SPATIAL DATA ANALYSIS

P.L.N. Raju
Geoinformatics Division
Indian Institute of Remote Sensing, Dehra Dun

Abstract : Spatial analysis is the vital part of GIS. Spatial analysis in GIS involves three types of operations- attribute query (also known as non-spatial), spatial query and generation of new data sets from the original databases. Various spatial analysis methods viz. single/multiplayer operations/overlay; spatial modeling; geometric modeling; point pattern analysis; network analysis; surface analysis; raster/grid analysis etc. are discussed in detail in this paper.

INTRODUCTION

Geographic analysis allows us to study and understand the real world processes by developing and applying manipulation, analysis criteria and models and to carryout integrated modeling. These criteria illuminate underlying trends in geographic data, making new information available. A GIS enhances this process by providing tools, which can be combined in meaningful sequence to reveal new or previously unidentified relationships within or between data sets, thus increasing better understanding of real world. The results of geographic analysis can be commercial in the form of maps, reports or both. Integration involves bringing together diverse information from a variety of sources and analysis of multi-parameter data to provide answers and solutions to defined problems.

Spatial analysis is the vital part of GIS. It 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 1980s, many government agencies have invested heavily in GIS installations, including the purchase of hardware and software and the construction of mammoth databases. Two fundamental functions of GIS have been widely realized: generation of maps and generation of tabular reports.

Satellite Remote Sensing and GIS Applications in Agricultural Meteorology
pp. 151-174

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.

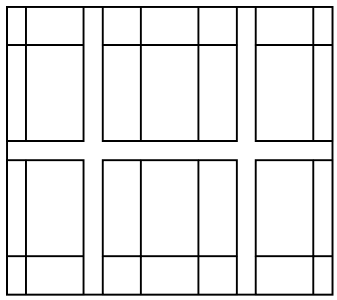
Making maps alone does not justify the high cost of building a GIS. The same maps may be produced using a simpler cartographic package. Likewise, if the purpose is to generate tabular output, then a simpler database management system or a statistical package may be a more efficient solution. It is spatial analysis that requires the logical connections between attribute data and map features, and the operational procedures built on the spatial relationships among map features. These capabilities make GIS a much more powerful and cost-effective tool than automated cartographic packages, statistical packages, or data base management systems. 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.

USING GIS FOR SPATIAL ANALYSIS

Spatial analysis in GIS involves three types of operations: Attribute Query-also known as non-spatial (or spatial) query, Spatial Query and Generation of new data sets from the original database (Bwozough, 1987). The scope of spatial analysis ranges from a simple query about the spatial phenomenon to complicated combinations of attribute queries, spatial queries, and alterations of original data.

Attribute Query: Requires the processing of attribute data exclusive of spatial information. In other words, it's a process of selecting information by asking logical questions.

Example: From a database of a city parcel map where every parcel is listed with a land use code, a simple attribute query may require the identification of all parcels for a specific land use type. Such a query can be handled through the table without referencing the parcel map (Fig. 1). Because no spatial information is required to answer this question, the query is considered an attribute query. In this example, the entries in the attribute table that have land use codes identical to the specified type are identified.



A sample parcel map

Parcel No.	*Size*	*Value*	*Land Use*
102	7,500	200,000	Commercial
103	7,500	160,000	Residential
104	9,000	250,000	Commercial
105	6,600	125,000	Residential

Attribute table of the sample parcel map

Figure 1: Listing of Parcel No. and value with land use = 'commercial' is an attribute query. Identification of all parcels within 100-m distance is a spatial query

Spatial Query: Involves selecting features based on location or spatial relationships, which requires processing of spatial information. For instance a question may be raised about parcels within one mile of the freeway and each parcel. In this case, the answer can be obtained either from a hardcopy map or by using a GIS with the required geographic information (Fig. 2).

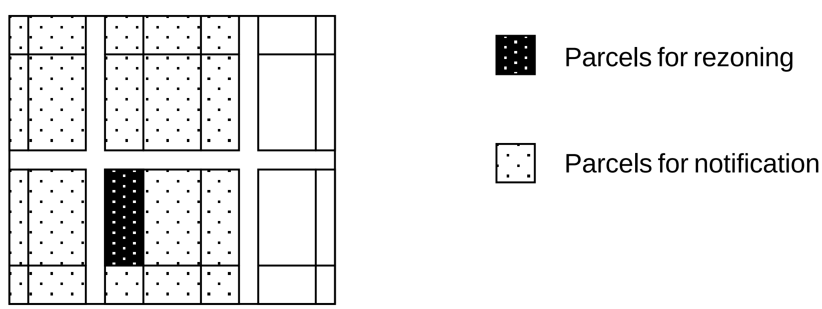


Figure 2: Land owners within a specified distance from the parcel to be rezoned identified through spatial query

Example: Let us take one spatial query example where a request is submitted for rezoning, all owners whose land is within a certain distance of all parcels that may be rezoned must be notified for public hearing. A spatial query is required to identify all parcels within the specified distance. This process cannot be accomplished without spatial information. In other words, the attribute table of the database alone does not provide sufficient information for solving problems that involve location.

While basic spatial analysis involves some attribute queries and spatial queries, complicated analysis typically require a series of GIS operations including multiple attribute and spatial queries, alteration of original data, and generation of new data sets. The methods for structuring and organizing such operations are a major concern in spatial analysis. An effective spatial analysis is one in which the best available methods are appropriately employed for different types of attribute queries, spatial queries, and data alteration. The design of the analysis depends on the purpose of study.

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

Single layer operations are procedures, which correspond to queries and alterations of data that operate on a single data layer.

Example: Creating a buffer zone around all streets of a road map is a single layer operation as shown in the Figure 3.

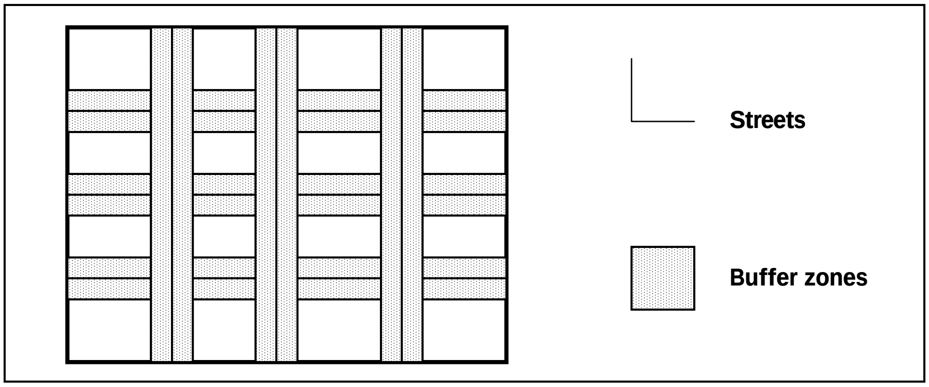


Figure 3: Buffer zones extended from streets

Multi layer operations: are useful for manipulation of spatial data on multiple data layers. Figure 4 depicts the overlay of two input data layers representing soil map and a land use map respectively. The overlay of these two layers produces the new map of different combinations of soil and land use.

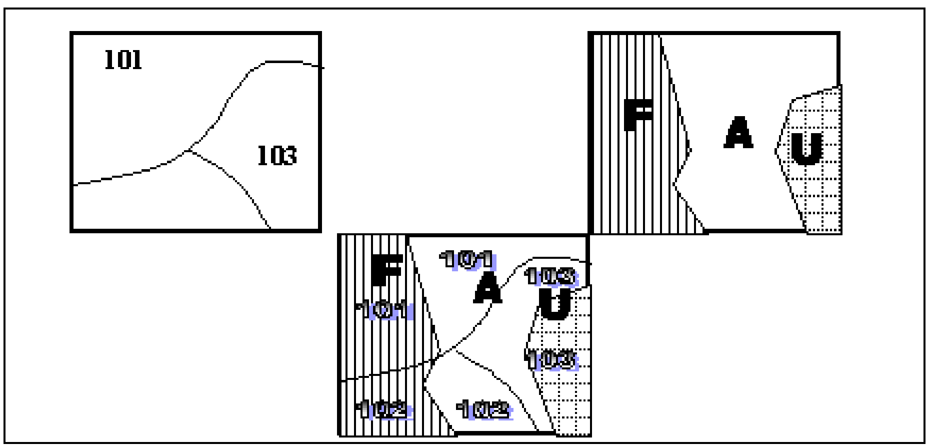


Figure 4: The overlay of two data layers creates a map of combined polygons

Topological overlays: These are multi layer operations, which allow combining features from different layers to form a new map and give new information and features that were not present in the individual maps. This topic will be discussed in detail in section of vector-based analysis.

Point pattern analysis: It deals with the examination and evaluation of spatial patterns and the processes of point features. A typical biological survey map is shown in Figure 5, in which each point feature denotes the observation of an endangered species such as big horn sheep in southern California. The objective of illustrating point features is to determine the most favourable environmental conditions for this species. Consequently, the spatial distribution of species can be examined in a point pattern analysis. If the distribution illustrates a random pattern, it may be difficult to identify significant factors that influence species distribution. However, if observed locations show a systematic pattern such as the clusters in this diagram, it is possible to analyze the animals' behaviour in terms of environmental characteristics. In general, point pattern analysis is the first step in studying the spatial distribution of point features.

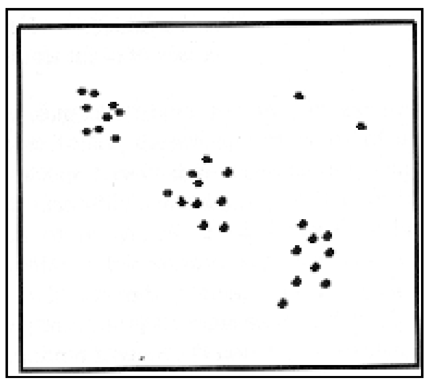


Figure 5: Distribution of an endangered species examined in a point pattern analysis

Network analysis: Designed specifically for line features organized in connected networks, typically applies to transportation problems and location analysis such as school bus routing, passenger plotting, walking distance, bus stop optimization, optimum path finding etc.

Figure 6 shows a common application of GIS-based network analysis. Routing is a major concern for the transportation industry. For instance, trucking companies must determine the most cost-effective way of connecting stops for pick-up or delivery. In this example, a route is to be delineated for a truck to pick up packages at five locations. A routing application can be developed to identify the most efficient route for any set of pick-up locations. The highlighted line represents the most cost-effective way of linking the five locations.

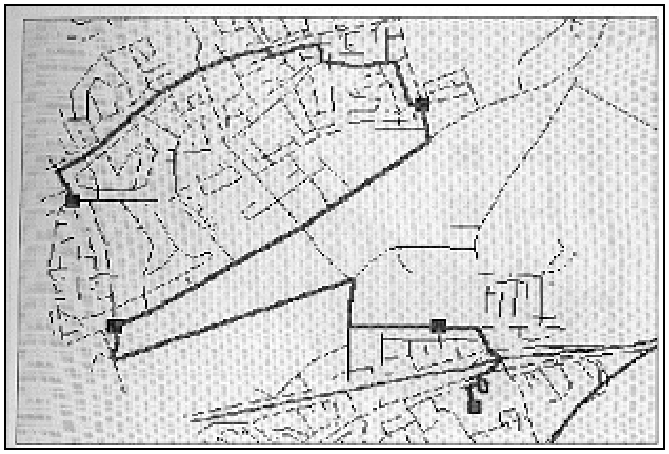


Figure 6: The most cost effective route links five point locations on the street map

Surface analysis deals with the spatial distribution of surface information in terms of a three-dimensional structure.

The distribution of any spatial phenomenon can be displayed in a three-dimensional perspective diagram for visual examination. A surface may represent the distribution of a variety of phenomena, such as population, crime, market potential, and topography, among many others. The perspective diagram in Figure 7 represents topography of the terrain, generated from digital elevation model (DEM) through a series of GIS-based operations in surface analysis.

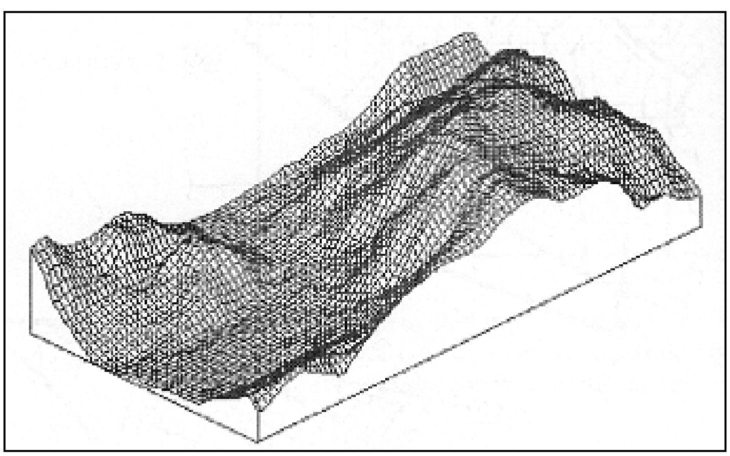


Figure 7: Perspective diagram representing topography of the terrain derived from a surface analysis

Grid analysis involves the processing of spatial data in a special, regularly spaced form. The following illustration (Figure 8) shows a grid-based model of fire progression. The darkest cells in the grid represent the area where a fire is currently underway. A fire probability model, which incorporates fire behaviour in response to environmental conditions such as wind and topography, delineates areas that are most likely to burn in the next two stages. Lighter shaded cells represent these areas. Fire probability models are especially useful to fire fighting agencies for developing quick-response, effective suppression strategies.

In most cases, GIS software provides the most effective tool for performing the above tasks.

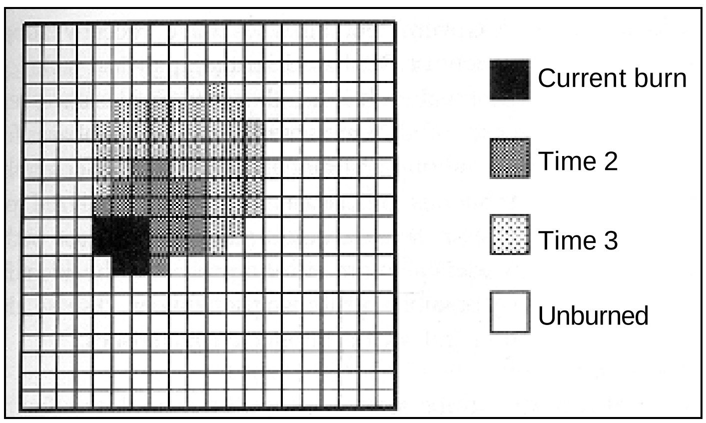


Figure 8: A fire behaviour model delineates areas of fire progression based on a grid analysis

Fuzzy Spatial Analysis

Fuzzy spatial analysis is based on Fuzzy set theory. Fuzzy set theory is a generalization of Boolean algebra to situations where zones of gradual transition are used to divide classes, instead of conventional crisp boundaries. This is more relevant in many cases where one considers ‘distance to certain zone’ or ‘distance to road’, in which case the influence of this factor is more likely to be some function of distance than a binary ‘yes’ or ‘no’. Also in fuzzy theory maps are prepared showing gradual change in the variable from very high to very low, which is a true representation of the real world (Bonhan-Carter, 1994).

As stated above, the conventional crisp sets allow only binary membership function (i.e. true or false), whereas a fuzzy set is a class that admits the possibility of partial membership, so fuzzy sets are generalization of crisp sets to situations where the class membership or class boundaries are not, or cannot be, sharply defined.

Applications

Data integration using fuzzy operators using standard rules of fuzzy algebra one can combine various thematic data layers, represented by respective membership values (Chung and Fabbri, 1993).

Example: In a grid cell/pixel if a particular litho-unit occurs in combination with a thrust/fault, its membership value should be much higher compared with individual membership values of litho-unit or thrust/fault. This is significant as the effect is expected to be "increasive" in our present consideration and it can be calculated by fuzzy algebraic sum. Similarly, if the presence of two or a set of parameters results in "decreasive" effect, it can be calculated by fuzzy algebraic product. Besides this, fuzzy algebra offers various other methods to combine different data sets for landslide hazard zonation map preparation. To combine number of exploration data sets, five such operators exist, namely the fuzzy AND, the fuzzy OR, fuzzy algebraic product, fuzzy algebraic sum and fuzzy gamma operator.

Fuzzy logic can also be used to handle mapping errors or uncertainty, i.e. errors associated with clear demarcation of boundaries and also errors present in the area where limited ground truth exists in studies such as landslide hazard zonation. The above two kinds of errors are almost inherent to the process of data collection from different sources including remote sensing.

GEOSTATISTICAL TOOLS FOR SPATIAL ANALYSIS

Geostatistics studies spatial variability of regionalized variables: Variables that have an attribute value and a location in a two or three-dimensional space. Tools to characterize the spatial variability are:

- Spatial Autocorrelation Function and
- Variogram.

A *variogram* is calculated from the variance of pairs of points at different separation. For several distance classes or lags, all point pairs are identified which matches that separation and the variance is calculated. Repeating this process for various distance classes yields a variogram. These functions can be used to measure spatial variability of point data but also of maps or images.

Spatial Auto-correlation of Point Data

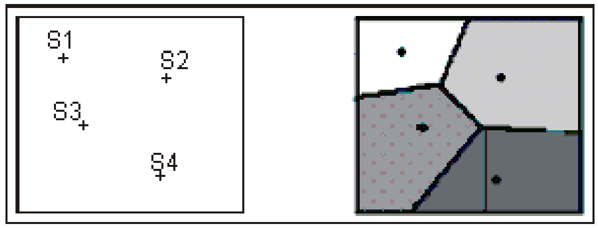
The statistical analysis referred to as spatial auto-correlation, examines the correlation of a random process with itself in space. Many variables that have discrete values measured at several specific geographic positions (i.e., individual observations can be approximated by dimensionless points) can be considered random processes and can thus be analyzed using spatial auto-correlation analysis. Examples of such phenomena are: Total amount of rainfall, toxic element concentration, grain size, elevation at triangulated points, etc.

The spatial auto-correlation function, shown in a graph is referred to as spatial *auto-correlogram*, showing the correlation between a series of points or a map and itself for different shifts in space or time. It visualizes the spatial variability of the phenomena under study. In general, large numbers of pairs of points that are close to each other on average have a lower variance (i.e., are better correlated), than pairs of points at larger separation. The auto-correlogram quantifies this relationship and allows gaining insight into the spatial behaviour of the phenomenon under study.

Point Interpolation

A point interpolation performs an interpolation on randomly distributed point values and returns regularly distributed point values. The various interpolation methods are: Voronoi Tesselation, moving average, trend surface and moving surface.

Example: Nearest Neighbor (Voronoi Tessellation)-In this method the value, identifier, or class name of the nearest point is assigned to the pixels. It offers a quick way to obtain a Thiessen map from point data (Figure 9).



(a) (b)

Figure 9: (a) An input point map, (b) The output map obtained as the result of the interpolation operation applying the Voronoi Tessellation method

VECTOR BASED SPATIAL DATA ANALYSIS

In this section the basic concept of various vector operations are dealt in detail. There are multi layer operations, which allow combining features from different layers to form a new map and give new information and features that were not present in the individual maps.

Topological overlays: Selective overlay of polygons, lines and points enables the users to generate a map containing features and attributes of interest, extracted from different themes or layers. Overlay operations can be performed on both raster (or grid) and vector maps. In case of raster map calculation tool is used to perform overlay. In topological overlays polygon features of one layer can be combined with point, line and polygon features of a layer.

Polygon-in-polygon overlay:

Output is polygon coverage.

Coverages are overlaid two at a time.

There is no limit on the number of coverages to be combined.

New File Attribute Table is created having information about each newly created feature.

Line-in-polygon overlay:

Output is line coverage with additional attribute.

No polygon boundaries are copied.

New arc-node topology is created.

Point-in-polygon overlay:

Output is point coverage with additional attributes.

No new point features are created.

No polygon boundaries are copied.

Logical Operators: Overlay analysis manipulates spatial data organized in different layers to create combined spatial features according to logical conditions specified in Boolean algebra with the help of logical and conditional operators. The logical conditions are specified with operands (data elements) and operators (relationships among data elements).

Note: In vector overlay, arithmetic operations are performed with the help of logical operators. There is no direct way to it.

Common logical operators include AND, OR, XOR (Exclusive OR), and NOT. Each operation is characterized by specific logical checks of decision criteria to determine if a condition is true or false. Table 1 shows the true/false conditions of the most common Boolean operations. In this table, A and B are two operands. One (1) implies a true condition and zero (0) implies false. Thus, if the A condition is true while the B condition is false, then the combined condition of A and B is false, whereas the combined condition of A OR B is true.

AND - Common Area/ Intersection / Clipping Operation

OR - Union Or Addition

NOT - (Inverter)

XOR - Minus

Table 1: Truth Table of common Boolean operations

A	B	A AND B	A OR B	A NOT B	B NOT A	A XOR B
0	0	0	0	0	0	0
0	1	0	1	0	1	1
1	0	0	1	1	0	1
1	1	1	1	0	0	0

The most common basic multi layer operations are union, intersection, and identify operations. All three operations merge spatial features on separate data layers to create new features from the original coverage. The main difference among these operations is in the way spatial features are selected for processing.

Overlay operations

The Figure 10 shows different types of vector overlay operations and gives flexibility for geographic data manipulation and analysis. In polygon overlay, features from two map coverages are geometrically intersected to produce a

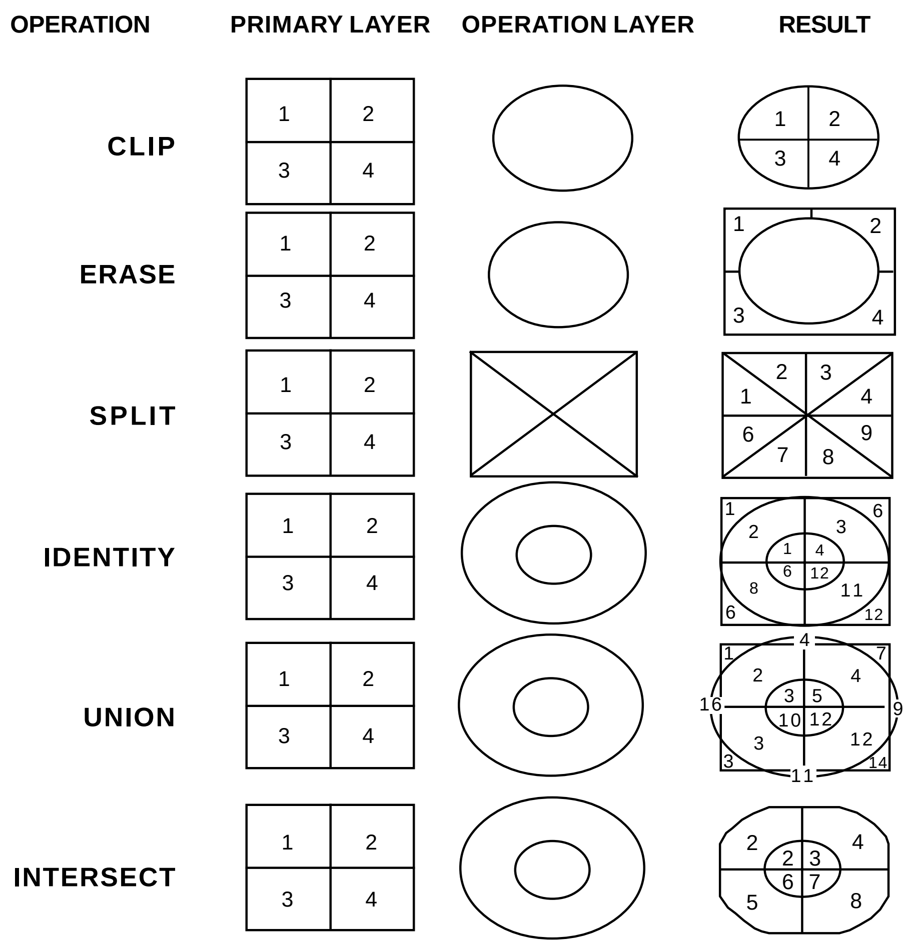
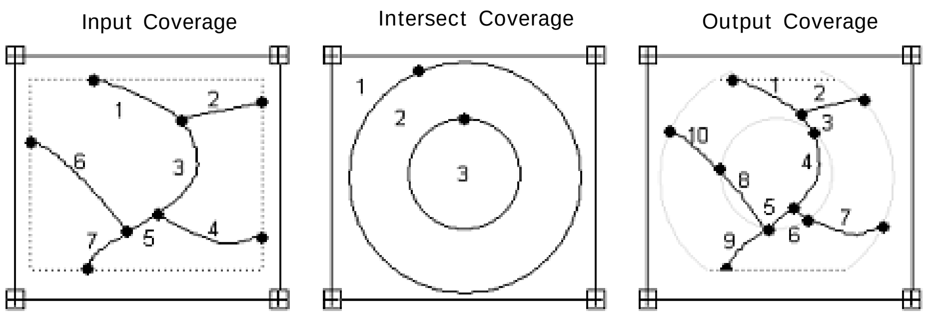


Figure 10 : Overlay operations

new set of information. Attributes for these new features are derived from the attributes of both the original coverages, thereby contain new spatial and attribute data relationships.

One of the overlay operation is AND (or INTERSECT) in vector layer operations, in which two coverages are combined. Only those features in the area common to both are preserved. Feature attributes from both coverages are joined in the output coverage.



INPUT COVERAGE	
#	ATTRIBUTE
1	A
2	B
3	A
4	C
5	A
6	D
7	A

INTERSECT COVERAGE	
#	ATTRIBUTE
1	
2	102
3	103

OUTPUT COVERAGE	INPUT COVERAGE		INTERSECT COVERAGE	
#	#	ATTRIBUTE	#	ATTRIBUTE
1	1	A	2	102
2	2	B	2	102
3	3	A	2	102
4	3	A	3	103
5	5	A	3	103
6	4	C	3	103
7	4	C	2	102
8	6	D	3	103
9	7	A	2	102
10	6	D	2	102

RASTER BASED SPATIAL DATA ANALYSIS

Present section discusses operational procedures and quantitative methods for the analysis of spatial data in raster format. 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.

Advantages of using the Raster Format in Spatial Analysis

Efficient processing: Because geographic units are regularly spaced with identical spatial properties, multiple layer operations 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 presentation unless 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.

Grid Operations used in Map Algebra

Common operations in grid analysis consist of the following functions, which are used in Map Algebra to manipulate grid files. The Map Algebra language is a programming language developed to perform cartographic modeling. Map Algebra performs following four basic operations:

- Local functions: that work on every single cell,

- Focal functions: that process the data of each cell based on the information of a specified neighborhood,

- Zonal functions: that provide operations that work on each group of cells of identical values, and

- Global functions: that work on a cell based on the data of the entire grid.

The principal functionality of these operations is described here.

Local Functions

Local functions process a grid on a cell-by-cell basis, that is, each cell is processed based solely on its own values, without reference to the values of other cells. In other words, the output value is a function of the value or values of the cell being processed, regardless of the values of surrounding cells. For single layer operations, a typical *example* is changing the value of each cell by adding or multiplying a constant. In the following example, the input grid contains values ranging from 0 to 4. Blank cells represent NODATA cells. A simple local function multiplies every cell by a constant of 3 (Fig. 11). The results are shown in the output grid at the right. When there is no data for a cell, the corresponding cell of the output grid remains a blank.

Input Grid

2	0	1	1
2	3	0	4
4		2	3
1	1		2

X 3 =

Output Grid

6	0	3	3
6	9	0	12
12		6	9
3	3		6

Figure 11: A local function multiplies each cell in the input grid by 3 to produce the output grid

Local functions can also be applied to multiple layers represented by multiple grids of the same geographic area (Fig. 12).

Input Grid

2	0	1	1
2	3	0	4
4		2	3
1	1		2

X

Multiplier Grid

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

=

Output Grid

2	0	2	2
2	6	0	12
8		6	9
2	3		8

Figure 12: A local function multiplies the input grid by the multiplier grid to produce the output grid

Local functions are not limited to arithmetic computations. Trigonometric, exponential, and logarithmic and logical expressions are all acceptable for defining local functions.

Focal Functions

Focal functions process cell data depending on the values of neighboring cells. For instance, computing the sum of a specified neighborhood and assigning the sum to the corresponding cell of the output grid is the "focal sum" function (Fig. 13). A 3 x 3 kernel defines neighborhood. For cells closer to the edge where the regular kernel is not available, a reduced kernel is used and the sum is computed accordingly. For instance, a 2 x 2 kernel adjusts the upper left corner cell. Thus, the sum of the four values, 2,0,2 and 3 yields 7, which becomes the value of this cell in the output grid. The value of the second row, second column, is the sum of nine elements, 2, 0, 1, 2, 3, 0, 4, 2 and 2, and the sum equals 16.

Input Grid

2	0	1	1
2	3	0	4
4	2	2	3
1	1	3	2

Focal Sum =

Output Grid

7	8	9	6
13	16	16	11
13	18	20	14
8	13	13	10

Figure 13: A Focal sum function sums the values of the specified neighborhood to produce the output grid

Another focal function is the mean of the specified neighborhood, the "focal mean" function. In the following example (Fig. 14), this function yields the mean of the eight adjacent cells and the center cell itself. This is the smoothing function to obtain the moving average in such a way that the value of each cell is changed into the average of the specified neighborhood.

Input Grid

2	0	1	1
2	3	0	4
4	2	2	3
1	1	3	2

Focal Mean =

Output Grid

1.8	1.3	1.5	1.5
2.2	2.0	1.8	1.8
2.2	2.0	2.2	2.3
2.0	2.2	2.2	2.5

Figure 14: A Focal mean function computes the moving average of the specified neighborhood to produce the output grid

Other commonly employed focal functions include standard deviation (focal standard deviation), maximum (focal maximum), minimum (focal minimum), and range (focal range).

Zonal Functions

Zonal functions process the data of a grid in such a way that cell of the same zone are analyzed as a group. A zone consists of a number of cells that may or may not be contiguous. A typical zonal function requires two grids – a zone grid, which defines the size, shape and location of each zone, and a value grid, which is to be processed for analysis. In the zone grid, cells of the same zone are coded with the same value, while zones are assigned different zone values.

Figure 15 illustrates an example of the zonal function. The objective of this function is to identify the zonal maximum for each zone. In the input zone grid, there are only three zones with values ranging from 1 to 3. The zone with a value of 1 has five cells, three at the upper right corner and two at the lower left corner. The procedure involves finding the maximum value among these cells from the value grid.

	Zone Grid				Value Grid					Output Grid			
	2	2	1	1	1	2	3	4		5	5	8	8
Zonal	2	3	3	1	5	6	7	8		5	7	7	8
Max [		3	2		1	2	3	4	]=		7	5	
	1	1	2	2	5	5	5	5		8	8	5	5

Figure 15: A Zonal maximum function identifies the maximum of each zone to produce the output grid

Typical zonal functions include zonal mean, zonal standard deviation, zonal sum, zonal minimum, zonal maximum, zonal range, and zonal variety. Other statistical and geometric properties may also be derived from additional zonal functions. For instance, the zonal perimeter function calculates the perimeter of each zone and assigns the returned value to each cell of the zone in the output grid.

Global Functions

For global functions, the output value of each cell is a function of the entire grid. As an example, the Euclidean distance function computes the distance from each cell to the nearest source cell, where source cells are defined in an input grid. In a square grid, the distance between two orthogonal neighbors is equal to the size of a cell, or the distance between the centroid locations of adjacent cells. Likewise, the distance between two diagonal neighbors is equal to the cell size multiplied by the square root of 2. Distance between non-adjacent cells can be computed according to their row and column addresses.

In Figure 16, the grid at the left is the source grid in which two clusters of source cells exist. The source cells labeled 1 are the first clusters, and the cell labeled 2 is a single-cell source. The Euclidean distance from any source cell is always equal to 0. For any other cell, the output value is the distance from its nearest source cell.

Source Grid

		1	1
			1
	2		

Euclidean distance =

Output Grid

2.0	1.0	0.0	0.0
1.4	1.0	1.0	0.0
1.0	0.0	1.0	1.0
1.4	1.0	1.4	2.0

Figure 16: A Euclidean distance function computes the distance from the nearest source cell

In the above example, the measurement of the distance from any cell must include the entire source grid; therefore this analytical procedure is a global function.

Figure 17 provides an example of the cost distance function. The source grid is identical to that in the preceding illustration. However, this time a cost grid is employed to weigh travel cost. The value in each cell of the cost grid indicates the cost for traveling through that cell. Thus, the cost for traveling from the cell located in the first row, second column to its adjacent source cell to the right is half the cost of traveling through itself plus half the cost of traveling through the neighboring cell.

Source Grid

		1	1
			1
	2		

Cost Grid

2	2	4	4
4	4	3	3
2	1	4	1
2	5	3	3

=

Output Grid

5.0	3.0	0	0
3.5	2.5	2.8	0
1.5	0	2.5	2.0
2.1	3.0	2.8	4.0

Figure 17: Travel cost for each cell is derived from the distance to the nearest source cell weighted by a cost function

Another useful global function is the cost path function, which identifies the least cost path from each selected cell to its nearest source cell in terms of cost distance. These global functions are particularly useful for evaluating the connectivity of a landscape and the proximity of a cell to any given entities.

SOME IMPORTANT RASTER ANALYSIS OPERATIONS

In this section some of the important raster based analysis are dealt:

- Renumbering Areas in a Grid File
- Performing a Cost Surface Analysis
- Performing an Optimal Path Analysis
- Performing a Proximity Search

Area Numbering: Area Numbering assigns a unique attribute value to each area in a specified grid file. An area consists of two or more adjacent cells that have the same cell value or a single cell with no adjacent cell of the same value. To consider a group of cells with the same values beside each other, a cell must have a cell of the same value on at least one side of it horizontally or vertically (4-connectivity), or on at least one side horizontally, vertically, or diagonally (8-connectivity). Figure 18 shows a simple example of area numbering.

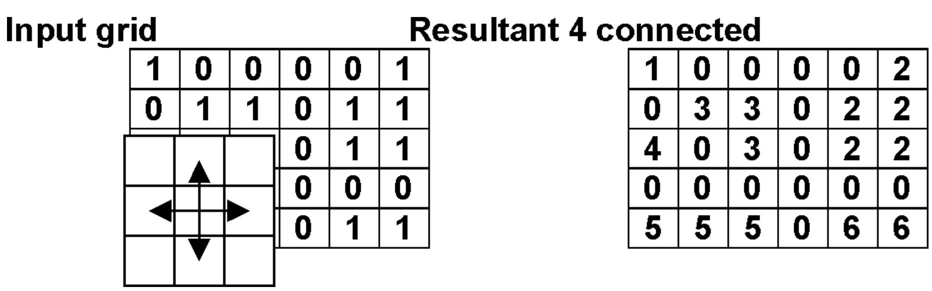


Figure 18. Illustrates simple example of Area numbering with a bit map as input. The pixels, which are connected, are assigned the same code. Different results are obtained when only the horizontal and vertical neighbors are considered (4-connected) or whether all neighbors are considered (8-connected)

One can renumber all of the areas in a grid, or you can renumber only those areas that have one or more specific values. If you renumber all of the areas, Area Number assigns a value of 1 to the first area located. It then assigns a value of 2 to the second area, and continues this reassignment method until all of the areas are renumbered. When you renumber areas that contain a specified value (such as 13), the first such area is assigned the maximum grid value plus 1. For example, if the maximum grid value is 25, Area Number assigns a value of 26 to the first area, a value of 27 to the second area, and continues until all of the areas that contain the specified values are renumbered.

Cost Surface Analysis: Cost Surface generates a grid in which each grid cell represents the cost to travel to that grid cell from the nearest of one or more start locations. The cost of traveling to a given cell is determined from a weight grid file. Zero Weights option uses attribute values of 0 as the start locations. The By Row/Column option uses the specified row and column location as the start location.

Optimal Path: Optimal Path lets us analyze a grid file to find the best path between a specified location and the closest start location as used in generating a cost surface. The computation is based on a cost surface file that you generate with Cost Surface.

One must specify the start location by row and column. The zeros in the input cost surface represent one endpoint. The specified start location represents the other endpoint.

Testing the values of neighboring cells for the smallest value generates the path. When the smallest value is found, the path moves to that location, where it repeats the process to move the next cell. The output is the path of least resistance between two points, with the least expensive, but not necessarily the straightest, line between two endpoints. The output file consists of only the output path attribute value, which can be optionally specified, surrounded by void values.

Performing A Proximity Search: Proximity lets you search a grid file for all the occurrences of a cell value or a feature within either a specified distance or a specified number of cells from the origin.

You can set both the origin and the target to a single value or a set of values. The number of cells to find can also be limited. For example, if you specify to find 10 cells, the search stops when 10 occurrences of the cell have been found within the specified distance of each origin value. If you do not limit the number of cells, the search continues until all target values are located.

The output grid file has the user-type code and the data-type code of the input file. The gird-cell values in the output file indicate whether the grid cell corresponds to an origin value, the value searched for and located within the specified target, or neither of these.

The origin and target values may be retained as the original values or specified to be another value.

GRID BASED SPATIAL ANALYSIS

Diffusion modeling and *Connectivity analysis* can be effectively conducted from grid data. Grid analysis is suitable for these types of problems because of the grid's regular spatial configuration of geographic units.

Diffusion Modeling: It deals with the process underlying spatial distribution. The constant distance between adjacent units makes it possible to simulate the progression over geographic units at a consistent rate. Diffusion modeling has a variety of possible applications, including wildfire management, disease vector tracking, migration studies, and innovation diffusion research, among others.

Connectivity Analysis: Connectivity analysis evaluates inter separation distance, which is difficult to calculate in polygon coverage, but can be obtained much more effectively in a grid.

The connectivity of a landscape measures the degree to which surface features of a certain type are connected. Landscape connectivity is an important concern in environmental management. In some cases, effective management of natural resources requires maximum connectivity of specific features. For instance, a sufficiently large area of dense forests must be well connected to provide a habitat for some endangered species to survive. In such cases, forest management policies must be set to maintain the highest possible level to connectivity. Connectivity analysis is especially useful for natural resource and environmental management.

CONCLUSIONS

GIS is considered as a decision making tool in problem solving environment. 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.

REFERENCES

Bonhan - Carter, G.F. 1994. Geographic Information Systems for Geoscientists. Love Printing Service Ltd., Ontario, Canada.

Burrough, P.A. 1987. Principles of Geographical Information System for Land Assessment. Oxford : Clardon Press.

Chung, Chang-Jo F. and Fabbri, A.G. 1993. The representation of Geoscience Information for data integration. *Nonrenewable Resources*, Vol. 2, No. 2, Oxford Univ. Press.