these elevations are preferred to be represented in metres, the cell values can be carried out locally with a simple, arithmetical transformation (original feet

elevation\*0.3048 i.e. new metre elevation). The output pixel is a weighted combination of the greyvalues of the pixels in the vicinity of the input pixel. The quarter size and the pixel weights determine the operator's action. On single or multiple rasters local operations can be performed. A local operation, if used on a single raster, usually takes the form of applying a mathematical transfer to each cell in the grid. A researcher, for example, could obtain a digital elevation model in which each cell value represents elevation in feet.

Using multiple rasters, analyses such as changes are possible over time. It is simple to remove these values and place the difference in the output raster that will show the variation in groundwater between these two times due to the two rasters with groundwater depth information on a parcel of land during 2000 and 2010. (Figure 8.5"Local Operation on a Raster Dataset"). However, as the number of input rasters grows, such local analyses can become slightly more complicated. In the Universal Soil Loss Equation, for example, a local mathematics formula for several rasters such as rainfall intensity, soil erodibility, and slope can be used (USLE).

Local operations are cell-by-cell operations, which form the core of raster data analysis. A local operation can create a new raster from one or more input rasters. A function relating to the input to the output is used or a classification table assigns the cell value of the new raster.

**Local Operations with a Single Raster** - In the case of single raster as an input, each cell value in the output raster is calculated by the local operation as a cell value in the input raster at the same place. As shown in Figure 6.4, large numbers of mathematical operators are available at GIS platform. The GIS function can be a computerised GIS tool or a mathematical operator.

Arithmetic	+, -, /, *, absolute, integer, floating-point
Logarithmic	exponentials, logarithms
Trigonometric	sin, cos, tan, arcsin, arccos, arctan
Power	square, square root, power

Table-: 6.1 Functions of Local Operation

For example, turning the floating point raster into an integer raster is a simple local operation which uses the integer operator to cell-by-cell truncates the cell value at the decimal point. Converting a slope raster measured in one degree is also a local operation, but a more

complicated mathematical expression is needed. In figure 8.5, slope  $p$  measured in % by slope dis converted to degrees measured in slope  $d$  by expression  $[\text{slope } d] = 57.296 \text{ os arctan } ([\text{slope } p]/100)$ . As computer packs usually use radial instead of trigonometric grade, the angular measure changes to degrees at constant 57,296 ( $360/2\pi$ ,  $\mu=3,1416$ ).

8.64	9.09	10.48	15.2	16.0	18.5
10.09	10.37	11.09	17.8	18.3	19.6
10.20	10.81	11.42	18.0	19.1	20.2

(a) (b)

Table:-6.2 Conversion of Slope from (a) percent to (b) degree

Local operation creates a new classification raster. Reclassification is also known as recoding or transforming by search tables (Tomlin 1990). There may be two methods of reclassification. A one-to-one change is the first method, which means that the output raster assigns a new cell value to the input raster. For example, the value of 1 in the output raster is allocated for irrigated croplands in a land-use raster. Three main purposes are used for reclassification. First, a reclassified raster can be made. For example, a raster may have 1 for a slope of 0 to 10 percent, 2 for 10 to 20 percent, and so on, rather than a continuous slope value. Second, a new raster can be created which contains a single category or value such as a 10-20 percent pitch. Thirdly, a new raster can be created, showing cell values in the raster input. A reclassified raster, for instance, can show the 1-5 ranking with 1 less adequate and 5 more appropriate.

Local Operations with Multiple Raster's - Compositing, overlaying, or superimposing maps are all terms used to describe local operations with multiple raster's (Tomlin 1990). Because local operations can work with multiple raster's, they are the vector-based overlay operations' equivalent. Multiple input raster's allow for a wider range of local operations than a single input raster. Aside from mathematical operators that can be applied to individual raster's, there are other measures that are based on raster's on the cell values or frequencies in the input raster's can also be calculated and stored in the output raster. Some of these measures, however, are restricted to raster's with numerical data. Statistics such as majority, minority, and number of unique values are also appropriate for raster's with numeric or categorical data. A majority output raster tabulates the most frequent cell value among the input raster's for each cell, while a minority raster tabulates the least frequent cell value. The number of different cell values is tabulated by a variety raster. There are no statistics or computation for

some local operations. A local operation known as Combine gives each single combination of input values a unique output value. Suppose there are three cell values for a slope raster (0, 20, 40 and 40 percent) and four cell values for an aspect raster (north, east, south, and west aspects). For every unique combination of slope and aspect, 1 creates a result raster with a value greater than 40 percent. Slope and southern aspect, 2 with 20 to 40% slope, southern aspect, etc. Local operations in GIS can easily be carried out through different software's. Arc GIS can be used to carry out these operations and result will be used for several other analyses.

### **Focal operations (Neighborhood Operations):-**

Focal or neighborhood operations produce a raster output dataset in which the outcome value for each cell location depends on the input value for the cells in a given neighborhood. The neighborhood is essentially a movable window, as each cell in the input is processed. The neighborhood's setup (size and shape) specifically determines what cells are to be used to calculate each output value surrounding the processing cell. The neighborhood is typically three by three cells, including the processing cell and its closest eight neighbor's. "Everything is related to everything else, but near things are more related than distant things," says Tobler's first geographic law. Neighborhood operations are a group of spatial analysis techniques that are frequently used and heavily rely on this concept. They examine the relationship between an object and similar objects on a point, line, or vector polygon. They can be used. Vector data sets are most commonly used for neighborhood analysis to carry out basic searches. For instance, a GIS could be used to establish the number of stores, given a point dataset that contains the convenience stores location within 5 Miles. When used with raster datasets, neighbourhood analyses are frequently more sophisticated. Moving windows, also known as filters or kernels, are used in raster analyses to calculate new cell values at every location across the extent of the raster layer. Depending on the type of output desired and the phenomenon being investigated, these moving windows can take many different forms. A rectangular, 3-by-3 moving window, Calculates the mean, standard deviation, sum, minimum, maximum, or range of values immediately surrounding a given "target" cell, for example. The target cell is the cell in the middle of the 3-by-3 moving window. Each cell in the raster is traversed by the moving window. As each central target cell passes through, the nine values in the 3-by-3 windows are used to calculate a new value. This new value is assigned to the same spot in the output raster as the previous one. You could expand the moving window to 5 by 5, 7 by 7, and so on if you wanted to investigate a larger sphere of influence around the target cells. In addition, the moving window does not have to be a rectangle. Other shapes used to calculate neighbourhood statistics

include the annulus, wedge, and circle.

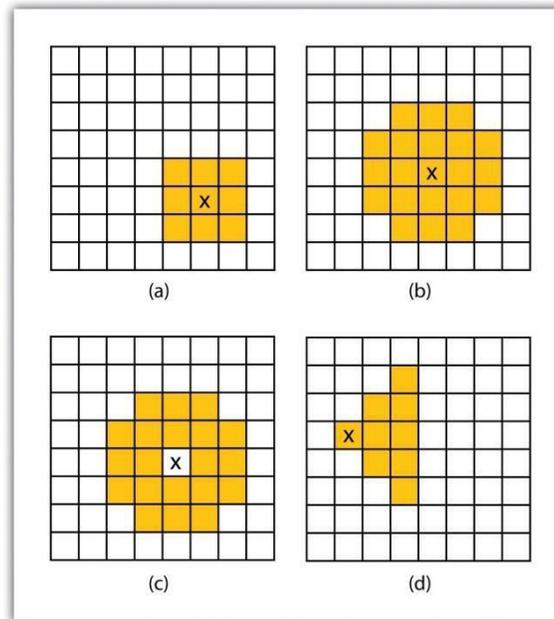


Figure:-6.4 Common Neighborhood Types around Target Cell “x”: (a) 3 by 3, (b)Circle, (c) Annulus, (d) Wedge

The cell values within the neighborhood are typically used in the calculation process and then the calculated value is assigned to the central cell. The focal cell is moved from one cell to another until all the cells are visited to complete a neighborhood operation on a grid. Different rules designed by the developers of GIS software are used in a raster-margin focal cell that does not have a district like a 3-by-3 rectangle. A simple rule is to use only cell values in the neighborhood available for computation (e.g. 6 rather than 9). Although an operation in the neighborhood works with a single raster, it does not work with multiple raster. Summary statistics including maximum, minimum, range, total, mean, median and standard deviation, as well as the table of measures such as majority, minority or diversion, can be found in the results of a neighborhood operation. These statistics and measures are the same as those for multiple raster local operations. On raster datasets, neighbourhood operations are commonly used for data simplification. Because the averaging process reduces the influence of outlying data values, an analysis that averages neighbourhood values would result in a smoothed output raster with dampened highs and lows. Neighborhood analyses, on the other hand, can be used to exaggerate

differences in a dataset. Edge enhancement is a type of neighbourhood analysis that looks at the moving window's range of values. A large range value indicates the presence of an edge within the window's boundaries, whereas a small range indicates the absence of an edge.

In raster data sets, neighbourhood operations are commonly used for data simplification. The effect of the outer data values is reduced through the average procedure, resulting in a smooth output raster with damped highs and lows. Alternatively, neighbourhood analysis can be used to exaggerate differences in a dataset. Edge enhancement is a type of quarter's analysis that examines the moving window's value spectrum. A large range indicates that there is an edge within the window's scope, whereas a small range indicates that there is no edge. The estimated value is assigned to the output raster block cells by a neighbourhood operation, which is a rectangle (block) operation.

Neighborhood operations can be important for studies requiring cells to select their neighbourhood features. For instance, the system of irrigation for gravity sprinklers requires information about a lifting drop within a circular cell neighborhood. Assume that a system requires 130 feet (40 metres) in height to make it financially viable within a distance of 0.5 mile (845 metres). By using a circle with a radius of 0.5 mile as a neighborhood and (high) as the data, an elevation raster operation in the neighborhood can answer the questions. A raster query can show which cells fulfill the criterion.

### **Zonal Operations:-**

An zonal operation is used in cell groups of similar value or features which are not surprisingly referred to as zones (for example, land parcels, political/municipal units, water bodies, types of soil/vegetation). The raster versions of the polygons could conceptualize such zones. Zonal raster is often reclassified into only a few categories by an input raster. A single raster or two overlaying rasters can be used for zonal operations. Given one input raster, the geometry of each raster area, like area, perimeter, space and centroid is measured by zonal operations given two zonal operation rasters, an input raster and a zonal raster, the output raster of the zonal operation summarizes the input raster cell values of each zone in the zonal raster.

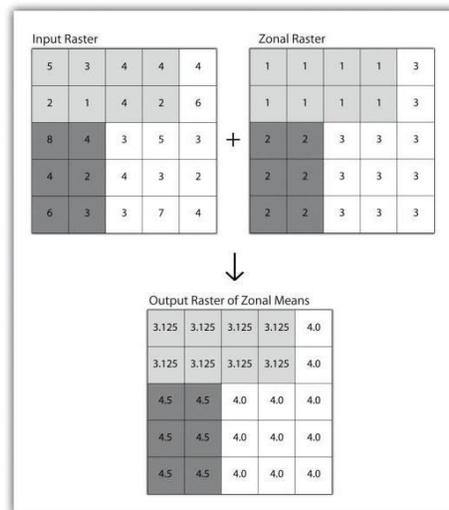


Figure 6.5 Zonal Operations on a Raster Dataset

An zonal operation is used in cell groups of similar value or features which are not surprisingly referred to as zones (for example, land parcels, political/municipal units, water bodies, types of soil/vegetation). The raster versions of the polygons could conceptualize such zones. Input raster is often reclassified in only a few categories by zonal raster's. A zonal operation is basically working with group of cells of same values or same feature present in any image. This operation is very useful when values in a raster are homogenous. A single raster or two overlays zonal operations can be used Raster's. Given one input raster, the geometry of each raster area, like area, perimeter, space and centroid is measured by zonal operations. Given two zonal raster's, an input raster and a zonal raster, an output raster is produced by the zonal operation which summarizes the cell values in each area in the input raster of a zonal raster.

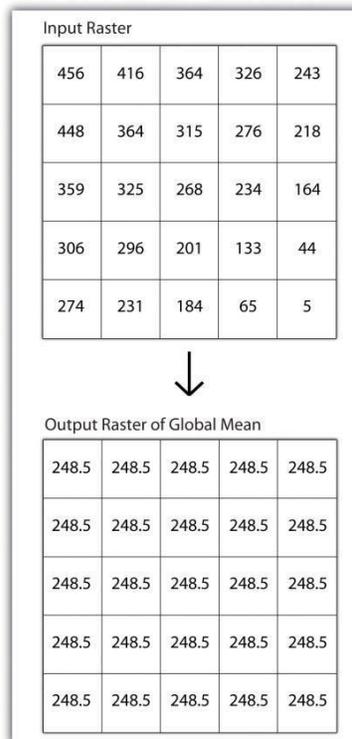
A zonal operation works with cell groups with the same value or similar characteristics. These groups are referred to as zones. Zones may be adjacent or non-adjacent. An adjoining area includes space-connected cells, whereas a non-contiguous area comprises separate cell areas. An example of an adjoining zone in a watershed raster is the spatially connection of cells belonging to the same watershed. An example of a noncontiguous zone is a land use raster in which a particular type of land may occur in various areas of the raster. A single raster or two rasters may be used for zonal operations. Given the single input raster, each zone's geometry such as area, perimeter, thickness, and centroid in the raster is measured by zonal operations. The area is the sum of cells falling within the zone times the size of the cell.

The perimeter of an adjacent zone is its length, while the perimeter of an adjacent area is the sum of each part's length. The thickness calculates the radius (in cells) of each zone of the largest circle. And the middle of a zone, at the intersection of the major axis and the small elliptical axis which is closest to the zone, is the geometric centre. Given two zonal rasters, an input raster and a zonal raster, an output raster is produced by the zonal operation which summarizes the cell values in each area in the input raster of a zonal raster. Summary statistics and measures cover areas, minima, sum, sum, range, median, minority, variety and standard deviations.

Zonal geometry measures such as the area, perimeter, thickness and centroid are especially useful for landscape ecology studies (Forman and Godron 1986; McGarigal and Marks 1994). Many other geometric measures from the area and perimeter can be derived. The perimeter/area ratio, for example, is a simple measure of the complexity of form in the landscape ecology used. In fields like landscape ecology, where geometry and spatial arrangement of habitat patches can have a significant effect on the type and number of species that may reside there, zonal operations and analyses can be valuable. In addition, zonal analyses can effectively quantify narrow habitat corridors which are important to regional flightless and migratory animal movement in densely urbanized areas.

**Global Operation:-**

Global operations are similar to zonal operations in which the whole raster data set is one zone. The basic statistical values of the raster as a whole are defined as typical world operations. For example, for the whole extent of the input raster the minimum, the maximum, the average, the range, and so on can be quickly calculated



**Figure 6.6** Global Operations on a Raster Dataset

---

## 6.4 SUMMARY

---

Raster analysis is similar to vector analysis in many ways, but there are some significant differences. In vector analysis, features in one layer are explicitly located in relation to other layer existing features. Containment and superposition are inherent relations between layers as a corollary to this. For instance, on one layer, a point is located in another layer on one side of an arc or on or off a polygon. Raster analysis reclassifies or recodes a data set often as one of the first steps. The reclassification is essentially a layer process by which all data pixels based on their original values have a new class or range value. This simplification makes it possible to achieve fewer unique values and cheaper storage. A zonal operation is used in cell groups of similar value or features which are not surprisingly referred to as zones. Given one input raster, the geometry of each raster area, like area, perimeter, space and centroid is measured by zonal operations. Zonal raster is often reclassified into only a few categories by an input rasters.

---

## 6.5 GLOSSARY

---

- **Raster Data-:** It is basically represent spatial data in the form of grid, cells, and pixels intabular format.
- **Raster Data Operation-:** Cell based analysis processed to extract unavailable raster information from satellite images.
- **Block Operation-:** It is a form of neighborhood operation which rectangular block to assign the eventual value to all block cells in the output raster.
- **Local Operation-:** It is basically cell by cell raster data analysis.
- **Zonal Operation-:** This form of raster analysis involves group of cells contains same values.
- **Reclassification-:** A local operation that reclassifies cell values of an input raster to create a new raster.

---

## 6.6 ANSWER TO CHECK YOUR PROGRESS

---

### Q-1-: What is Single Layer Analysis?

Ans-: Raster analysis reclassifies or recodes a data set often as one of the first steps. The reclassification is essentially a layer process by which all data pixels based on their original values have a new class or range value.

### Q-2-: What is Multiple Layer Analysis?

Ans-: The process of a raster clip leads to a raster that is identical to the raster but shares the scope of the polygon clip layer.

### Q-3-: What is Combine Raster?

Ans-: A local operation known as Combine gives each single combination of input values a unique output value.

### Q-4-: What is Edge Enhancement?

Ans-: Edge enhancement is a type of neighborhood analysis that examines the range of values in the moving window.

---

## 6.7 REFERENCES

---

1. Begueria, S., and S. M. Vicente- Serrano. 2006. Mapping the Hazard of Extreme Rainfall by Peaks over Threshold Extreme Value Analysis and Spatial Regression Techniques. *Journal of Applied Meteorology and Climatology* 45:108–24.
2. Brennan, J., and E. Martin. 2012. Spatial proximity is more than just a distance measure.

- International Journal of Human-Computer Studies 70:88–106. Che, V. B., M. Kervyn, C. E. Suh, K. Fontijn, G. G. J. Ernst,
3. M.-A. del Marmol, P. Trefois, and P. Jacobs. 2012. Landslide Susceptibility Assessment in Limbe (SW Cameroon): A Field Calibrated Seed Cell and Information Value Method. *Catena* 92:83–98.
  4. Congalton, R. G. 1991. A Review of Assessing the Accuracy of Classification of Remotely Sensed Data. *Photogrammetric Engineering & Remote Sensing* 37:35–46.
  5. Crow, T. R., G. E. Host, and D. J. Mladenoff. 1999. Ownership and Ecosystem as Sources of Spatial Heterogeneity in a Forested Landscape, Wisconsin, USA. *Landscape Ecology* 14:449–63.
  6. Forman, R T. T., and M. Godron. 1986. *Landscape Ecology*. New York: Wiley. Guan, Q., and K. C. Clarke. 2010. A General-Purpose Parallel Raster Processing Programming Library Test Application Using a Geographic Cellular Automata Model. *International Journal of Geographical Information Science* 24:695–722.
  7. Herr, A. M., and L. P. Queen. 1993. Crane Habitat Evaluation Using GIS and Remote Sensing. *Photogrammetric Engineering & Remote Sensing* 59:1531–38.
  8. Heuvelink, G. B. M. 1998. *Error Propagation in Environmental Modelling with GIS*. London: Taylor and Francis. Kar, B., and M. E. Hodgson. 2008.
  9. A GIS-based model to determine site suitability of emergency evacuation shelter. *Transactions in GIS* 12:227–48.
  10. Lillesand, T. M., R. W. Kiefer, and J. W. Chipman. 2007. *Remote Sensing and Image Interpretation*, 6th ed. New York: Wiley.
  11. Bhatta, (2008) *Remote Sensing and Gis* Oxford University Press
  12. Chang, Kang- tsung *Introduction to Geographic Information Systems* 5th edition 2009 McGraw- Hill.
  13. Lillesand, Thomas M., Ralph W. Kiefer, and Jonatham W. Chipman, 2004
  14. *Textbook of Remote Sensing and Geographical Information System*, M. Anji Reddy, Second Edition, Pp 1-23.
  15. C. P. Lo and Albert K. W. Yeung (2002) *Concepts and Techniques of Geographic Information Systems*, Upper Saddle River, New Jersey: Prentice Hall, 2002).

---

## 6.8 *TERMINAL QUESTIONS*

---

- Q-1 What is Raster Data Analysis?
- Q-2 What is reclassification?
- Q-3 What do you mean by focal operation with Raster Images?
- Q-4 Differentiate between Local and Zonal operations.
- Q-5 Explain Significance of Raster Data Operation.

---

**UNIT 7 - RASTER DATA ANALYSIS- ARITHMETIC OPERATIONS AND DECISION RULE BASED**

---

**7.1 OBJECTIVES**

**7.2 INTRODUCTION**

**7.3 RASTER DATA ANALYSIS- ARITHMETIC OPERATIONS AND DECISION RULE BASED**

**7.4 SUMMARY**

**7.5 GLOSSARY**

**7.6 ANSWER TO CHECK YOUR PROGRESS**

**7.7 REFERENCES**

**7.8 TERMINAL QUESTIONS**

---

## 7.1 OBJECTIVES

---

After reading this unit you should be able to understand:

- I. Various functions of raster data analysis.
- II. Functions of arithmetic operation and its implications with respect to GIS data.
- III. Various arithmetic operators in GIS
- IV. Significance of decision rule based analysis for raster data.

---

## 7.2 INTRODUCTION

---

Operational procedures and quantitative methods for the analysis of spatial data in raster format is always important to understand and discuss. In raster analysis, geographic units are regularly spaced, and the location of each unit is referenced by row and column positions. Because geographic units are of equal size and identical shape, area adjustment of geographic units is unnecessary and spatial properties of geographic entities are relatively easy to trace. All cells in a grid have a positive position reference, following the left-to-right and top-to-bottom data scan. Every cell in a grid is an individual unit and must be assigned a value. Depending on the nature of the grid, the value assigned to a cell can be an integer or a floating point. When data values are not available for particular cells, they are described as NODATA cells. NODATA cells differ from cells containing zero in the sense that zero value is considered to be data. The regularity in the arrangement of geographic units allows for the underlying spatial relationships to be efficiently formulated. For instance, the distance between orthogonal neighbors (neighbors on the same row or column) is always a constant whereas the distance between two diagonal units can also be computed as a function of that constant. Therefore, the distance between any pair of units can be computed from differences in row and column positions. Furthermore, directional information is readily available for any pair of origin and destination cells as long as their positions in the grid are known.

The mathematics that underpins all geographical analysis involves the application of rules, most of which are straightforward. Mathematics makes use of symbols. We will add and explain other symbols later. In this chapter we will consider arithmetic, which is concerned with numerical calculations such as adding, subtracting, multiplying, and dividing.

The whole of arithmetic is based essentially on seven axioms, as shown in Box 2.1. Outside arithmetic, these axioms may not apply, for instance, when two rain drops running down a window pane come together to make one rain drop so that in symbolic form:  $1+1=1$ .

Furthermore, computer programmers often write “ $N = N + 1$ ,” meaning “Take the number in the box labeled N, add one to that number and put it back in the box labeled N”; although partially an arithmetic operation, the use of the “=” sign has a different meaning from that which we are considering here.

Arithmetic map operations are very common procedures used in GIS to combine raster maps resulting in a new and improved raster map. It is essential that this new map be accompanied by an assessment of uncertainty. This paper shows how we can calculate the uncertainty of the resulting map after performing some arithmetic operation. Actually, the propagation of uncertainty depends on a reliable measurement of the local accuracy and local covariance, as well. In this sense, the use of the interpolation variance is proposed because it takes into account both data configuration and data values. Taylor series expansion is used to derive the mean and variance of the function defined by an arithmetic operation. We show exact results for means and variances for arithmetic operations involving addition, subtraction and multiplication and that it is possible to get approximate mean and variance for the quotient of raster maps.

#### **Advantages of using the Raster Format in Spatial Analysis:-**

- **Efficient processing:** Because geographic units are regularly spaced with identical spatial properties, multiple layer operation can be processed very efficiently.
- **Numerous existing sources:** Grids are the common format for numerous sources of spatial information including satellite imagery, scanned aerial photos, and digital elevation models, among others. These data sources have been adopted in many GIS projects and have become the most common sources of major geographic databases.
- **Different feature types organized in the same layer:** For instance, the same grid may consist of point features, line features, and area features, as long as different features are assigned different values.

#### **Grid Format Disadvantages:**

- **Data redundancy:** When data elements are organized in a regularly spaced system, there is a data point at the location of every grid cell, regardless of whether the data element is needed or not. Although, several compression techniques are available, the advantages of gridded data are lost whenever the gridded data format is altered through compression. In

most cases, the compressed data cannot be directly processed for analysis. Instead, the compressed raster data must first be decompressed in order to take advantage of spatial regularity.

- **Resolution confusion:** Gridded data give an unnatural look and unrealistic present at ionone less the resolution is sufficiently high. Conversely, spatial resolution dictates spatial properties. For instance, some spatial statistics derived from a distribution may be different, if spatial resolution varies, which is the result of the well-known scale problem.
- **Cell value assignment difficulties:** Different methods of cell value assignment may result in quite different spatial patterns

Map algebra is an informal and commonly used scheme for manipulating continuously sampled (i.e. raster) variables defined over a common area. It is also a term used to describe calculations within and between GIS data layers, according to some mathematical expression, to produce a new layer; it was first described and developed by Tomlin. Map algebra can also be used to manipulate vector map layers, sometimes resulting in the production of a raster output. Although no new capabilities are brought to GIS, map algebra provides an elegant way to describe operations on GIS datasets. It can be thought of simply as algebra applied to spatial data which, in the case of raster data, are facilitated by the fact that a raster is a geo-referenced numerical array.

Map Algebra models the surface of the earth as a multitude of independent, coincident layers or themes. The layers interact according to mathematical models and are typically based on real world observations. Planners develop layers on development and population (Steinitz et al. 1976). Social scientists develop layers on demographics, ethnicity, and economic factors (Mc Harg 1969). Applying Map Algebra model to input layers produces a new layer, which maybe a physical map sheet, a vision perceived through a stack of my lars on a light table, or an electronic dataset displayed on a computer screen. Regardless of mechanism, the result allows its users to explain complex phenomena, predict trends, or make adjustments to the model.

However it is the mechanism which bounds usability of Map Algebra. How easy it is for scientists to perform simple tasks? Can complex models be developed and tested? Historically layers were plotted on individual transparent maps which, when super imposed and registered

Provide visually integrated view of the data. The manual process of map overlay is slow and tedious.

The ability to express problems in a formal mannerism necessary part of solving problems with computers. A Geographic Information System is the best computer tool for solving geographical problems (GIS). Basic functionality for visualizing, managing, and manipulating spatially referenced data is provided by such systems. A computer language is used to express problem solving, either one provided by the system or one that interoperates with it. Not only for advanced spatial analysis problems, but also for many ad hoc queries, GIS users are faced with the task of writing programmes as a concrete formulation of their particular problem. Given that the majority of GIS users are land-related professionals rather than programmers, it is critical to make the language interface to a GIS as simple and intuitive as possible. Writing programmes can be difficult for inexperienced users. Even if the problem being solved has a systematic or scientific approach, expressing it in a programme is a difficult intellectual task. The following are two primary reasons for this:

a) Representation mismatch between the object level representation of spatial data in the programming language and the application view. When an application presents information in one way but the programming environment for accessing and manipulating that information is different, this problem occurs. A map organized into thematic layers is a popular way to present information in GIS, whereas in a programming environment, the user is presented with tables containing records. This has an impact on how queries are phrased. For example, the application interface can provide associative access by querying feature properties, whereas the programming language can provide data access by retrieving records from a table based on relative record numbers.

b) A mismatch between how a user expresses a problem and how programming languages implement the solution exists. When a user expresses a problem in set theoretic terms, but is forced to solve it in an application programme as a series of operations on individual records, this happens. A programmatic approach to solving a problem involving detailed actions in strict operational order is required by many 16 Proceedings of Geo Computation programming languages.

---

### **7.3 RASTER DATA ANALYSIS- ARITHMETIC OPERATIONS AND DECISION RULE BASED**

---

Data Types (Arithmetic Operations)-:

This section focuses on map algebra operations available for gridded datasets, such as those implemented in ESRI's Spatial Analyst extension for Arc View. Similar operations are available in other grid-based GIS as, such as GRASS (Geographic Resources Analysis Support System).

The data associated with any grid cell can be of any type whatsoever. It is conceptually useful to divide data types into several classes, however. These include:

- **Categorical data:** These are non-numerical data. Grids that classify land use or land cover exemplify this category. Other examples are **proximity grids** (values identify then nearest object) and **feature grids** (only two values are possible: one value for cells where features occur, another value—typically zero or No Data—where features do not occur).
- **Integral data:** These data may be relative **ranks** or preferences or they may be counts of occurrences or observations, for example. Thus, what they measure is in here only integral.
- **Floating-point ("real" data).** These typically represent a real surface, such as elevation, or the values of a scalar function (a "conceptual surface," if you will). Examples of such functions would be temperature, slope, amount of sunlight received per year, distance to the nearest feature, population density.
- **Vector data:** These are ordered tuples of real values that represent fields of directions. For example, hydraulic gradients (for two-dimensional groundwater models), wind velocities (again for two-dimensional models), and ocean currents are two-dimensional vector fields. Vector data may have more than two dimensions, even though they are defined over a strictly two-dimensional domain. For example, models using

astronomical data, such as climate models, may make use of information about the three-dimensional location (on the earth's surface) of each grid point.

(Scientific visualization systems usually have built-in support for vector data, whereas most GIS es require the modeler to represent vector data as an ordered collection of floating-point grids).

### **Working with null data:**

An essential part of map algebra or spatial analysis is the coding of data in such a way as to eliminate certain areas from further contribution to the analysis. For instance, if the existence of low-grade land is a prerequisite for a site selection procedure, we then need to produce a layer in which areas of low-grade land are coded distinctively so that all other areas can be removed. One possibility is to set the areas of low-grade land to a value of 1 and the remaining areas to 0. Any processes involving multiplication, division or geometric mean that encounters the zero value will then also return a zero value and that location (pixel) will be removed from the analysis. The opposite is true if processing involves addition, subtraction or arithmetic mean calculations, since the zero value will survive through to the end of the process. The second possibility is to use a null or No Data value instead of a zero. The null is a special value which indicates that there is no digital numerical value. In general, unlike zero, any expression will produce a null value if any of the corresponding input pixels have null values. Many functions and expressions simply ignore null values, however, and in some circumstances this may be useful, but it also means that a special kind of function must be used if we need to test for the presence of (or to assign) null values in a dataset. For instance, within ESRI's Arc GIS, the function ISNULL is used to test for the existence of null values and will produce a value of 1 if null, or 0 if not. Using ER Mapper's formula editor, null values can easily be assigned, set to other values, made visible or hidden. Situations where the presence of nulls is disadvantageous include instances where there are unknown gaps in the dataset, perhaps produced by measurement error or failure. Within map algebra, however, the null value can be used to great advantage since it enables the selective removal or retention of values and locations during analysis.

Table 7.1 Operations categorized according to their spatial or non-spatial nature.

Output	Spatial attributes involved?	
	Yes	No(not necessarily)
Map or image	Neighborhood processing (filtering), zonal and focal operations, mathematical morphology	Reclassification, rescaling (unary operations), overlay (binary operations), thresholding and density slicing.
Tabular	Spatial autocorrelation and variograms	Various tabular statistics (aggregation, variety) and tabular modeling (calculation of new fields from existing ones), scatter graphs

(Source: After Bonham-Carter, 2002)

**Logical and conditional processing:**

These two processes are quite similar and they provide a means of controlling what happens during some function. They allow us to evaluate some criterion and to specify what happens next if the criterion is satisfied or not. Logical processing describes the tracking of true and false values through a procedure. Normally, in map algebra, a non-zero value is always considered to be a logical true, and zero, a logical false. Some operators and functions may return either logical true values (1) or logical false values (0), for example relational and Boolean operators. The turn of a true or false value acts as a switch for one or other consequence within the procedure. Conditional processing allows that a particular action can be specified, according to the satisfaction of various conditions; if the conditions are evaluated as true then one action is taken, and an alternative action is taken when the conditions are evaluated as false. The conventional if-then-else statement is a simple example of a conditional statement:

If  $i < 16$  then 1 else null where  $i = \text{input pixel } d_n$

Conditional processing is especially useful for creating analysis „masks“. In Fig. 24.1, each input pixel value is tested for the condition of having a slope equal to or less than 15°. If the value tests true (slope angle is 15° or less), a value of 1 is assigned to the output pixel. If it tests false (exceeds 15°), a null value is assigned to the output pixel. The output could then be used as a mask to exclude areas of steeper slopes and allow through all areas of gentle slopes, such as might be required in fulfilling the prescriptive criteria for a sites election exercise.

### **Other types of operator:**

Expressions can be evaluated using arithmetic operators (addition, subtraction, logarithmic, trigonometric) and performed on spatially coincident pixel DN values within two or more input layers (Table 7.2). Generally speaking, the order in which the input layers are listed denotes the precedence with which they are processed; the input or operator listed first is given top priority and is performed first, with decreasing priority from left to right.

A relational operator enables the construction of logical functions and tests by comparing two numbers and returning a true value (1) if the values are equal or false (0) if not. For example, this operator can be used to find locations within a single input layer with DN values representing a particular class of interest. These are particularly useful with discrete or categorical data.

A Boolean operator, for example AND, OR or NOT , also enables sequential logical functions and tests to be performed. Like relational operators, Boolean operators also return true (1) and false (0) values. They are performed on two or more input layers to select or remove values and locations from the analysis. For example ,to satisfy criteria with in a slopest ability model, Boolean operators could be used to identify all locations where values in one input representing slope are greater than 40° AND where values in an elevation model layer are greater than 2000m(asin Fig. 7.2a).

Logical operators involve the logical comparison of the two inputs and assign a value according to the type of operator. For instance, for two inputs (A and B)ADIFFB assigns the value from A to the output pixel if the values are different or a zero if they are the same. An expression AOVER B assigns the value from A if a non-zero value exists; if not then the value from B is assigned to the output pixel. A combinatorial operator finds all the unique combinations of values among the attributes of multiple input rasters and assigns a unique value to each

combination in the output layer. The output attribute will contain fields and attributes from all the input layers.

All these operators can be used, with care, alone or sequentially, to remove, test, process, retain or remove values (and locations) selectively from datasets alone or from within a spatial analysis procedure.

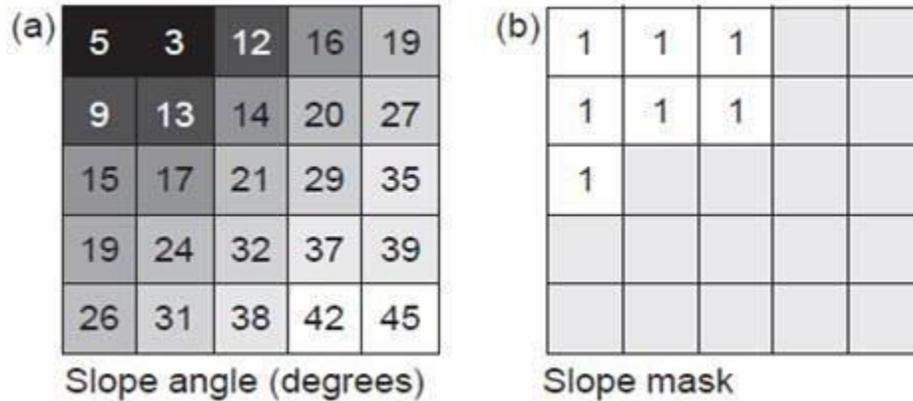


Fig.7.1.Logical test of slope angle data, for the condition of being no greater in value than 15°:  
 (a) slope angle raster and (b) slope mask (pale grey blank cells indicate null values).  
 (Source: Liu, and Mason, 2009)

Table 7.2. Summary of common arithmetic, relational, Boolean, power, logical and combinatorial operator.

Arithmetic	Relational (return true/false)	Boolean (return true/false)
+, Addition	==, EQ Equal	^, Not complement
-, Subtraction	^=, <>, NE No tequal	Logical AND
*, Multiplication	<=, LT Less than/equal to	OR
/,	<=, LE Less than/equal to	OR
	>, GT Greater than	!, XOR
		Logical

Division MOD, Modulus	>=, GE greaterthan/equalto	XOR
Power	Logical	Combianational
Sqrt, Square rootSqr, Square Pow, Raisedtoapo wer	DIFF, Logicaldifference IN{list}, Contained inlist OVER, Replace	CAND, CombinationalANDC OR, Combinational ORCXOR, CombinationalXO R

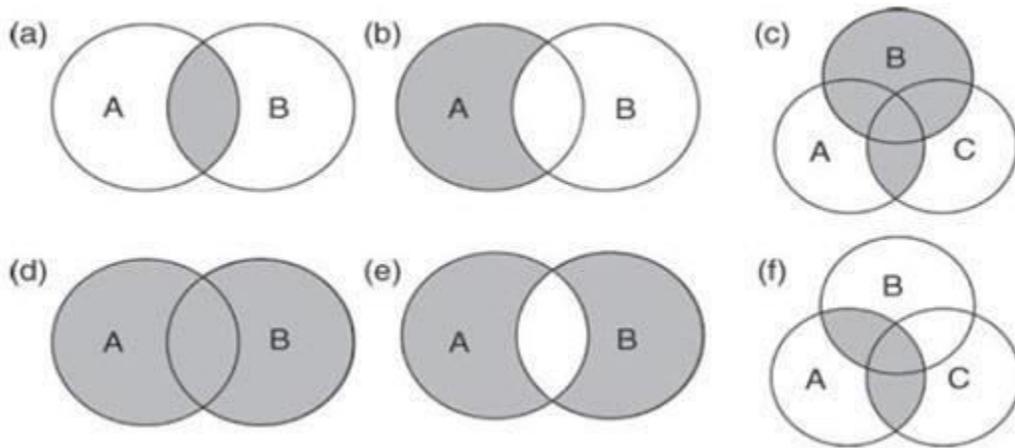


Fig. 7.2. Use of Boolean rules and set theory within map algebra; here the circles represent the feature classes A, B and C, illustrating how simple Boolean rules can be applied to geographic data sets, and especially rasters to extract or retain values, to satisfy a series of criteria:(a)A AND B (intersection or minimum ); (b) A NOT B; (c) (A AND C) OR B; (d) A OR B (union or maximum);(e) A XOR B; and (f)A AND (B OR C).(Source: Liu, and Mason, 2009)

Table7.3. Summary of local operations.

Type	Includes	Example
Primary	Creation of a layer from nothing	Rasters of constant value or containing randomly generated values
Unary	Conversion of units of measurement and as intermediary steps of spatial analysis	Rescaling, negation, comparing or applying mathematical functions reclassification
Binary	Operations on ordered pairs of numbers in matching pixels between layers	Arithmetic and logical combinations of rasters
N-ary	Comparison of local statistics between several rasters (many to one or many to many)	Change or variety detection

(Source: Liu and Mason, 2009)

### Local Operations:

A local operation involves the production of an output value as a function of the value(s) at the corresponding locations in the input layer(s). These operations can be considered point operations when performed on raster data, i.e. they operate on a pixel and its matching pixel position in other layers, as opposed to groups of neighbouring pixels. They can be grouped into those which derive statistics from multiple input layers (e.g. mean, median, minority), those which combine multiple input layers, those which identify values that satisfy specified criteria or the number of occurrences that satisfy specified criteria (e.g. greater than or less than), or those which identify the position in an input list that satisfies a specified criterion. All types of operator

previously mentioned can be used in this context. Commonly they are subdivided according to the number of input layers involved at the start of the process. They include primary operations where nothing exists at the start, to n-ary operations where n layers may be involved; they are summarized in Table 24.3 and illustrated in Fig. 24.3.

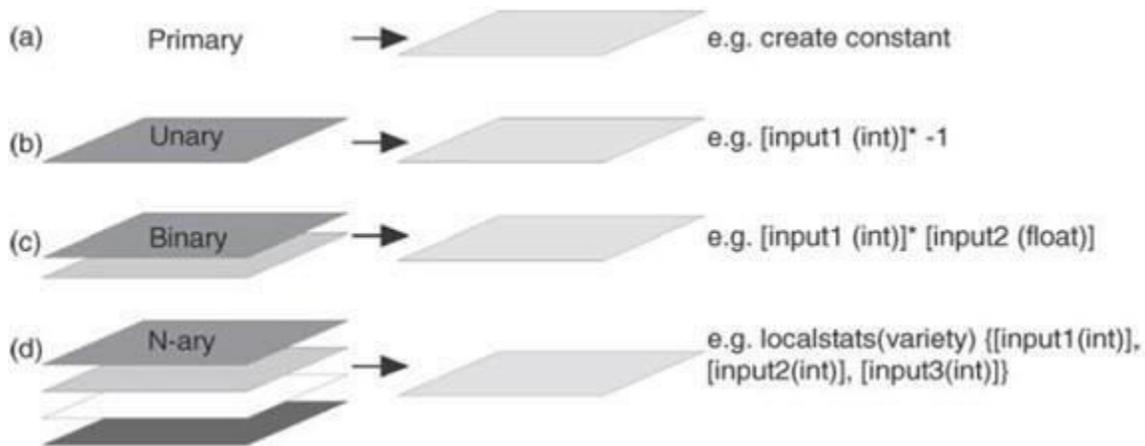


Fig. 7.3. Classifying map algebra operations in terms of the number of input layers and some examples.(Source: Liu and Mason, 2009)

### Primary operations:

This description refers primarily to operations used to generate a layer, conceptually from nothing, for example the creation of a raster of constant value, or containing randomly generated numbers, such as could be used to test for error propagation through some analysis. An output pixel size, extent, data type and output DN value (either constant or random between set limits) must be specified for the creation of such a new layer.

### Unary operations:

These operations act on one layer to produce a new output layer and they include tasks such as rescaling, negation, comparison with other numbers, application of functions and reclassification. Rescaling is especially useful in preparation for multi-criteria analysis where all the input layers should have consistent units and value range: for instance, in converting from byte data, with 0 to 255 value range, to a percentage scale (0-100) or a range of between 0 and 1, and vice versa.

Negation is used in a similar context, in modifying the value range of a dataset from being entirely positive to entirely negative and vice versa. Comparisons create feature grids: the places where the comparison is true can be considered features on the earth's surface. They map the regions where a logical condition (the comparison) holds. These could be regions where, say, ozone concentrations exceed a threshold, ocean depths are below a certain target, or land use equals a given code. Mathematical functions are useful for changing the visualization of a grid. An equal interval classification using the square roots of the values will differ from an equal interval classification of the values themselves, for instance. Functions are also important as intermediate steps in many models. Reclassification is especially significant in data preparation for spatial analysis, and so deserve rather more in-depth description, but all these activities can be and are commonly carried out in image processing systems.

To illustrate different applications succinctly, suppose that three grids appear in the current view: "Integer" is an integer grid, "Float" is a floating-point grid, and "Indicator" is an integer grid containing only 0, 1, and No Data values. A value of 0 can be interpreted as a logical "false" and a value of 1 as a logical "true". In practice, of course, we will replace these names by the names of our themes.

- **Rescale** a grid: that is, Multiply all its values by a constant value.

[Float] \* 3.1415927 Multiply all values by

Pi[Integer].Float\*(39.37/12)Convert meters to feet[Integer]\* (-1)Negate

all values

Not [Indicator] Negate all logical values:0becomes1,1becomes0

- **Compare** a grid to a constant value. The result of a comparison is true, 0 where the comparison is false, and No Data where the original value is No Data

[Float] < 1 Returns 1 where values are less than 1, otherwise returns

0[Integer]=0Converts all zeros to ones and all other values to zeros

- Apply a mathematical (or logical) **function** to a grid, cell by cell.

[Float].Cos Computes the cosine of each value (interpreted as radians). [Float].IntRounds all values and converts the result to an integer grid.

[Float].Sqrt Computes the square root of each value. Negative values return No Data (because the square root is not defined for negative values).

[Float].IsNull Returns 1 at all cells with No Data values, otherwise returns 0.

### **Binary operations:**

Binary numeric operations act on ordered pairs of numbers. Likewise, binary grid operations act on the pairs of numbers obtained in each set of matching cells. The resulting grid is defined only where the two input grids overlap.

Suppose there are several floating-point grids represented by themes named "Float", "Float1", "Float2", and soon; with a similar supposition for integer and logical grids.

- **Mathematical operators**

[Float] + [Integer] Converts the values in [Integer] to floats, then performs the additions.

- **Logical operators**

[Float1]

< [Float2] Returns 1 in each cell where [Float1]'s value is less than [Float2]'s value; otherwise, returns 0.

[Indicator1] And [Indicator2] Returns 1 where both values are non-zero otherwise returns 0.

This description refers to operations in which there are two input layers, leading to the production of a single output layer. Overlay refers to the combination of more than one layer of data, to create one new layer. The example shown in Fig. 7.4 illustrates how a layer representing average rainfall, and another representing soil type, can be combined to produce a simple, qualitative map showing optimum growing conditions for a particular crop. Such operations are equivalent to the application of formulae to multiband images, to generate ratios, differences and other inter-band indices and as mentioned in relation to point operations on multi-spectral

images, it is important to consider the value ranges of the input bands or layers, when combining their values arithmetically in some way. Just as image differencing requires some form of stretch applied to each input layer, to ensure that the real meaning of the differencing process is revealed in the output, here we should do the same. Either the inputs must be scaled to the same value range, or if the inputs represent values on an absolute measurement scale then those scales should have the same units.

The example shown in Fig. 7.4 represents two inputs with relative values on arbitrary nominal or ordinal (Fig. 7.4a) and interval (Fig. 7.4b) scales. The resultant values are also given on an interval scale and this is acceptable providing the range of potential output values is understood, having first understood the value ranges of the inputs, since they may mean nothing outside the scope of this simple exercise.

Another example could be the combination of two rasters as part of a cost-weighted analysis and possibly as part of a wider least cost pathway exercise. The two input rasters may represent measures of cost, as produced through reclassification of, for instance, slope angle and land value, cost here being a measure of friction or the real cost of moving or operating across the area in question. These two cost rasters are then aggregated or summed to produce an output representing total cost for a particular area (Fig. 7.5).

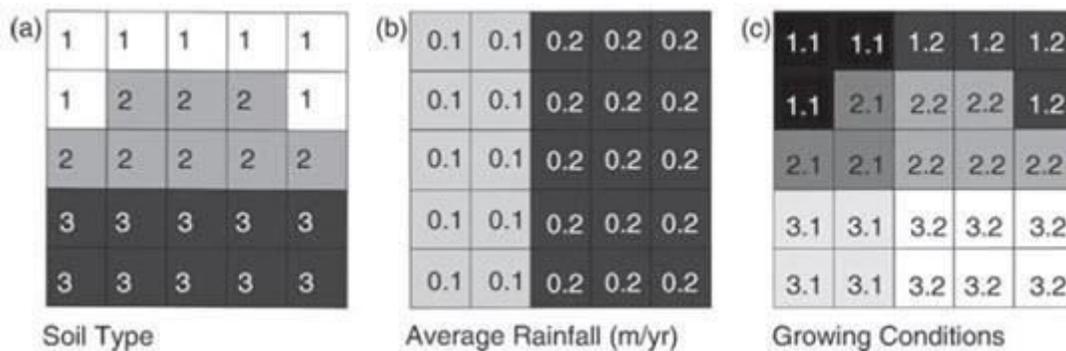


Fig. 7.4. An example of a simple overlay operation involving two input rasters: (a) an integer raster representing soil classes (class 2, representing sandy loam, is considered optimum); (b) a floating-point raster representing average rainfall, in metres per year (0.2 is considered optimum); and (c) the output raster derived by addition of a and b to produce a result representing conditions for a crop; a value of 2.2 (2 + 0.2), on this rather arbitrary scale,

represents optimum growing conditions and it can be seen that there are five pixel positions which satisfy this condition.(Source:Liu and Mason, 2009)



Fig. 7.5 (a) Slope gradient in degrees; (b) ranked (reclassified) slope gradient constituting the first cost or friction input; (c) ranked land value (produced from a separate input land-use raster) representing the second cost or friction input; and (d) total cost raster produced by aggregation of the input friction rasters ( $f_1$  and  $f_2$ ). This total cost raster could then be used within a cost-weighted distance analysis exercise.(Source: Liu and Mason, 2009)

### Local statistics:

When we have many related grids defined in the same region, we often want to assess change: at each cell, how varied are the grid results? How large do they get? How small? What is the average? These questions make sense for numerical data.

For grids with ordinal data--that is, values that can be ordered, but which may not have any absolute meaning--you can still ask about order statistics. These are the relative rankings of values within the ordered collections of values observed at each cell. For grids with categorical data, you might want to know at each cell whether one category predominates throughout the collection of grids and how many different categories actually appear at the cell's location. (Liu, and Mason,2009)

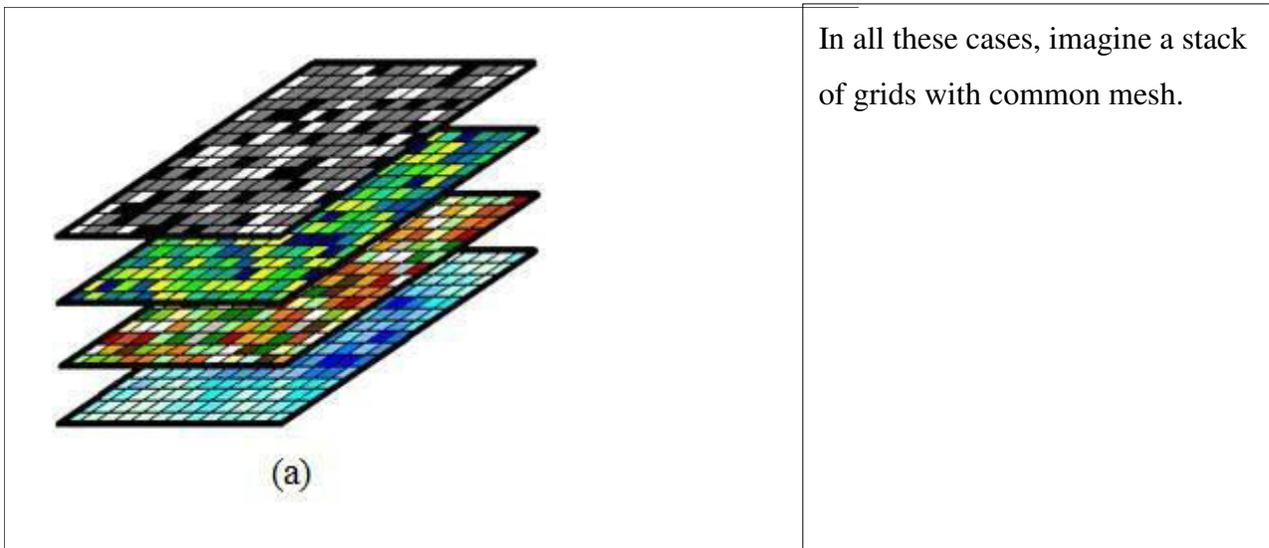


Fig.7.6(a)stack of grids with common mesh

The Spatial Analyst syntax for some of these requests is strange, because it wants to force expressions into the form "a Grid. Request (list of other grids)". This is inherently asymmetric because it singles out one grid in the collection to play the role of the object ("a Grid") to which the calculation is applied and leaves the other grids in the role of a list of arguments ("list of other grids"). Despite this syntax, for some requests, such as the local statistics, there is no asymmetry in the calculation itself: all the grids are equivalent. For some other requests, there is an asymmetry in the calculation: one grid plays a special role.

Spatial Analyst constructs lists with curly braces { } and separates the elements by commas.

- **Compute local statistics**

[Float]. Local Stats (#GRID\_STATATYPE\_MAX, {[Float1], [Float2], [Float3]}) Computes the largest value among four grids.

[Float]. Local Stats (#GRID\_STATATYPE\_MEDIAN, {[Float1], [Float2]}) Computes the median of three values.

[Integer]. Local Stats (#GRID\_STATATYPE\_MAJORITY, {[Integer1], [Integer2]}) Computes the value occurring the most times (out of the three input values at each cell). If two or more values occur an equal number of times, Spatial Analyst returns No Data.

The Majority statistic evidently is not very useful when many ties occur: that is, when there are many cells where two or more values occur equally often.

Compare one grid (a "base" grid) to many others simultaneously

[Float]. Grids Greater than ({[Float1], [Float2], [Float3], [Float4]}) For each base cell in [Float], computes the number of times corresponding cells from [Float1], ..., [Float4] exceed (and do not equal) the base cell's value. There is a corresponding Grids Less Than operator.

Combine the values of two grids based on values at a third grid

[Indicator].Con([Float1],[Float2]) Creates a grid with the values of [Float1] where [Indicator] is non zero and with the values of [Float2] where [Indicator] is zero.

The Con request is especially useful. The result of Con, by default, is the second grid ([Float2] or [Mosaic] in the examples). However, at cells where [Indicator] is true, the values of the first grid ([Float1] or [Average]) are "painted" over the default values. Thus the Con request is a natural vehicle for selectively editing grids.

### **Decision Based Rule-Programming:**

Different types of data abstractions and procedural abstractions are used in programming languages. Different languages place a greater emphasis on one characteristic over another. There are four major paradigms to consider: Geo Computation Proceedings

- Logic Programming
- Functional Programming
- Rule-Based Programming
- Object-Oriented Programming

### Logic Programming:

To solve problems, logic programming employs rules of exact logic, or to be more precise, rules of first order predicate logic. Problems are expressed as statements to represent beliefs about the world that we hold. A set of logical terms and logical connectors make up the statements. A truth table contains the rules for evaluating statements.

In first order predicate logic all objects belong to a single universe. This leads to a characteristic of “flatness” in pure logical languages. All objects are universal and so are the axioms by which they are related. There is no procedural abstraction in first order predicate logic.

In practice, logic programming languages use some procedural mechanisms to interpret logical statements. The most popular of these programming languages is PROLOG (Bratko, 1990). A logical statement is expressed as a Horn clause consisting of a conclusion head “C” and several conditional terms “B” in the body. They have the form:

“B1 and B2 and B3 ... and BN implies C”

Different combinations of a head and body create three types of clauses: queries, rules and facts. The fundamental form of programming control is a query that is answered by searching for matching facts, or rules whose heads match the query and whose body may be proven. This ability to search through a set of facts and to further deduce relations from rules gives PROLOG its deductive capability. Terms AND connector OR Connector Implication p q p q p q p q true true true true true false false true false false true false true false false false false true

Figure 1: Truth Table.  $\square \vee \rightarrow$  The power of PROLOG-like languages to express both spatial queries and spatial models has been well demonstrated. LOBSTER is an early example of a prototype system that used PROLOG as the language interface to query a spatial DBMS (Egenhofer, 1990). The prototype provided a high level language to manipulate symbolic representations of spatial features. This was possible because the DBMS was able to handle complex record structures, and user defined functions could be programmed as built ins to the PROLOG interpreter. Spatial data types for points, lines, areas, and surfaces were defined in the DBMS and manipulated at a semantic level by the rules and facts expressed in Horn clauses. All low level access to spatial data and spatial manipulation is handled by the built in functions. This ability to include declarative expressions of spatial queries within a logic language is viewed as a key requirement by other researchers (Abdelmoty et al., 1993)

**Functional Programming:**

Functional programming is based upon mathematical concepts of mapping functions. A function maps object values from one domain to another. This is expressed formally  $f: X \rightarrow Y$ , the function of maps object values from the domain X to the domain Y. The object returned by a function depends only on its arguments. In addition functions do not induce any side effects so all state information evolves in an explicit and controlled way. This trait is known as referential transparency. Any transformations on objects are handled by explicitly returning new objects. This has a bearing on the data and procedural abstractions used by functional languages. Both rely upon mapping functions to express structural and behavioural relationships.

A GIS database perceived and manipulated by a functional language is viewed as a collection of objects together with a collection of functions. This has not proven to be a very attractive quality for feature-based GIS applications as there is not sufficient selective distinction between the different operations permitted on various types of spatial features (ie. point, linear, and area features). However, GIS applications that use a simple image-based structure are more predisposed to this type of manipulation. Map algebra is an example of a function-oriented language used in GIS for manipulating and analysing surface data (Tomlin, 1991). Map algebra uses a set of conventions to provide finer interpretation of the geographic locations (ie. local, neighbourhood, zonal) but these are still manipulated by functional transformations. Map algebra has the advantage of a straight forward notation and is very useful for developing models of spatial interpretations.

**Rule-Based Programming:**

Rule-based programming is a special case of logic programming. The language is based on a procedural scheme with the canonical condition-action form:

IF condition-pattern THEN actions the left-hand side consists of several conditions that return a logical result. The right-hand side consists of several actions. Actions can fire other rules, establish new facts, and perform procedural operations. Rules express relationships and meta-information. Rules are grouped in rule sets known to the inference engine. The engine works in a continuous loop, at each cycle a rule that matches some condition- pattern is chosen and the related actions are fired. The execution stops when no more rules are fire able. Rule-based programming uses a simple procedural abstraction to search for goals that satisfy the condition-pattern and then subsequently firing the action clauses. Queries are solved as proofs computed from the facts and rule set. Rule-based programming does not directly support data abstractions but relationships can be expressed by meta rules. Rule-based programming provides a model of the decision process that suits arrange of problems used for spatial reasoning (Scarponciniet al.,1995). The techniques have been used in

several ad hoc system developments for decision support (Lowes and Bellamy,1994)(Davis and McDonald, 1993).

### **Object-Oriented Programming:**

Object-oriented programming (OOP) is based on concepts for objects, classes, and the inheritance mechanism between classes. An object is an instance of a class to hold all related state information. Since objects can reference other objects, it is possible to build compositions of more complex objects. The classes in a program define categories of objects which share the same state information and procedural interfaces. Inheritance provides a relationship between classes based upon a taxonomy hierarchy. These organizing principles are formally based upon classification theory.

OOP has become very popular as it provides a framework for designers to encapsulate the structure and behaviour of design problems as objects. Data abstraction is supported through associative references to express structural relationships between objects, and class inheritance. Procedural abstractions are provided in two ways. The permissible actions on an object, and a configuration of objects, are integrated as part of the object class description. But the final implementation code still uses low level procedural mechanisms to perform operations in sequence, by conditional branching, or within an iteration. A disadvantage is that these control constructs involve the introduction of state variables to hold computational values between operations and procedures.

Writing a program in an OOP language does not necessarily make the program object-oriented. But in general programs incorporate object-oriented design principles. Proceedings of Geo Computation '97 & SIRC '97 19 (Rumbaugh et al., 1991). OOP is especially suited to problems where there is a large number of entities to be modelled, each with complex structural relationships and operational semantics. In recent years OOP has made a significant impact on graphical user interfaces (GUI's) and the application programming environment. Desktop GIS's often use object-oriented concepts in the user interface and application programming environment. But in most cases spatial data handling is still based upon a geo-relational model, and so data abstractions such as association and inheritance are not applied to the spatial data. Morehouse(1990)discusses the implications and difficulty of having true object-oriented modelling semantics for spatial databases. The Open GIS Specification (OGC, 1997) incorporates object-oriented geo-processing concepts. The full development of models to allow user defined schemas will require information representation specified by data dictionaries, schematic catalogues, geometry rules, etc. This technology specification will have an important

impact on the adoption of object-oriented data abstractions within GIS programming languages.

### Decision Tables:

Rule-sets are difficult to interpret for any reasonably sized knowledge base. An alternative technique for representing decision rules is as decision trees (Giarratano and Riley, 1994) or decision tables (Reilly et al., 1987). The different forms for representing rules can be shown by example. The example describes rules for choosing the best wine to have with a meal

Given the following rule-set:

IF (main\_course is beef) THEN (wine is red) IF (main\_ course is fish)

THEN (wine is white)

IF (main\_ course is poultry) AND (meat is light) THEN (wine is

white )

Some of the advantages of decision tables include compactness, self-documentation, modifiability and completeness checking (Reilly et al., 1987). Given that information is stored and viewed in a tabular for min geo-relational databases, it seems fortuitous to represent the rules in a similar form. This presents the user with a very consistent representation of data and procedures.

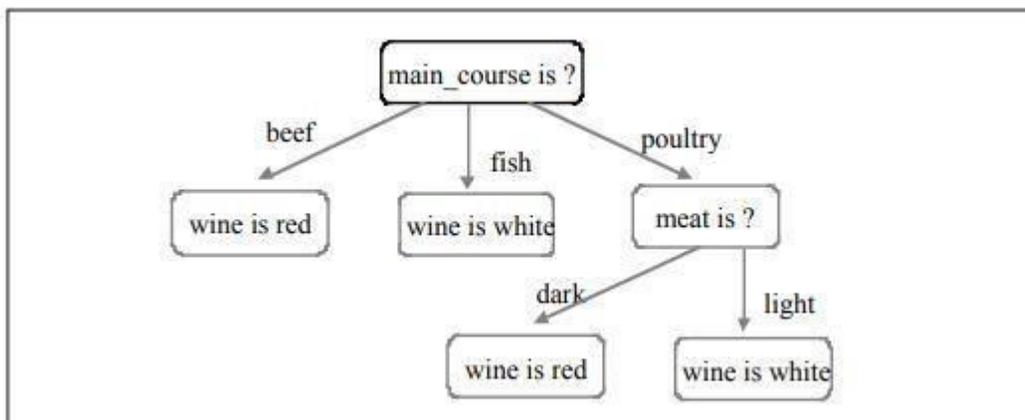


Fig.7.7 Decision Table

---

## 7.4 SUMMARY

---

In raster analysis, geographic units are regularly spaced, and the location of each unit is referenced by row and column positions. All cells in a grid have a positive position reference, following the left-to-right and top-to-bottom data scan. The regularity in the arrangement of geographic units allows for the underlying spatial relationships to be efficiently formulated. An arithmetic map operations are very common procedures used in GIS to combine raster maps. It is essential that this new map be accompanied by an assessment of uncertainty. We show exact results for means and variances for arithmetic operations involving addition, subtraction and multiplication. The use of the interpolation variance is proposed because it takes into account both data configuration and data values. Novice users often find writing programs to be a daunting task. Users of GIS are faced with the task of writing programs as a concrete formulation of their particular problem. GIS is a necessary part of solving problems with computers is to express them in a formal way. The appropriate computer tool to solve geographical problems is a Geographic Information System (GIS). Such systems provide the basic functionality for visualising, managing and manipulating spatially referenced data. It's important to make the language interface to a GIS as easy to use and intuitive as possible. It can be thought of simply as algebra applied to spatial data which, in the case of raster data, are facilitated by the fact that a raster is a georeferenced numerical array. This problem occurs when an application presents information in one way but the programming environment to access and manipulate that information is different. A popular way to present information in GIS is as a map organised into the thematic layers. In a programming environment the user is presented with tables containing records.

---

## 7.5 GLOSSARY

---

- **Map algebra-:** It is the most common scheme for manipulating continuously sampled (i.e. raster) variables defined over a common area.
- **Primary operations-:** Operations used to generate a layer, conceptually from nothing.
- **Binary operations-:** Binary grid operations action the pairs of numbers obtained in each set of matching cells
- **Local statistics-:** When many related grids are accumulated over same region.
- **Integral Data-:** These data may be relative ranks or preferences of the main databases.

---

## 7.6 ANSWERTO CHECKYOURPROGRESS

---

### Q-1:-Discuss one advantage of using data in Raster format?

Ans:- Efficient processing: Because geographic units are regularly spaced with identical spatial properties, multiple layer operations can be processed very efficiently.

### Q-2:-What are categorical data?

Ans:- These are non-numerical data. Grids that classify land use or landcover exemplify this category.

### Q-3:- What is relational operator?

Ans:- A relational operator enables the construction of logical functions and tests by comparing two numbers and returning a true value or false value.

### Q-4:- What is decision table?

Ans:- Rule-sets are difficult to interpret for any reasonably sized knowledge base. An alternative technique for representing decision rules is as decision trees or table.

---

## 7.7 REFERENCES

---

1. Abdelmoty A.I., Williams M.H. and Paton N.W. (1993) Deduction and Deductive Databases for Geographic Data Handling. 3rd International Symposium on Large Spatial Databases, SSD'93, Singapore, pp.443-464
2. Arentze T.A., Borgers A. and Timmermans H. (1995) The Integration of Expert Knowledge in Decision Support Systems for Facility Location Planning. Computers, Environment and Urban Systems 19(4), pp.227-247
3. Bratko Ivan (1990) Prolog Programming for Artificial Intelligence. Addison Wesley.  
Davis J.R. and McDonald G. (1993) Applying a Rule-Based Decision Support System to Local Government Planning. In: Expert Systems in Environmental Planning, Editors J.R. Wright, et. al. Springer-Verlag, pp.23-45
4. ESRI (1994) Avenue – Customization and Application Development for ArcView. Environmental System Research Institute Inc, Redlands, CA.
5. Egenhofer M. and Frank A. (1990) LOBSTER: Combining AI and Database Techniques in GIS. Photogrametric Engineering & Remote Sensing 56(1), pp.919-926
6. Frank A.U. and Kuhn W. (1995) Specifying Open GIS with Functional Languages. Advances in Spatial Information Systems, Proceedings SSD'95, Portland, pp.184-195
7. Giarratano J. and Riley G. (1994) Expert Systems - Principles and Programming. PWS Publ. Co., Boston.
8. Jian Guo Liu, Philippa J. Mason, "Essential Image Processing and GIS for Remote Sensing," Imperial College London, UK, 261-280 (2009).
9. Lowes, D. and Bellamy J.A. (1994) Object Orientation in Spatial Decision Support System for Grazing Land Management. AI Applications 8(3), pp.55-66
10. Morehouse S.D. (1985) ARC/INFO - A Geo-Relational Model for Spatial Information. Proceedings Auto-Carto 7, Washington, pp.388-397
11. Morehouse S.D. (1990) The Role Of Semantics In Geographic Data Modelling. Proceedings 4th International Symposium on Spatial Data Handling, Zurich, pp.689-698
12. Newell A. and Simon H. (1972) Human Problem Solving, Prentice-Hall.
13. Paulson L.C (1996) ML For Working Programmers. Cambridge University Press.

14. OGC (1997) The Open GIS Consortium, [http:// www.opengis.org](http://www.opengis.org)
15. Reilly K.D., Salah A., and Yang C. (1987) A Logic Perspective on Decision Table Theory and Practice. Data and Knowledge Engineering (2), pp.191-210
16. Rumbaugh J., Blaha M., Premerlani W., Eddy F., and Lorensen W. (1991) Object-Oriented Modeling and Design, PrenticeHall.
17. Scarponini P., Clair D., and Zobrist G. (1995) An Inferencing Language for Automated Spatial Reasoning About Graphic Entities. Advances in Spatial Information Systems, Proceedings SSD'95, Portland, pp.259-278
18. Tomlin C.D. (1991) Cartographic Modelling. In: Geographical Information Systems - Vol. 1, Editors D. MacGuire, M. Goodchild, and D. Rhind, Longman Scientific & Technical, pp.361-374
1997. User Interfaces for Map Algebra. Journal of the Urban and Regional Information Systems Association, Vol.9, No. 1, pp. 44-54.
19. Bhatta,(2008)RemoteSensingandGisOxford UniversityPress
20. Chang, Kang tsung Introduction to Geographic Information Systems 5th edition 2009 Mcgraw Hill.
21. Lillesand,Thomas M.,RalphW. Kiefer,and JonathamW.Chipman, 2004  
C.P.Loandalbert
22. k.W.Yeung(2002)ConceptsandTechniquesofGeographicInformationSystems, UpperSaddle River,New Jersey:PrenticeHall,2002)

---

## ***7.8 TERMINAL QUESTIONS***

---

- Q-1** What is Arithmetic Operation?
- Q-2** What are the disadvantages of raster data format?
- Q-3** What do you mean by Binary operators?
- Q-4** What is Logical programming?
- Q-5** What is Unary operation?

---

## **UNIT 8 - RASTER DATA FORMATS**

---

**8.1 OBJECTIVES**

**8.2 INTRODUCTION**

**8.3 RASTER DATA FORMATS**

**8.4 SUMMARY**

**8.5 GLOSSARY**

**8.6 ANSWER TO CHECK YOUR PROGRESS**

**8.7 REFERENCES**

**8.8 TERMINAL QUESTIONS**

---

## 8.1 OBJECTIVES

---

After going through this unit the learner will able to:

1. Learn about elements of raster data model.
2. Learn about Raster Data Structure and Data compression.
3. Get to know about the format of raster data.
4. Understand what the raster data in a GIS is and how it can be used.

---

## 8.2 INTRODUCTION

---

We represent Earth surface features and man-made features in the GIS environment, and this form of data is known as spatial data in remote sensing processes. Raster and Vector are the two data formats used to generate spatial data in geographic information systems and remote sensing. Point, line, and area geometric objects are used to represent spatial features in vector data. Vector data is ideal for isolated features with well-defined positions and forms, but it is ineffective for spatial patterns that shift over time, such as soil erosion, precipitation, and elevation. For representing continuous phenomena, raster data is the best choice. To cover space, raster data use a regular grid. The value of each grid cell is a spatial attribute at a certain point of the cell. The cell value variation reflects the phenomena of spatial variation.

GIS software used to be either raster-based or vector-based, but now, in the current situation, the majority of GIS software can handle data in both formats. In GIS applications, advancement in computing science has effectively eliminated the distinction between raster and vector data.

Many opportunities exist in the combined raster and vector data working area, where one may incorporate mathematical and simulation approaches suitable for each of the formats in the review. We will study about Raster data format in this chapter.

---

## 8.3 RASTER DATA FORMATS

---

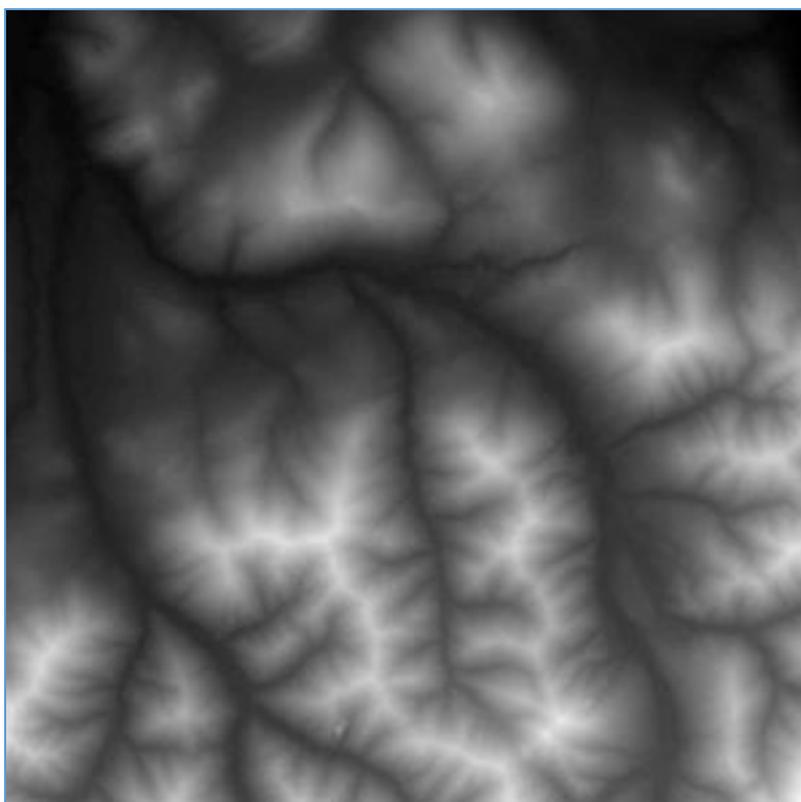
### **Raster data:**

Raster data are increasingly used in a number of GIS applications and has become the primary source of spatial data in geographic databases. Raster data is defined as a grid/cell format that represents a function of the earth's surface, both natural and man-made. This ensures that all raster data is represented by image, cell, and grid formats. All satellite images are recorded in raster format, which is the most important feature. The analysis of raster data can be defined as follows: the research area is divided into normal cells with unique dimensions, and each cell's measurement or attribute is expressed by a digital code, i.e. (DN). The locations of raster cells are inferred from their positions in the image rather than being specifically recorded. Raster data is typically represented as a matrix (2D array) with row and column numbers indexed into each cell.

**Raster data representation of Features:**

- **Points features by single cells**
- **Line by sequence of neighbouring cells**
- **Polygons by collections of contiguous cells**

In the raster format, each cell has a value that is either an integer or a floating point number. Integers are commonly used to display discrete statistics, such as temperature, average annual precipitation, and elevation, while floating point numbers are typically used to represent continuous data, such as forest area, agricultural land or built up area.



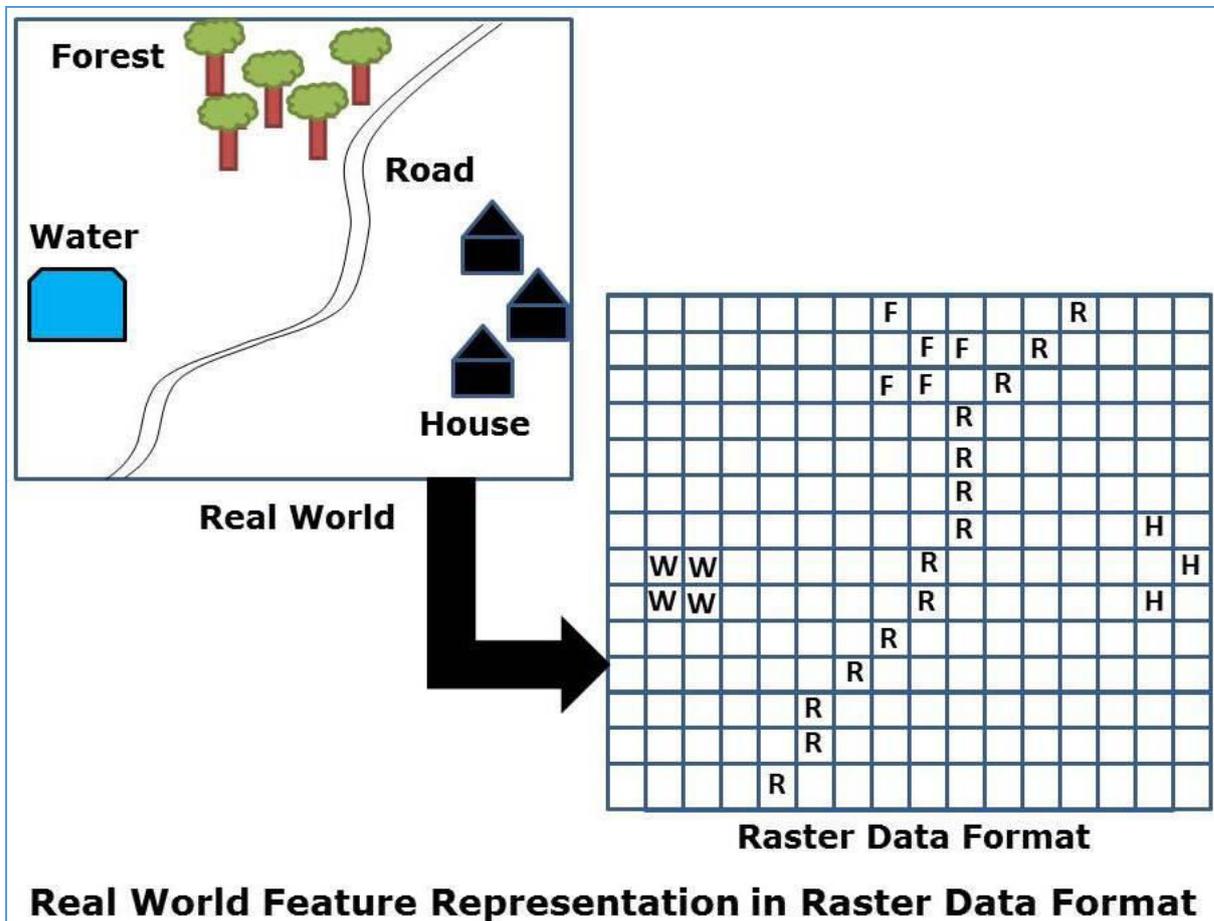
**Fig. 8.1** A continuous raster with darker shades for higher altitude

**Organization of Raster Data:**

Raster data are usually organised into layers, which are also called as bands, themes, or grid. Each layer has a feature-based theme, such as irrigation, soil type, topography, land use, and vegetation cover, etc. The raster data models are better suited to continuous phenomena, but they can also be used to describe discrete functions.

## Elements of Raster Data Model:

The raster data model is often referred to as the grid, raster map, or picture in a geographic information system. The grid has a continuous surface, but is split into squares, columns and cells for data storage and analysis. Cells are also referred to as pixels with images.



Source: gisoutlook.com

**Fig 8.2** Illustration of line, point, and area feature: on the right side we see raster format whereas vector format on the left side.

Raster data represents a point with a single cell, lines with a series of adjacent cells, and areas with a set of contiguous cells (Fig. 8.2). While it lacks the precision of the vector data model in describing the position of spatial features, the raster data model has the distinct advantage of having fixed cell locations. A raster can be analysed as a matrix with rows and columns in computational algorithms, and the cell value in a two-dimensional array can be stored. Arrayed variables are conveniently handled by all widely used programming languages. As a result, raster data is significantly easier to process, aggregate, and analysis than vector data.

### Cell Value:

Each raster cell contains a value that corresponds to the spatial characteristics of the location of the column and row indicated. Based on the coding of their cell numbers, the Raster Data can be either an integer or a floating point raster. The integer value does not have decimals, whereas floating point values have it. Usually, integer cell values reflect the categorical data that may be or may not be ordered. The land cover raster can be set to 1 for urban built-up areas, 2

for forestland areas, 3 for water bodies, and so on. The continuous numerical data are seen by the floating point cell value. For instance, a raster of precipitation could have values such as 20.15, 12.23 (millimetre), etc.

A raster floating point needs more memory, or we can say more working space, than an integer raster, and this disparity can be significant in GIS projects that cover a large area. We can view the cell values of an integer raster from an attribute table, which is one of the few variations we can address here. Due to the vast number of data, a floating-point raster typically does not have an attribute table.

### **Cell Size:**

The raster data model's resolution is determined by the cell size. A cell size of 30 metres equals 900 square metres per cell (30 X 30 meters). On the other side, a cell size of 10 metres means that each cell is 100 square metres in size (10 X 10 meters). As a result, we might assume that a raster of 10 metres has a higher resolution than a raster of 30 metres. A larger cell can't reflect the exact position of spatial features, so mixed properties like woodland, grassland, and water are more often in a cell. The most popular approach is to enter the division that takes up the greatest percentage of the cell area, but where a raster uses a smaller cell size, these issues are minimised. A small cell size, on the other hand, increases data volume and processing time.

### **Raster Bands:**

A raster can have one or more bands i.e. multi-band. In a raster with many bands, each cell has several values associated with it. A satellite image of five, seven, or more bands at each cell position is an example of a multi-band raster. In the other hand, only one cell value is present in a single-band raster. An example of a single-band raster is a height raster, where each cell position has one height value.

### **Spatial Reference:**

In order to coordinate spatially with other data sets in a GIS, raster data must provide spatial reference information. For instance, we must first ensure that two data sets are in the same coordinate system to superimpose an elevation raster on a vector-based forest cover. A Georeferenced raster is usually named, that has been processed to match a projected coordinating system.

In association with the raster, two modifications are required. First, at the top-left corner is the base of the proposed coordinate scheme or we can say projected coordinate. Second, the coordinate projected must match the raster rows and columns.

### **Types of Raster Data:**

A wide range of data we use in GIS are raster encoded. The raster data model is composed of the same fundamental components. There are many types of Raster data that we are going to discuss one by one.

### **Satellite Imagery:**

The user of remotely sensed satellite data is very similar to a GIS user. The satellite image's spatial resolution is equal to the ground pixel size. A spatial resolution of 30 metres, for example, means that it covers 900 square metres on the earth. The pixel value, also known as the

brightness value, represents the amount of light energy produced by the earth's surface. Light energy is measured using spectral bands from the electromagnetic spectrum, which is a continuous range of wavelengths. Multispectral images are made up of several bands, while a single spectral band is made up of panchromatic images.

### **USGS Digital Elevation Model (DEMs):**

A digital elevation model includes a range of evenly separated data on elevation. Due to the point-based DEM, it is easy to convert to raster by putting any elevation point in the centre of the cell. The 7.5-minute DEM, 30-minute DEM, 1-degree DEM, and Alaska DEM are all USGS DEMs.

The **7.5-minute** DEMs have elevation data at 30-meter and 10-meter intervals on a grid plotted in UTM coordinates and referenced to either the (NAD27) or (NAD83) North America Datum's of 1927 and 1983, respectively. Each DEM corresponds to a USGS 1:24,000 scale quadrangle and covers a 7.5-by-7.5-minute block. By data consistency and production process, the USGS divides the 7.5-minute DEM into two stages, with level 1 providing the least accurate data. The vertical accuracy of Level 1 is  $\pm 15$  metres, while Level -2 is  $\pm 7$  metres.

On the spatial grid, the **30-minute DEMs** have elevation data at a 2 arc-second spacing. Each DEM corresponds to the east or west half of a USGS 30-by-60-minute 1:100,000 scale quadrangle and covers a 30-by-30-minute block. The 30-minute DEMs have a vertical accuracy of  $\pm 25$  metres or greater.

On the spatial grid, the **1-degree DEMs** have elevation data at a 3 arc-second spacing. Each DEM corresponds to the east or west half of the USGS 1-by-2-degree 1:250,000 scale quadrangle and covers a 1-by-1-degree block. The defence map originally produced 1-degree DEMs by interpolation from digitized contour lines. The vertical accuracy of elevation data is about  $\pm 30$  meters.

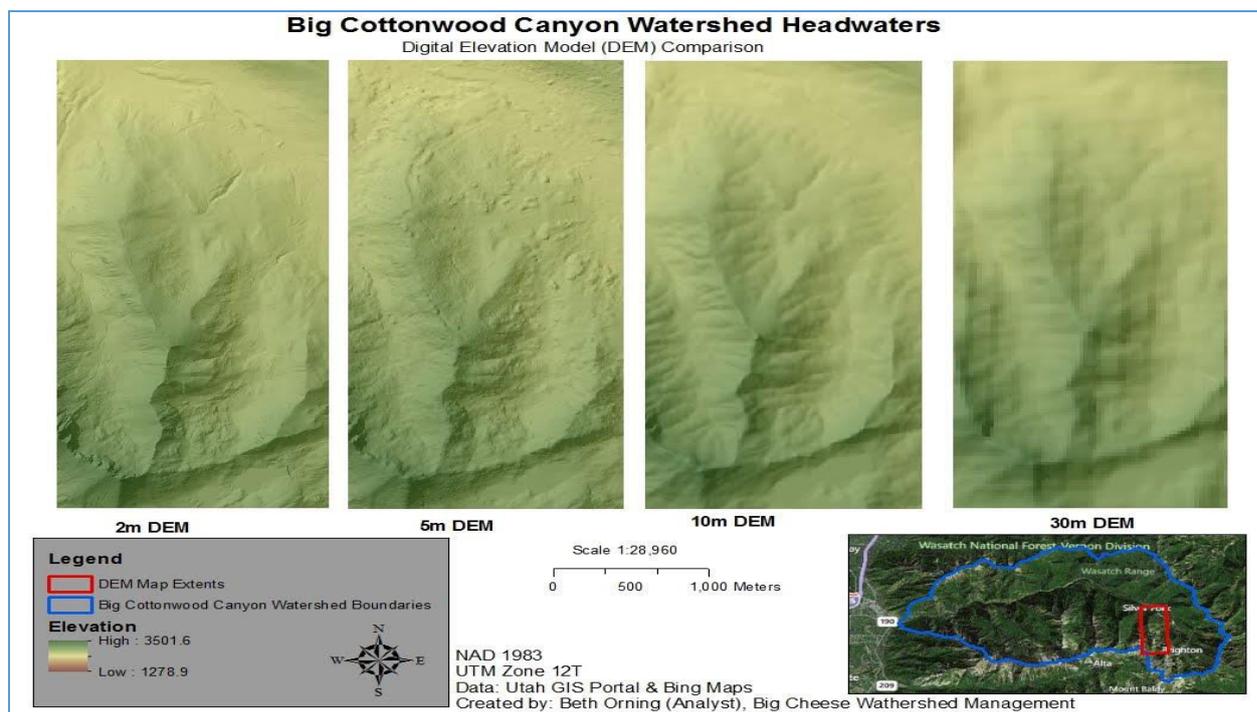
### **Non-USGS Digital Elevation Model (DEMs):**

A stereo-plotter and aerial photographs of over lapped areas are basic methods for creating DEMs. The stereo-plotter generates a three-dimensional model that the operator can use to compile elevation data. While this method will produce DEM data with a more accurate resolution than USGS DEMs and, it is more costly for wider areas.

There are many other options besides using a stereo-plotter. The generation of DEMs from satellite imaging, such as the SPOT stereo model, is a common alternative among GIS users. Commercially available software packages for collecting elevation data from SPOT photos on a personal computer are available. Ground control points, which can be calculated in the field using GPS with differential correction, are required for the data extraction process in addition to imagery data. However, the software package used and the quality of the input data shall decide the accuracy of these DEMs.

Radar data can also be used to create a DEM. Radar is an active remote sensor that sends out microwave signals and measures the energy returned by ground sources. Radar can see through cloud cover, allowing data to be collected at any time of day or night, in sunny or dark conditions. LIDAR is a modern method of creating DEM. A laser scanner mounted on an aircraft, a GPS, and an Inertial Measuring Unit are the main components of a LIDAR device. A LIDAR sensor emits rapid laser pulses over a large area and measures distance using the pulse's

time lapse. At the same time, GPS and the Inertial Measurement Unit (IMU) decide the position and direction of the laser source.



**Source:** <https://sites.google.com/site/bethorninggis6920/labs/working-with-dems>

**Fig 8.3** DEMs at four resolutions: 30m, 10m, 5m, 2m. The DEM of 2 metres, containing more topographical information than the other three.

## Global DEMs:

DEMs with different resolutions are now available on a global scale. SRTM DEMs, with a coarser spatial resolution of approximately 3-arc seconds, are available outside the USA with a spatial resolution of 90 metre at the equator). These global-scale DEMs are referred to as SRTM DTED Level 1 (digital terrain elevation data), as opposed to DTED Level 2 for the US and its territories. Although the values of SRTM DTED level 1 are determined from the values of the elevation of SRTM DTED level 2, at coincident points of less than 16 metres they have the same vertical accuracy.

With a grid spacing of 5 minutes of latitude by 5 minutes of longitude, ETOPO5 (Earth Topography-5 Minute) data cover both the ground surface and the ocean floor of the Earth. Global DEMs with a horizontal grid spacing of 30 arc-seconds (approximately 1 kilometre) are available from both GTOPO30 and GLOBE. GTOPO30 and GLOBE were created using raster data from satellite imagery and vector data from the Digital Map of the World's contour lines. GLOBE's vertical accuracy is measured to be within 30 metres when using raster sources and 160 metres when using vector sources.

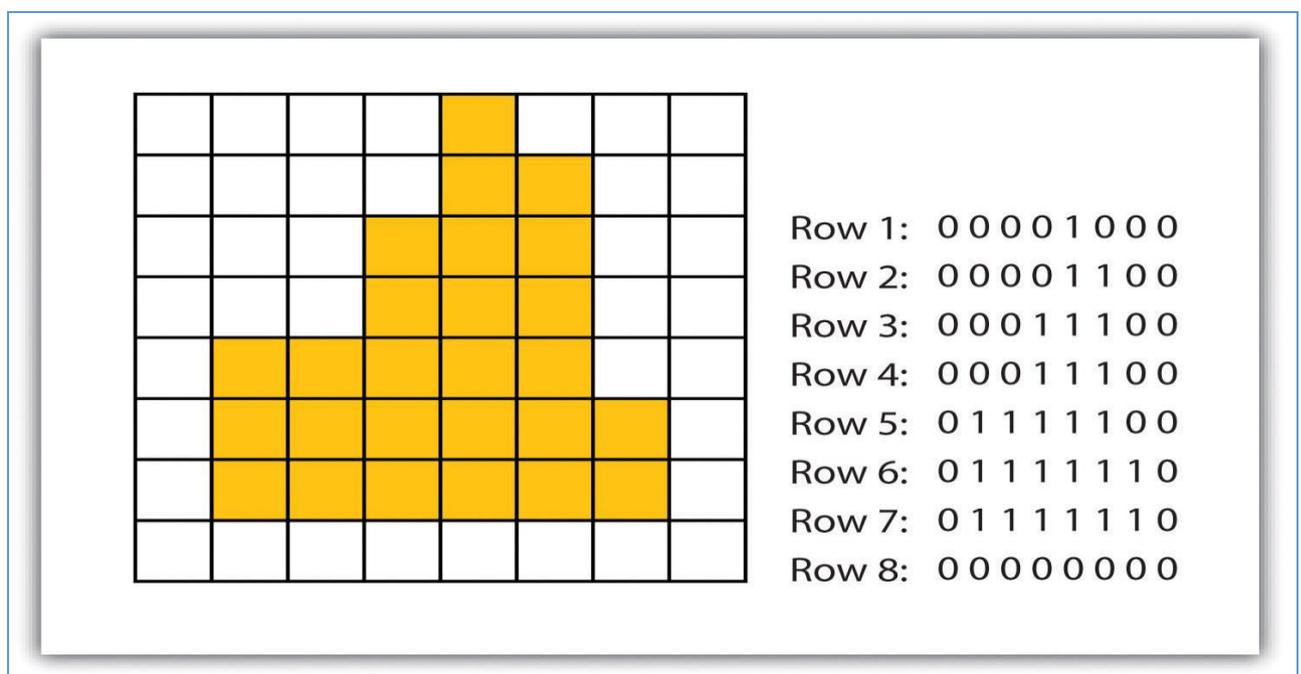
## Raster Data Structure:

The storage of raster data that can be used and analysed by the computer, is referred to as raster data structure. Several raster data models have been proposed over the last three decades. The **cell-by-cell encoding, run-length encoding, and quad tree** are discussed in this section.

### Cell-by-Cell Encoding:

The shortest raster data structure is available from the cell-by-cell encoding system. A raster is saved as a matrix and the cell values are entered into a file in rows and columns (Fig 8.4). This technique operates at the cell level and is perfect if a raster's cell value also varies.

DEMs are using the data structure cell by cell so the values of the neighbouring elevation are rarely identical. The cell-by-cell encoding process is often used to store data in satellite images.

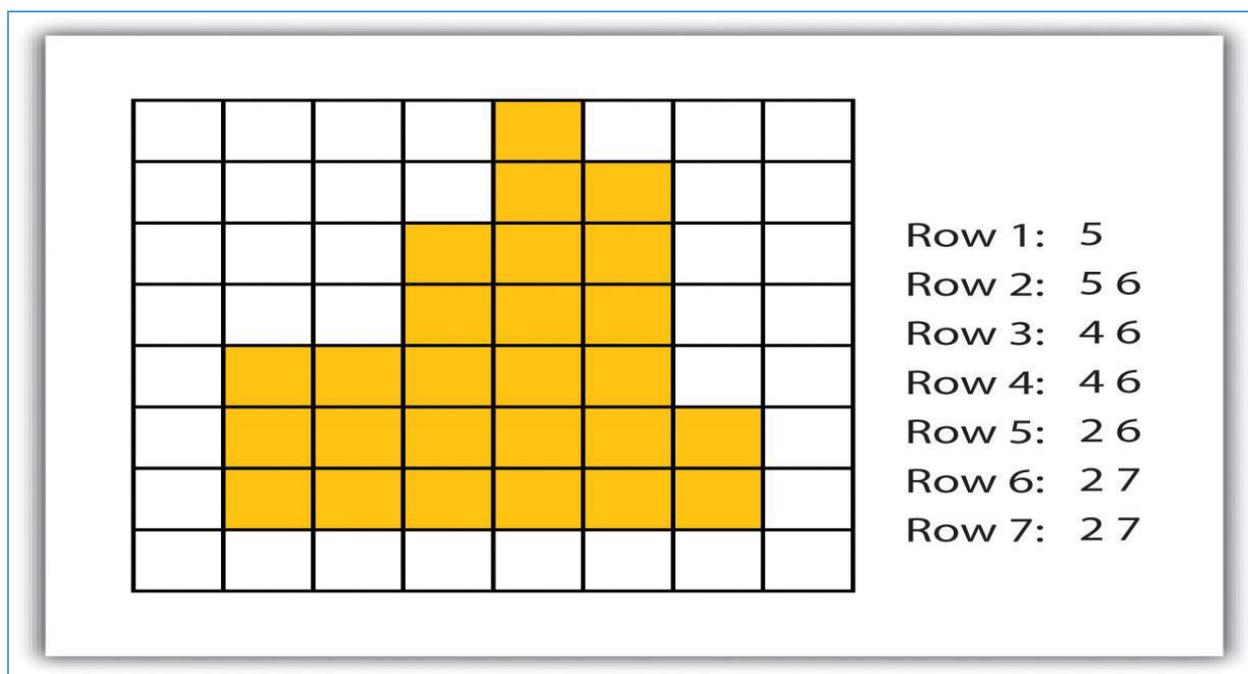


**Source:** [https://saylordotorg.github.io/text\\_essentials-of-geographic-information-systems/s08-01-raster-data-models.html](https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s08-01-raster-data-models.html)

**Fig 8.4** Each cell value is recorded by row and column in the cell-by-cell data structure. The cell value of the yellow cells is 1

### Run-Length Encoding:

Whenever a raster consists of multiple redundant cell values, the mechanism of cell-by-cell encoding is inefficient. For instance, a scanned bi-level file on the soil map contains many 0s representing non-inked whereas 1s representing inked soil lines. The RLE process, which records cell values by row and category, can be used to store raster models with many repeated cell values more effectively. A group is made up of cells that have the same cell value. The polygon's run-length encoding is shown in yellow in Fig. 8.5. The length of the group ("run") that falls inside the polygon is indicated by the beginning cell and the end cell for each row.



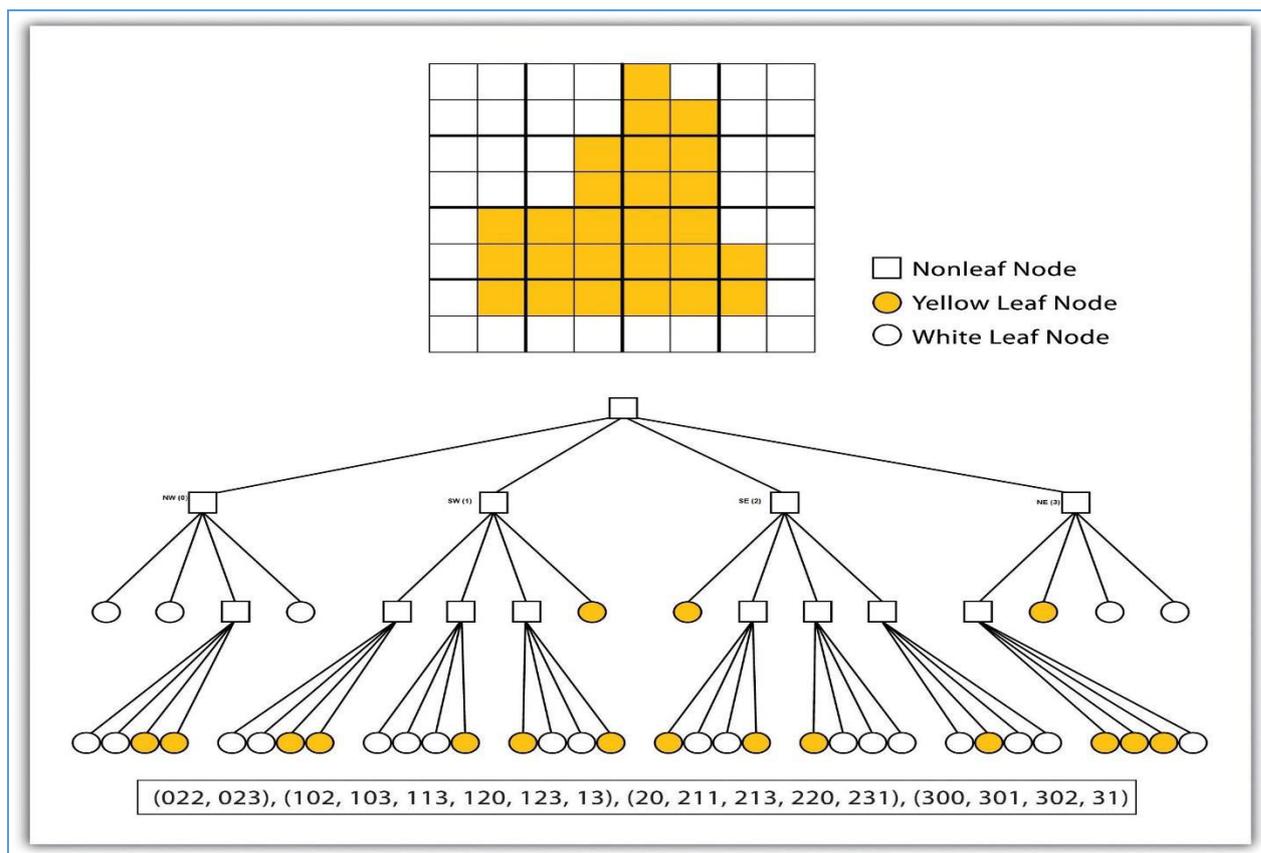
**Source:** [https://saylordotorg.github.io/text\\_essentials-of-geographic-information-systems/s08-01-raster-data-models.html](https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s08-01-raster-data-models.html)

**Fig 8.5** The run-length encoding method records the yellow cells by row. Row 2 has two adjacent yellow cells in columns 5 and 6. Row 2 is therefore encoded with one run, beginning in column 5 and ending in column 6. The same method is used to record other rows.

### Quad Tree:

Quad tree divides a raster into a hierarchy of quadrants using recursive decomposition rather than operating along one row at a time. Recursive decomposition is a subdivision process that continues until each quadrant in a quad tree has just one cell value. Figure 8.6 shows a raster with a yellow polygon and a quad tree which holds the element. Nodes and branches make up the quad tree. A quadrant is represented by a node. A node may be a non-leaf node or a leaf node depending on the cell value in the quadrant. A quadrant with different cell values is represented by a non-leaf node. As a result, a non-leaf node is a branch point, where the quadrant is subdivided while leaf node, is the point at which the same quadrant value can be coded. The depth of the quad tree or the number of levels in the hierarchy will differ depending on the complexity of the 2-D feature.

After the subdivision is over, the 2-D feature is coded using a quad tree and a spatial indexing tool. Figure 8.6 shows two yellow leaf nodes in the level-1 NW quadrant (with a spatial index of 0). 022 refer to the level-2 SE quadrant, while 023 refers to the level-3 SE quadrant of the level-2 NE quadrant. The coding of the two-dimensional function is completed by the string of (022, 023) and others for the other three level-1 quadrants. The regional quad tree is a good way to store area data, particularly if there are few categories. This method is also efficient for data processing. Quad trees can also be used in GIS. For sorting, indexing, and displaying global data, researchers suggest using a hierarchical quad tree structure. Quad trees may also be used as a spatial indexing technique. Spatial indexing makes it simple and easy to find raster and vector spatial data. Oracle spatial, for example, uses quad tree as a way of indexing spatial data.



**Source:** [https://saylordotorg.github.io/text\\_essentials-of-geographic-information-systems/s08-01-raster-data-models.html](https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s08-01-raster-data-models.html)

**Fig 8.6** A raster is divided into a hierarchy of quadrants using the regional quad tree system. The split ends where a quadrant consists of cells of equal value (Yellow or white). A leaf node is a quadrant that cannot be subdivided.

### Data Compression:

Raster data sets typically contain large amount of data and require considerable memory space. Approximate file sizes are 1.1 megabytes (MB) for a 30-meter DEM, 9.9 MB for a 10-meter DEM, 5 to 15 MB for a 7.5- minute digital raster graphic (DRG), and 45 MB for a 3.75- minute quarter DOQ in black and white. The capacity requirements for an uncompressed 7–band TM scene are nearly 200 MB. The memory requirement becomes even higher for high-resolution satellite images.

The reduction of data volume is referred to as data compression, and it is a subject that is especially important for data distribution and internet mapping. We're all familiar with data compression applications like WinZip for Windows and gzip for UNIX. These programmes can operate on any kind of data file while preserving the original file and folder structure. This part, on the other hand, is about image compression.

Image compression can be accomplished using a number of methods. Lossless or lossy compression methods are available. The original image can be precisely restored using a lossless compression technique. Lossless compression is exemplified by RLE. Pack Bits, a more versatile variant of RLE, and LZW (Lempel-Ziv-welch) are two other lossless compression techniques. Despite being functionally lossless, the GIF format does not maintain the accuracy of the original image due to its maximum colour depth of 256 or 8 bits. Pack Bits and LZW image compression are available in the TIFF data format. TIFF is used by the USGS to deliver DRGs and DOQs.

A lossy compression system cannot completely recreate the original file, but it can achieve high compression ratios. Lossy compression is used in the popular JPEG format. The procedure divides an image into 64 (8 X 8) blocks and processes each block separately. To minimise the amount of data decoding, the colours in each block are changed and condensed. The “blocky” look is normally the product of this block-based processing. The image distortion due to lossy compression will hamper the GIS related activity such as the collection of ground controls from aerial images or satellite imaging for georeferencing.

Image compression strategies, including GIS techniques, are continually evolving. MrSID is a relatively modern method. MrSID (Multi-resolution Seamless Image Database) is a compression technique developed at the Los Alamos National Laboratory and licenced to Lizardtech Inc. in the late 1900s. MrSID's ability to recall image data at various resolutions or sizes is referred to as multi-resolution. MrSID can compact a big image with sub blocks, such as in a satellite image or DOQ, the compression process eliminates the arbitrary borders of blocks. Significant GIS businesses fund MrSID, which is used by federal agencies including the USGS and NGA to deliver DOQs. While MrSID is a proprietary format, it is well known that it compresses data using the wavelet transform. The wavelet transform is also used in JPEG 2000, a modified version of the existing open format that is likely to become common with GIS users. Therefore, the wavelet transformation seems to be the new compression option.

The wavelet converts a picture into a wave and gradually breaks down the wave into basic wavelets. The transformer uses a wavelet (mathematical) to record the variations between the original and the average pixels and repeats average classes of neighbouring pixels (e.g. 2,4,6,8 or higher). Differences may be 0, more than 0, or less than 0, also known as wavelet coefficient. Many pixels have coefficients of 0 or nearly 0 in sections of an image which have few important variations. To store these image sections in lower resolution, low coefficients can be rounded up to 0. Except for areas of the same image that have major differences, higher resolution storage is essential.

It can be both a lossless and a lossless encoding by MrSID and JPEG 2000. A lossless compression saves and re-built the original image using wavelet coefficients. In the other hand, a failure compression just saves averages and coefficients which have not been reduced to 0. Commercial reports have shown that for images with colour and from 10:1 to 50:1 for images with grey sized images, JPEG 2000 will reach compression ratios from 20:1 to 300:1, with no significant change in the image quality. As compression JPEG 2000 is at or ground control points of aerial photos or georeferencing satellite photos.

## **Raster Data Format:**

There are only two storage mechanisms in the geographical information system for all data mapping references, i.e. raster and vector data. Here we will talk about different formats of GIS Raster data files. Raster data represents the world as a surface divided into regular grid of cells. So, we can say that raster data is very helpful for analysis.

Raster data models are useful for the continuous storage of data, such as in an aerial photograph, a satellite image or a surface elevation. There is continuous and discrete two forms of raster data. Raster stores data in a digital image type that is defined by reduced and extendable grids, which holds continuous and discrete information, such as temperature data and land use or soil data

Raster data provides the cell matrix with coordinate values and is often added to the attributes table, and with certain layer combinations it is much easier. Due to the basic data structure, raster data is very easy to modify or to programme. Now the question is how many Raster data formats have? And here, as Raster file format lists shown in GIS with all data format lists.

### **1) Portable Network Graphics (PNG)**

This provides a well-compressed and lossless raster file encoding. It supports a wide range of bit depths from monochrome to sixty four bits. Colour-indexed images up to 256 colours and 100% accurate pixel lossless images up to 16 bits are included in its capability.

### **2) Joint Photographic Experts Group (JPEG2000)**

JPEG 2000 is an imaging coding method using state-of-the-art wavelet-based compression methods with an incredibly high degree of scalability and usability. Content can be coded at any rate once, without failure, but can be accessed and decoded at a very high level with other values and resolutions, or with no significant coding performance penalty by area of interest. The norm supports up to 16384 modules, with dimensions of thousands of treatment pixels and accurate 38 bits/sample, tiling and random data progressions and access capabilities. The JPEG 2000 architecture is suited for a wide variety of applications including compact digital cameras, advanced pre-press, medical imaging, geospatial applications and other core applications.

### **3) Multi-resolution Seamless Image Database (MrSID)**

A compressed wavelet format, allows both lossy and lossless compression. MrSID is a proprietary format of geo-express imagery compression applications for Lizard Tech, which is used for orthography. Most of the greyscale TIFF-images are either compressed to 10:1 or 15:1 with MrSID. In general, 30:1 or 40:1, colour pictures are compressed. Often used in the construction of picture mosaics is GeoExpress. Most modern GIS applications like the ArcGIS, without extra plugins, is able to read MrSID compressed files. Nevertheless, ArcView 3.x requires MrSID picture control extension. Plug-ins may or may not be available for other applications like AutoCAD or Photoshop.

### **4) Network Common Data Form (netCDF)**

Binary storage optionally compressed in open format. Allows direct web access by means to OpeNDAP protocol of subsets/compilations of maps. It supports the development, access and

sharing of array-oriented science data in machinery-independent data formats. Multiple binary formats support netCDF libraries for netCDF files:

- In the first version of the netCDF the classic format was used and the file generation format remains standard.
- In version 3.6.0 a 64-bit offset format has been implemented which supports larger variable and file sizes.
- In version 4.0, netCDF-4/HDF5 format has been adopted, with certain constraints.
- For read-only access, the HDF4 SD format is supported.
- In cooperation with the concurrent netcdf initiative, the CDF5 format is supported.

### **5) Digital raster graphic (DRG)**

A DRG is a digital image that results in scanning a topographical USGS paper chart that can be seen on a monitor. USGS DRGs are typically scanned at 250 dpi for TIFF saving. The raster picture normally contains the initial detail about the boundary, known as the map collar. The map file is UTM projected and georeferenced to the surface of the earth. In GIS applications DRG's are regularly used. In 1995, DRGs were developed for the first time.

### **6) ARC Digitized Raster Graphic (ADRG)**

ADRG is a digital product, developed in 1989-1990, by the National Imaging and Mapping Agency (NIMA) for applications with raster map background displays. ARC Digitized Raster Graphics were digitised maps and charts converted into a particular framework for georegistration accompanied by ASCII encoded files of support. The raster scans and the georeferencing of the map image using an equal arc second raster-chart/map (ARC) method was used to transform maps/charts into digital data in which the globe is divided into 18 latitudinal bands or zones. Typically the source graphic is a map sheet. Data from a single graph / graphical map sequence can be retained as a seamless worldwide database of raster graphic data with each pixel having a distinct geographic location.

### **7) Raster Product Format (RPF)**

The raster product format (RPF) is a common data structure that was introduced as the U.S. military standard in 1994 for geospatial databases consisting of rectangular pixel arrays in a compressed or uncompressed format (e.g., in digitised maps or images). Its aim was to design a family of digital data exchange products which consists, for military applications, of digital maps, images and other geographic data.

It was developed to provide raster items in a compressed or uncompressed format in a general, adaptable format. It was intended to allow application software, without further manipulation or

transformation. The file format is distinguished by a hierarchical directory hierarchy that contains a content table file, often referred to as an A.TOC file.

#### **8) Enhanced Compressed Wavelet (ECW)**

It's a compressed, sometimes lossy wavelet format. ECW is an ER Mapper proprietary image compression format. It is more recent than MrSID, but is becoming more popular as free compression utilities are available on the website of ER Mapper. ECW is an exclusive format. It's like a JPEG 2000 wavelet-based, lossy compression. The encoding format for aerial and satellite photographs is a proprietary one.

This format can be used for desktop applications, but for printing or publication you need an ArcGIS Extension License ECW.

#### **9) Extensible N-Dimensional Data Format (NDF)**

For storing n-dimensional arrays of numbers, such as images, the following format is used. The data objects are handled by container files (directories containing files and directories).

#### **10) Tagged Image File Formats (TIFF)**

Scanners are linked to this format. It stores and reads the scanned pictures. TIFF may use run-length and compression schemes for other images. The colours like a GIF are not confined to 256. This is commonly used in the field of desktop printing. It serves as an interface to many scanners and packages in graphic arts. TIFF supports black and white images as well as pseudo colour, which can also be stored in both compressed and decompressed formats.

#### **11) Geo Tagged Image File Formats (GeoTIFF)**

GeoTIFF is a metadata standard in the public domain, which enables the embedding of geographical information into a TIFF file. The possible additional material includes a map, coordinates, ellipsoids, datums, and any other information required to determine the exact file geographical relation. A programme that does not read and decode the advanced metadatas will still be able to open a GeoTIFF format file, which is fully compatible with the TIFF 6.0 format.

#### **12) Graphic Interchange Format (GIF)**

An animated GIF is a GIF file that includes multiple images or "frames." This pictures are played series by opening or displaying the file on a web browser. The effect is a short film or an

animation clip. The GIF format consists of an extension to graphics control (or "GCE block"), which allows several frames in a single GIF file. This section also defines the interval between frames, which can be used at some points in the animation to adjust the frame rate or place pauses. Another part, the Netscape Application Block (NAB), indicates the repeated animations (a setting of "0" is used for infinite repetitions).

### 13) BMP

Short for "Bitmap." You might say it as "bump," "B-M-P," or just "bitmap." The BMP format is widely used to store image files as a raster graphic format. It is now recognisable among several applications on all macbooks and pcs, but has been released on the Windows platform. The BMP format saves colour data without any encoding for each pixel in the image. For example, a BMP image of 10x10 pixels would have 100 pixel colour data. This approach allows crisp, high-quality graphics to be saved, but still creates massive file sizes. The JPEG and GIF formats are bitmaps, but use algorithms for compression images which can reduce file size considerably. For this reason, JPEG and GIF images are used on the Web, while BMP images are often used for printable images.

---

## 8.4 SUMMARY

---

In this unit we have discussed about raster data and its type which includes Satellite Imagery, USGS Digital Elevation Model, and Non-USGS Digital Elevation Model & Global DEMs. We have also discussed about Raster Data Structure which includes Cell-by-Cell Encoding, Run-Length Encoding and Quad Tree. We have also learned how to organise raster data and elements of raster data model such as Cell value, Cell size, Raster bands etc. The chapter also explains the process of Data Compression, Raster Data Format which explains Portable Network Graphics (PNG), Joint Photographic Experts Group (JPEG2000), JPEG File Interchange Format (JFIF), Graphic Interchange Format (GIF), Geo Tagged Image File Formats (GeoTIFF), etc.

---

## 8.5 GLOSSARY

---

- **Bi-level scanned file:** A scanned file containing values of 1 or 0.
- **Cell-by-cell encoding:** A raster data structure that stores cell values in a matrix by row and column.
- **Data compression:** Reduction of data volumes, especially for raster data.
- **Digital Elevation model (DEM):** A digital model with an array of uniformly spaced elevation data in raster format.
- **Digital raster graphic (DRG):** A scanned image of a USGS topographic map.
- **ESRI grid:** A proprietary ESRI format for raster data.

- **Floating-point raster:** This raster are used to store what is considered a continuous data with floating point cell values.
- **Georeferenced raster:** A raster that assigning information about geospatial positioning to raster data based on a given coordination scheme.
- **Landsat:** An orbiting satellite that provides the images of the Earth's surface with repeated photos.
- **Lossy compression:** A method of data compression capable of achieving high compression ratios but which cannot completely rebuild the original image.
- **Quad tree:** A system that divides a raster into a quadrant hierarchy.
- **Raster data model:** A data model using spatial characteristics by rows, columns and cells.
- **Rasterization:** Rasterization is a process in which conversion of vector data to raster data is to be done.
- **Run length encoding (RLE):** A structure for raster data, recording cell values in rows and groups. An encoded run-length file can also be considered a compressed run-length file (RLC).
- **Vectorization:** Rasterization is a process in which conversion of raster data to vector data is to be done.
- **Wavelet transform:** A modern technique of image compression that deals image like a wave and gradually breaks the wave into simpler wavelet.

---

## 8.6 ANSWER TO CHECK YOUR PROGRESS

---

Q.1 The vector data model is based on which of the following?

- A) Collections of points joined by straight lines.
- B) Pixels or grid cells.
- C) Cartesian coordinate system.

Q.2 which of the following is NOT a raster data structure?

- A) Block encoding
- B) Run-Length encoding
- C) Spaghetti.
- D) Quad tree.

Q.3 DTMs can be created by digitizing from a paper map.

- A) True
- B) False

Q.4 Run length encoding reduces the size of a raster data set on a row by row basis.

- A) True
- B False

Q.5 Write a short note on Digital elevation model?

---

## 8.7 REFERENCES

---

- Bhatta, (2008) Remote Sensing and GIS Oxford University Press, Pp: 442,121,129,135, 144.
- Floyd F. Sabins, (1996/1997) Remote Sensing Principles and Interpretation, W.H. Freeman And Company New York 3rd Edition, Pp: 29, 69,105,177,236.
- Kalicharan Sahu, (2008), Text Book of Remote Sensing and GIS, Atlantic Publications, Pp: 1-2,127-198.
- Textbook Of Remote Sensing and Geographical Information System, M.Anji Reddy, Second Edition, Pp 1-23.
- [http://en.wikipedia.org/wiki/Raster\\_data](http://en.wikipedia.org/wiki/Raster_data)
- [http://geospatial.referata.com/wiki/Raster\\_Data\\_Model](http://geospatial.referata.com/wiki/Raster_Data_Model)
- <http://gis.stackexchange.com/questions/57142/what-is-the-difference-between-vector-and-raster-data-models> Websites Books
- Bhatta, (2008) Remote Sensing and Gis Oxford University Press
- Chang, Kang-tsung Introduction to Geographic Information Systems 5 th edition 2009 Mcgraw-Hill.
- Lillesand, Thomas M., Ralph W. Kiefer, and Jonatham W.Chipman, 2004
- Textbook Of Remote Sensing and Geographical Information System, M.Anji Reddy, Second Edition, Pp 1-23.

---

## 8.8 TERMINAL QUESTION

---

- Q.1 Explain the difference between lossless and lossy compression method?
- Q.2 Explain the relationship between cell size, raster data resolution, and raster representation of spatial feature?
- Q.3 what are the basic elements of the raster data model?
- Q.4 Describe three new data sources for producing DEMs?
- Q.5 Use a diagram to explain how the run-length encoding methods work?

## **BLOCK 3: SPATIAL DATABASE VECTOR ANALYSIS**

---

### **UNIT 9 - OVERLAY ANALYSIS- UNION, INTERSECTION**

---

**9.1 OBJECTIVES**

**9.2 INTRODUCTION**

**9.3 OVERLAY ANALYSIS- UNION, INTERSECTION**

**9.4 SUMMARY**

**9.5 GLOSSARY**

**9.6 ANSWER TO CHECK YOUR PROGRESS**

**9.7 REFERENCES**

**9.8 TERMINAL QUESTIONS**

---

## 9.1 OBJECTIVES

---

After reading this unit learner will be able:

1. To understand overlay analysis.
2. To know details and difference between raster and vector overlay.
3. Gain Knowledge about application of Overlay analysis

---

## 9.2 INTRODUCTION

---

Overlay operations are part of most spatial analysis processes and generally form the core of GIS projects. These operations combine several maps and thus give new information that was not present in the individual maps. In overlay operations new spatial elements are created on the basis of multiple input maps.

---

## 9.3 OVERLAY ANALYSIS- UNION, INTERSECTION

---

Overlay operations are part of most spatial analysis processes and generally form the core of GIS projects. These operations combine several maps and thus give new information that was not present in the individual maps. In overlay operations new spatial elements are created on the basis of multiple input maps. The principle of the overlay of spatial data can be illustrated through the example of producing maps. When printing a map, different layers of spatial information are printed one after the other onto white paper. Each color represents new information on the paper: green represents vegetation, blue for hydrography, and red for the signatures representing residential areas. All the colors together give a very detailed model of reality: the map. This is the main basis of overlays in GIS.

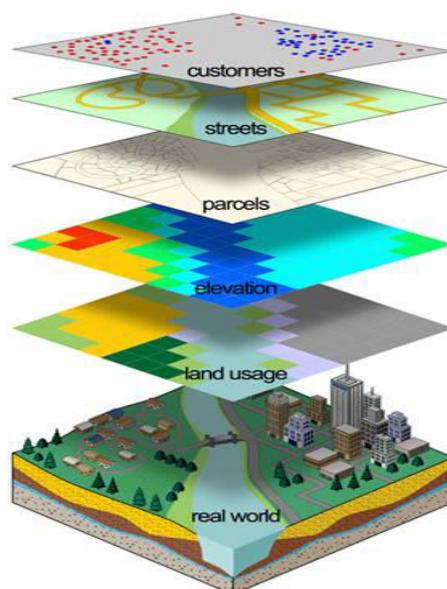


Figure 9.1:

First, two or more layers of information from the same area are overlaid onto each other. Then, the topology of the new layer is updated: if a point now lies within a polygon, it gets assigned this information as a new attribute. If two lines intersect ("arcs") a new node will be added at their intersection. If two polygons intersect, a unique identification number is given to the intersecting set and so forth. Ultimately, the overlay results in an information gain. In order for this integration to make sense, all input layers must have the same reference system and scale. The map will only be legible if all the layers fit together exactly in regards to position and scale. The process itself is independent of whether a raster or a vector model is used. With a raster model, the overlay operation is rather an overlay than an intersection. The integration of information from various sources through overlay is one of the most important functions of a GIS.

Finally, Overlay analysis gives us: "what's within what?"

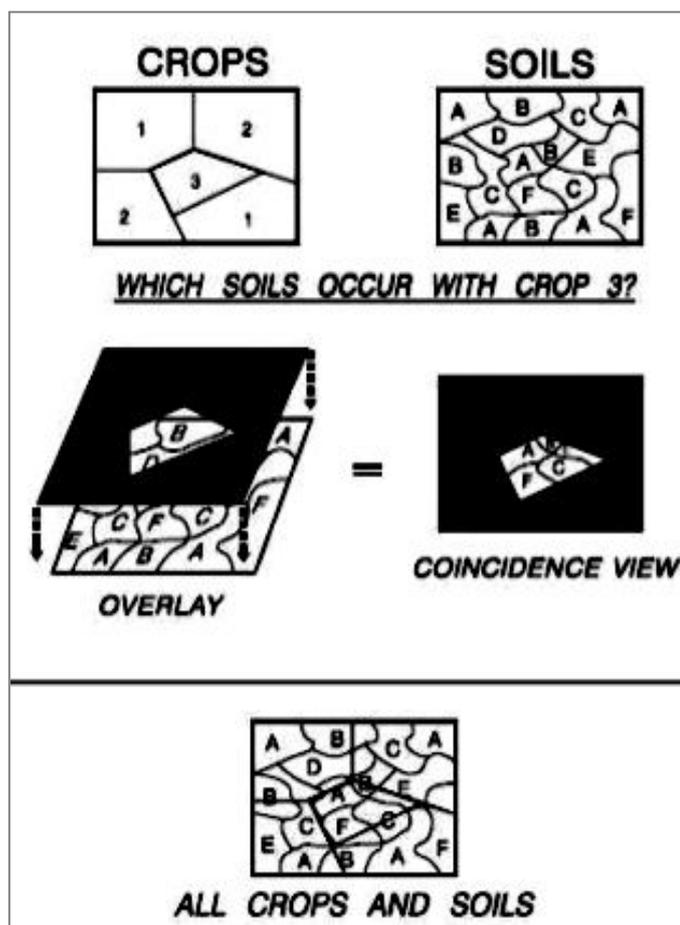


Figure 9.2: How Overlay Analysis Help Users in Different Applications

Overlay creates an output by combining geometries and attributes from different layers (either vector or raster). Overlay output: combines two different layers to form a new layer (different geometry and attribute table).

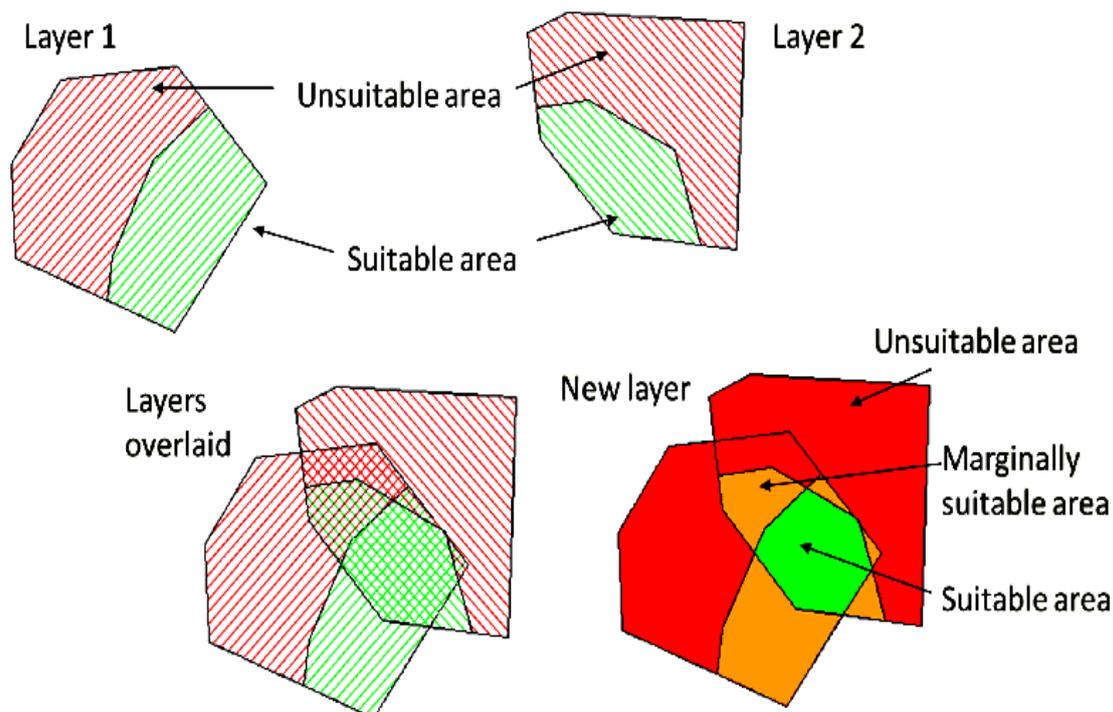


Figure 9.3:

### OVERLAY OPERATIONS IN GIS:

The hallmark of GIS is overlay operations. The capability to overlay multiple data layers in a vertical fashion is the most required and common technique in geographic data processing. In fact, the use of a topological data structure can be traced back to the need for overlaying vector data layers. With the advent of the concepts of mathematical topology, polygon overlay has become the most popular geo-processing tool, and the basis of any functional GIS software package. Topological overlay is predominantly concerned with overlaying polygon data with polygon data, e.g., soils and forest cover. However, there are requirements for overlaying point, linear, and polygon data in selected combinations, e.g., point-in-polygon, line-on-polygon, and polygon-on-polygon are the most common. Vector- and raster-based software differs considerably in their approach to overlay.

Raster-based software is oriented towards arithmetic overlay operations, e.g., the addition, subtraction, division, and multiplication of data layers. The nature of the one-attribute map approach, typical of the raster data model, usually provides a more flexible and efficient overlay

capability. The raster data model affords a strong numerically modelling (quantitative analysis) capability. Most sophisticated spatial modelling is undertaken within the raster domain.

In vector-based systems, topological overlay is achieved by the creation of a new topological theme from two or more existing themes. This requires the rebuilding of topological tables, e.g., arc, node, polygon, and therefore can be time consuming and CPU-intensive. The result of a topological overlay in the vector domain is a new topological theme that contains attributes of the original input data layers. In this way, selected queries of the original layer can then be undertaken, e.g., soils and forest cover, to determine where specific situations occur, e.g., deciduous forest cover where drainage is poor.

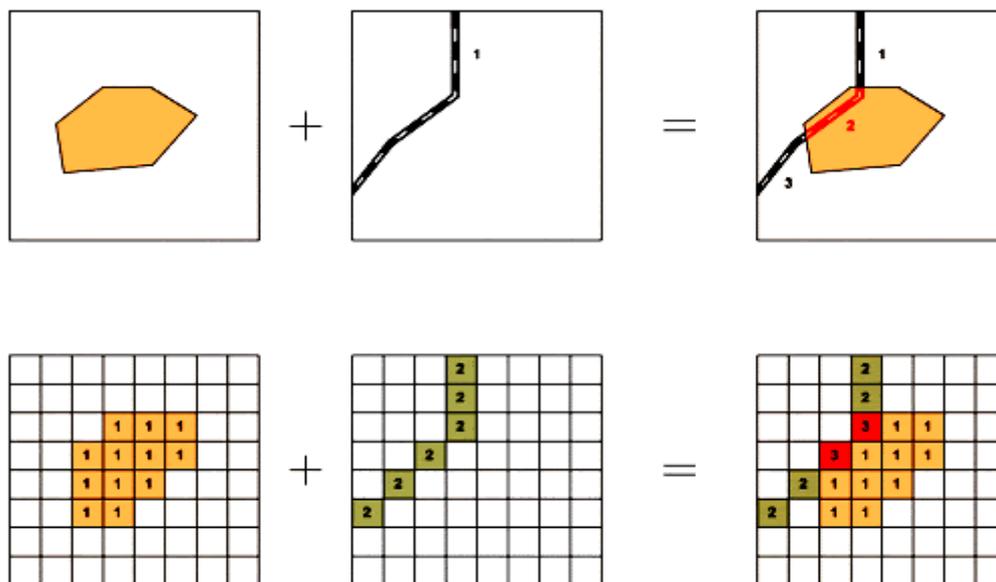


Figure 9.4: Raster and Vector Overlay Analysis in GIS

Most GIS software makes use of a consistent logic for the overlay of multiple data layers. The rules of Boolean logic are used to operate on the attributes and spatial properties of geographic features. Boolean algebra uses the operators AND, OR, XOR, and NOT to see whether a particular condition is true or false. Boolean logic represents all possible combinations of spatial interaction between different features. The implementation of Boolean operators is often transparent to the user.

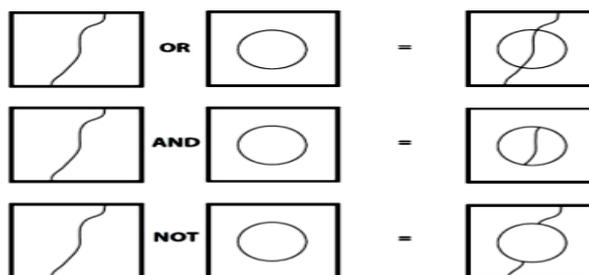


Figure 9.5: AND/OR/NOT in Overlay Analysis

Till date, the primary analysis technique used in GIS applications, vector and raster, is the overlay of selected data layers.

### 1. Union

- It reserves all features from the input and overlay layers
- The area extent of the output combines the area extents of both layers
- Input layers have to be polygons

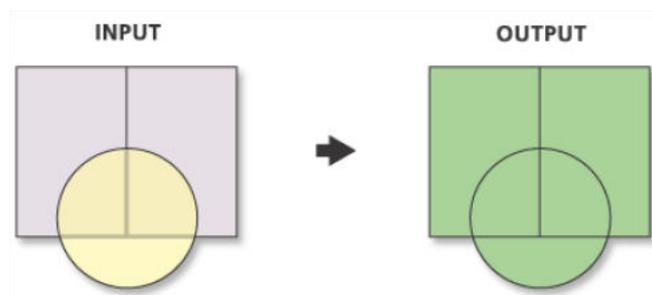


Figure 9.6: Union in Overlay Operation

### 2. Intersect

- It preserves only those features that fall within the area extent common to both layers.
- The inputs can take any geometry but the overlay layer is a polygon
- The attribute table contains only data from both layers

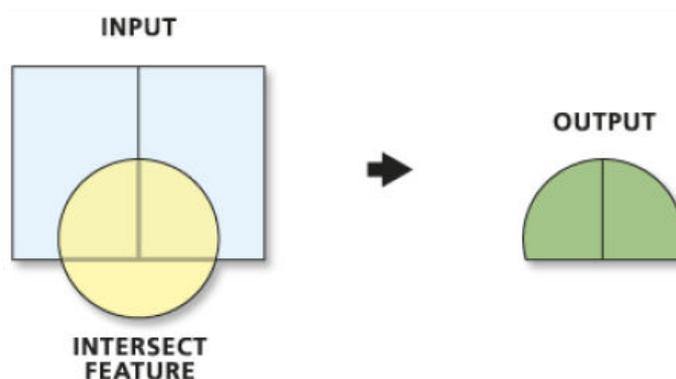


Figure 9.7: Intersect in Overlay

### 3. Symmetrical difference

- It preserves features common to either the input layer or overlay layer but not both
- The geometry of the overlay layer as to be the same as the input

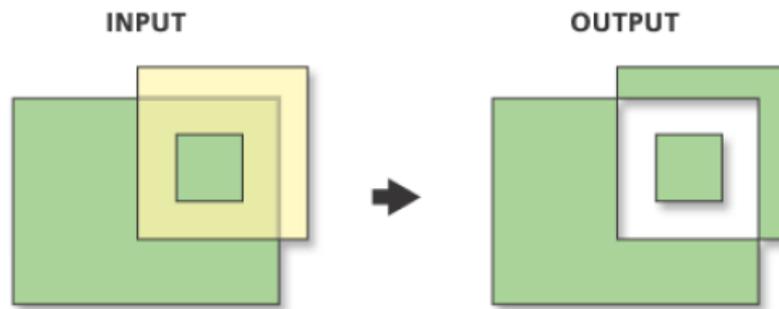


Figure 9.8: Symmetrical Difference in Overlay

#### 4. Identity

- It preserves only features that fall within the area extent of the input layer
- The overlay layer has to be a polygon or the same geometry as the input

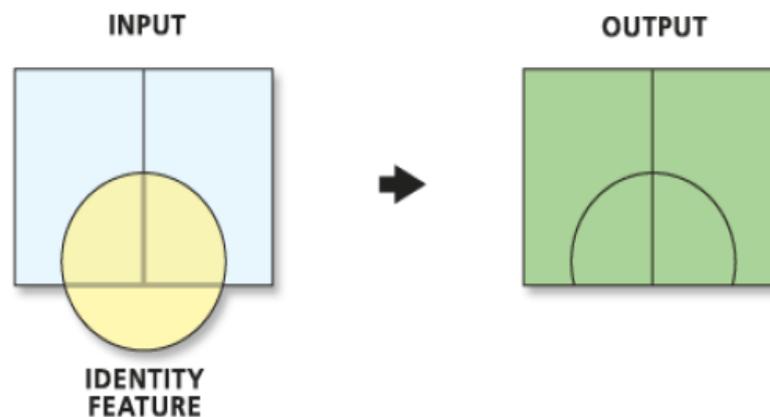


Figure 9.9: Identity Features in Overlay Operation

It is also termed as spatial overlay because it is accomplished by joining and viewing together separate data sets that share all or part of the same area. The result of this combination is a new data set that identifies the spatial relationships. Before the use of computers, a similar effect was developed by Ian McHarg and others by drawing maps of the same area at the same scale on clear plastic and actually laying them on top of each other. Map overlay is used in both model overlay of vector data and overlay of raster data.

#### VECTOR OVERLAY

Overlay of vector data is slightly complicated because it must update the topological tables of spatial relationships between points, lines, and polygons. During the process of overlay, the attribute data associated with each feature type are merged. The resulting table contains both the attribute data. The process of overlay depends upon the modelling approach the user needs. Generally, GIS software implements the overlay of different vector data layers by combining the spatial and attributes data files of the layers to create a new data layer. Again, different GIS software utilizes varying approaches for the display and reporting of overlay results. Some

systems require that topological overlay occur on only two data layers at a time, creating a third layer. One might need to carry out a series of overlay procedures to arrive at the conclusion, which depends upon some criteria.

A union overlay combines the geographic features and attribute tables of both inputs into a single new output. An intersect overlay defines the area where both inputs overlap and retains a set of attribute fields for each. A symmetric difference overlay defines an output area that includes the total area of both inputs except for the overlapping area. Using these operations, new spatial elements are created by the overlaying of maps.

There are three types of vector overlay; point-in-polygon, line-on-polygon, and polygon-on-polygon.

- 1. Point in Polygon Overlay:** Points are overlaid on polygon map as shown in Figure 9.10. Topology of point in polygon is 'is contained in' relationship. Point topology in the new data layer is a new attribute of polygon for each point.

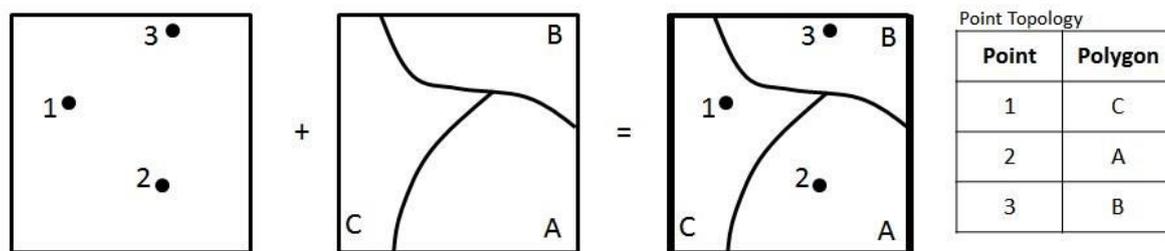


Figure 9.10: Point in Polygon Overlay

- 2. Line on Polygon Overlay:** Lines are overlaid on polygon map with broken line objects as shown in Figure 9.11. Topology of line on polygon is 'is contained in' relationship. Line topology in new data layer is the attribute of old line ID and containing area ID.

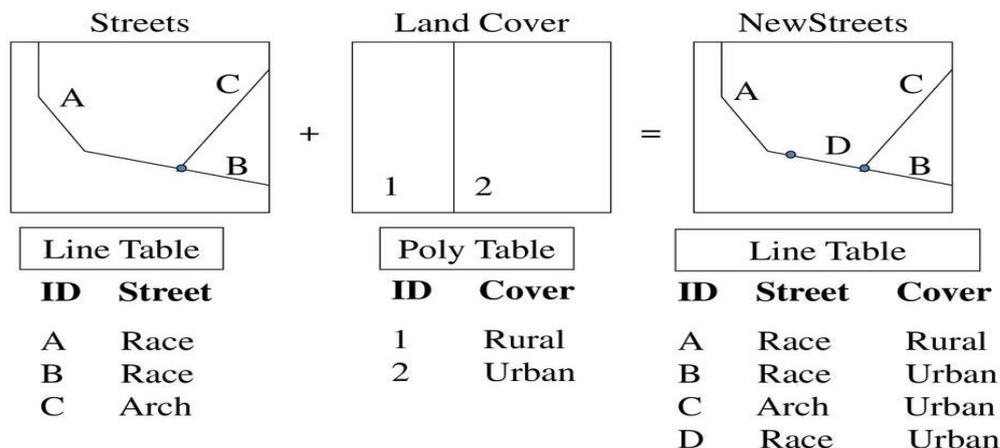


Figure 9.11: Line in Polygon Overlay

**3. Polygon on Polygon Overlay:** Two layers of area objectives are overlaid, resulting in new polygons and intersections as shown in Figure 9.12. The number of new polygons is usually larger than that of the original polygons. Polygon topology in the new data layer is a list of original polygon IDs.

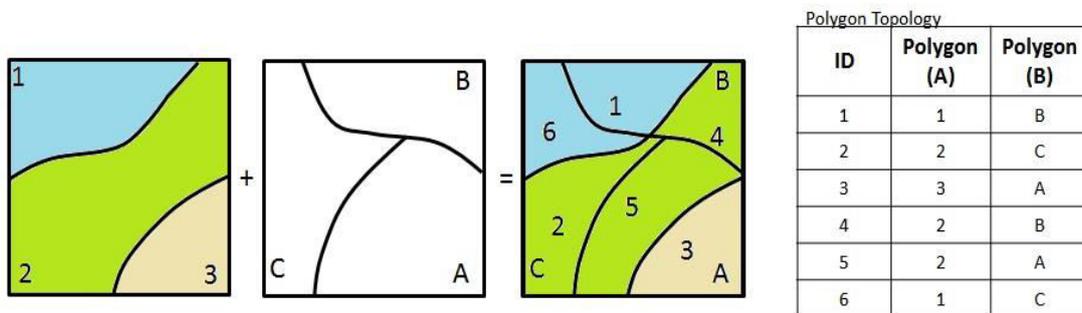


Figure 9.12: Polygon in Polygon Overlay

**RASTER OVERLAY:**

Overlay of raster data with more than two layers is rather easier as compared with overlay of vector data, because it does not include any topological operation but only pixel by pixel operations. In raster data analysis, the overlay of datasets is accomplished through a process known as ‘local operation on multiple rasters’ or ‘map algebra’, through a function that combines the values of each raster’s matrix for mathematical calculations. This function may weigh some inputs more than others through use of an ‘index model’ that reflects the influence of various factors upon a geographic phenomenon (Figure 9.13).

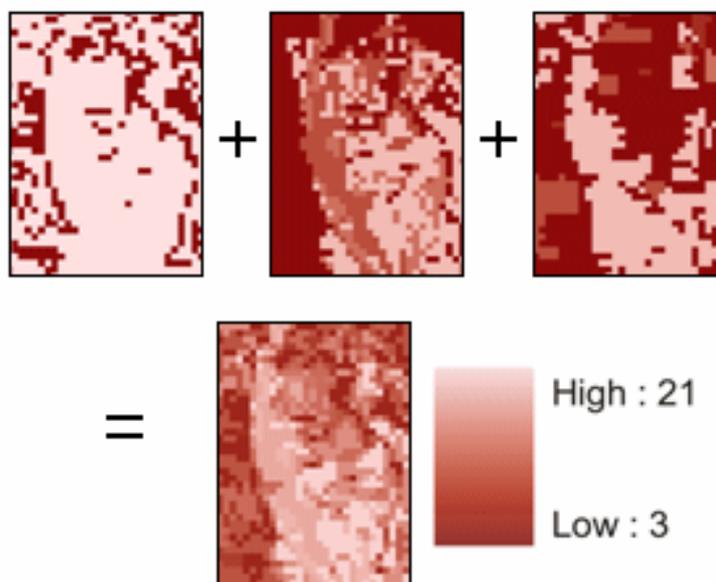


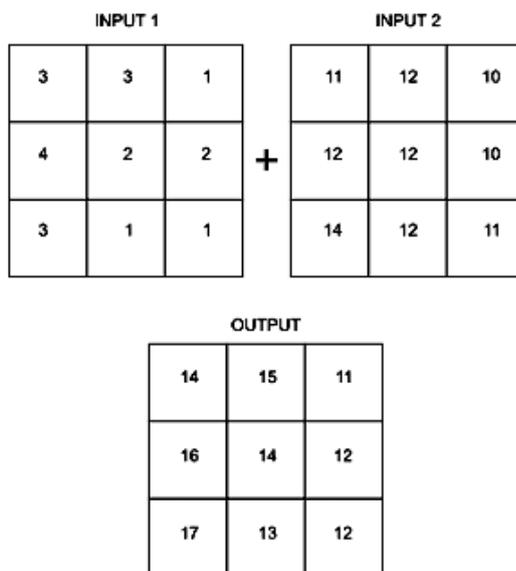
Figure 9.13: Raster Overlay



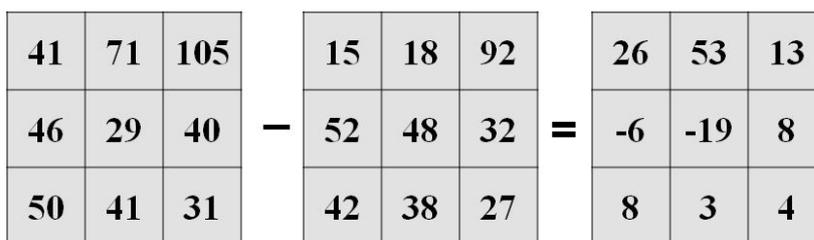
introduces the idea of map algebra. It means in the raster data processing, some analysis use individual cells only and some rely on neighboring or regional associations. Thus the raster data processing methods can be classified into the following categories:

1. Local operations
2. Neighborhood operation
3. Regional operations

**Local operations** are based on point-by-point or cell-by-cell analysis. The most important of this group is the **overlay analysis**. In the raster based analysis either the logical or arithmetic operators are used. The logical overlay methods use operators AND, OR, and XOR (exclusive OR). Mathematically AND multiplies the individual cells whereas logical OR and XOR add individual values of corresponding cells. The most important consideration in raster overlay is the appropriate coding of the features in the input layers. The raster overlay is affected by the resolution (cell size) and scale of measurement (nominal, ordinal, interval or ratio). It is advised that the resolution and the scale of measurement of both the input and analysis layer should be compatible. Basic arithmetic operators in raster overlay operations are **ADDITION, SUBTRACTION, DIVISION, and MULTIPLICATION**. All these operators are explained here with the self-explanatory diagrams.



**Figure 9.16: Addition**



**Figure 9. 17: Subtraction**

20	18	25	36	12
15	32	20	30	18
10	20	30	32	24
36	10	35	32	6
24	20	35	32	18

 $\div$ 

5	6	5	6	2
3	8	4	5	6
5	5	6	8	4
5	5	7	4	2
6	4	5	2	6

 $=$ 

4	3	5	6	6
5	4	5	6	3
2	4	5	4	6
6	2	5	8	3
4	5	7	16	3

Figure 9.18: Division

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

 $\times$ 

5	6	5	6	2
3	8	4	4	6
5	5	6	8	4
5	5	2	4	2
6	4	5	2	6

 $=$ 

15	18	15	18	2
6	32	8	4	24
15	20	18	32	8
20	15	2	8	2
12	4	15	2	18

Figure 9.19: Multiplication

The cell-by-cell operations are termed as location specific overlay. But, assigning values to the entire thematic regions creates a new layer, and termed as category- wide overlay.

**Local neighborhood operations** are also known as focal operations. It uses the topological relationship of adjacency between cells in the input raster layer to create a new layer. This operation assumes that the value of particular cell is affected by the value of the neighboring cell. Hence a moving window of 3\*3 cells is generally applied on the input raster layer. The value of the output cell may be either the average of all the cells of the moving window, or the central cell of the window or the median value of the window.

**Operators on regions (Regional operators)** are also known as zonal operations. Generally a region is defined as the area with homogeneous characteristics. In raster model it has been defined as the collection of cells that exhibits the same attribute characteristics. Though it also uses two input layer but the mode of operation and the purpose is altogether different from the local operations as it use the boundary of one raster layer to extract the cell values from the other raster layer. The major purpose of this kind of operation is to obtain relevant data from an existing layer for further spatial analysis. There is no uncertainty in the location of the region boundaries because they are in perfect registration.

### RECLASSIFICATION:

Reclassification is method of changing the attribute values without altering the geometry of the map. In fact it is a database simplification process that aims at reducing the number of categories of attribute data layer. Accordingly, features adjacent to one another that have a common value,

will be treated and appear as one class. Reclassification is an attribute generalization technique. Typically this function makes use of polygon patterning techniques such as crosshatching and/or colour shading for graphic representation. It usually uses either logical or arithmetic operators for raster data or arithmetic operator for vector data. After reclassification, the common boundaries between polygons with identical attribute values are dissolved. Consequently the topology will be rebuilt.

### **WEIGHTED OVERLAY :**

The objective behind area-on-area overlay on a vector data model, “to identify one or more parts of the new geometry that met simple criteria. Areas that did not meet the criteria were discarded”. This was processed as a single task. The function of weighted overlay is to determine a new set of values for the complete coverage based on a combination of input values. There are two task to perform that working with a vector data model.

1. Create a new set of geometries for the entire area, and
2. Compute a new set of attributes for those geometries.

After performing the above mention task is a matter of describing a mathematical equation to process the input values. In the first task requires you to extend the basic polygon overlay operation to consider every intersection between all polygons in every data layer. As you can imagine this can be computationally demanding, especially if the GIS you are using computes topology 'on the fly' and does not store it in the data structure. As we shall see this is one of the reasons why weighted overlay is more frequently applied to a raster data model. However, there are requirements for overlaying point, linear, and polygon data in selected combinations, e.g. point in polygon, line in polygon, and polygon on polygon are the most common.

The arcs of the input layers are split at their intersection with arcs of the union layer. Thus the number of polygons in the output layer will be larger than the input layer. It is the Boolean operation that uses OR. Therefore the output map corresponds to the area extent of input layer or analysis layer or both. UNION requires both the input and analysis layer be polygon. This operator is generally used for querying and analysis of urban sprawl.

INTERSECT operator performs the intersection of two input layers. The resultant layer will keep those portions of the first input layer features which fall within the second input layer polygon. That is, features that lie in common area of both the input layers. It uses the Boolean operator AND. The point of caution is that the input layer may either be a point or, line or polygon but the analysis layer should always be a polygon.

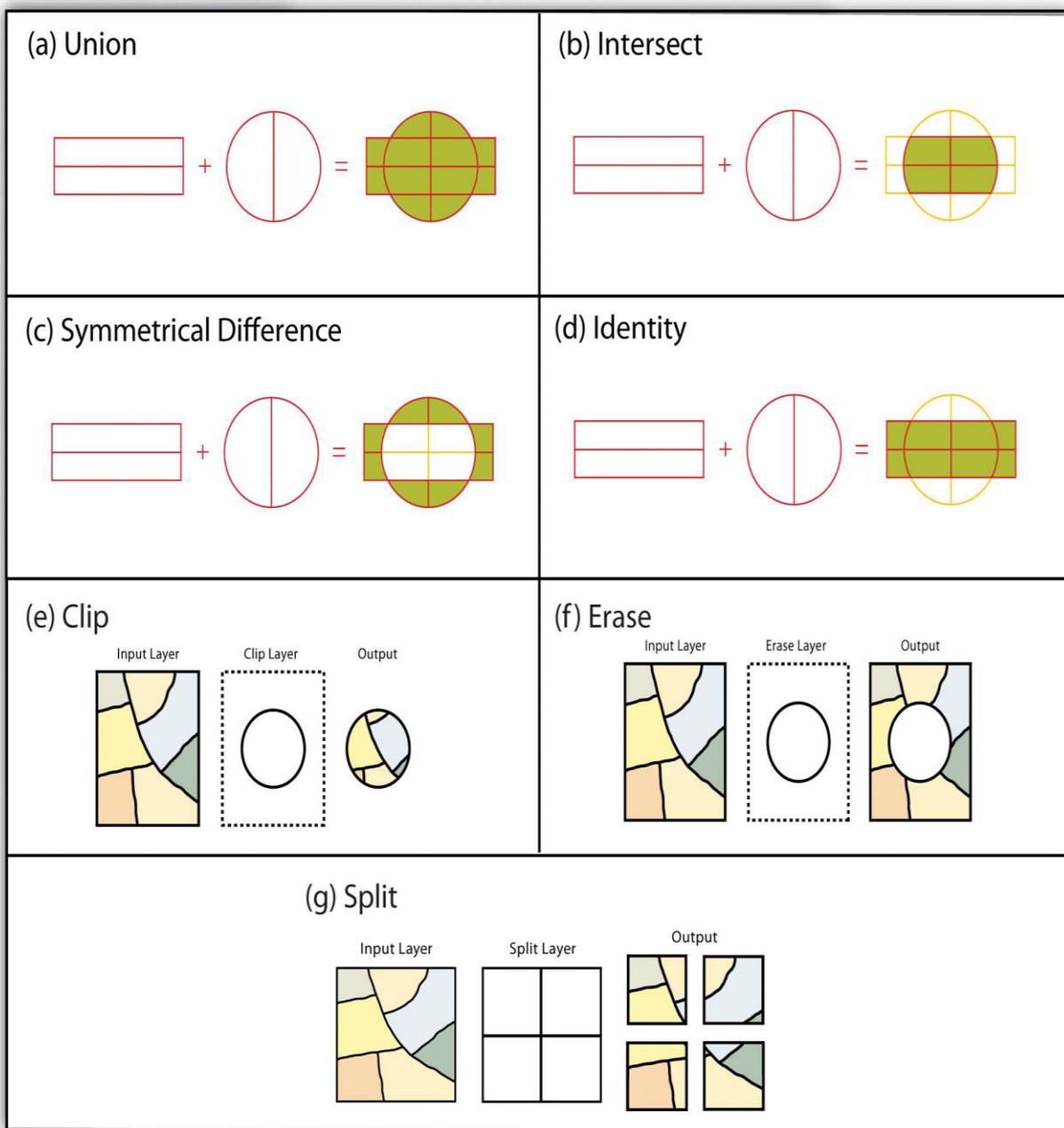


Figure 9.20: Weighted Overlay Operation

Some important overlay operations are discussed here.

1. **CLIP** creates a new map that includes only those features of the input layer that falls within area extent of the clip map. The input layer may be points, lines, or polygons but the analysis layer (clip layer) must be polygon layer. This operator is used to extract a smaller dataset from a larger dataset. For example, the extraction of a city from the state maps. The city boundary is considered as the analysis layer.
2. **ERASE** is a reverse process of Clip where the features of the input layer that fall within the boundary of the analysis are erased and those fall outside the boundary of the analysis

layer are retained. The retained layer will have all attributes from the input layer only. The input layer may be point, line and polygon but the analysis layer is always a polygon.

3. **SPLIT** divides the input coverage into two or more coverages. For this a series of clip operation is performed. Each resultant layer contains only those portions of the input layer that are overlapped by the polygon satisfying the specified criteria. For example, a national forest cover can use SPLIT to divide vegetation coverage by district so that each district can have its own vegetation coverage.
4. **UPDATE and IDENTITY UPDATE** uses a cut and paste operation to replace the input coverage and its map features.
5. **IDENTITY** operation overlays polygons and keeps all input layer features and only those features from the analysis layer that overlap the input layer. The resultant layer will have the same spatial features as that of the input layer. In case of polygon overlays the number of polygon in the output layer will always be larger in number than the input layer.

It is expressed in a Boolean operation (input map) AND (overlay map) OR (input map). The input map may contain points, lines or polygons. The word of caution is that this operation can only be ideally applied if the map boundary is precisely maintained. Beside these operators a number of other operators such as MERGE, Append, ELIMINATE, and RESELECT DISSOLVE etc are also used in the vector overlay operation

#### **APPLICATION:**

1. Delimitation of protected zones around features like defining buffer zones along river streams to restrict urban developments.
2. Creating restrictions criteria for the location of an industrial site based on buffers along conservation areas, river streams, and residential areas.
3. Definition of areas of influence like generating a buffer zone centred on a school to estimate the number of potential students
4. Delimitation of protected zones around features
5. GIS spatial analysis using buffer to identify riparian land use
6. GIS spatial analysis using buffer to define a search radius centred in one specific feature.

A suitability model can be used to find the best location to construct a new school, hospital, police station, industrial corridors etc. Certain land uses are more conducive than others for building a new school for example, forest and agriculture were more favorable than residential housing in this model. It was desired to locate the school on flat slopes, near recreation sites, and far from existing schools.

---

## 9.4 SUMMARY

---

We have to remember that analysis through GIS begins with overlay analysis; let it be the examples given above or other decision making analysis undertaken using diverse spatial data sets. Today the analysts are working with real time data sets, voluminous big data component and speedy processing of layered data with ready output. GIS is considered as a decision making tool in problem solving almost all environmental concerns. Spatial analysis is a vital part of GIS and can be used for many applications like site suitability, natural resource monitoring, environmental disaster management and many more. Vector, raster based analysis functions and arithmetic, logical and conditional operations are used based on the recovered derivations. With the technology expanding and the output tool readily available in the hands of individuals and experts the challenges have increased and expanded their dimensions.

---

## 9.5 GLOSSARY

---

1. **Overlay:** Overlay is a GIS operation that superimposes multiple data sets (representing different themes) together for the purpose of identifying relationships between them.. An overlay creates a composite map by combining the geometry and attributes of the input data sets.
2. **Raster Overlay:** Raster overlay involves two or more different sets of data that derive from a common grid. The separate sets of data are usually given numerical values. These values then are mathematically merged together to create a new set of values for a single output layer.
3. **Vector Overlay:** A vector overlay involves combining point, line, or polygon geometry and their associated attributes. All overlay operations create new geometry and a new output geospatial data set. The clip function defines the area for which features will be output based on a “clipping” polygon.
4. **Weighted Overlay:** Weighted overlay is one method of modeling suitability. ArcGIS uses the following process for this analysis. Multiplying each layer's weight by each cell's suitability value produces a weighted suitability value. Weighted suitability values are totaled for each overlaying cell and then written to an output layer.
5. **Reclassification:** **Reclassification** operations merely repackage existing information on a single map. **Overlay** operations, on the other hand, involve two or more maps and result in the delineation of new boundaries.
6. **Boolean:** Binary (two-valued) system of variables and operations for logical operations developed by George Boole in the mid-nineteenth century.

7. **Decision Support System:** An interactive, computer based system that supports decision making.
8. **Digitizing:** The process of converting analog spatial information from sources like paper maps to digital data.
9. **Projection (Map):** A method to transform the Earth's curved surface onto a plane.
10. **Topology:** The geometric relationship between points, lines, and geometric forms that remains consistent throughout spatial operations in a digital mapping environment.

## 9.6 ANSWER TO CHECK YOUR PROGRESS

1. Find out the differences between Raster and Vector Overlay.
2. Explain the application of Overlay.

## 9.7 REFERENCES

1. Carver, S. J. (1991). Integrating multi-criteria evaluation with geographical information systems. *International Journal of Geographic Information Systems* 5, 321--339.
2. Chrisman, N. (2002). *Exploring geographic information systems* (2nd edn.). New York: Wiley. Eastman, J. R. (2005).
3. Multi-criteria evaluation and GIS. In Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (eds.) *Geographical information systems – principles and technical issues* (2nd edn.). Hoboken, NJ: Wiley.
4. Herbertson, A. J. (1905). The major natural regions: An essay in systematic geography. *The Geographical Journal* 25, 300--310.
5. Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities*. Washington, DC: US Federal Housing Administration.
6. McHarg, I. (1965). *Design with nature*. Garden City, NY: Natural History Press.
7. Tomlin, D. (1990). *Geographic information systems and cartographic modelling*. Engelwood Cliffs, NJ: Prentice-Hall.
8. Tomlinson, R. (1967). *An introduction to geographic information system of the Canada land information inventory*. Ottawa: Department of Forestry and Rural Development.

## 9.8 TERMINAL QUESTIONS

1. Define Overlay. What are the major types of Overlay?
2. Carry out some spatial operations in vector layer using any open source GIS Software.
3. Explain major operations, we can do in overlay.

---

## **UNIT 10 - PROXIMITY ANALYSIS- BUFFERING**

---

***10.1 OBJECTIVES***

***10.2 INTRODUCTION***

***10.3 PROXIMITY ANALYSIS- BUFFERING***

***10.4 SUMMARY***

***10.5 GLOSSARY***

***10.6 ANSWER TO CHECK YOUR PROGRESS***

***10.7 REFERENCES***

***10.8 TERMINAL QUESTIONS***

---

## 10.1 OBJECTIVES

---

After reading this unit learner will be able to understand:

- Proximity analysis.
- Buffers in GIS.
- Application of buffering

---

## 10.2 INTRODUCTION

In geographic information systems and spatial analysis, proximity analysis is the determination of a zone around a geographic feature containing locations that are within a specified distance of that feature, the buffer zone. A buffer is likely the most commonly used tool within the proximity analysis methods. Let's discuss, what are buffers in GIS?

---

## 10.3 PROXIMITY ANALYSIS- BUFFERING

---

In geographic information systems and spatial analysis, proximity analysis is the determination of a zone around a geographic feature containing locations that are within a specified distance of that feature, the buffer zone. A buffer is likely the most commonly used tool within the proximity analysis methods.

Let's discuss, what are buffers in GIS? A buffer is a reclassification based on distance: classification of within/without a given proximity. Buffering involves measuring distance outward in directions from an object. Buffering can be done on all three types of vector data: point, line, area. The resulting buffer is a polygon file.

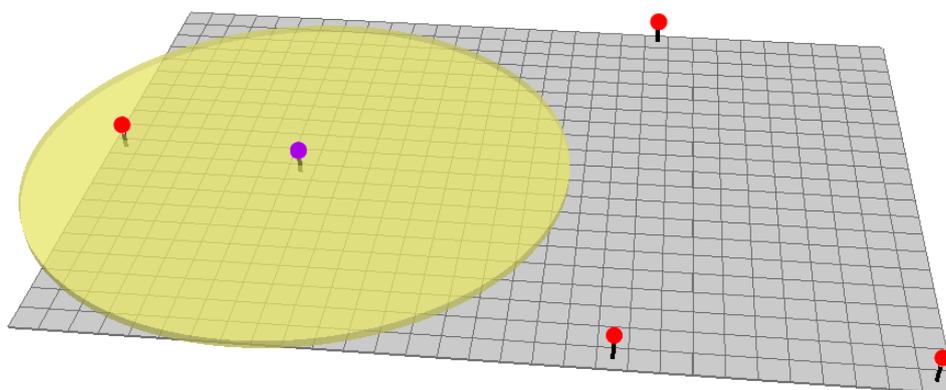


Figure 10.1: Proximity Analysis

Proximity usually creates two areas: one area that is within a specified distance to selected real world features and the other area that is beyond. The area that is within the specified distance is called the buffer zone.

A buffer zone is any area that serves the purpose of keeping real world features distant from one another. Buffer zones are often set up to protect the environment, protect residential and commercial zones from industrial accidents or natural disasters, or to prevent violence. Common types of buffer zones may be greenbelts between residential and commercial areas, border zones between countries eg. noise protection zones around airports, or pollution protection zones along rivers.

### **IMPORTANCE OF PROXIMITY ANALYSIS:**

You should note that much of the analysis using GIS is performed utilizing information about the distance (proximity) between features. For example, if a bank is interested in opening its ATM branch in city, an important criterion will be its distance from the places such as business offices, shops, markets, railway station, bus stand etc. this kind of analysis is known as proximity analysis which is a way of analysing locations of features by measuring the distance between certain features in an area. There are more than two ways of proximity analysis as the distance between point 1 and point 2 may be measured either as a straight line or by following a networked path, such as a road network. For the two types of proximity analysis, buffer and networked path approaches are performed.

You just need to provide the points locations representing the site and the railway station or bus stand or market in order an approximate distance measure. Once the distance are determined, other relevant information such as availability of required resources, prices of real state, etc. can be analysed from the database. Proximity analysis can be performed from point in a layer to point in another layer, or from each point in a layer to its nearest point or line in another layer.

For Example: Finding Shortest Path Activity

The shortest path can be found using road graph plug-in of Q-GIS. It calculates the shortest path between two points on any line layers and plots this path over the road network. Some of the main features of this plugin are:

- Calculate path, its length and travel time
- Optimise by length or by travel time
- Export path to a vector layer and

- Highlight road directions

In order to bind the start and stop points of the route to the road network, road graph selects the nearest point or arc of the graph. In fact, it can bind to any part of the road network. Nevertheless, their route and its characteristics do not take into account the distance from the starting point to the road network to the shopping point.

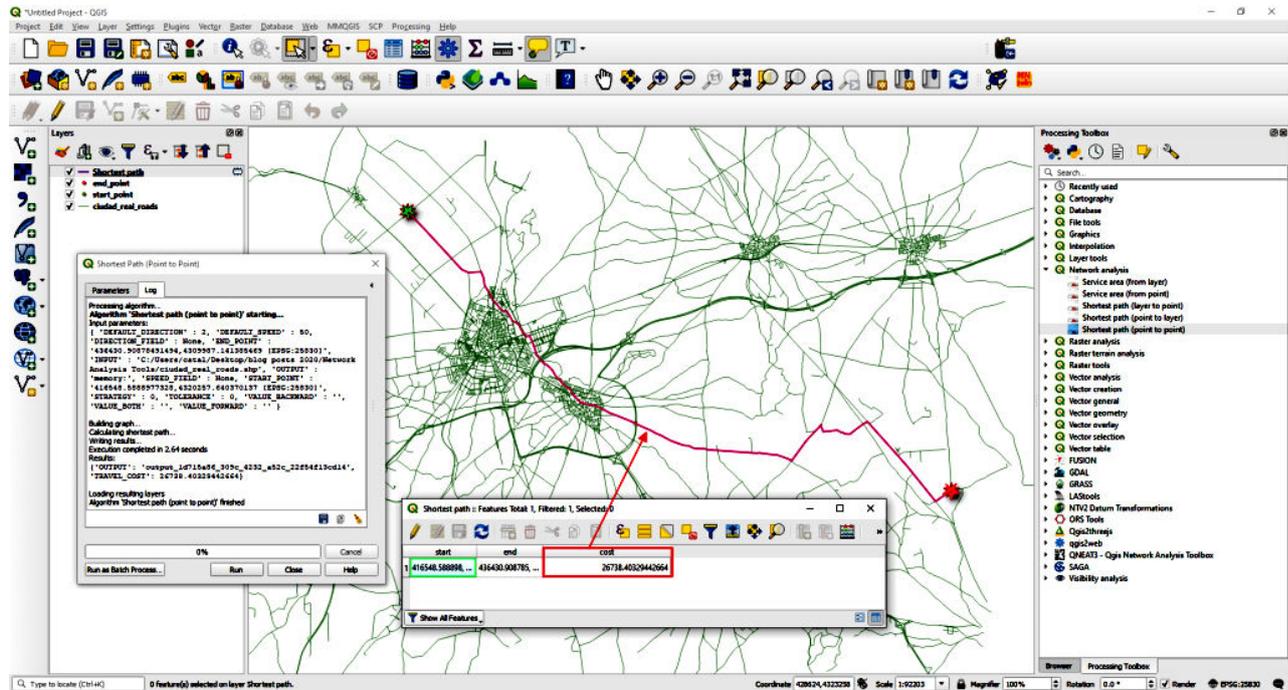


Figure 10.2: Finding Shortest Path in Q-GIS

## CREATING BUFFERS :

Tools plugin gives geo-processing tools which create buffer around features based on distance field. Follow the steps given below to create buffer in Q-GIS.

1. Open Road Layer and Places Layer on Q-GIS (to simplify, clip both the layers to features within India)
2. Select Buffer from Geoprocessing Tools sub menu item of Vector Menu. This opens buffer dialogue box which appears as in figure below.



Figure 10.3: Selecting Geoprocessing Tool and Buffering Option in Q-GIS

3. Select major points as input layer, buffer distance to be 5 km.
4. Define an output file and click ok which appears to be in figure 10. 5.

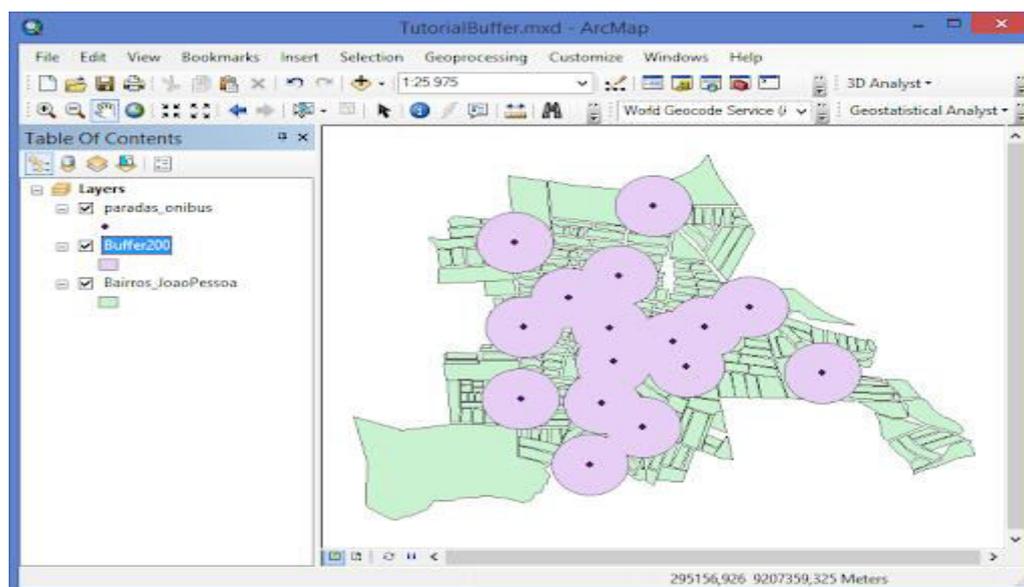


Figure 10.4: Buffering Exercise in Q-GIS

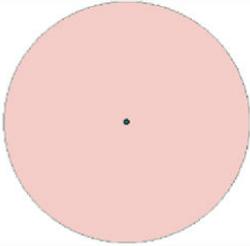
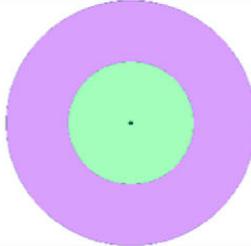
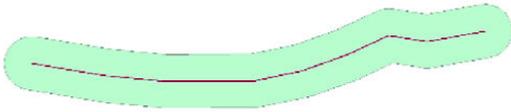
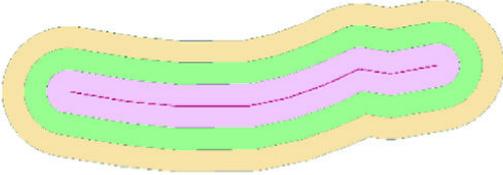
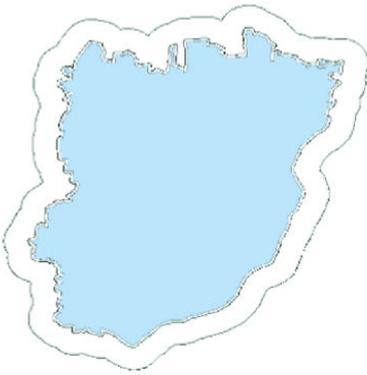
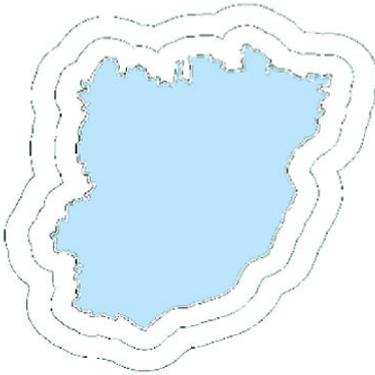
Feature	Single buffer	Multiple-ring buffer
Point		
Line		
Polygon		

Figure 10.5: Buffering Types in Vector Layer

### Two Types of Buffers:

Two main types of buffers are fixed width buffers, and variable width buffers.

#### 1. Fixed Width Buffers:

The fixed width buffer is one of the most common buffers. A fixed width buffer is exactly as its name implies; it is a buffer that has a uniform, unchanging width all the way around the object. This type of buffer is used with the assumption that the impact zone of the buffered object has an equal impact all the way around itself. An example of a fixed width buffer is how far away houses are allowed to be placed from a stream or river. Since laws may govern this, the distance does not change anywhere along the path of the stream or river.

#### 2. Variable Width Buffers:

Variable width buffers allow a varied width between the outside of the buffer and the object being buffered. This type of buffer takes into account variables that would

cause the buffered zone around an object to be inconsistent. An example of this type of buffer is mapping the fallout zone around a nuclear reactor, while the fallout zone is being blown by the wind. In the event of a nuclear fallout, the wind could be blowing from east to west. The wind blowing the radiation to the west would cause the area to the west to have a much higher radiation hazard than the area to the east of the reactor. Buffering this area would show a very narrow buffer zone on the east side of the reactor and a very elongated buffer zone to the west.

### VARIATIONS IN BUFFERING:

There are several variations in buffering. The buffer distance or buffer size **can vary** according to numerical values provided in the vector layer attribute table for each feature. The numerical values have to be defined in map units according to the Coordinate Reference System (CRS) used with the data. For example, the width of a buffer zone along the banks of a river can vary depending on the intensity of the adjacent land use. For intensive cultivation the buffer distance may be bigger than for organic farming.

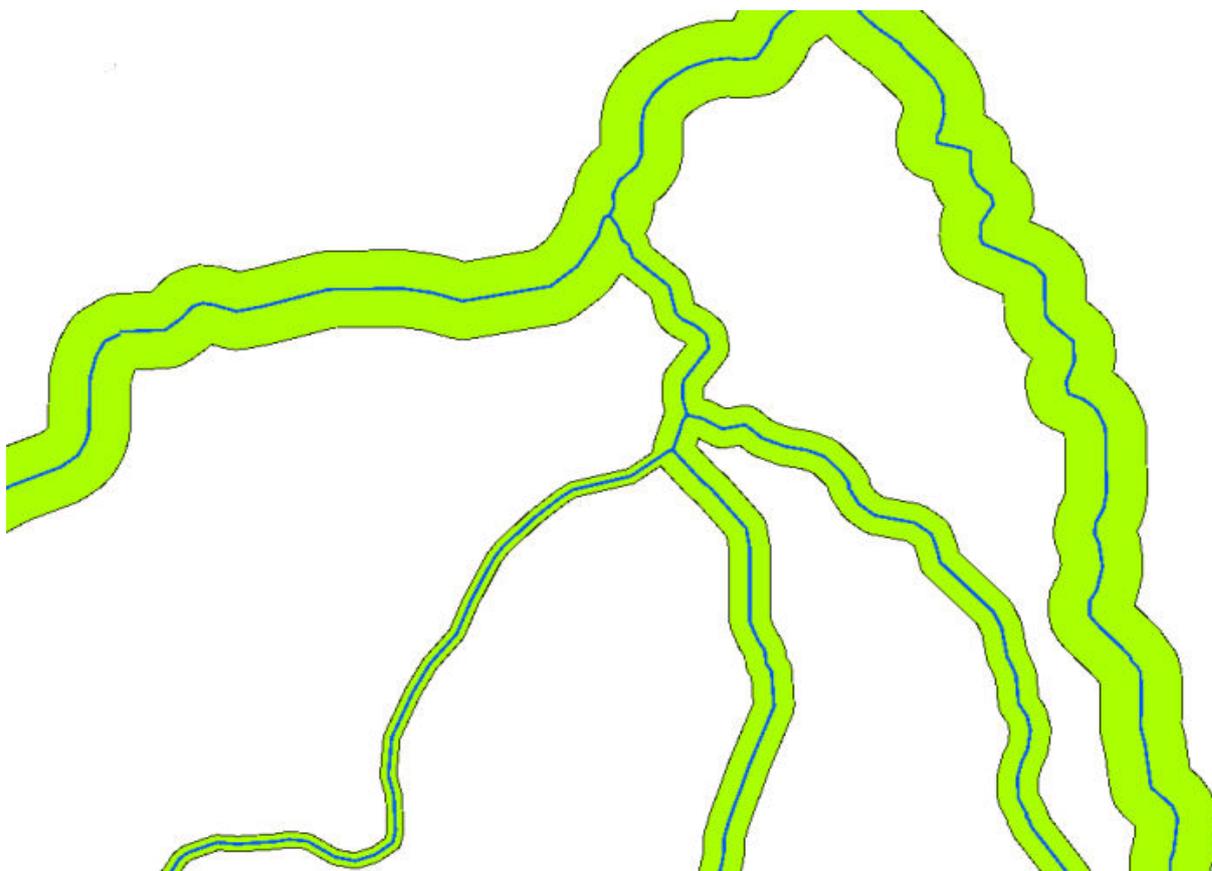


Figure 10.6: Variations in Buffering

Table 10.1: buffering Rivers with different buffer distances.

River	Adjacent land use	Buffer distance (meters)
Breede River	Intensive vegetable cultivation	100
Komati	Intensive cotton cultivation	150
Oranje	Organic farming	50
Telle river	Organic farming	50

Buffers around polyline features, such as rivers or roads, do not have to be on both sides of the lines. They can be on either the left side or the right side of the line feature. In these cases the left or right side is determined by the direction from the starting point to the end point of line during digitizing.

#### MULTIPLE BUFFER ZONES:

A feature can also have more than one buffer zone. A nuclear power plant may be buffered with distances of 10, 15, 25 and 30 km, thus forming multiple rings around the plant as part of an evacuation plan.

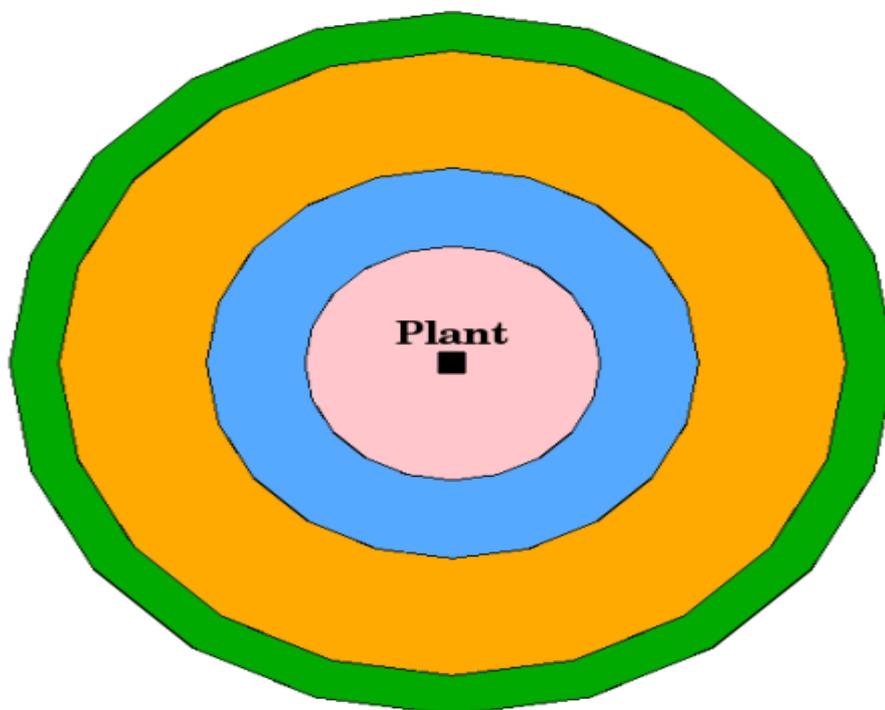


Figure 10.7: Multiple buffering a point feature with distances of 10, 15, 25 and 30 km.

**BUFFERING WITH INTACT OR DISSOLVED BOUNDARIES:**

Buffer zones often have dissolved boundaries so that there are no overlapping areas between the buffer zones. In some cases though, it may also be useful for boundaries of buffer zones to remain intact, so that each buffer zone is a separate polygon and you can identify the overlapping areas.

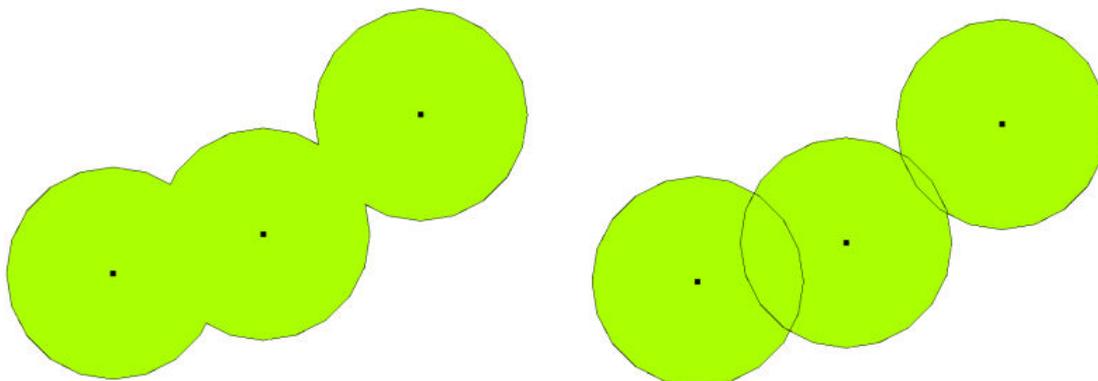


Figure 10.8: Buffer zones with dissolved (left) and with intact boundaries (right) showing overlapping areas.

**BUFFERING OUTWARD AND INWARD:**

Buffer zones around polygon features are usually extended outward from a polygon boundary but it is also possible to create a buffer zone inward from a polygon boundary. Say, for example, the Department of Tourism wants to plan a new road around Robben Island and environmental laws require that the road is at least 200 meters inward from the coast line. They could use an inward buffer to find the 200 m line inland and then plan their road not to go beyond that line.

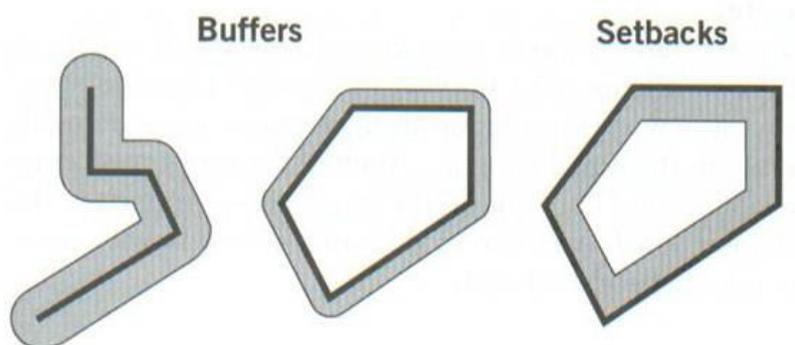


Figure 10.9: Buffering Inward and Outward

## COMMON PROBLEMS IN BUFFERING :

Most GIS Applications offer buffer creation as an analysis tool, but the options for creating buffers can vary. For example, not all GIS Applications allow you to buffer on either the left side or the right side of a line feature, to dissolve the boundaries of buffer zones or to buffer inward from a polygon boundary. A buffer distance always has to be defined as a whole number (integer) or a decimal number (floating point value). This value is defined in map units (meters, feet, decimal degrees) according to the Coordinate Reference System (CRS) of the vector layer.

## APPLICATION OF BUFFERING :

**Buffer zones** are areas created to enhance the protection of a specific conservation area, often peripheral to it. Within **buffer zones**, resource use may be legally or customarily restricted, often to a lesser degree than in the adjacent protected area so as to form a transition **zone**.

1. Buffering creates a buffer zone data set.
2. A buffer zone often treated as a protection zone and is used for planning and regulatory purposes.
3. A city requires a buffer zone of 550m for alcohol trading from school.
4. A 30m buffer zone along bank may be needed to protect a river.
5. Buffering operations include, for example, identifying protected zone, around lakes and streams, zone of noise pollution, around highway, service zone around bus route and ground water pollution zone around waste site.

## EXAMPLE:

### Geodesic Buffer Example:

The goal of this example is to compare 1,000 kilometer geodesic and Euclidean buffers of a number of select world cities. Geodesic buffers were generated by buffering a point feature class with a geographic coordinate system, and Euclidean buffers were generated by buffering a point feature class with a projected coordinate system (in both the projected and unprojected datasets the points represent the same cities).

When working with a dataset in one of the common projected coordinate systems for the whole world, such as Mercator, projection distortion may be minimal near the equator, but significant near the poles. This means that for a Mercator projected dataset, distance measurements and buffer offsets should be quite accurate near the equator and less accurate away from the equator.

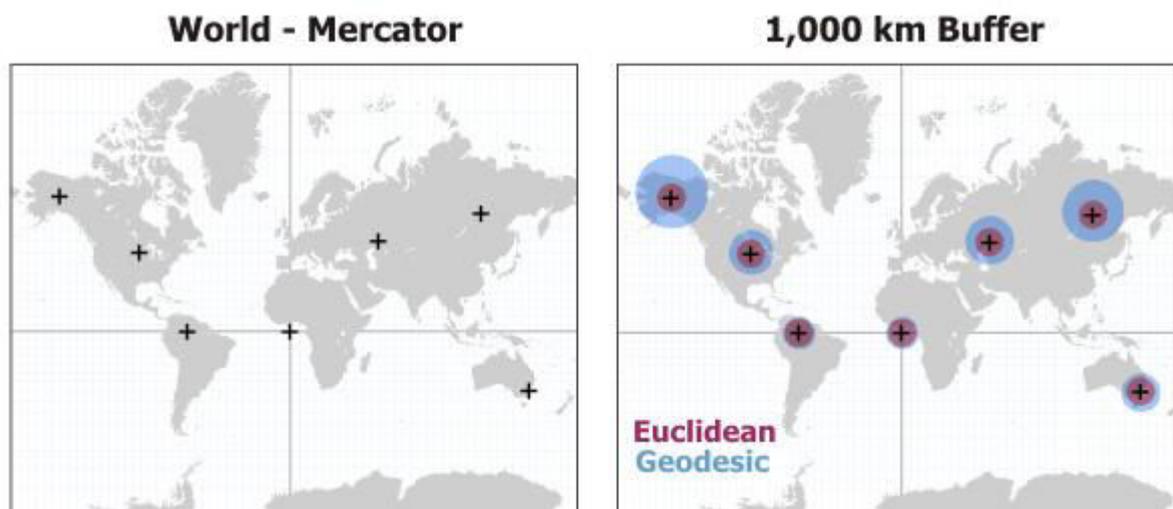


Figure 10.10: Buffering of Major Cities for 1000km

The graphic on the left shows the input point locations. The equator and prime meridian are shown for reference. Both graphics are displayed in the Mercator (World) projection.

In the graphic on the right, the points near to the equator have geodesic and Euclidean buffers that are coincident. For points near to the equator, the Mercator projection does a good job of producing accurate distance measurements. However, the buffers of points far from the equator show considerably more distance distortion, as their Euclidean buffers are much smaller than the geodesic buffers; this occurs with the Mercator projection because at the poles areas are stretched (land masses close to the poles, such as Greenland and Antarctica have enormous areas in comparison with the land masses close to the equator). All 1,000 kilometer Euclidean buffers are the same size since the Euclidean buffer routine assumes that map distances are the same everywhere in the projection (1,000 kilometers in Brazil is the same as 1,000 kilometers in central Russia); this is not true since away from the equator the projection's distances become more and more distorted. With any type of analysis of distance on a global scale geodesic buffers should be used as they will be accurate in all areas while Euclidean buffers will not be accurate in high distortion areas.

---

## 10.4 SUMMARY

---

What have we learned?

Let's wrap up what we covered in this worksheet:

- Buffer zones describe areas around real world features.
- Buffer zones are always vector polygons.
- A feature can have multiple buffer zones.

- The size of a buffer zone is defined by a buffer distance.
- A buffer distance has to be an integer or floating point value.
- A buffer distance can be different for each feature within a vector layer.
- Polygons can be buffered inward or outward from the polygon boundary.
- Buffer zones can be created with intact or dissolved boundaries.
- Besides buffering, a GIS usually provides a variety of vector analysis tools to solve spatial tasks.

---

## 10.5 GLOSSARY

---

1. **Proximity:** Proximity analysis is a class of spatial analysis tools and algorithms that employ geographic distance as a central principle. Proximity analysis is a crucial tool for business marketing and site selection. Marketers analyze demographics and infrastructure to determine trade areas.
2. **Buffer:** A buffer is a reclassification based on distance: classification of within/without a given proximity. Buffering involves measuring distance outward in directions from an object. Buffering can be done on all three types of vector data: point, line, area.

---

## 10.6 ANSWER TO CHECK YOUR PROGRESS

---

1. Explain various types of buffering.
2. Find out how you will perform proximity analysis on raster data.

---

## 10.7 REFERENCES

---

1. Jensen, John R.; Jensen, Ryan R. (2013). *Introductory Geographic Information Systems*. Pearson. 149.
2. Jump up↑ *Buffer (Analysis)*, ArcGIS Desktop 10 online help, Accessed 4 March 2010
3. Jump up↑ Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2011). *Geographic Information Systems & Science*. Danvers, Massachusetts: John Wiley & Sons.
4. Jump up↑ Bolstad, Paul (2008). *GIS Fundamentals: A First Text on Geographic Information Systems, Third Edition*. White Bear Lake, Minnesota: Eider Press.

5. Jump up↑ Bolstad, Paul (2008). *GIS Fundamentals: A First Text on Geographic Information Systems, Third Edition*. White Bear Lake, Minnesota: Eider Press.
6. Jump up↑ Lo, C.P., Young, Albert K.W. (2002). *Concepts and Techniques of Geographic Information Systems*. Upper Saddle River, New Jersey: Prentice-Hall inc.
7. Jump up↑ Lo, C.P., Young, Albert K.W. (2002). *Concepts and Techniques of Geographic Information Systems*. Upper Saddle River, New Jersey: Prentice-Hall inc. pg 207.

---

## ***10.8 TERMINAL QUESTIONS***

---

1. Check and note the differences between raster and vector based proximity operations.

---

## **UNIT 11 - NETWORKING ANALYSIS: OPTIMAL PATH & NEIGHBORHOOD**

---

***11.1 OBJECTIVES***

***11.2 INTRODUCTION***

***11.3 NETWORKING ANALYSIS: OPTIMAL PATH & NEIGHBORHOOD***

***11.4 SUMMARY***

***11.5 GLOSSARY***

***11.6 ANSWER TO CHECK YOUR PROGRESS***

***11.7 REFERENCES***

***11.8 TERMINAL QUESTIONS***

---

## 11.1 OBJECTIVES

---

After studying this unit you will be able to:

- Know about the concept of network and different networking elements
- Methods of network analysis
- Process of optimal path finding and neighborhood analysis
- Application context of network analysis using GIS

---

## 11.2 INTRODUCTION

---

A network is a system of connected set of lines representing some geographic phenomenon. It identifies through forms through which resources are transported or communication is achieved. The “goods” transported can be almost anything: people, cars and other vehicles along a road network, commercial goods along a logistic network, phone calls along a telephone network, or water pollution along a stream/river network.

---

## 11.3 NETWORKING ANALYSIS: OPTIMAL PATH & NEIGHBORHOOD

---

A *network* is a system of connected set of lines representing some geographic phenomenon. Network Analysis provides tools to;

- Find the shortest or minimum impedance (resistance to movement) through a network (PATH)
- Most efficient path in the series of locations (TOUR)
- Assign a portion of network to a location (ALLOCATION)
- Whether the location is connected to the other (TRACING)
- Model the accessibility of location and interaction between location based on some cost related to travel (SPATIAL INTERACTION)

### Network Elements:

Network consists of different elements each of which can be associated with an attribute defining the characteristics of an element (refer fig.11.1). They are:

- a) **Links/Lines** – Links are the basic element of network as it serves as the conduit for the movement. There are two terms which should be understood in this regard: **Resistance**, which describes the amount of impedance it involves for the free flow of resources, example may be cost of transportation, condition of road, time... etc. It is user defined and depends on the direction of flow, hence can be categorized as, “from-to resistance” or “to-from resistance”. **Negative link** impedance signifies that the link cannot

be traversed in that line. **Resource Demand** is the associative attribute on the link; example may be number of households dependent along each water pipeline for potable use in an area.

- b) **Nodes** – These are end points. Links are always connected at nodes.
- c) **Turns** – It is the direction of flow from one link to the other connected through node (point of location). The resource flow can be regulated by the turn, example, as no U-turn at specific traffic intersection to reduce the inflow of traffic in a specific direction.
- d) **Stops** -These represent location where resources can be picked and dropped in a link. A classic example would be a bus stop in a bus route where passenger can be picked and dropped. The demand of resource is an attribute of stop. Positive resource demand would indicate resource picked up where as negative demand means resource drop off.
- e) **Centers** – These are location which have specific attribute as points which has supply of resources to distribute it further in the link of a network. Resource capacity is an important attribute.
- f) **Barriers** - These represent location through which there are no resource flows. These are generally visualized as obstacle in the link.

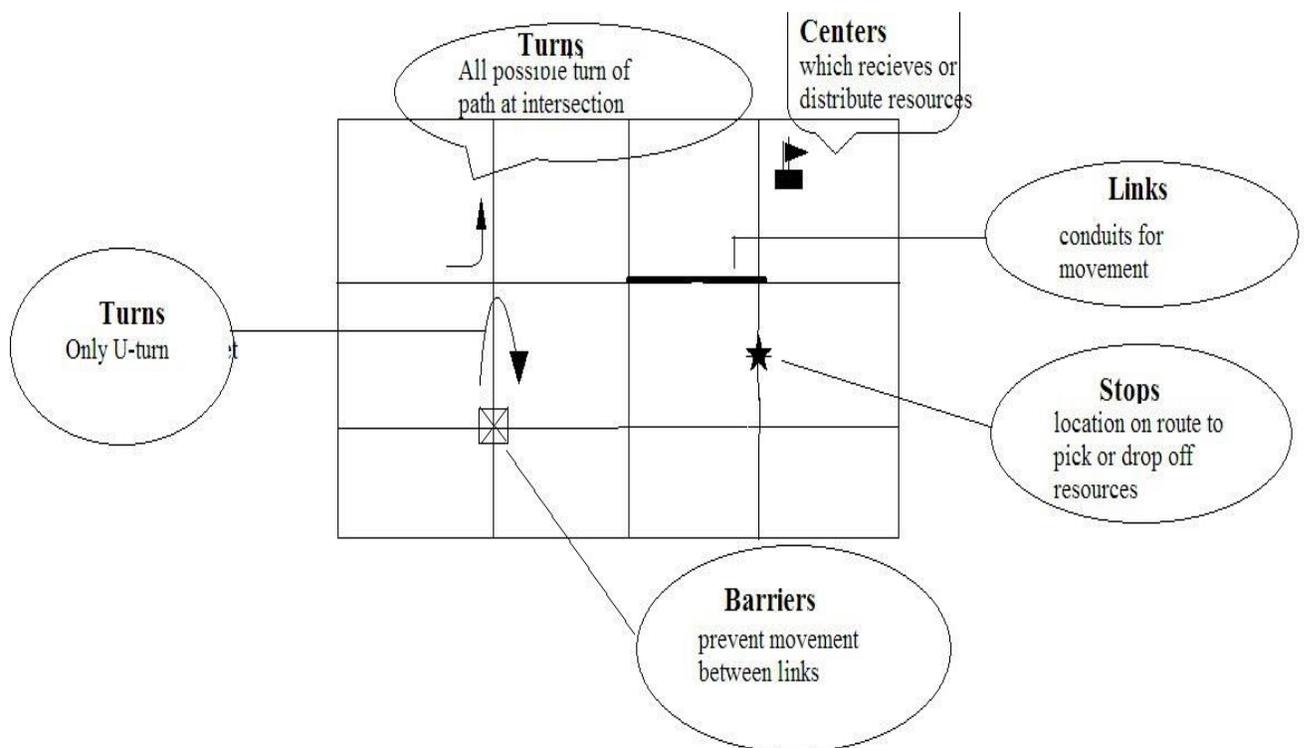


Figure11.1. Elements of Network

## Network Analysis:

The real world is full of set of network which consistently facilitates movement of resources, people and all sources of communications for utilities under concern. Network Analysis is the study of representation, management and manipulation of such network features. Each utility service does have certain requirements and some optimum desired levels of services are required. Connectivity functions represent spatial linkages between features. Analysis of such networks may entail shortest path computations (in terms of distance or travel time) between two points in a network for routing purposes. Other forms are the find all points reachable within a given distance or duration from a start point for allocation purposes, or determination of the capacity of the network for transportation between an indicated source location and sink location.

Network analysis is done to achieve any of the following requirements of services dependent on the utilities under concern:

- a) Path determination –It is the process of calculating the optimal path through series of point in a network to simulate the flow of resources through them. Depending on the application the path determination can be categorized under two heads:
  - Source-Destination Path, which is the optimal path from pre defined source to pre defined destination. The path of least resistance is determined from source to the destination by evaluating the link, turn and resistance for the links.
  - Optimal Cyclic Path, where optimal path is determined from evaluating resistance for each pair of links in the network. It can be worked out for multiple stops in a network by optimizing the order of visit depending upon distance between each stops evaluating impedance in the network.
- b) Resource allocation – It is associated with links as resource centers. In order to meet the demand of the link the principal of least resistance is followed and all possibility of turns and links at those points are being analyzed.
- c) Utility location – It is a search for facility point location in a network unlike allocation of resources in an already located point of variation. It is determine by evaluating set of constraints defined by the facility points and also the flow demand of each link.
- d) Finding the closest facility – For a known event the closet facility to a given location can be estimated. Number of facility provider can be established to give choice to the customer.

## Check Your Progress I

Q1. What are the characteristics of networks?

.....

.....

.....

.....

Q2. What are the different types of analysis that can be performed on a network?

.....

.....

.....

## Optimal Path Finding:

The aim of optimal path finding is optimal selection of nodes that will enable to achieve a high performance in the network. *Optimal-path finding* techniques are used when a least-cost path between two nodes in a network must be found. The two nodes are called origin and destination. The aim is to find a sequence of connected lines to traverse from the origin to the destination at the lowest possible cost.

In Optimal-path finding, the cost function can be simple: for instance, it can be defined as the total length of all lines of the path. The cost function can also be more elaborate and take into account not only length of the lines but also their capacity, maximum transmission (travel) rate and other line characteristics, for instance to obtain a reasonable approximation of travel time. There can even be cases in which the nodes visited add to the cost of the path as well. These may be called turning costs, which are defined in a separate turning-cost table for each node, indicating the cost of turning at the node when entering from one line and continuing on another. This is illustrated in Figure 11.1 of the examples.

Problems related to optimal-path finding may require *ordered optimal path finding* or *unordered optimal-path finding*. Both have as an extra requirement that a number of additional nodes need to be visited along the path. In ordered optimal-path finding, the sequence in which these extra nodes are visited matters; in unordered optimal-path finding it does not.

In the following illustration it will be noticed that it is possible to travel on line b in Figure 11.2, make a U-turn at node N, and return along as to where one came from. The question is whether doing this makes sense in optimal-path finding. After all, to go back to where one came from will only increase the total cost. In fact, there could be situations where it would be optimal to do so to go to a new node depending on utility of services and accessibility.

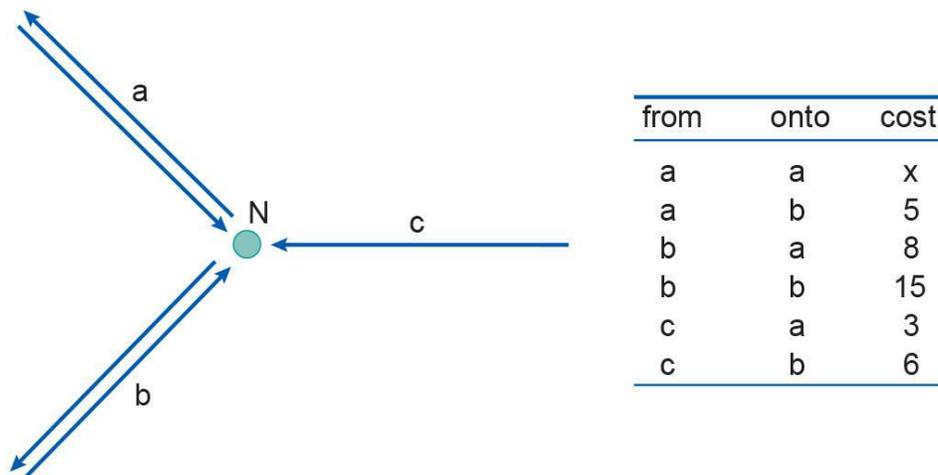


Figure 11.2: Network neighborhood of node N with associated turning costs at N. Turning at N onto c is prohibited because of its direction, so no costs are mentioned for turning onto c. A turning cost of infinity ( $\infty$ ) also means that the turn is prohibited.

An illustration of ordered and unordered path finding types is provided in Figure 11.3. Here, a path is found from node A to node D, via nodes B and C. Obviously, the length of the path found under non-ordered requirements is at most as long as the one found under ordered requirements. Some GISs provide support for these more complicated path-finding problems.

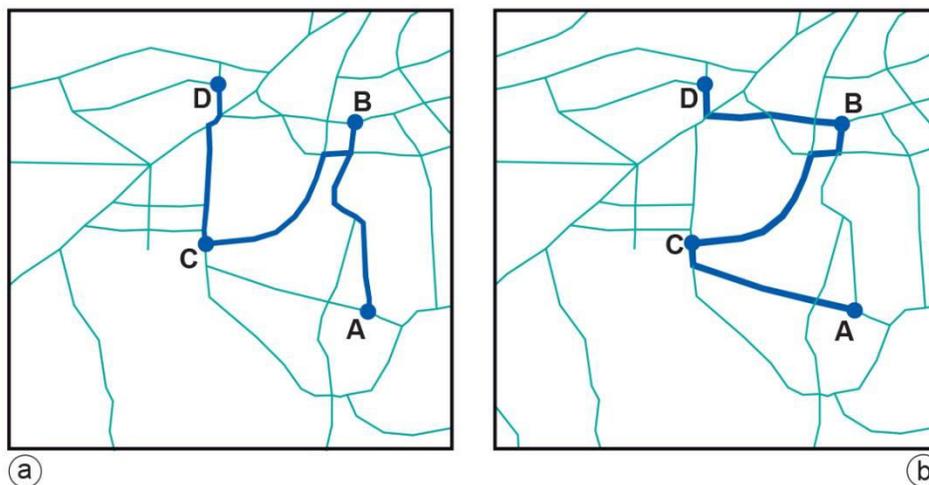


Figure 11.3: Ordered (a) and unordered (b) optimal-path finding. In both cases, a path had to be found from A to D: in (a) by visiting B and then C; in (b) also by visiting both nodes, but in arbitrary order.

But when the network is very big, then it becomes inefficient since a lot of computations need to be repeated. There are many optimization techniques for finding optimal path and these are defined as below:

### a) Ant Colony Optimization(ACO):

Ant colony optimization technique is used to find the shortest path finding algorithm in spite of GPS(global position satellite) or any other method. ACO is a class of optimization algorithms modeled on the actions of an ant colony. Ant Colony Optimization (ACO) studies artificial systems that take inspiration from the behavior of real ant colonies and which are used to solve discrete optimization problems.

Different steps of a simple ant colony system algorithm are as follows:

- **Initialization:** To create a arena by creating image of certain pixels and introducing noise randomly acting as hurdles for the path. This noise is known as salt and pepper noise, it will be introduced in R, G, and B format of image.
- **Random Points:** Taking two random points in image to define the path and making straight line between them using loop. Straightline defines the possible hurdles coming in way.
- **Routing:** Defines the possible routes to reach the destination.
- **Infilling:** Each object will be taken separately in a dummy image and will be dilated and eroded to fill the holes .It will be repeated for every object.
- **Finding Nearest Points:** Nearest points are taken by forming the matrices defining rows and column. Then loop starts for finding the shortest path through the hurdles and finally nearest points are attached to their startup and end point.

The disadvantage related is of problems of stagnation and premature convergence and the convergence speed of ACO being always slow.

### b) Particles Swarm Optimization (PSO):

PSO is an optimization algorithm which has been applied to finding shortest path in the network. However, it might fall into local optimal solution. In this algorithm, the flow starts with a population of particles whose position that represents the solutions for the problem, and velocities are randomly initialized in the search space.

The search for optimal position is performed by updating the particle velocities, hence positions, in each iteration in a specific manner as follows: in every iteration, the fitness of each particle's position is determined by fitness measure and the velocity of each particle is updated by keeping track of two "best" positions.

- **Pbest:** The first one is the best position a particle has traversed so far, this value is called "pbest".
- **Nbest:** Another best value is the best position the any neighbor of a particle has traversed so far, this best value is a group best and is called "nbest".
- **Gbest:** When a particle takes the whole population as its neighborhood, the neighborhood best becomes the global best and it is accordingly called "gbest".

In the PSO algorithm, the potential solutions, called as particles, are obtained by "flowing" through the problem space by following the current optimum particles. Generally speaking, the

PSO algorithm has a strong ability to find the most optimistic result, but it has a disadvantage of easily getting into a local optimum. The PSO algorithm's search is based on the orientation by tracing  $P_b$  that is each particle's best position in its history, and tracing  $P_g$  that is all particles' best position in their history; therefore, it can rapidly arrive around the global optimum. However, because the PSO algorithm has several parameters to be adjusted by empirical approach, if these parameters are not appropriately set, the search will become very slow.

### c) Tabu Search:

It is an iterative search that starts from some initial feasible solution and attempts to determine the best solution in the manner of a hill-climbing algorithm. The algorithm keeps historical local optima for leading to the near global optimum fast and efficiently. During these search procedures the best solution is always updated and stored aside until the stopping criterion is satisfied. The two main components of the tabu search algorithm are the tabu list restrictions and the aspiration criterion. TS use short-term and/or long-term memory while making moves between neighboring solutions. It is essential for a local search to be balanced in terms of quality of solutions and computing time of these solutions. In that sense, a local search does not necessarily evaluate all neighborhood solutions. Generally, a subset of solutions is evaluated. If the optimal score is unknown (which is usually the case), it must be told when to stop looking (for example based on time spend, user input, etc...).

### d) Dijkstra's Algorithm:

Dijkstra's algorithm is a graph search algorithm that solves the single-source optimal path problem for a graph with nonnegative edge path costs, producing an optimal shortest path tree. This algorithm is often used in routing and as subroutine in other graph algorithms. It can also be used for finding costs of shortest paths from a single vertex to a single destination vertex by stopping the algorithm once the optimal path to the destination vertex has been determined. Traffic information systems use Dijkstra's algorithm in order to track the source and destinations from a given particular source and destination. The computation is based on Dijkstra's algorithm which is used to calculate the shortest path tree inside each area of the network.

Dijkstra's labeling method is a central procedure in shortest path algorithms. An out-tree is a tree originating from the source node to other nodes. The output of the labeling method is an out-tree from a source node  $s$ , to a set of nodes  $L$ . Three pieces of information are required for each node  $i$  in the labeling method while constructing the shortest path tree:

- the distance label,  $d(i)$ ,
- the parent-node/predecessor  $p(i)$ ,
- the set of permanently labeled nodes  $L$

Where  $d(i)$  stores an upper bound on the optimal shortest path distance from  $s$  to  $i$ ;  $p(i)$  records the node that immediately precedes node  $i$  in the out-tree. By iteratively adding a temporarily labeled node with the smallest distance label  $d(i)$  to the set of permanently labeled nodes  $L$ , Dijkstra guarantee so ptimality. The algorithm canbe terminated when the destination node is permanently labeled.

The major disadvantage of the algorithm is the fact that it does a blind search there by consuming a lot of time waste of necessary resources. Another disadvantage is that it cannot handle negative edges. This leads to acyclic graphs and most often cannot obtain the right shortest path.

## Check Your Progress II

Q1. What is Optimal Path Finding in Network Analysis?

.....  
 .....  
 .....

Q.2. Discuss various methods to overcome problem of optimal path finding in a network.

.....  
 .....  
 .....

## Neighborhood Analysis:

There are numerous ways to represent the structure in a network but finding the proper account to convey the desired network information is always not an easy task. However, like any large data set, summary statistics (e.g. graph invariants) are one way to help succinctly describe certain aspects of the networks. Another approach is to break up the network into smaller, easier to manage components and study the properties of the sub-networks. The local regions are defined as the neighborhoods around the vertices (i.e. ego networks). Neighborhood analysis can reveal certain aspects of the network that are concealed when only aggregate global network measures are considered. This allows the small patterns, anomalies, and features (as might be relevant to crime and terrorism networks) to be discovered that would be missed in a more global analysis. For example, identifying all the local leadership changes or increased activity regions can help identify terrorist cells.

**Defining Neighborhood:** The breakdown of a network into neighborhoods has several advantages. First, the smaller order neighborhood can disclose important information about the individual vertices' "position" in the network, allowing an analyst to identify the vertices that

suit a certain position. In addition, the neighborhoods also specify a community from which sub-graph level statistics can be ascertained. Community characteristics, like density, can be used to identify tightly coupled regions of the network.

**Neighborhood Matrix:** There are various metrics that can be calculated for networks. Some are specific to vertices (e.g. degree) and others describe an aspect of the entire network (e.g. density). However, all metrics are dependent on the specification of the network and their values can change if the network composition changes. This makes the specification of the network a very important task, and one that is often crucial to the success of network analysis. To minimize the effects of network selection, the neighborhood representation will allow evaluation of smaller parts of the network, allowing smaller scale effects to be captured, in addition to being sufficiently large to also capture the large scale effects.

**Anomaly Detection:** For detecting anomalies in networks, we adopt an approach similar to the network scan statistics and we calculate a discrepancy measure for each neighborhood (over all sizes). The discrepancy is a measure of the noise of an observation.

Based on the neighborhood statistic employed, we can define a discrepancy measure  $D_t(B)$  describing how unusual the sub-graph, given by the vertex and neighborhood size,  $B = (v, k)$  appears at time  $t$ . These measures should be suitably standardized to allow direct comparison between all neighborhood sizes and times. However, unless there are very large abrupt changes, this may not detect the change with sufficient power.

### Check Your Progress III

Q1. What is Neighborhood? How is it identified?

.....  
 .....  
 .....

Q2. Write a detailed note on Neighborhood Analysis.

.....  
 .....  
 .....

### Application Areas of Network Analysis:

- Traffic routing for transportation planners
- For planning demand-capacity ratio of resources
- To define service area for each facility
- Facility locating, sighting and locating different facility.

---

## ***11.4 SUMMARY***

---

In this unit you have learnt the following:

- The decision making management facilities desires optimal and shortest path to allocate resources on the basis of demand and capacity. Networks are framework through which resources flow.
- In order to facilitate the desired utility services the network systems are studied in detail to represent, manipulate and manage the association of linear features.
- Optimal Path determination and neighborhood analysis are methods of network analysis.
- Network analysis applications are oriented towards planning, administrating and operational management of resource facility in a system.

---

## ***11.5 GLOSSARY***

---

1. Optimal Path finding- Optimal path finding is optimal selection of nodes that will enable to achieve a high performance in the network.
2. Network analysis- Network Analysis is the study of representation, management and manipulation of such network features.

---

## ***11.6 ANSWER TO CHECK YOUR PROGRESS***

---

- Q1. What is Optimal Path Finding in Network Analysis?
- Q2. What is Neighborhood? How is it identified?

---

## ***11.7 REFERENCES***

---

1. Chang, Kangtsung Introduction to Geographic Information Systems 5 th edition 2009
2. McgrawHill. • Lillesand, Thomas M., Ralph W. Kiefer, and JonathamW.Chipman, 2004

---

## ***11.8 TERMINAL QUESTIONS***

---

- Q1. Write a detailed note on Neighborhood Analysis.
- Q2. Write application areas of Network Analysis.

---

## **UNIT 12 - MAP MANIPULATION**

---

***12.1 OBJECTIVES***

***12.2 INTRODUCTION***

***12.3 MAP MANIPULATION***

***12.4 SUMMARY***

***12.5 GLOSSARY***

***12.6 ANSWER TO CHECK YOUR PROGRESS***

***12.7 REFERENCES***

***12.8 TERMINAL QUESTIONS***

---

## ***12.1 OBJECTIVES***

---

After going through this unit the learner will able to learn:

1. Understand the meaning of Maps and their types.
2. Learn about Map Manipulation tools.
3. Capable to work on GIS formats.

---

## ***12.2 INTRODUCTION***

---

The word "map" is derived from the Latin word "mappa", which means napkin or paper. A map is a symbolic depiction of the quality of a selected location, usually drawn on a flat surface, or simply, we can say that the map is a model of the world depicted on a flat surface. The map displays information about the world in a simple and intuitive way. They inform about the world by showing the sizes and shapes of countries, distance between places, and the location of features. But in present time GIS maps go far beyond the static maps of years Past.

---

## ***12.3 MAP MANIPULATION***

---

### **What is map?**

The word "map" is derived from the Latin word "mappa", which means napkin or paper. A map is a symbolic depiction of the quality of a selected location, usually drawn on a flat surface, or simply, we can say that the map is a model of the world depicted on a flat surface. The map displays information about the world in a simple and intuitive way. They inform about the world by showing the sizes and shapes of countries, distance between places, and the location of features. But in present time GIS maps go far beyond the static maps of years Past.

GIS is an acronym for geographic information system. It is a particular sort of mapping that permits you to layer data tied to geographic points. Instead of seeing only a couple of key features on a static map, GIS mapping allows you to see adjustable combination of data layers in a dynamic tool. GIS mapping enables to imagine and recognize patterns that are difficult to be seen if the data elements are in table format. It also helps to identify patterns that emerge when you see two are more dataset together.

## Types of Map:

Each and every map shows different kind of information. Function and symbolization both play significant role in map making. By function maps can be general reference or thematic and by symbol, maps can be qualitative or quantitative. There are many types of maps that we are going to discuss below: -

### According to Function:

Maps can be classified according to their functions. For example soil map shows different soil types in a particular region and there are many other examples that we are going to discuss below. Based on functions, maps are classified in the physical and cultural maps.

**Physical Map-** Physical maps are prepared to show natural features such as relief, soil, rocks, vegetation, and Climate etc. These maps are further sub – divided into following types:

- a) **Astronomical Maps:** -These type of maps are prepared to show heavenly bodies, like stars, moon and planets in our solar system. These maps have both large and small scales.
- b) **Relief Maps:** Relief map are made to show actual topography feature of earth surface like mountains, plateau, river system etc.
- c) **Geological Maps:** These types of maps are made to show various geological features such as rocks, minerals and surficial deposits, as well as location of geologic structure such as faults and folds.
- d) **Climatic Maps:** Climatic maps are drawn to show the geographic distribution of Monthly or annually an average values of climatic variables i.e. Temperature, precipitation, humidity, atmospheric pressure over a particular region, in simple words we can say that climatic Maps depict different types of climate zones of an area.
- e) **Weather Maps:** Weather maps are made to show the average condition of elements of weather (Temperature, Pressure, direction and velocity of winds etc.) over a short period i.e. on day-to-day basis.
- f) **Soil Maps:** Soil maps is a geographical representation of soil types or soil properties in the area of interest by using different shades and colours.

**Cultural Map -** Cultural maps are drawn to represent man – made features such as canals, dams, buildings, rails and road network etc.

- a) **Political Maps:** Political map that represent the political sub – division of the world, of continent, or of major Geographic regions. For example, Political map of India shows 29 states and 9 union territories of India.



Source: <https://www.mapsofindia.com/>

Fig. 12.1 Political Map of India

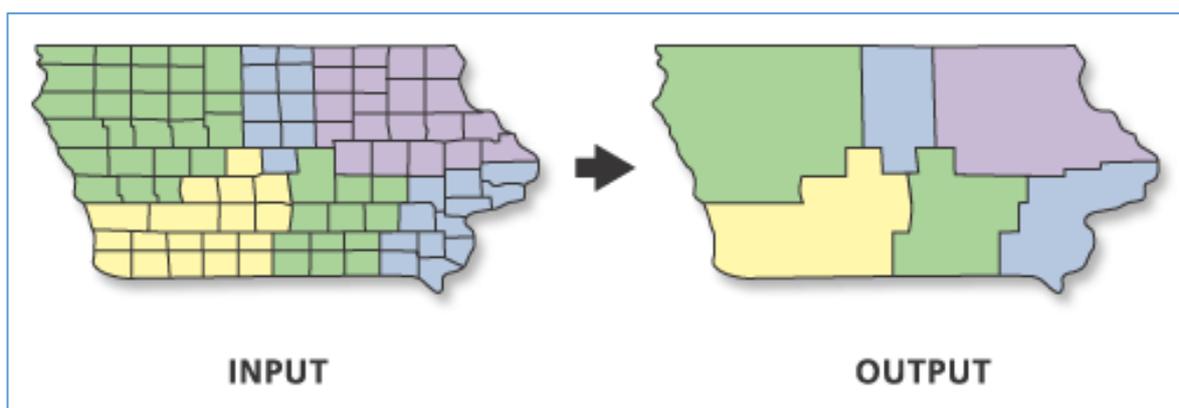
- b) **Population Map:** These maps are drawn to show distribution and density of Population.
- c) **Agricultural Map:** These maps are drawn to represent production and distribution of different types of crop in a particular area.

## Map Manipulation:

As we know, Maps are very important for us because they provide us very useful information in a very lucid way or in simple words we can say that a map simplifies complex information. But, it is not necessary that a single map is useful to every person, because different type of map will contain different type of information, he/she wants to seek from that particular map. Now, when every person gathers information according to their need then concept of map manipulation comes into play.

Map manipulation is done in GIS environment. There are various tools for processing and managing maps in the database in the GIS software package. Like overlay and buffering, these two tools are very basic tool that are frequently used to data pre-processing and data analyses. Map manipulation is easy to follow graphically, even though terms describing the various tools may differ between GIS packages. There is alots of tools that are used in Map Manipulation such as Dissolve, Clip, Append, Select, Eliminate, Update, Erase and split. All these tools are discussed one by one:

**Dissolve** - Dissolve is a tool that aggregates features and thus referred to as 'Merge' or 'Amalgamation'. In this process, a new map feature is created by merging adjacent polygons with common values of specified attributes. In GIS, dissolve is one of the Data Management tool that is used for generalizing features. For instance in choropleth mapping, dissolve can delete boundaries with common values and draw larger areas with the same common values, as shown in Figure 12.2.

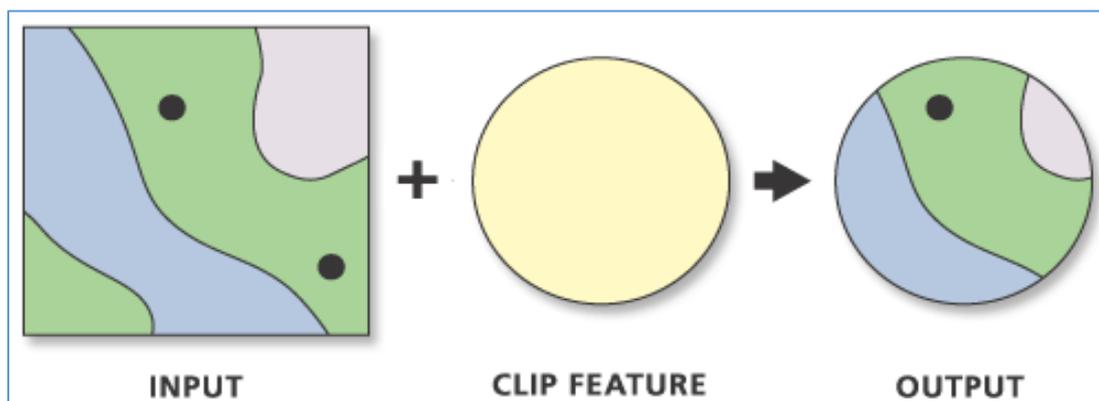


Source: <https://pro.arcgis.com>

**Fig. 12.2**

Thus, using the dissolve operator on multiple Polygon with same value will yield one new polygon, combining the dimension of original, dissolved polygons.

**Clip** - In geographic information system, clipping is to superimpose a polygon on one or more layers, and extract features from the layer located in the area outlined by the clipping polygon. In other words, we can say that the boundary of the second polygon is superimposed on the first polygon. All other areas will be discarded and will no longer be part of the first polygon feature. The clipped data becomes a new feature.

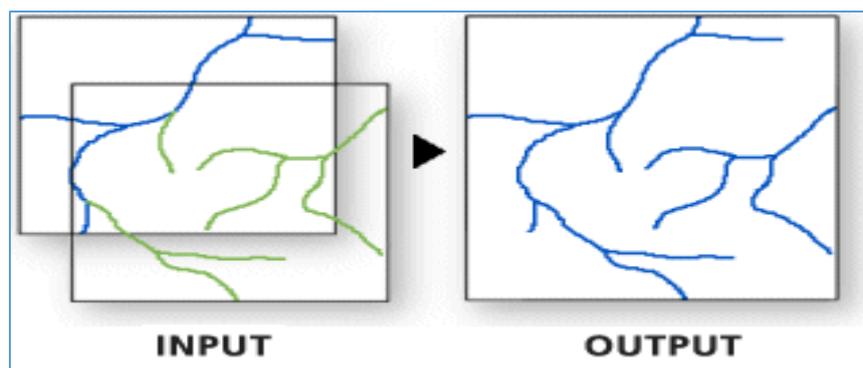


Source: <https://pro.arcgis.com>

**Fig. 12.3**

Clipping to form new layer creates specific area of interest, which is an important function when working in GIS. This becomes very advantageous when the analyst only needs to work in a specific area. He/she can easily discard unnecessary spatial information and does not affect the original data. An example of the use of clip tool is to analyse traffic patterns in a Central business district (CBD). The analyst does not need road data outside of the CBD. So they easily cut the road data of CBD boundaries of the particular area. A clip operation can be processed using both Raster and Vector data. In figure 12.3 we can see that how the clipping tool clips certain area from a larger area. The first frame is an original frame that is base for further processing, followed by the Polygon that is circular in shape clipped from original frame. Finally, the last image shows the new layer, clipped from the original and that is our study area.

**Append** - This creates a new layer by piecing two or more layers together (Figure 12.4). This tool can append point, polygon and linefeature class, raster catalogues, and tables into existing dataset. For example, several raster can be appended to an existing raster dataset, for example, you can attach multiple raster's to an existing raster dataset, but you cannot attach a line feature class to a point feature class.



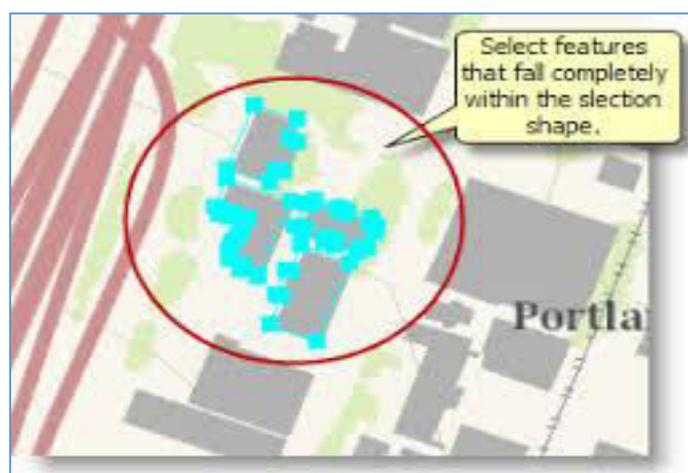
Source: <https://pro.arcgis.com>

Fig. 12.4

### Limitation of Append tool:

- I. This tool does not perform edge-matching. It will not adjust the geometry of the feature.
- II. The map layer can be used as an input dataset. If the layer has selections, the "Append" tool uses only the selected records.
- III. The tool cannot use multiple input layers with similar names.

**Select** - The select by location tool select features based on their location relative to feature in another map layer In simple words, Select can create a new layer that contains attributes selected from user-defined query expressions.(Figure 12.5).



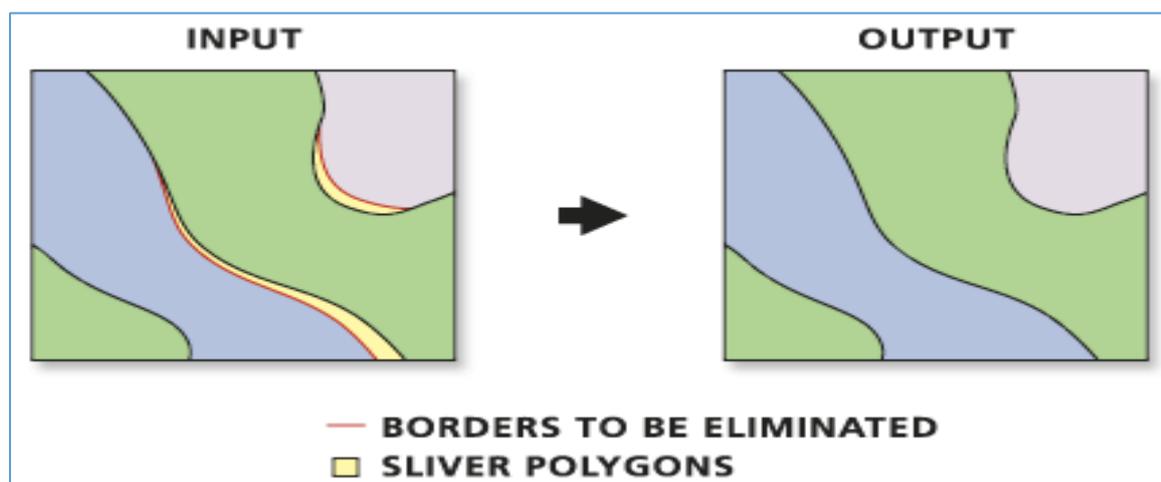
Source: <https://pro.arcgis.com>

Fig. 12.5

For example, if we want to know how many houses were affected by a recent flood, we simply select all houses (layer 1) that fall within the flood boundary (Layer 2). A variety of

selection methods are available to select point, line, and polygon feature in one layer that overlaps the features in same or another layer.

**Eliminate** - This tool creates a new layer by removing a user- defined Polygon by merging with neighbouring polygon with the largest area. Eliminate, as a GIS tool, eliminates the small polygons that are usually used to remove the result of overlapping operations (Figure 12.6). For example, intersection or union.



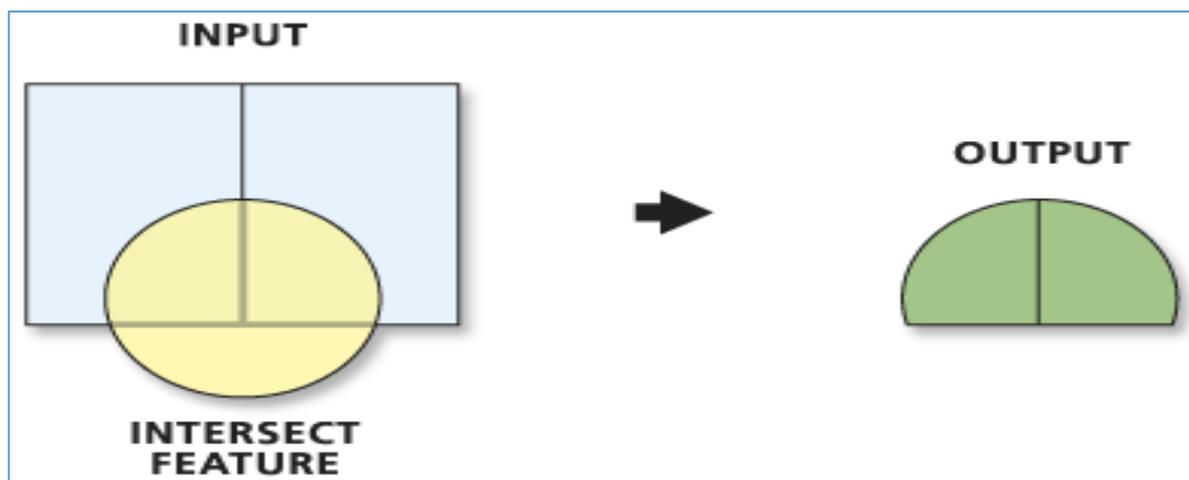
Source: <https://pro.arcgis.com>

**Fig. 12.6**

Eliminate operator can also be referred to as 'omit'. As it can be implemented to remove feature when they become unnecessary at a certain scale. It is evolved from omission operator, instituted by Raisz in 1962.

**Note: The input layer must include a selection, otherwise, Eliminate will fail.**

**Intersect** - The intersection tool calculates the geometric intersection of any number of feature layers and feature classes common to all inputs (that is, they intersect), and writes these features to the output feature class (Figure 12.7). Input feature must be simple feature i.e. Point, multipoint, line or Polygon. They cannot be complex feature such as annotation features or Network feature. There is one more thing if input feature is Point, line or Polygon, The output function will be the same as the input function by default. For example if one or more of the input are Polygon, the default output will be Polygon, if one or more of the input is line, the default output will be line, if one or more of the input is of type point, the default output will be point.



Source: <https://pro.arcgis.com>

Fig.12.7

**Union** - Union is an analytical process in which the features from two or more map layers are combined into a single, composite layer. In simple words we can say that the output feature class will contain union of all the inputs from all the input feature classes. There is one important thing to remember that all Input feature classes must be Polygon.



Source: <https://pro.arcgis.com>

Fig.12.8

Union includes the data from all the included layers, meaning both overlapping and non-overlapping areas are included in new polygon.

**Update** - In this tool, the attributes and geometry of the input features will be updated through the updated features in the feature class, or simply we can say that 'cut and Paste' operation to replace the input layer with updated layer as we see in Fig. 12.9

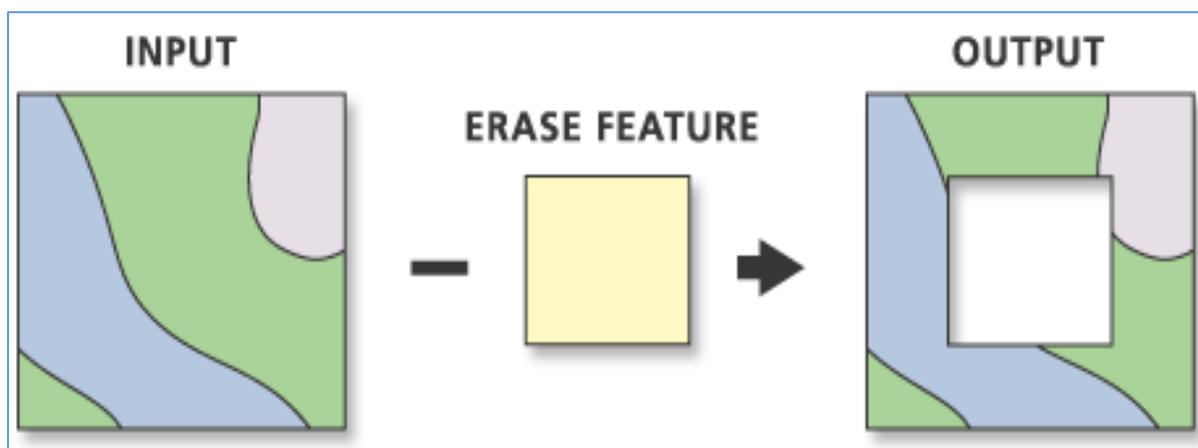


Source: <https://pro.arcgis.com>

**Fig. 12.9**

As the name suggests update is very useful for updating an existing layer rather than redrawing that layer once again. So, we can say that update is a better option than re-digitizing the entire map.

**Erase** - The GIS operation deletes those features that fall within the area of the erased layer from the input layer (Figure 12.10).



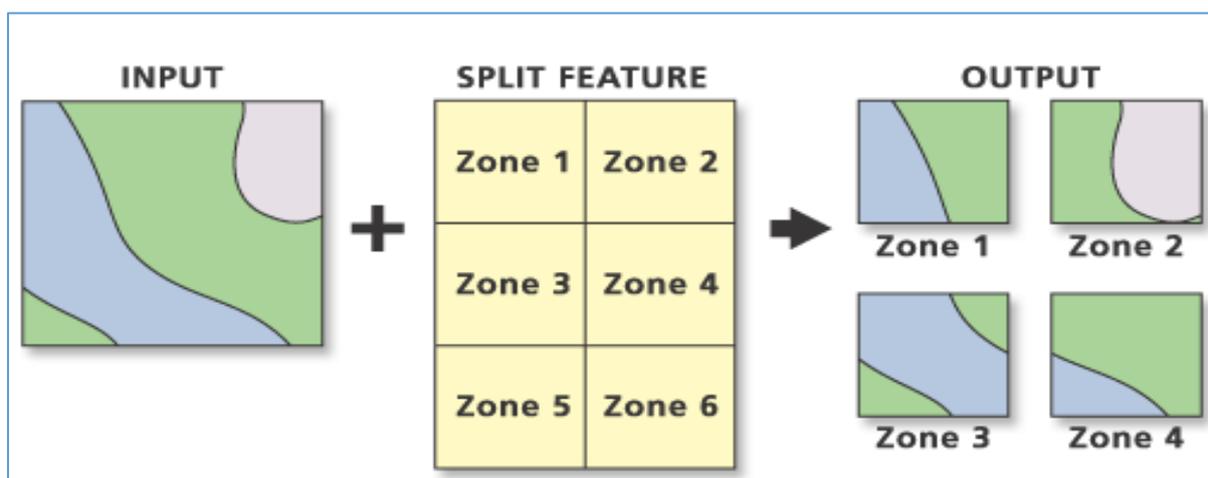
Source: <https://pro.arcgis.com>

**Fig.12.10**

In simple words, it can be said that the feature class is created by superimposing the input feature with the polygon of the erased feature, and only those parts of the input feature that are beyond the boundary of the erased feature are copied to the output feature class. As long as the input features are of the same or smaller order, the "erased features" can be points, lines or polygons. The polygon erase feature can be used to erase polygons, lines or points in

the input feature. The line erase feature can be used to erase lines or points from the input feature; the point erase feature can be used to erase points from the input feature.

**Split** - Split divides the input layer into two or more layers (Fig 12.11). The split layer of the sub-units of the display area is used as a template for dividing the input layer. For example, a national forest can split a stand layer by district so that each district office can have its own layer.



Source: <https://pro.arcgis.com>

Fig. 12.11

In ArcGIS, clip and split are also editing tools. These editing tools work with features rather than layers. For example, the editing tool of Split splits a line at a specified location or a polygon along a line sketch. The tool does not work with layers. It is therefore important that we understand the function of a tool before using it.

## 12.4 SUMMARY

In this unit we have discussed about maps and their types we have also seen that there are maps that are classified based on functions i.e. Physical and Cultural maps in which physical maps are prepared to show natural features such as relief, soil, rocks, vegetation and climate etc. while cultural maps are drawn to represent man-made features such as canals, dams, buildings, rail and road network.

Further we learned about Map manipulation and their tools that are used in GIS Software which are dissolve, clip, append, select, eliminate, update, erase, and split. These all tools are used to manipulate the maps. So, the theoretical understanding of these tools is the

prerequisite condition for carrying out map manipulations in order to extract specific information from a map.

---

## 12.5 GLOSSARY

---

- **Append:** A GIS operation that creates a new layer by merging two or more layers together.
- **Buffering:** A GIS operation used to create an area consisting of areas within a specified distance of selected features.
- **Clip:** A GIS operation that creates a new layer that includes only those features in the input layer that belong to the region of the clip layer.
- **Dissolve:** A GIS operation that deletes the boundary between polygons with the same attribute value.
- **Eliminate:** The GIS operation of creating a new layer by deleting features satisfying a user-defined logical expression from the input layer.
- **Erase:** A GIS operation that deletes those features within the area of the erased layer from the input layer.
- **Identity:** An overlay method that retains only features belonging to the domain defined by the input layer.
- **Intersect:** An overlay method that preserves only features falling within the area extent common to the input layers.
- **Minimum mapping unit:** the smallest unit of area managed by a government agency or organization.
- **Nearest neighbour analysis:** A spatial statistic that determines if a point pattern is random, regular, or clustered.
- **Overlay:** A GIS operation that combines the geometries and attributes of the input layer to create the output.
- **Select:** A GIS operation that uses a logical expression to select features from the input layer for the output layer.
- **Slivers:** Very small polygon found along the shared boundary of the two input layers in overlay.
- **Split:** A GIS operation that divides the input layer into two or more layers.

- **Union:** A polygon-on-polygon overlay method that preserves all feature from the input layers.
- **Update:** A GIS operation that replaces the input layer with the update layer and its features.

---

## ***12.6 ANSWER TO CHECK YOUR PROGRESS***

---

Q.1 Mappa is Greek word? (T/F)

Q.2 What is Physical map?

Q.3 What is cultural map?

Q.4 World Map is a small scale map? (T/F)

Q.5 Map feature is created by merging adjacent Polygon that have a common value for a specified attribute is called?

- a) Dissolve
- b) Clip
- c) Append
- d) Erase

Q.6 What are the basic difference between Climate and Weather map?

---

## ***12.7 REFERENCES***

---

- Chang, Kang-tsung Introduction to Geographic Information Systems 5 th edition 2009 Mcgraw-Hill.
- ArcGIS Desktop Help 9.1,  
<http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?TopicName=welcome>
- <https://www.mapsofindia.com/>
- <https://geology.com/maps/types-of-maps/>
- <https://www.nationalgeographic.com/science/article/151022-data-points-how-make-maps-influence-people>
- Chang, Kang-tsung Introduction to Geographic Information Systems 5 th edition 2009 Mcgraw-Hill.
- Lillesand, Thomas M., Ralph W. Kiefer, and JonathamW.Chipman, 2004

---

## ***12.8 TERMINAL QUESTIONS***

---

- Q.1 What is map Manipulation and discuss any four tools?
- Q.2 Describe a scenario in which Intersect is preferred over union for an overlay operation.
- Q.3 What does a Split operation accomplish?
- Q.4 Define slivers from an overlay operation.
- Q.5 Provide an example of Polygon-on-Polygon overlay operation from your discipline.

---

## **UNIT 13 - VECTOR DATA FORMATS**

---

***13.1 OBJECTIVES***

***13.2 INTRODUCTION***

***13.3 VECTOR DATA FORMATS***

***13.4 SUMMARY***

***13.5 GLOSSARY***

***13.6 ANSWER TO CHECK YOUR PROGRESS***

***13.7 REFERENCES***

***13.8 TERMINAL QUESTIONS***

## 13.1 OBJECTIVES

After reading this unit learner will be able to:

1. Describe vector GIS data models;
2. Discuss advantages and disadvantages of vector data models;
3. Explain topology, topological and non-topological data structures.

## 13.2 INTRODUCTION

Vector model uses discrete points, lines, and areas corresponding to discrete entity, and can be defined by the coordinate geometry. Vectors are graphical objects that have geometrical primitives such as points, lines, and polygons to represent geographical entities in the computer graphics. Vectors have a precise direction, length, and shape.

## 13.3 VECTOR DATA FORMATS

The simple vector data structure that can be used to represent spatial entity is a file containing coordinate pairs (X, Y) that represent the location and is known as spaghetti model. In this concept, each point is stored by its location (X, Y). Each line is stored by the sequence of first and last point. Sequentially connected multiple lines are known as polyline. Polygon is represented by a closed sequence of lines. Unlike line or polyline (sequence of lines), polygons are always closed. This implies that the first point is equal to the last point. A polygon can be represented by a sequence of points where the last point is equal to the first point. Figure below describes how the real-world can be represented by vectors.

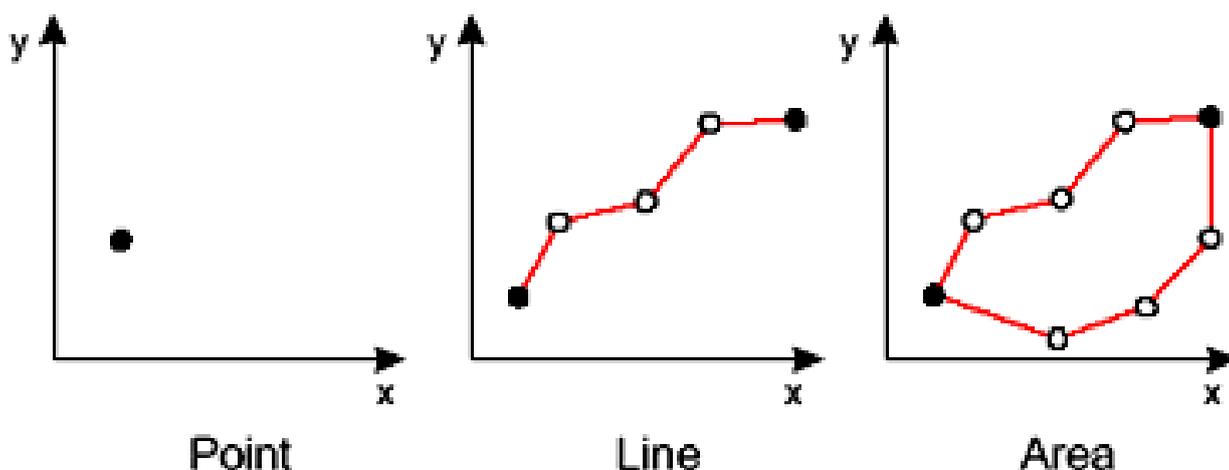


Figure 13.1: Vector Data Structure

Though vectors are ideal for representing discrete objects, it can also be used for field-based or continuous data such as elevation, temperature, etc. Mass-points, contour lines/isolines, and

TINs are used to represent elevation or other continuously changing values. Mass-point is a technique to represent surfaces using several points in a very dense manner. Contour is an imaginary line of constant elevation on the ground surface. The corresponding line on a map is called a contour line, a line on a map that joins places of the same elevation (height) above sea level. Contour interval is the difference in elevation between two contour lines. Isoline is a line on a surface, connecting points of equal value such as temperature, rainfall, etc. TINs record values at point locations, which are connected by lines to form an irregular mesh of triangles. The faces of the triangles represent the terrain surface. However, it should be borne in mind that raster like continuity cannot be obtained by any of the aforementioned models.

### THINGS THAT PRERESENTS POINT, LINE AND POLYGON:

Point is simpler to input and analyze. Points are required to represent entities whose areas are negligible or not important; such as electric poles, postbox, and tube well, etc. Lines, defined by two points, are used to represent features that are linear in nature, for example, roads or pipelines. They can also be used to represent linear features that do not have any physical existence (they are abstract), such as a line showing international border. It is often not possible to represent real-world linear entities by a single line. Multiple sequentially connected segments of line are used to create an object. These multiple lines are collectively called as polyline. Polylines have different structure than lines. Multiple lines are multiple objects, but each of the polylines having multiple linear segments is a single object. Arc is another term used in ARC/INFO synonymously with polyline.

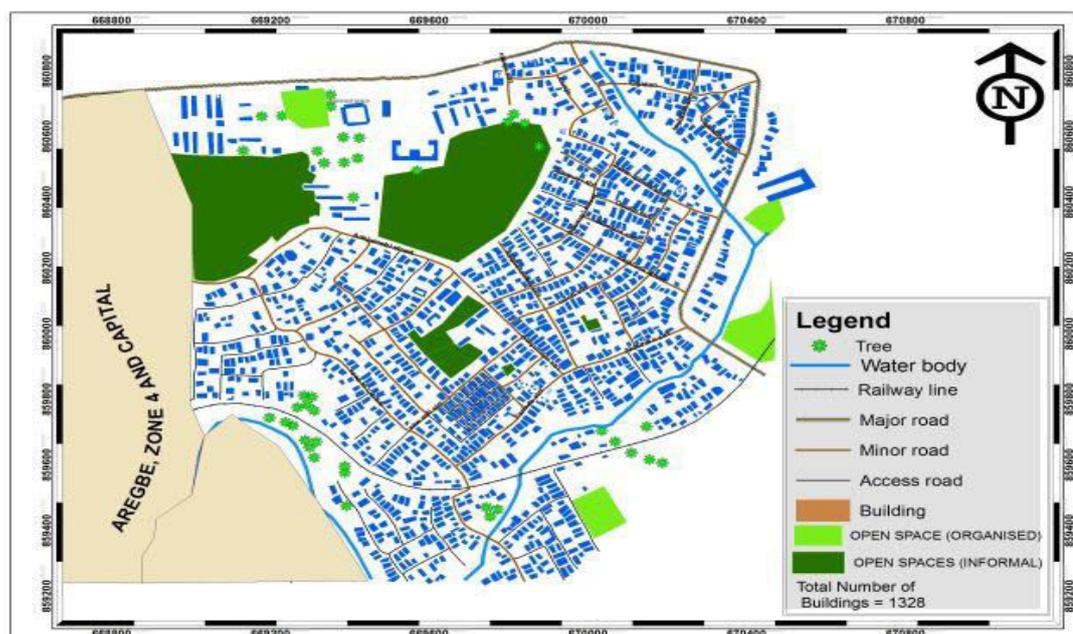


Figure 13.2: Earth Features Mapping through Vector in GIS

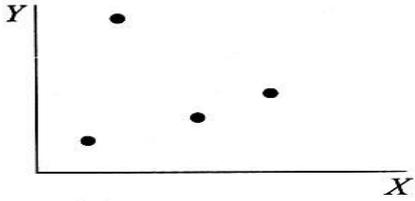
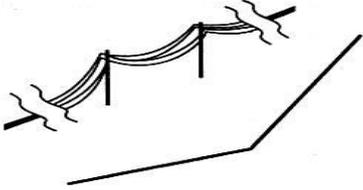
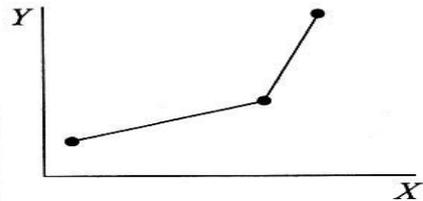
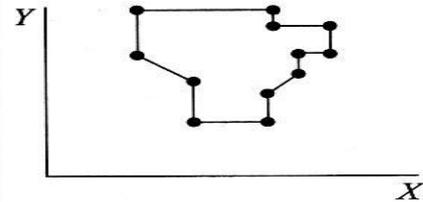
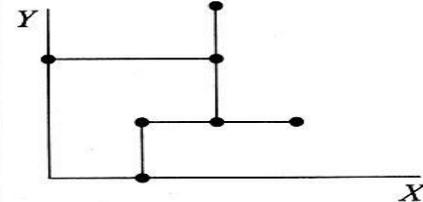
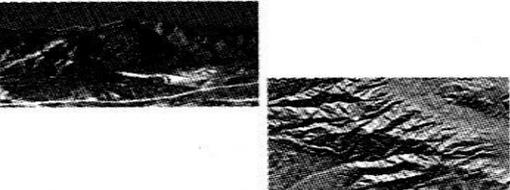
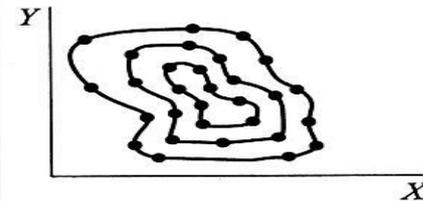
Real world	Vector representation
 <p data-bbox="379 443 724 472">Points: Electric Towers</p>	
 <p data-bbox="339 689 740 719">Lines: Electric power lines</p>	
 <p data-bbox="459 936 628 965">Area: Pond</p>	
 <p data-bbox="427 1182 676 1211">Network: Roads</p>	
 <p data-bbox="411 1429 692 1458">Surface: Elevation</p>	

Figure 13.3: Representation of real-world features by Vector Format

Areas are represented by a closed set of lines and are used to define features such as fields, building, or administrative areas. These closed set of lines are referred to as polygons or regions. As with line features, some of these polygons exist (physically) on the ground, while others are imaginary (abstract). Polygons need only points to input but the area, perimeter, and other geometric attributes may be computed by the GIS software rather than by the manual input. Regions are similar to polygons but it may contain a hole within an area, or one region may contain multiple polygons which are not adjacent. For example, private land lots scattered within a national forest should be subtracted from the forest to get the exact coverage and area

of the forest. Another example is that a district having many small islands requires all those island areas to be converted in a single object.

Vector data representation using points, lines, and areas is not always straightforward because it depends on map scale, functions we wish to perform in our later analysis, and occasionally, on the criteria established by government mapping agencies. (Map scale is the ratio of the map distance to the corresponding distance on the ground.)

It can be difficult for a GIS user to decide when a feature should be represented by a line. Whether a road be represented by a single line along its centre, or are two lines required, one for each side of the road. GIS requires a single line, and not two lines, along with the centre. A stream may be represented using lines near its headwaters but as an area along its lower reaches. In this case, the width of the river and the scale of the map should be considered to take a decision. Government mapping agencies have some standards to make this task easier; for example, a river having width less than 40 ft wide should be represented as lines on 1:24,000 scale maps.

The things that are represented as line (or polyline) may be easy to guess such as road, pipeline, water line, bus route, and so on that have their basic shape similar to line or combination of lines. However, in the city map with a scale of approximately 1:25,000 or 1:10,000, we may represent buildings, parks, bus terminus, and so on as points. If we need more detailed map, however, for instance, in the scale of 1:1000, the aforelisted infrastructure may be better to be represented as polygons, rather than as points. On a district map, cities are required to be represented as an area but on a map of large country such as India, it is not possible to represent cities as areas.

The simplest vector data model stores and organizes the data without establishing relationships among the geographic features are generally called as spaghetti model. In this model, lines in the database overlap but do not intersect, just like spaghetti on a plate. The polygon features are defined by lines which do not have any concept of start and end node or intersection node. However, the polygons are hatched or coloured manually to represent something. There is no data attached to it and, therefore, no data analysis is possible in the spaghetti model.

### **VECTOR DATA STRUCTURES:**

As you know description of geographical phenomena explained in the form of point, line, or polygons is called as vector data structure. Vector data structures are now widely used in GIS and computer cartography. This data structure has an advantage in deriving information from digitisation, and is more exact in representation of complex features such as administration boundaries, land parcels, etc. In early GIS, vector files were simply lines and were having only

starting and ending points. The vector file consists of a few long lines, many short lines, or even a mix of the two. The files are generally written in a binary or ASCII (American Standard Code for Information Interchange) code which refers to a set of codes used to represent alpha numerical characters in computer data processing. Therefore, a computer programmer needs to follow the line from one place to another in the file to enter the data in system. This unstructured vector data are called as cartographic spaghetti. Vector data in the spaghetti data model may not be usable by GIS. However, most of the systems still use this basic data structure because of their standard format (e.g., mapping agency's standard linear format). To express the spatial relationships more accurately between the features, the concept of topology has evolved. Topology can explain the spatial relationships of adjacent, connectivity and containment between spatial features. Topological data are useful for detecting and correcting digitizing errors e.g., two streams do not connect perfectly at an intersection point. Therefore, topology is necessary for carrying out some types of spatial analysis such as network and proximity. There are commonly two data structures used in vector GIS data storage viz. topological and non-topological structures. Let us now discuss about the two types of data structure.

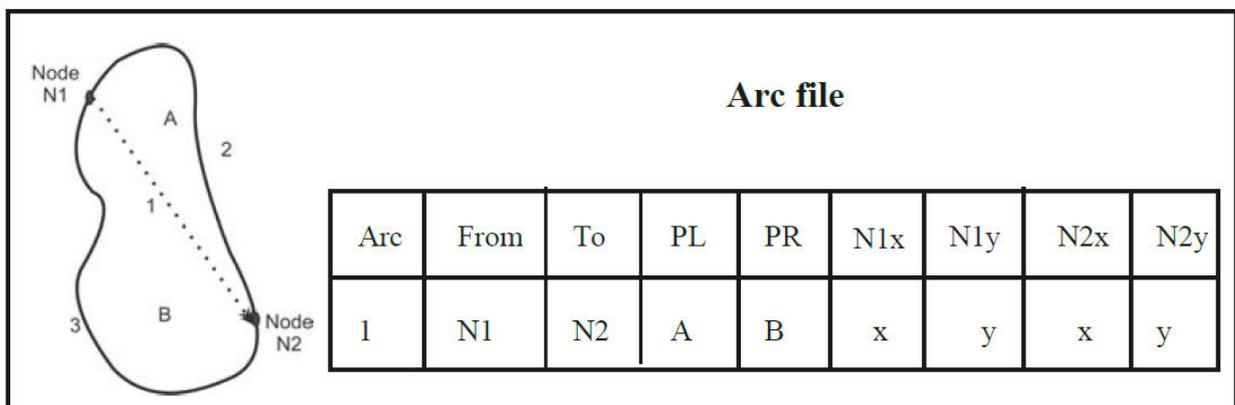


Figure 13.4: Topological Structure of the arc

#### a) Topological Data Structure

Topologic data structure is often referred to as an intelligent data structure because spatial relationships between geographic features are easily derived when using them. Because of this reason topological vector data structure is important in undertaking complex data analysis. In a topological data structure, lines cannot overlap without a node whereas lines can overlap without nodes in a nontopological data structure (e.g., spaghetti). The arc-node topological data structure is now used in most of the systems. In the arc-node data structure, the arc is used for the data storage and it also works when it is needed to reconstruct a polygon. In file of arcs, point data is stored and linked to the

arc file. Node refers to the end points of the line segment. The arc has information not only related to that particular arc but also to its neighbours in geographic space. It includes the arc number of the next connecting arc and the polygon number i.e. A: the left polygon (PL) and B: the right polygon (PR). The arc forms areas or polygons, and the polygon identifier number is the key for constructing a polygon. Some important vector data structures are such as Topologically Integrated Geographic Encoding and Referencing (TIGER) and Coverage Data Structure.

i) **Topologically Integrated Geographic Encoding and Referencing (TIGER):**

It is an early application of topology in preparing geospatial data created by US Bureau of Census as an improvement to the Geographic Base File/Dual Independent Map Encoding (GBF/DIME) data structure. This data structure or format was used in the 2000 census by US Bureau of the Census. In the TIGER database, points are called 0-cells, lines 1-cells, and areas 2-cells. Each 1-cell represents a direct line which starts from one point and ending at another point. The line comprises both sides of the data. Each 2 and 0-cells share of the information of the 1-cells associated with it. The main advantage of this data structure is that the user can easily identify an address on either the right side or the left side of a street or road.

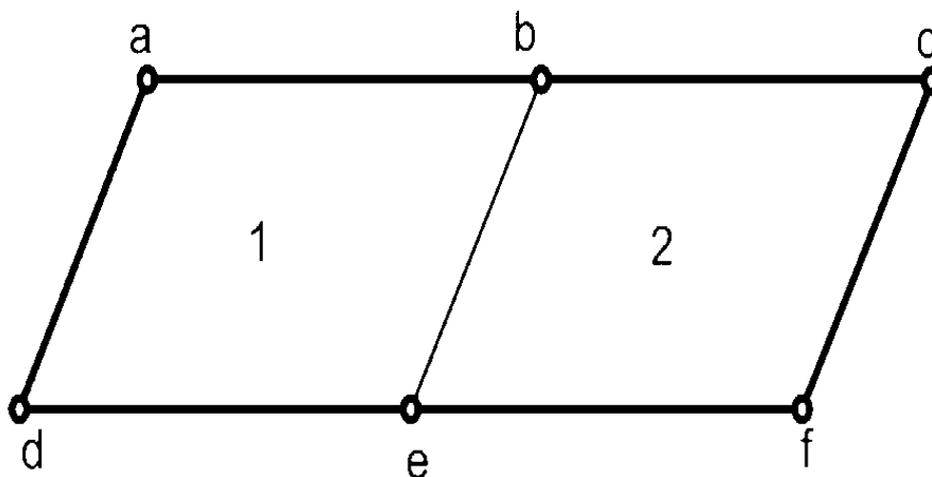


Figure 13.5: TIGER Model

ii) **Coverage Data Structure:** Coverage data structure was practiced by many GIS companies like ESRI, in their software packages in 1980s to separate GIS from CAD (Computer Aided Design). A coverage data structure is a topology based vector data structure that can be a point, line or polygon coverage. A point is a simple spatial entity which can be represented with topology. The point

coverage data structure contains feature identification numbers (ID) and pairs of x, y coordinates, as for example A (2, 4). The starting point of the arc is called from node (F-Node) and where it ends to node (T-Node). The arc-node list represents the x, y coordinates of the nodes and the other points (vertices) that generate each arc. For example, arc C consists of three line segments comprising F-Node at (7, 2), the T-Node at (2, 6) and vertex at (5, 2). Figure below shows the relationship between polygons and arcs (polygon/arc list), arcs and their left and right polygons (left poly/right poly list), and the nodes and vertices (arc-coordinate list). Polygon 'a' is created with arcs A,B,G,H and I. Polygon 'c' surrounded by polygon 'a' is an isolated polygon and consists of only one arc, i.e. 8. 'o' is the universal polygon which covers outside the map area. Arc A is a directed line from node 1 to node 2 and has polygon 'o' as the polygon on the left and polygon 'a' as right polygon. The common boundary between two polygons (o and a) is stored in the arc coordinate list once only, and is not duplicated (Chang, 2010).

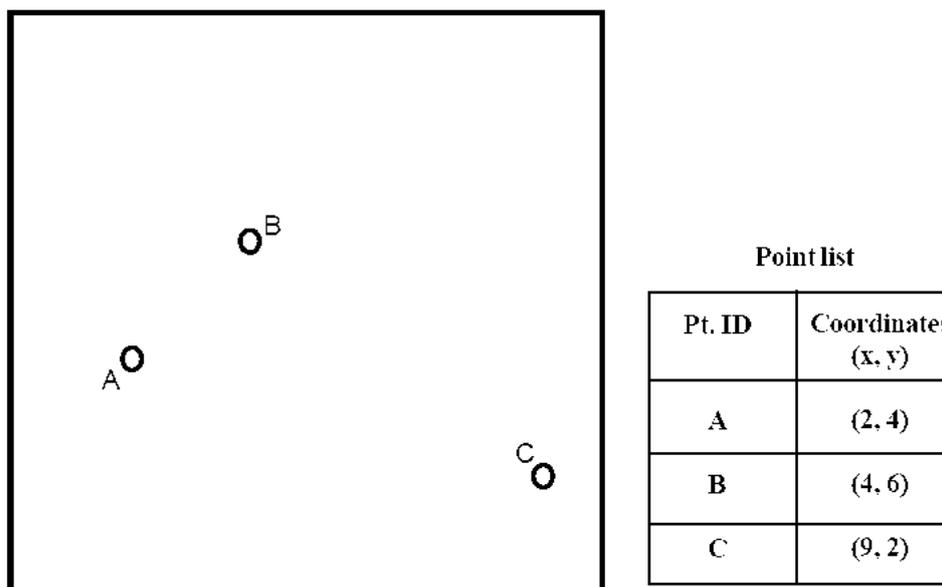
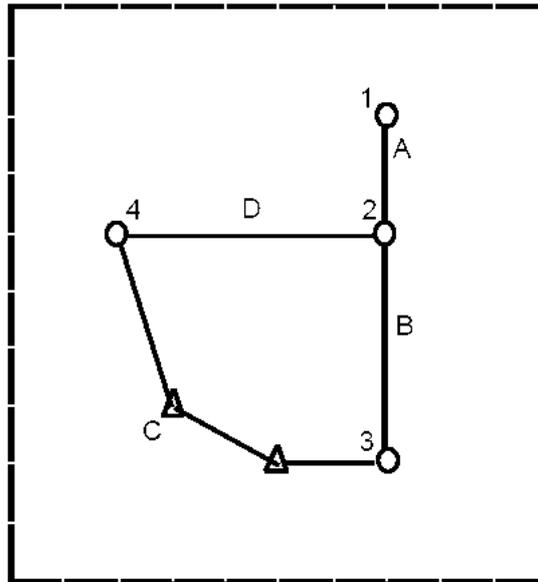


Figure 13.6: Point Coverage Data Structure

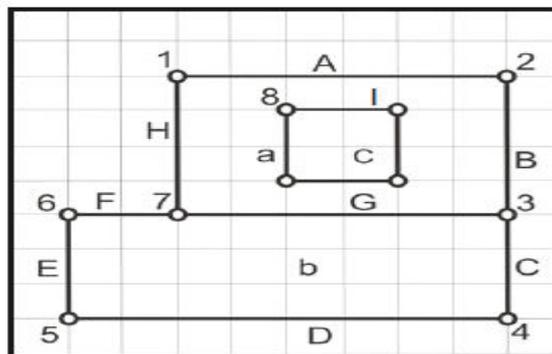


Arc-node-coordinate list

Arc ID	F-node	T-node
A	1	2
B	2	3
C	3	4
D	4	2

Arc ID	Coordinates (x, y)
A	(7,8) (7,6)
B	(7,6)(7,2)
C	(7,2)(5,2)(3,3)(2,6)
D	(2,6) (7,6)

Figure 13.7: Line Coverage Data Structure



Arc coordinate list

Arc ID	Coordinates (x,y)
A	(3,8)(9,7)
B	(9,7)(9,4)
C	(9,4)(9,1)
D	(9,1)(1,1)
E	(1,1)(1,4)
F	(1,4)(3,4)
G	(3,4)(9,4)
H	(3,4)(9,8)
I	(5,7)(7,7)(7,5)(5,5)

Arc/polygon list

Arc ID	Poly ID
A,B,G,H,I	a
C,D,E,F,G,H	b
I	c

Left poly/right poly list

Arc ID	Left poly	Right poly
A	o	a
B	o	b
C	o	b
D	o	b
E	o	b
F	o	b
G	o	b
H	o	b
I	a	c

Figure 13.8: Polygon Coverage Data Structure

**b) Non-Topological Data Structure :**

Vector data structure that is common among GIS software is the Computer Aided Design (CAD) data structure. Drawing Exchange Format (DXF) is used in the CAD package (e.g., AutoCAD) for transferring of the data files. DXF does not support topology and arrange the data as individual layers. This structure consists of listing elements, not features, defined by strings of vertices, to define geographic features, e.g., points, lines, or areas. There is considerable redundancy with this data model since the boundary segment between two polygons will be stored twice, once for each feature. This format allows user to draw each layer by using different line symbols, colours and texts. In this structure, polygons are independent and difficult to answer about the adjacency of features. The CAD vector model lacks the definition of spatial relationships between features that is defined by the topological data model.

Since 1990s almost all commercial GIS packages such as ArcGIS, MapInfo, Geomedia have adopted non-topological data structure. Shape file (.shp) is a standard non-topological data format used in GIS packages. In ArcInfo coverage, the geometry of shape file is stored into two extension types such as .shp and .shx. Shape file (.shp) stores the feature geometry and .shx file maintains the spatial index of the feature geometry. The advantage of nontopological data structure, i.e. shape file, lies in quick display on the system than the topological data. Many software packages such as ArcGIS, MapInfo uses the .shp file format.

**Table 13.1: Some vector file formats**

File format	Full form	Software
EOO	Arc Export	ARC/INFO
Coverage	ARC/INFO Coverage	ARC/INFO
MDB	Personal Geodatabase	ArcGIS
DWG	AutoCAD Drawing File	AutoCAD/Autodesk Map
DXF	Data Interchange (Exchange) File	Many
DLG	Digital Line graphs	Many
HPGL	Hewlett-Packard Graphic Language	Many
MIF/MID	MapInfo Data Transfer Files	MapInfo
TAB	MapInfo Table	MapInfo
DGN	MicroStation Design Files	MicroStation
SDTS	Spatial Data Transfer System	Many
TIGER	Topologically Integrated Geographic Encoding and	

	Referencing Many	
VPF	Vector Product Format	Military mapping systems
SHP	Arc View Shape	Arc View

### **RASTER VERSUS VECTOR:**

There are advantages and disadvantages of using a raster or vector data model to represent reality. Raster data represents a graphic object as a pattern of dots (or pixels), whereas vector data represents the object as points or a set of lines drawn between specific points. Let us consider a line drawn diagonally on a piece of paper. A raster file represents this image by subdividing the paper into a matrix of small squares—similar to a sheet of graph paper—called cells. Each cell is assigned a position in the data file and given a value based on the colour at that position. This data representation allows the user to easily reconstruct or visualize the original scene.

A vector representation of the same diagonal line records the position of the line by just recording the coordinates of its starting and ending points. Each point is expressed as two or three numbers (depending on whether the representation is 2D or 3D, often referred to as X,Y or X,Y,Z coordinates). The vector line is formed and displayed by joining the measured points. Each entity in a vector file appears as an individual data object. It is easy to record information about an object or to compute characteristics such as its exact length or surface area. It is much harder to derive this kind of information from a raster file because they contain little (and sometimes no) discrete geometric information.

Some applications can be handled much more easily with raster techniques than with vector techniques. Raster works best for surface modelling and for applications where distinct features are not important. Terrain elevations can be recorded in a raster format and used to construct digital elevation models (DEMs). Some land-cover/use information may come in raster format (as classified thematic images).

Raster files are often larger (in data volume) than vector files. The raster representation of the line in the earlier example required a data value for each cell on the page, whereas the vector representation only required the positions of two points.

The size of the cells in a raster file is an important matter. Smaller cells improve image quality because they increase details. As the cell size increases, image definition decreases or blurs. However, there is a trade-off between the cell size and the file size, dividing the cell size in half increases file size by a factor of four (approximately).

Cell size in a raster file is referred to as a resolution. For a given resolution value, the raster cost does not increase with image complexity. This implies that any scanner can quickly make

a raster file. It takes no more effort to scan a map of a dense urban area than to scan a sparse rural one. On the other hand, a vector file requires careful measuring and recording of each point, so an urban map is much more time-consuming to draw than a rural map. Unlike raster, the process of making vector maps is not fully automated, and thus the cost increases with map complexity.

Raster data can be compressed more easily than vector data because it is often more repetitive and predictive. Many raster formats, such as TIFF have compression options that drastically reduce image sizes, depending upon image complexity and variability. Raster data are most often used for digital representations of aerial photographs, satellite images, scanned paper maps, and other applications with very detailed images. Raster data are used when costs have to be reduced or when the map does not require analysis of individual map features or when 'backdrop' maps are required.

In contrast, vector data are appropriate for highly precise applications, when file sizes are important, when individual map features require analysis, and when descriptive information (attribute) must be stored.

Additional non-spatial (attribute) data can also be stored besides the spatial data represented by the coordinates of the vector geometry or the position of a raster cell. In the vector data, the additional data are attributes of the object. For example, a forest polygon may also have an identifier value and information about tree species. In raster data, the cell value can store attribute information, which can also be used as an identifier that can relate to records in another table, but it maintains a complex structure and has several limitations.

Raster and vector maps can also be combined visually. For example, a vector street map could be overlaid on a raster aerial photograph. The vector map provides discrete information about individual street segments; the raster image provides a backdrop of the surrounding environment. Table 13.2 summarizes the advantages and disadvantages of raster and vector.

Table 13.2: Advantages and disadvantages of Raster and Vector

<b>Raster model</b>	<b>Vector model</b>
<p><b>Advantages</b></p> <ul style="list-style-type: none"> <li>• Simple data structure</li> <li>• Easy and efficient overlaying</li> <li>• Compatible with remote sensing imagery</li> <li>• High spatial variability is efficiently represented</li> <li>• Efficient to represent continuous data</li> </ul>	<p><b>Advantages</b></p> <ul style="list-style-type: none"> <li>• Smaller file size</li> <li>• Individual identity for discrete objects like line, polygon, etc.</li> <li>• Efficient for topological relationship</li> <li>• Efficient projection transformation</li> <li>• Accurate map output</li> </ul>

	<ul style="list-style-type: none"> <li>• Easy to edit</li> </ul>
<p><b>Disadvantages</b></p> <ul style="list-style-type: none"> <li>• Larger file size</li> <li>• All the objects are series of pixels, no identity for discrete objects other than points/pixels</li> <li>• Difficult to build topological relationship</li> <li>• Inefficient projection transformations</li> <li>• Loss of information when using large cells</li> <li>• Difficult to edit</li> </ul>	<p><b>Disadvantages</b></p> <ul style="list-style-type: none"> <li>• Complex data structure</li> <li>• Difficult overlay operations</li> <li>• High spatial variability is inefficiently represented</li> <li>• Not compatible with remote sensing imagery</li> <li>• Not appropriate to represent continuous data</li> </ul>

---

### 13.4 SUMMARY

---

You have learnt the following in this unit:

- Real world features such as temples, parks, roads, railways, crop land, and forest land are represented as point, line/polyline and polygon. Spatial information of features or objects can be stored in a GIS using vector or raster models. Spatial database of real world features need to be translated into simplified representations which can be stored and updated in a system.
- Two data models, namely, vector data model which is used to symbolize discrete features, and the raster data model, which is most often used to represent continuously varying phenomena currently dominate the commercial GIS software.
- Main advantage of vector model is easy access and complex analysis, while raster model is useful for overlaying and spatial analysis.
- The raster data structure represents the information in the form of grid cells or pixels which stands for picture element. Important raster data structures viz. cell-by-cell encoding, run length encoding, and quadtree give an idea to store the raster data information.
- The data structures are mainly topological, i.e. TIGER, coverage and non-topological data structures under vector models.
- Database management system organizes the spatial data in a systematic pattern.

---

### ***13.5 GLOSSARY***

1. **Data:** Data are units of information, often numeric, that are collected through observation. In a more technical sense, data are a set of values of qualitative or quantitative variables about one or more persons or objects, while a datum is a single value of a single variable.
2. **Vector:** Vector is a data structure, used to store spatial data. Vector data is comprised of lines or arcs, defined by beginning and end points, which meet at nodes. A vector based GIS is defined by the vectorial representation of its geographic data.
3. **Point:** A point feature is a GIS object that stores its geographic representation an X and Y coordinate pair as one of its properties (or fields) in the row in the database. Some point features, such as airplane locations need to also include a z-value, or height, to correctly locate itself in 3D space.
4. **Line:** A line is one of three features with which most vector data is represented in GIS maps. The others are point and polygon. Lines are used to represent the shape and location of geographic objects, such as street centerlines and streams, too narrow to depict as areas. A line is formed by connecting two data points.
5. **Polygon:** A polygon feature is a GIS object that stores its geographic representation, a series of x and y coordinates pairs that enclose an area—as one of its properties (or fields) in the row in the database.

---

### ***13.6 ANSWER TO CHECK YOUR PROGRESS***

1. Explain the advantage and disadvantages of vector GIS models.

---

### ***13.7 REEFERENCES***

1. Burrough, P. A. and McDonnell, R. A., (1998), *Principles of Geographical Information Systems*, Oxford University Press, New York.
2. Chang, K.-t., (2010), *Introduction to Geographic Information Systems*, Tata McGraw-Hill, New Delhi.
3. Lo, C. P. and Yeung, K. W., (2009), *Concepts and Techniques of Geographic Information Systems*, PHI Learning Pvt. Ltd, New Delhi.
4. Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rhind, D. W., (2005), *Geographic Information Systems and Science*, John Wiley and Sons, West Sussex.
5. Rolf, A. D. B., (ed.). (2001), *Principles of Geographical Information Systems ?An Introductory Text Book*, ITC, The Netherlands.

6. Anjireddy, M., (2002), *Textbook of Remote Sensing and Geographical Information Systems*, B. S. Publications, Hyderabad.

---

### ***13.8 TERMINAL QUESTIONS***

---

2. What are geospatial models? Explain vector GIS model in detail.
3. What is vector GIS model? Give an example of objects representing point, line or polygon.