GIS -06 : SPATIAL ANALYSIS (1)-_ *Overlay Operations &*

Analysis in GIS

SUBJECT	GEOGRAPHY
PAPER	REMOTE SENSING, GIS and GPS
MODULE	SPATIAL ANALYSIS (2) :Overlay Operations & Analysis in GIS
Module ID	RS/GIS- 24

Development Team

Principal Investigator	Prof. Masood Ahsan Siddiqui Jamia Millia Islamia, New Delhi
Paper Coordinator	Dr M.P. Punia BITS , Jaipur
Content Writer	Dr Seema Mehra Parihar Kirori Mal College, University of Delhi, Delhi
Content Reviewer	Prof. Masood Ahsan Siddiqui Jamia Millia Islamia, New Delhi

GIS-06: SPATIAL ANALYSIS (1)- Overlay Operations & Analysis in GIS

Learning objective

In this chapter you will learn the following:

- Understand the concept of Spatial Anaysis in Geospatial Environment;
- Conceptualise Overlay analysis in GIS;
- Comprehend the importance, uses and essential characteristics of overlay analysis; and
- Be able to visualise the role of vector and raster data model in these analysis.

1. INTRODUCTION

Spatial analysis is the vital part of GIS. We have learned in the previous chapters that these can be done in two ways. One is the vector-based and the other is raster based analysis. Since the advent of GIS in the 1960s, many government agencies have invested heavily in GIS installations, including the purchase of hardware and software and the construction of mammoth databases. As aptly stated by Raju, P.L ,2015, "two fundamental functions of GIS have been widely realized: generation of maps and generation of tabular reports". Indeed, GIS provides a very effective tool for generating maps and statistical reports from a database. However, GIS functionality far exceeds the purposes of mapping and report compilation. In addition to the basic functions related to automated cartography and data base management systems, the most important uses of GIS are spatial analysis capabilities. As spatial information is organized in a GIS, it should be able to answer complex questions regarding space. Indeed, functions required for performing spatial analyses that are not available in either cartographic packages or data base management systems are commonly implemented in GIS.

2. USING GIS FOR SPATIAL ANALYSIS

Spatial analysis in GIS involves three types of operations:

i. Attribute Queryalso known as non-spatial (or spatial) query,
ii. Spatial Query and
iii. Generation of new data sets from the original database (Bwozough, 1987).

3. GIS Usage in Spatial Analysis

GIS can interrogate geographic features and retrieve associated attribute information, called identification. It can generate new set of maps by query and analysis. It also evolves new information by spatial operations. Here are described some analytical procedures applied with a GIS. GIS operational procedure and analytical tasks that are particularly useful for spatial analysis include:

- ☐ Single layer operations
- ☐ Multi layer operations/ Topological overlay
- ☐ Spatial modeling
- ☐ Geometric modeling
- ☐ Calculating the distance between geographic features
- ☐ Calculating area, length and perimeter
- ☐ Geometric buffers.
- ☐ Point pattern analysis
- ☐ Network analysis
- ☐ Surface analysis
- ☐ Raster/Grid analysis
- ☐ Fuzzy Spatial Analysis
- ☐ Geostatistical Tools for Spatial Analysis

4. Overlay Analysis

Overlay analysis is operation in GIS for superimposing the multiple layer of datasets that representing different themes together for analysing or identifying relationship of each layer. Overlay analysis represent the composite map by the combination of different attribute and geometry of datasets or entity. Overlay is the operations of comparing variables among

multiple coverages. In the overlay analysis new spatial data sets are created by merging data from two or more input data layers. Overlay analysis is one of the most common and powerful GIS technique. It analyses the multiple layer with common coordinate systems and determine what is on the top layer. Overlay operations combine the data from same entity or different entities and create the new geometries and new unit of change entity (figure 1).
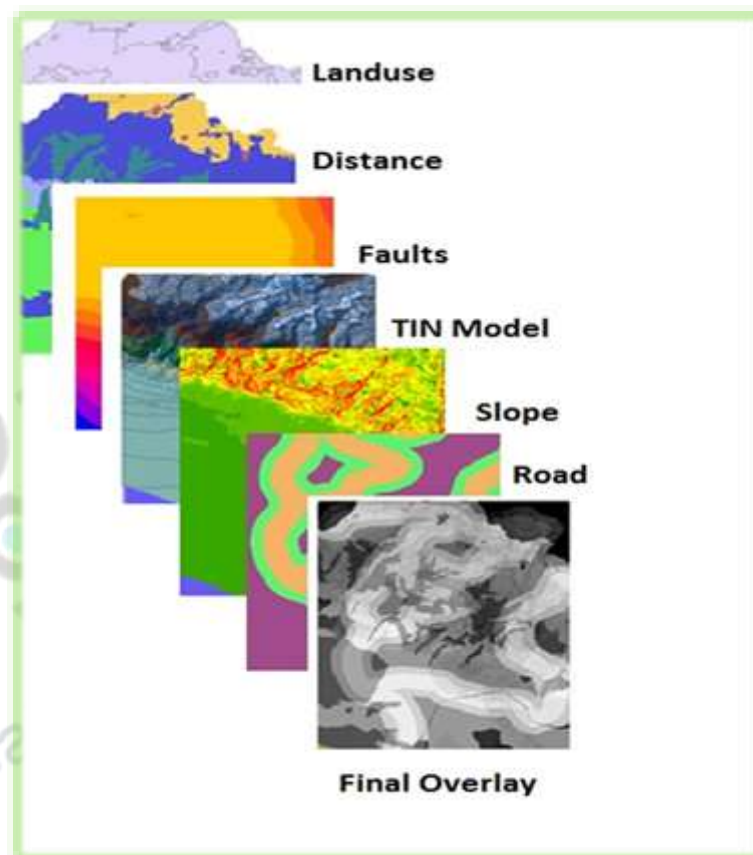

Figure 1 : Overlay Analysis

Overlay operations performs many type of analysis for example cropping pattern in the field, dominance of particular ethnic population in a region, age and sex composition of region, physical landforms of the surface. Suppose I have to find the potential park sites of population so I will create the map of density of population and the map of distance to parks of

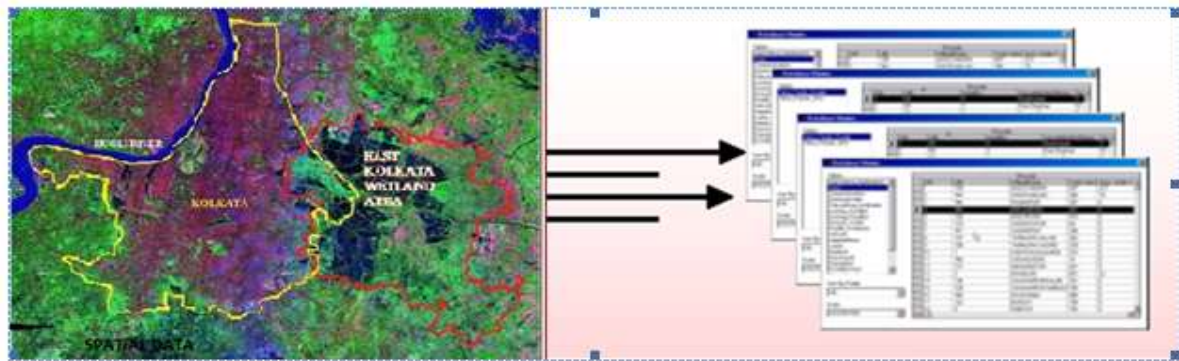populated area then I will superimpose the both layer and find the suitable location.

Figure 2 : **The data structure in overlay analysis comprises the connection of thematic and geometric information**

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 (figure 2). 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.
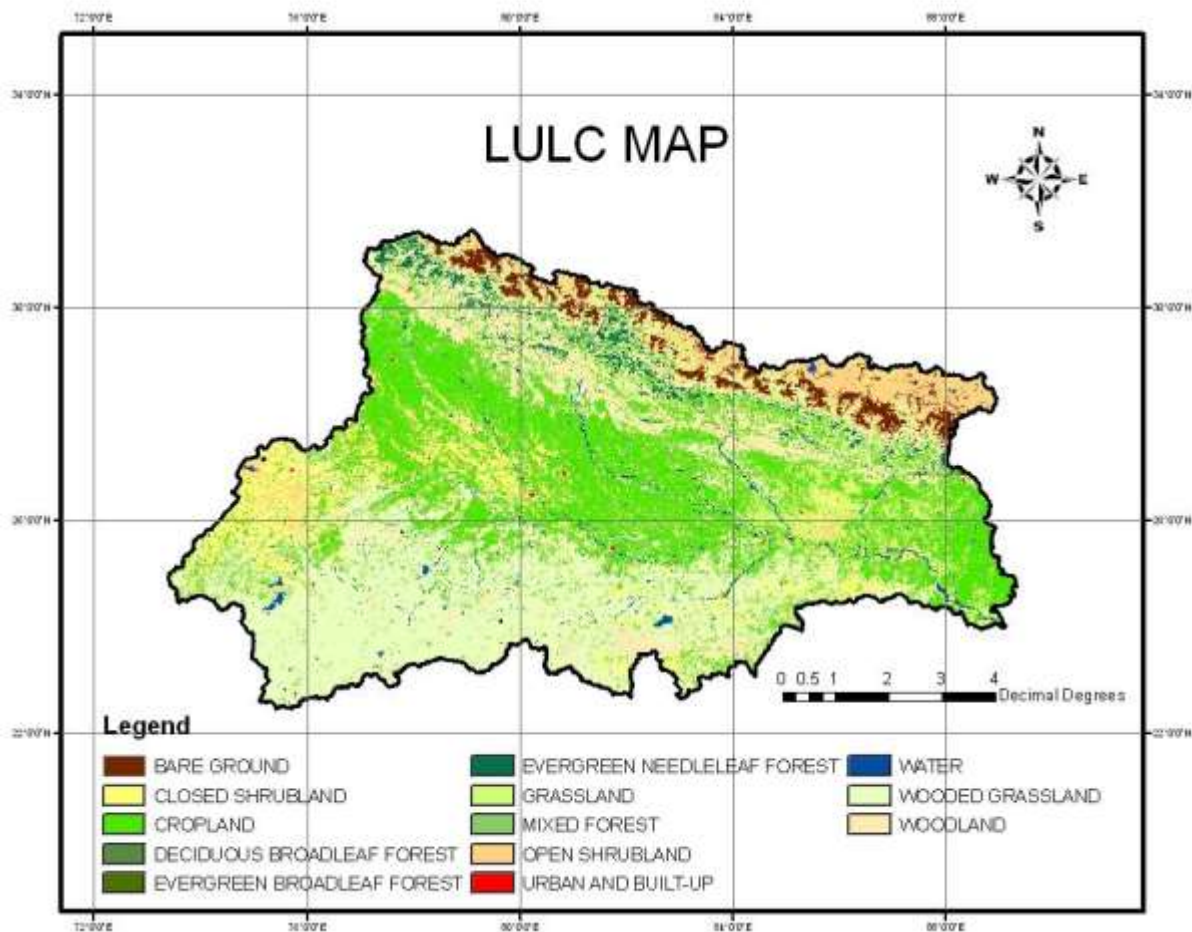
Figure 3: Overlays of multiple Layers resulting in LULC MAP

There are four overlay operators in common use:

1. Point-in-area (also known as point in polygon)
2. Line-in-area (also known as line in polygon)
3. Area-on-area (also known as polygon on polygon)
4. Weighted overlay

Vector based overlay

Overlay of vector data combine point, line, and polygon features. In this data model operations rely on geometry and topology of surface. Vector based overlay is time consuming, complex and computationally expensive. For example taking the ordering network layer of Ganga Watershed and laying

over it with the layer of village. The result would be which orders of stream of Ganga flow in which village.

Point-in-polygon

Point in Polygon Overlay (figure 4) operation will also generate combinative properties of point attributes of one layer and the polygon attribute of the analysis layer. It is a spatial operation in which one point coverage is overlaid with polygon coverage to determine which points falls within the polygon boundaries. After overlay operation, points assume the attributes of the polygons within which they fall. It helps in formulating hypothesis about the spatial relationships between the occurrence of points and the attributes of the polygons. This kind of overlay operation can also be used to calculate number of points located in each of the polygon.
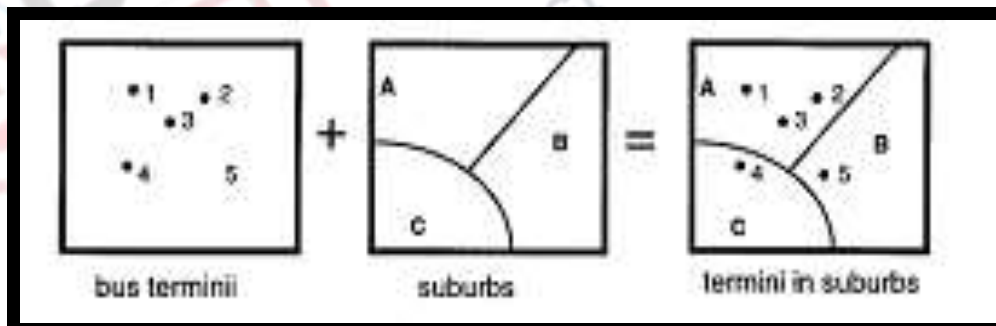

Figure 4:Point in polygon

Polygon on polygon overlay operations

In the polygon on polygon overlay operations (figure 5) I need to check before starting the input layer it should topologically correct. If it is correct output map will also correct. In the polygon overlay it is essential to add new

intersections and create polygon for new topology. The overlay of 2 layers of polygon will produce large number of polygons and increase the number of intersections zone and arcs. If the new polygon, arcs and set of nodes have been shaped then meaningful set of layer can be extract. It is necessary to keep in mind that area should be common to both input features.
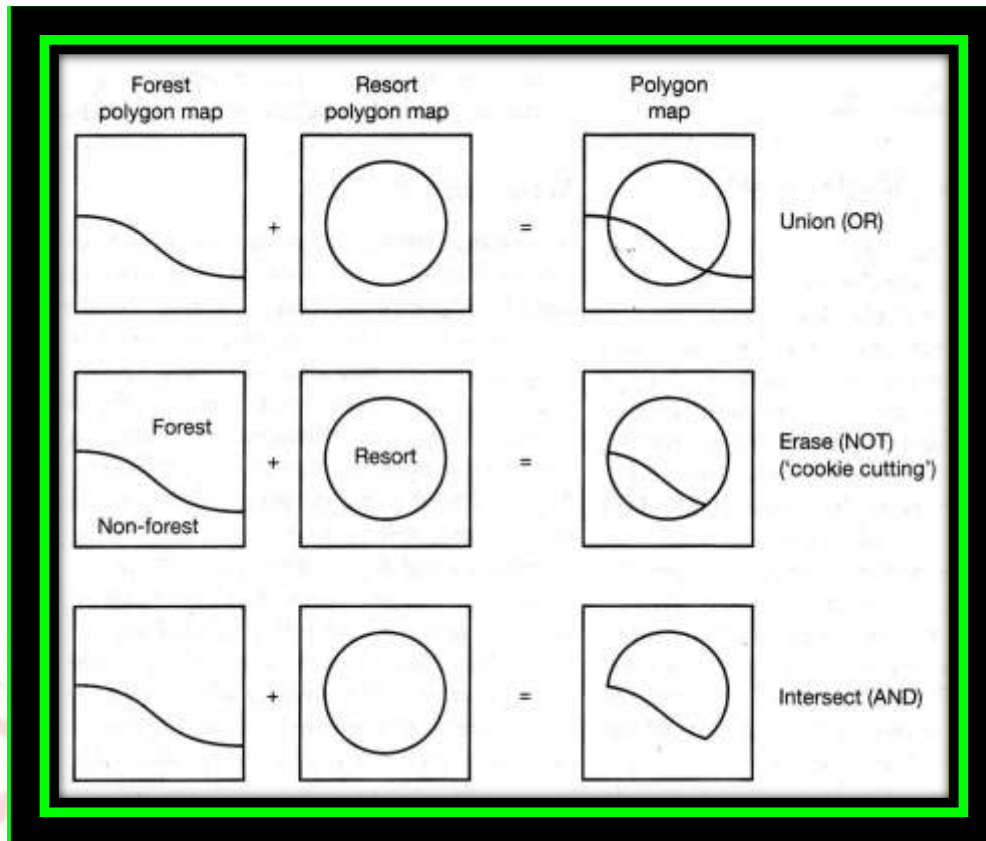

Figure 5: **Polygon on polygon overlay operations**

For example, a farmer wants to know which part of field has loam soil for the cultivation of crops. The farmer will overlay the map of loam soil polygons on field polygon to extract a feature that meets both criteria 'loam soil and in-field' for the cultivation of crops. The variables is processed by the farmer, both are categorical or nominal data type. Mathematician have developed set of algorithms, termed as Boolean operators for handling with this type of data and GIS analysts exploit in area on area overlay analysis. Boolean operators is the base of polygon overlay when it apply for vector data due to discrete area objects and nominal attributes.

Line in area Overlay operations

Line in area overlay operations (figure 6)need to check linear object or attribute which will combine or meagre with area layer. It should be also topological correct. Suppose I have to know about which road is going in forest are or non- forest. Simply I need to overlay the line layer on polygon, remaining extracted layer will be result.
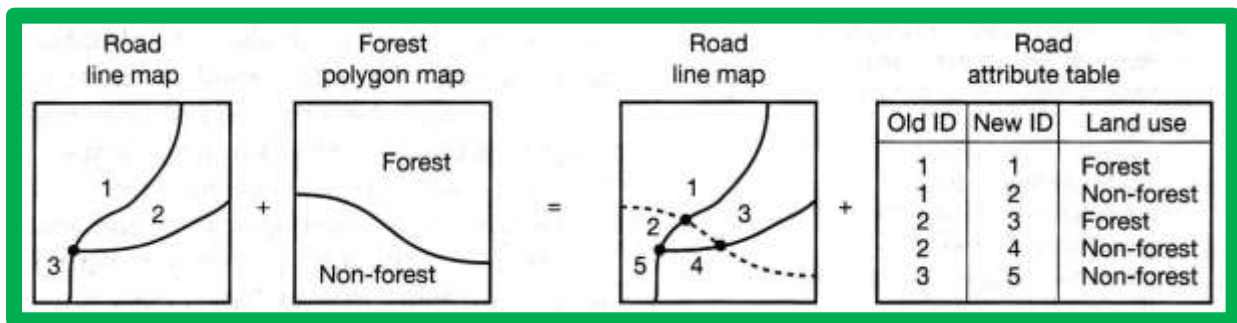

Figure 6: **Line in area Overlay operations**

Point in area overlay operations

In this type of operations (figure 7) need to check the point attribute or object that superimpose on area. For example I have to know location of met station in forest or non-forest area. It is simply done by overlaying the point layer on area layer.
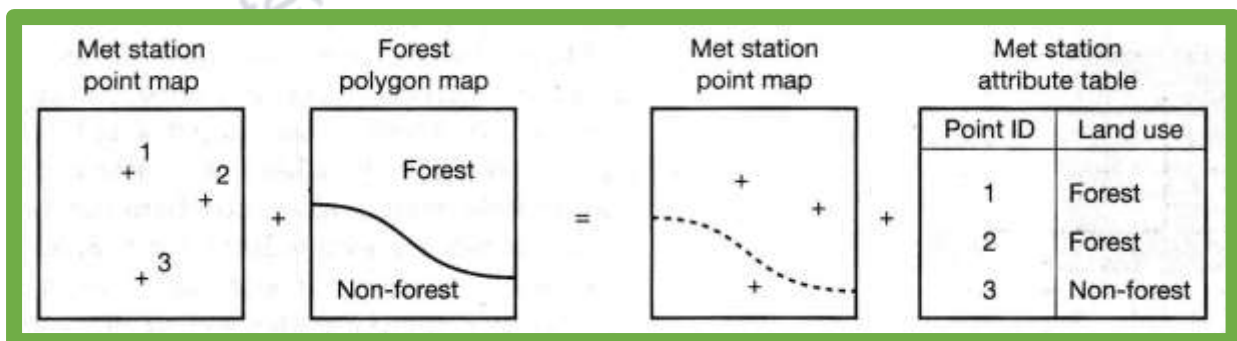

Figure 7: Point in area overlay operations

Weighted overlay and the vector data model

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. UNOIN 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.
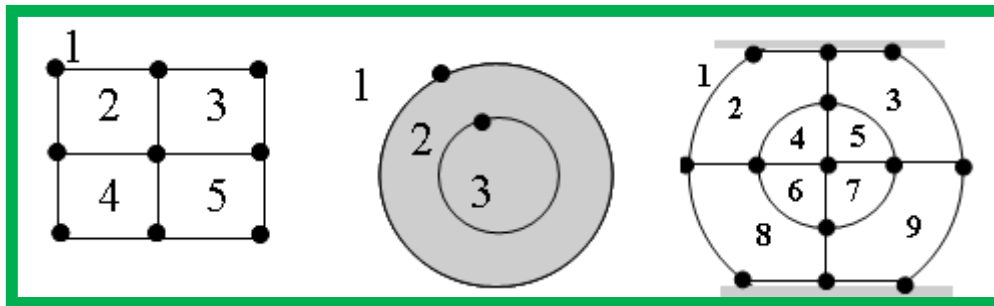
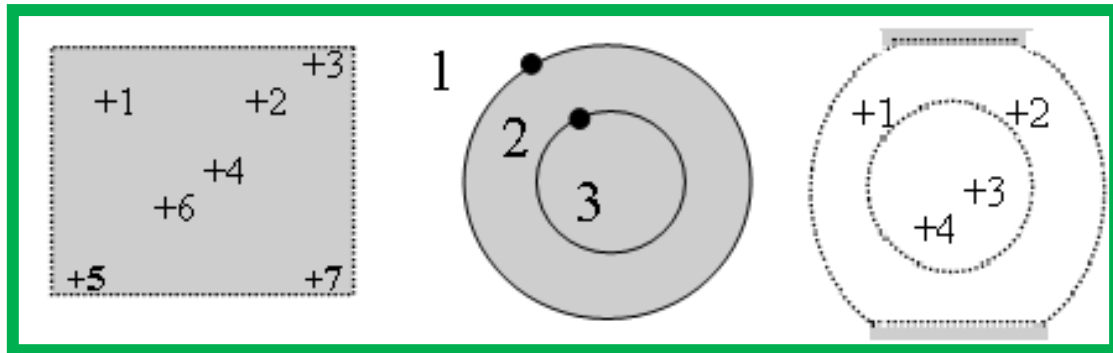(Polygon in Polygon)

INTERSECT
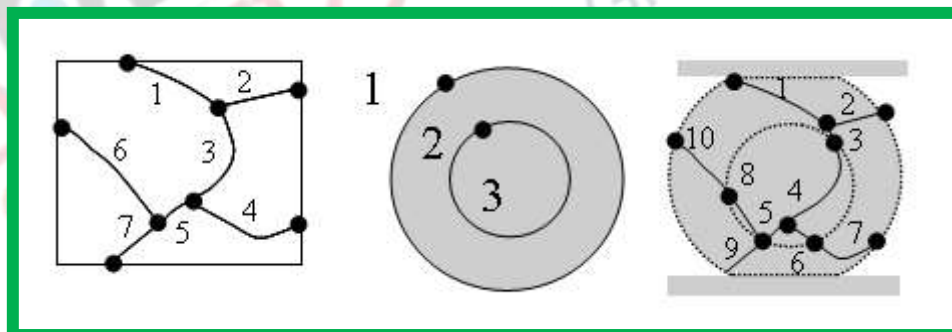
Input coverage   Intersect coverage   output coverage

(Line in Polygon)

INTERSECT

Input coverage   Intersect coverage   output coverage

(Point in Polygon)

INTERSECT

Figure 8: Intersect Operations during Overlay Analysis

Polygon overlay is often used for site selection or suitability analysis. One problem with vector overlay is the possible generation of sliver polygons. These sliver polygons arise from inconsistencies and inaccuracies in the digitized data.

Some important overlay operations are discussed here.

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 map. The city boundary is considered as the analysis layer.
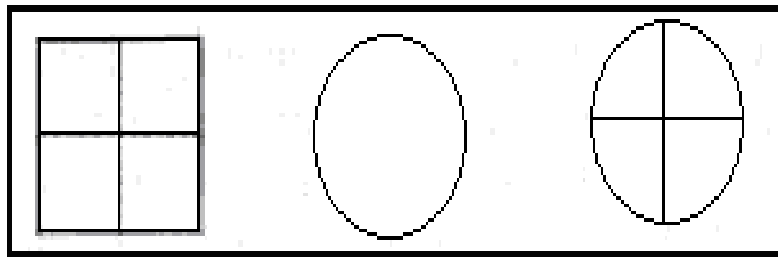
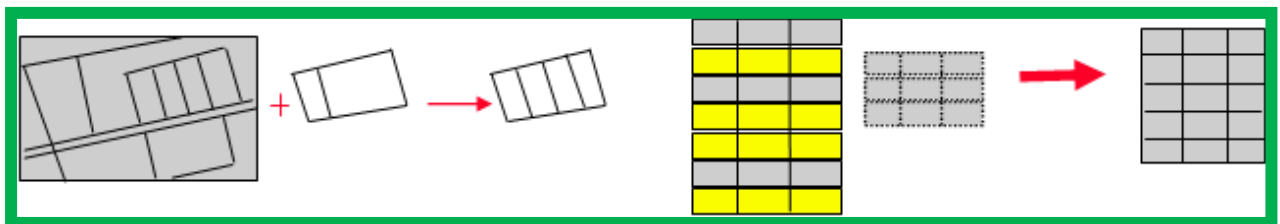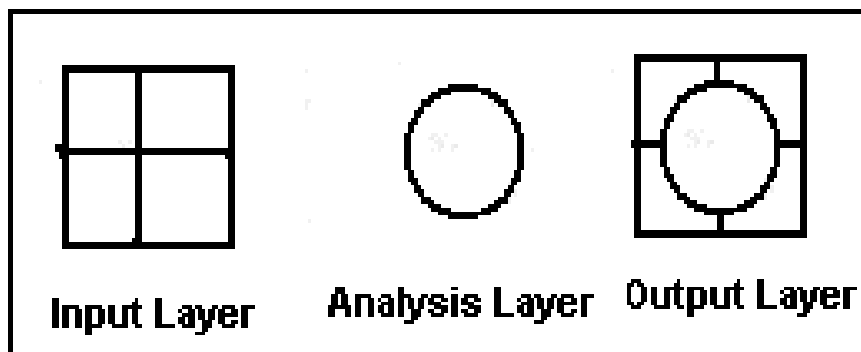Input layer   Analysis layer   Output layer

Figure 9: CLIP

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.

Figure 10: ERASE

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**.**
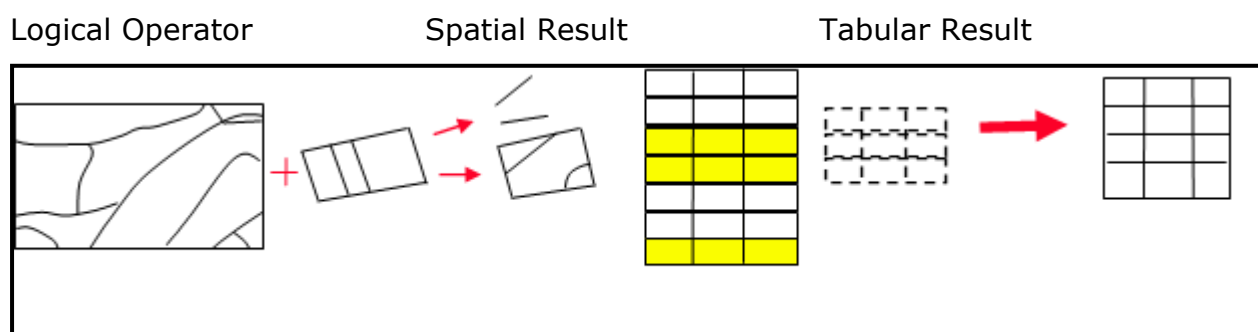
Logical Operator   Spatial Result   Tabular Result

Figure 11: SPLIT

UPDATE and IDENTITY

UPDATE uses a cut and paste operation to replace the input coverage and its map features.

Logical Operator   Spatial Result   Tabular Result
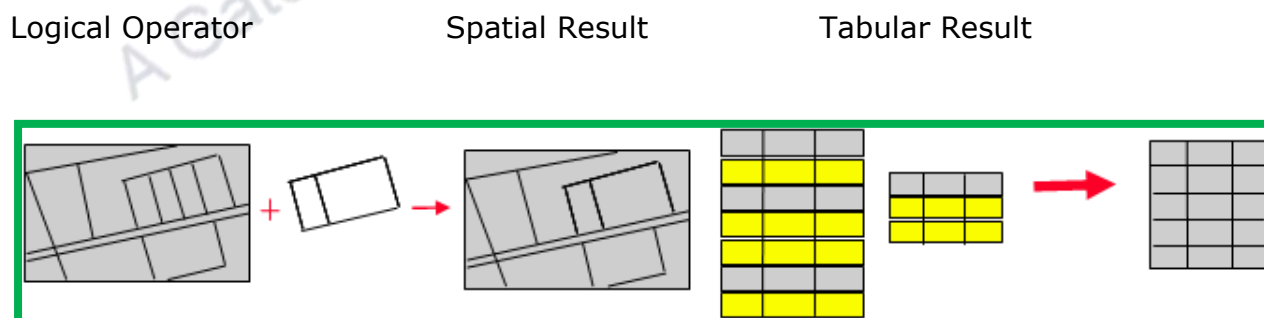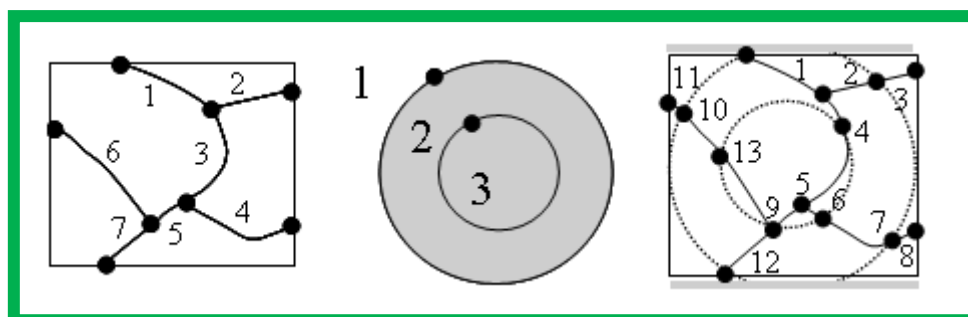

Figure 12: UPDATE

The input layer will be updated (replaced) by the analysis layer for generating the output layer. It means the resultant layer will have features from the input layer in the non-overlapping region and features from the analysis layer for the overlapping region. Both, the input and analysis layer have polygon topology. This operation is useful for updating the existing coverage with new map features in limited areas of coverage. FOR example, we have the land use map of 1990 and the built-up area of 2000 then the map of 1990 will be used as input layer and the map of 2000can be used as the analysis layer to get the updated result. The resultant layer will generate the urban sprawl. It saves the user from re-digitizing the coverage

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.
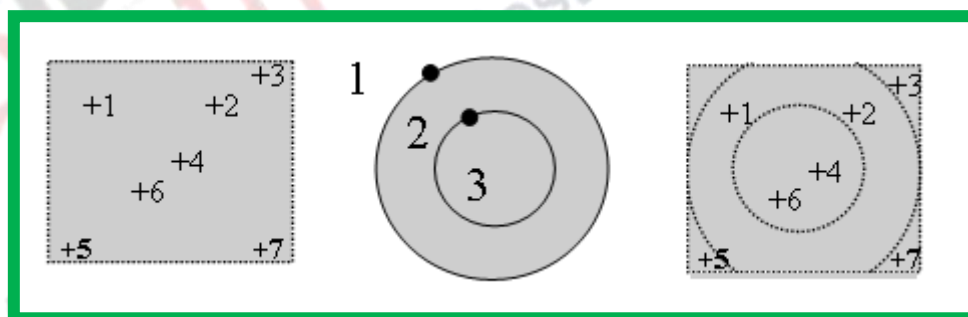
Input coverage   Identity coverage   Output coverage
                      (Point in Polygon)
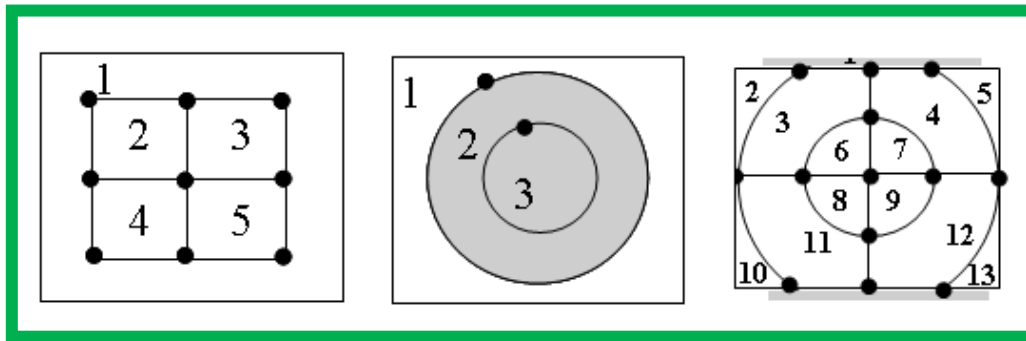
(Polygon in Polygon)
Figure 13: IDENTITY

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, RESELECT DISSOLVE etc are also used in the vector overlay operation (figure 14).
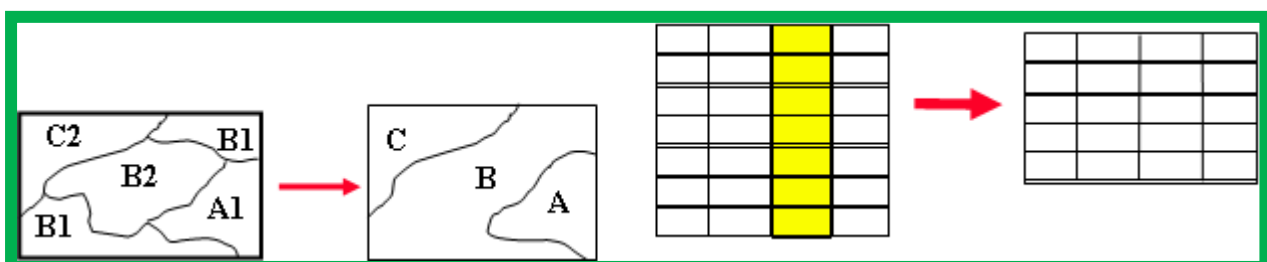
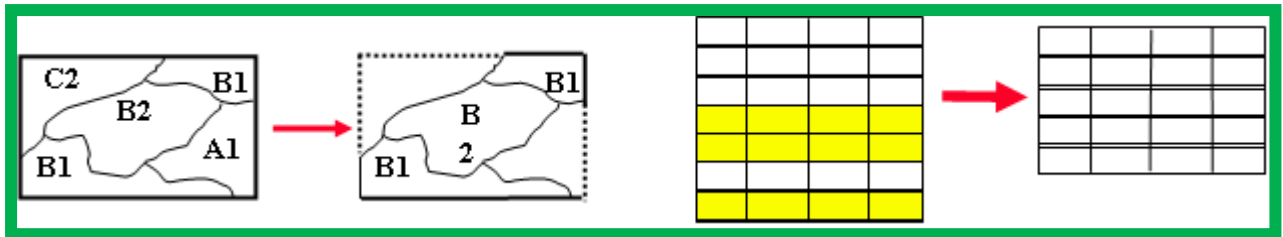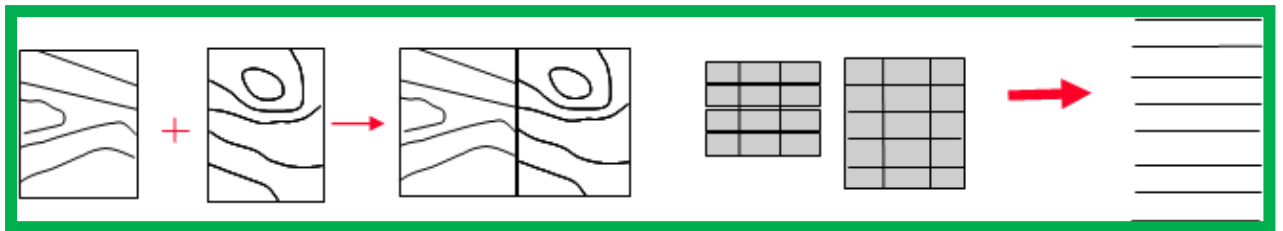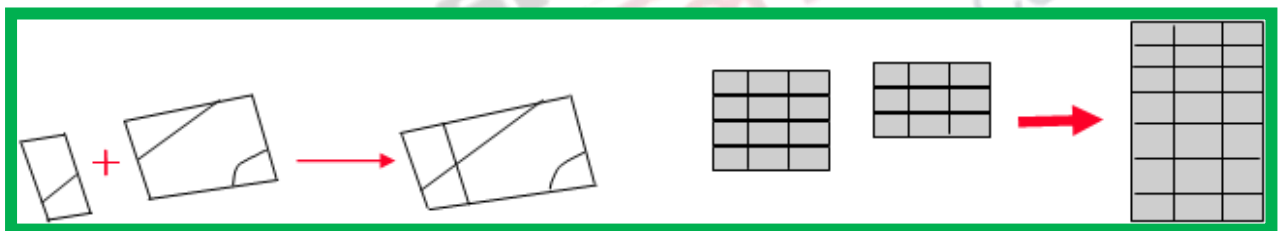Logical Operator   Spatial Result   Tabular Result
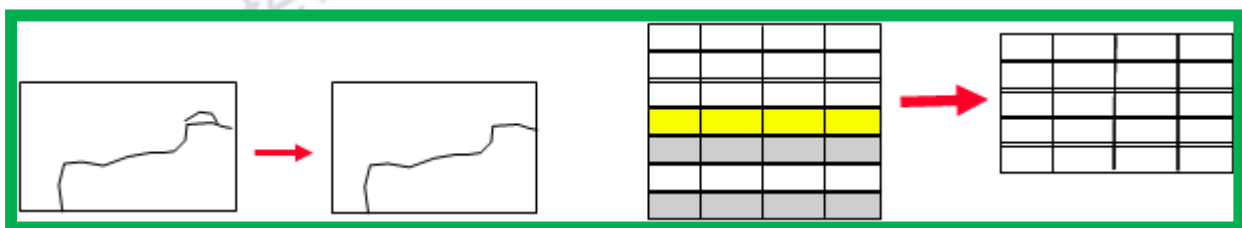
ELIMINATE

DISSOLVE

RESELECT

APPEND

MAP JOIN

Figure 14: SOME OTHER OVERLAY OPERATIONS

Raster based Overlay

Overlay of Raster datasets combine Pixel-based calculations or Map Algebra. It is quick, straight forward and efficient data sets for operations. It is known as cell by cell combination or operation. It is computationally less demanding. Overlay in raster datasets include two or more different sets of data that derive from a common grid. Each separate sets of data are usually specified numerical values. In the raster these values are mathematically merged together to create a new set of values for a single output layer.

The raster based overlay is done for create risk surfaces, sustainability assessments, value assessments, and other procedures. For example, raster based overlay divide the habitat of an endangered species into a grid, and then getting data for multiple factors that have an effect on the habitat and then creating a risk surface to illustrate what sections of the habitat need protecting most. If two grids are aligned and have the same grid cell size then it is relatively easy to perform overlay operations. A new layer of values is produced from each pair of coincident cells.
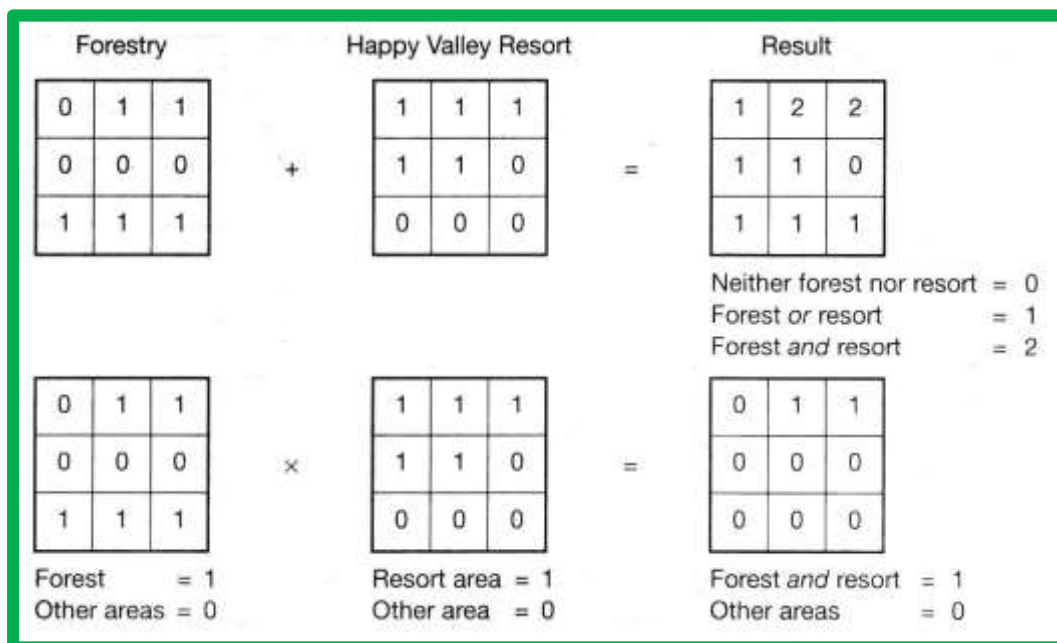

**Figure 15: Raster based Overlay**

The values of these cells can be added, subtracted, divided or multiplied, the maximum value can be extracted, mean value calculated, a logical expression computed and so on (figure 15). The output cell simply takes on a value equal to the result of the calculation.

Most sophisticated spatial modeling is undertaken within the raster domain. Each point can be addressed by as a part of a *neighborhood* of surrounding values. If all the neighboring points having the same attribute value are grouped together, is termed as *region.* Raster map 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:

- Local operations
- Neighborhood operation
- 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.
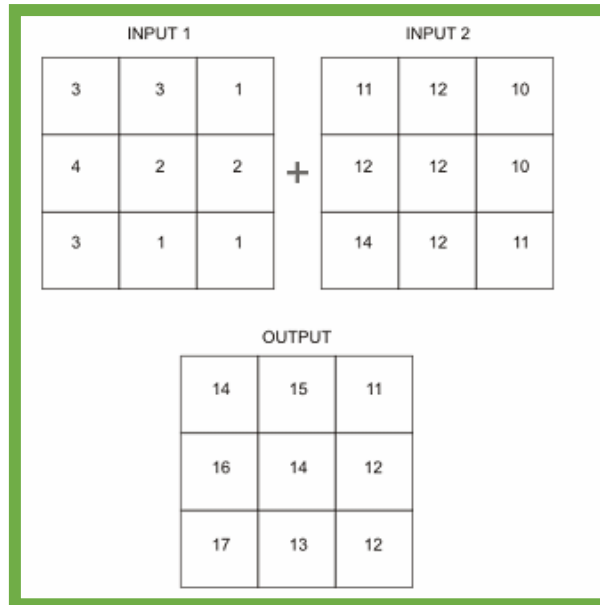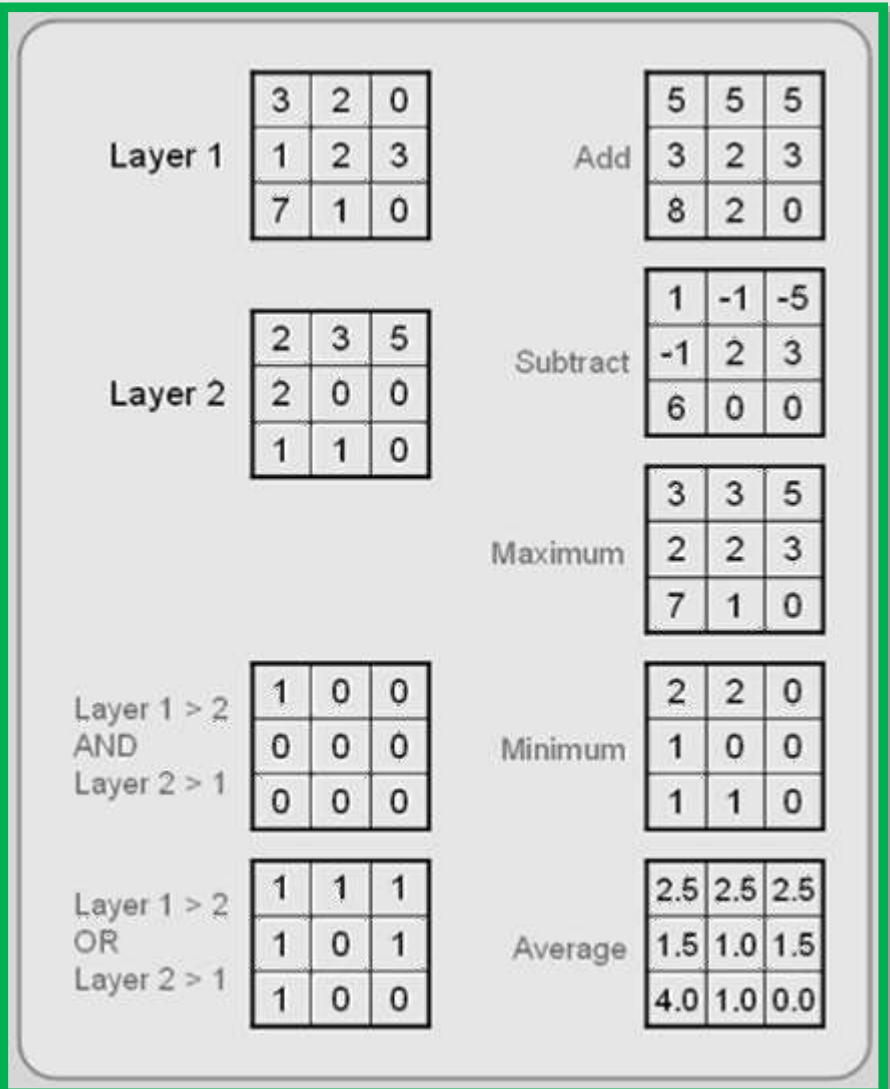
Add

Figure 16:Addition
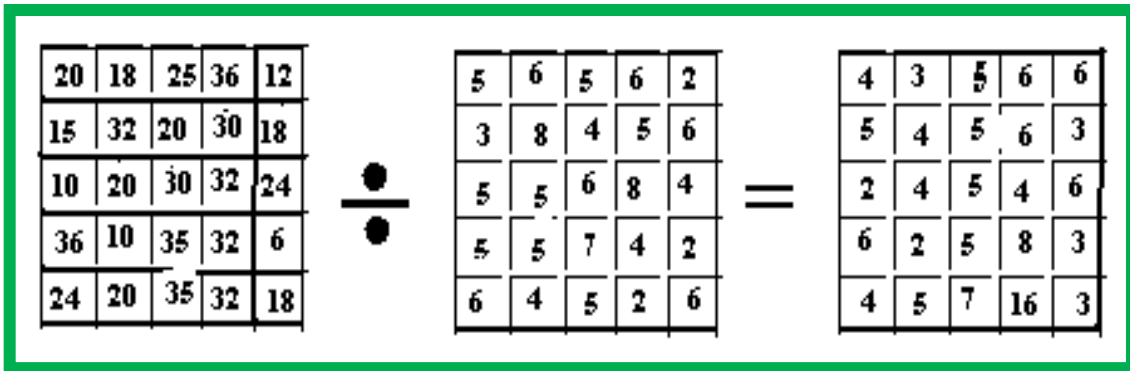
Figure 17: Subtraction
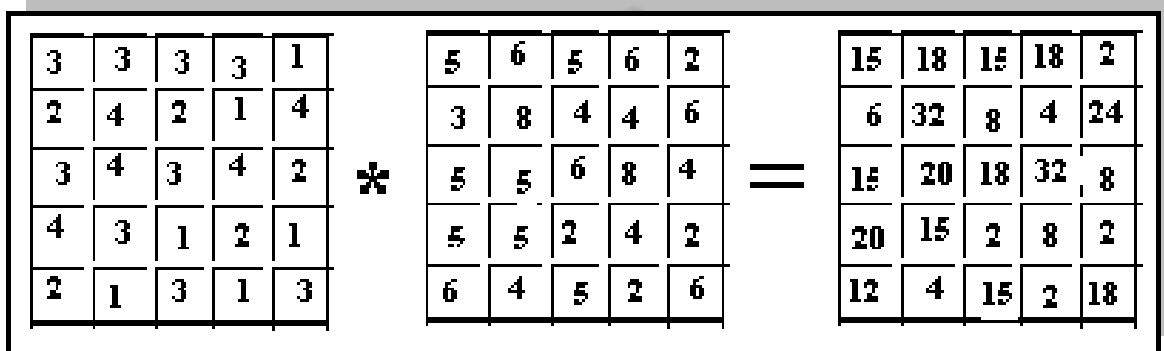
Figure 18:DIVISION

Figure 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.

5. 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.

In a vector based GIS, boundaries between polygons of common reclassed values are dissolved to create a homogeneous map. The dissolving of map boundaries based on a specific attribute value often results in a new data layer being created. Almost all GIS software provide the capability to easily dissolve boundaries based on the results of a reclassification. Some systems allow the user to create a new data layer

for the reclassification while others simply dissolve the boundaries during data output. The exact process for undertaking a reclassification varies greatly from GIS to GIS. Some will store results of the query in *query sets* independent from the DBMS, while others store the results in a newly created attribute column in the DBMS.

6.Application of Overlay Operations

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 favourable than residential housing in this model. It was desired to locate the school on flat slopes, near recreation sites, and far from existing schools.

For the site suitability overlay keep in mind some point

- Selection of criteria
- Reclassify the data as per the criteria
- Overlay (Boolean or Map Algebra)
- Extraction and representation of suitable site.

For example criteria of nuclear waste repository, site should be an area of suitable geology, site must not be easily accessible, site must be away from high population, and site must be outside the area of conservation.
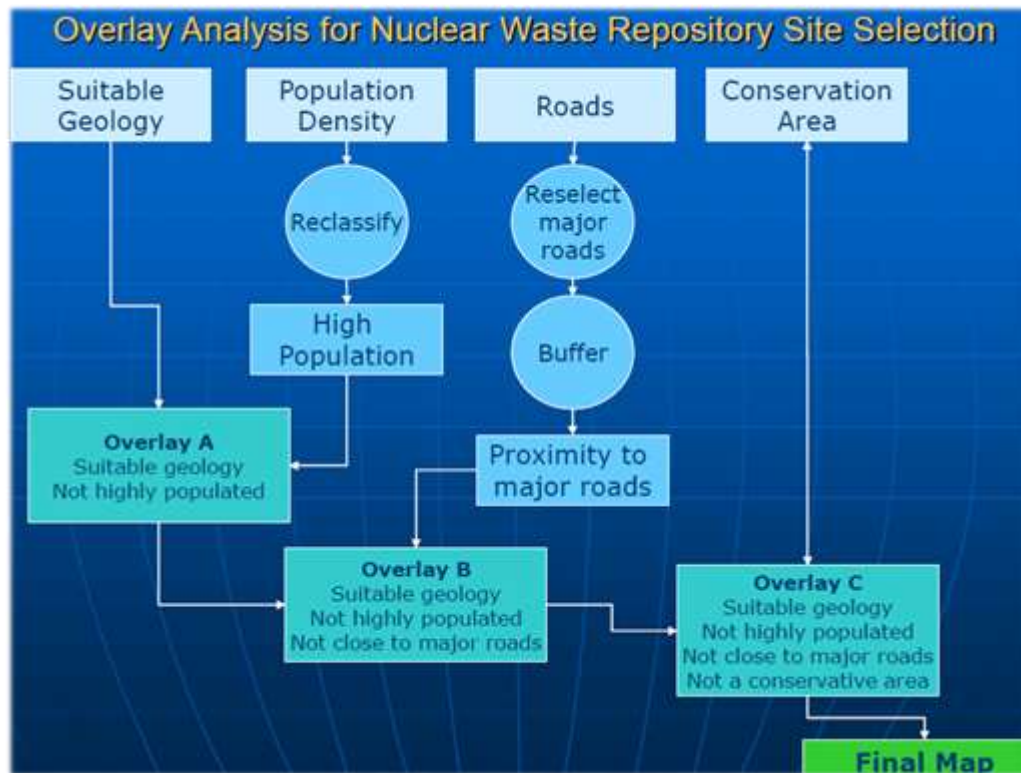
Figure 20: An Example of Overlay Analysis

Overlay Analysis if done appropriately could shape the Infrastructure of Future. The urban development and planning can certainly be implemented more efficiently with the help of GIS. The few factors, which we consider for choosing the location for urban establishments, are generally that land should be vacant; accessibility to water resources; area of vacant land and the vegetation demographic.

Another example could be of "Underground Water Utilities. Pipeline networks can be better planned with the GIS data as it allows better visualization and hence effective management. The available water resources, assets, networks, and pipelines can be simulated with accuracy through the GIS maps. Using the digital database, prepared through thispractice, water distribution can be undergrounded. Not only

water, oil and gas distribution processes can be also enhanced using this GIS,2017"

One of the example using real time data sets is 'In Making Metro Ecosystem Perfect' project where tools such as Image Station by Hexagon Geospatial allow high-volume photogrammetry and speedy processing of data. Certainly, sharing the accurate data in real-time, planning the metro stations and analyzing the moving trains is a challenge but certainly GIS technology is still growing and excelling.

7.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.