SPATIAL DATA ANALYSIS OF VECTOR AND RASTER DATA

Prachi Misra Sahoo

I.A.S.R.I., New Delhi -110012

1. Introduction
The basic GIS functionality includes: Querying, Integrating and Manipulating Spatial Data. The core functionality of GIS is that it combines computer mapping functionality that handles and displays spatial data, with database management system functionality to handle attribute data. The basic tools that GIS provides to the user that are not available in other types of software packages are its basic functionality, which are as follows:

1. **Querying** both spatially and through attribute
2. **Manipulating** the spatial component of the data: for example, through changing projections, rubber sheeting to join adjacent layers of data together, and calculating basic statistics such as areas and perimeters of polygons
3. **Buffering** where all locations lying within a set distance of a feature or set of features are identified
4. **Data integration** which includes various single layer and multiple layer operations including overlay operation.

2. Querying Data
As with all database systems, one of the core parts of GIS functionality is the ability to query the data. With GIS software there are two basic forms of querying: spatial and attribute. Spatial querying asks the question ‘what is at this location’? This is often done by simply clicking on a feature and then listing its attributes. More complex spatial queries could select all the features within a box or a polygon, or ask ‘what is near to this feature”? These more complex queries often require the use of buffering or overlay techniques as described later in the chapter. Attribute querying asks the question ‘where does this occur”? If a user has a layer consisting of the locations of churches with various information about each church, an attribute query could select all the churches whose denomination is Catholic and then draw them with a certain symbol. The user could then query the database to select all Protestant churches and draw these with a different symbol to compare the patterns.

3. Manipulating and Measuring Spatial Data
Most GIS software packages come with a suite of options that allow the user to manipulate spatial data. One of the most basic of these is simply to change the projection system used. This can significantly alter the appearance of maps, particularly maps of the world, and can also make it possible to integrate data from layers that use different projection systems. In Britain, this could be used to take a variety of early maps on different projections and re-project them onto the National Grid. This would allow comparisons between different early maps, as well as with modern ones. Putting adjacent map sheets onto the same projection allows their digital representations to be joined to form a single layer. Where there are distortions to the sheets the maps will have to be rubber-sheeted (or edge-matched) to ensure that the edges of the two sheets
make a perfect join. This involves telling the software where certain key points are on the layer and where they actually should be. The entire layer will then be distorted using these references. Figure 1 shows an example of this.

![Input layers](image)

![Output layer](image)

**Figure 1. Rubber sheeting to join two layers together**

A user wants to join the two input line layers together, but due to inaccuracy in one or both of the layers the lines do not match exactly. As part of the joining process the lines on the right-hand layer are systematically distorted to allow them to join to the corresponding line on the left. This distortion is at its maximum at the end points of the line and reduces as we move away from the join. The layers can then be seamlessly joined.

Most GIS software packages will also calculate basic statistics about their spatial features. Typical examples include calculating the length of lines, the area and perimeter of polygons, and the distances between points. There are many examples of why these basic measures can be useful. These include measuring distances along a transport network, the use of areas to calculate population densities, and calculating the distances between settlements.

**4. Buffering, Thiessen Polygons and Dissolving**

There are times when rather than simply being interested in the locations of a type of feature, a user is interested in the locations within a set distance of a feature. Examples of this might include wanting to know all areas within (or outside of) 1km of a hospital, or areas within 10km of a railway line, or within 5km of an urban area. Where information of this type is required a buffering operation is used. Buffering takes a point, line, or polygon layer as input and produces a polygon layer as its output as is shown in Figure 2.
5. Data Integration

Manipulating the spatial component of a single layer of data is useful, but the full potential of GIS lies in its ability to integrate data from a variety of layers. At a basic level this merely involves combining layers on-screen to compare patterns. This might be as simple as taking a raster scan of a map and placing a vector layer over the top. The raster layer provides a spatial context for the features in the vector layer. Another option is to lay one vector layer on top of another, for example to compare the pattern of roads with the location of farms to see which farms lies near the major roads. Field boundaries might be a third layer added to this. This approach goes beyond basic mapping, as querying the underlying attribute database allows a detailed understanding of a multi-faceted study area to be developed. In this way an integrated understanding of the problem can be derived from many (possibly highly disparate) sources. In this section various single and multiple layer operation are mentioned.

5.1 Overlay Operations for Vector data

In addition to simply combining layers, querying them and comparing them, layers can be combined to produce new layers through geometric intersections. This is called overlay. Any of the three types of vector data can be overlaid with any of the others. An overlay operation combines not only the spatial data but also the attribute data. This has many potential uses. Overlay is one of the most important function of GIS. These involve combining different feature type (point, line, area) from different layers to produce a new map containing features and attributes of user interest. Thus it of three types:

i. **Point-in-polygon Overlay**

ii. **Line-in-polygon Overlay**

iii. **Polygon-in-polygon Overlay**

Figure 2. Buffers around points, lines and a polygon
5.1.1 Point-in-polygon Overlay
Overlays point coverages on an polygon coverage. It computes contained in relationship and the resulting point coverage contains new attributes (point features assumes the polygon attributes they lie within). For e.g. combining wells (point coverage) and planning districts (polygon coverage) will help in queries like how many wells are there in each district.

5.1.2 Line-in-polygon Overlay
Overlays line coverage onto a polygon coverage. The line features in the output coverage assumes the attribute of the polygon they lie within. The lines are broken at each area object boundary.

5.1.3 Polygon-in-polygon Overlay
Polygon overlay is a spatial operation which overlays one polygon coverage onto another to create a new polygon coverage. The spatial locations of each set of polygons and their polygon attributes are joined to derive new data relationships. The output coverage of such overlay operation is a polygon coverage. Polygon-in-polygon overlay is of three types: Union, Intersection, Identity.

Union The UNION procedure is equivalent to the boolean operator "OR" in which two or more data layers are overlaid to produce a combined coverage. Every polygon in the output coverage carries the attribute information of both the input and union coverage. The input and the union coverages must be a polygon coverage.

Intersect The INTERSECT procedure is equivalent to the boolean AND operation. When the two coverages are overlaid, only portion of the input coverage that falls inside the intersect coverage will remain in the output coverage. While the intersect coverage must be a polygon, the input coverage can be a line, polygon or point coverage. If the input coverage is a line coverage the output coverage will also be a line coverage.

Identity The IDENTITY overlay function everything located within the boundaries of the input coverage is collected in the output coverage. The boundary of the output coverage is identical to the input coverage. The identity procedures applies to point, line and polygon coverages. If the input coverage is a point feature the output coverage is also a point coverage.

5.2 Connectivity Analysis
Connectivity analysis is to analyze the connectivity between points, lines and areas in terms of distance, area, travel time, optimum path etc. Connectivity analysis consists of the following analyses.

i. Proximity Analysis
ii. Neighbourhood analysis
iii. Network Analysis

5.2.1 Proximity Analysis
Proximity analysis is measurement of distances from points, lines and boundaries of polygons. One of the most popular proximity analysis is based on "buffering", by which a buffer can be generated around a point, line and area with a given distance. Buffering is easier to generate for raster data than for vector data. Proximity analysis is not always based on distance but also time.

III.156
For example, proximity analysis based on access time or travel time will give the distribution of time zones indicating the time to reach a certain point.

5.2.2 Neighbourhood Analysis

Some kind of spatial associations are made in case of neighbourhood analysis. It evaluates the characteristics of an area surrounding a specified location. Neighbourhood functions include the calculation of a value representing a weighted average, maximum value, minimum value, measure of diversity or rate of change of a part of the statistical surface represented by the overlay in an area around the point, and so on. The various methods of identification of neighbours includes: Contiguity based neighbours and Distance based neighbours.

5.2.3 Network Analysis

Network analysis includes determination of optimum paths using specified decision rules. The decision rules are likely based on minimum time or minimum distance, maximum correlation occurrence or capacity, shortest path and so on.

6. Spatial Analysis of Raster Data

There are four basic functions of raster data according to which analysis of raster data is done. These are:

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

6.1 Overlay Operation in Raster Data

Overlay can also be performed on raster datasets providing they use the same pixel sizes. This is sometimes referred to as map algebra as two or more input layers are used to create an output layer whose cell values are calculated based on a mathematical operation between the input layers. An example of this is shown in Figure 9 where cell values on the two input layers are added to calculate values on the output layer. Other mathematical operations such as subtraction and multiplication can also be used.

<table>
<thead>
<tr>
<th>Input 1</th>
<th>Input 2</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1 1</td>
<td>0 0 0 2</td>
<td>1 1 1 3</td>
</tr>
<tr>
<td>1 1 1 1</td>
<td>0 1 3 2</td>
<td>1 2 4 3</td>
</tr>
<tr>
<td>2 3 3 1</td>
<td>0 1 3 2</td>
<td>2 4 6 3</td>
</tr>
<tr>
<td>2 2 2 1</td>
<td>1 1 2 2</td>
<td>3 3 4 3</td>
</tr>
<tr>
<td>2 2 2 1</td>
<td>1 1 1 2</td>
<td>3 3 3 3</td>
</tr>
</tbody>
</table>

**Figure 9. Map algebra with raster data**

In raster overlay the values of cells in the output layer is calculated from the results of a mathematical operation on the input layers. In this example the two input layers have been added. Other operations such as multiplication and subtraction can also be used. When two
layers are combined using an overlay operation, the resulting layer will be at best as accurate as the less accurate layer. Raster based overlay operation tools are:

i. **Arithmetic functions** (+, -, *, /)

ii. **Relational functions** (<, >, =)

iii. **Logical operations** (and, or, xor, not)

iv. **Conditional functions** (if, then, else)

### 6.2 Classification

Based on the number of classes before and after the classification, three types of classifications can be differentiated:

a) one to one (1:1): The number of classes before is the same as the number of classes after the classification process: there are no changes in the geometry of the spatial objects, they have been re-assigned.

b) many to one (M:1): The number of classes after the classification is smaller than the number of classes before the process: generalization, aggregation, merging

c) one to many (1:M): The number of classes after the classification process is more than the those before the classification: in vector format spatial objects are split in different objects; in raster format e.g. unique ID’s are assigned to each pixel in the output map

### 6.3 (re)Classification

(re)Classification involves the selection and presentation of a selected layer of data based on the classes or values of a specific attribute. It involves looking at an attribute, or a series of attributes, for a single data layer and classifying the data layer based on the range of values of the attribute. For examples: Reclassify a soil map into a PH map, classify an elevation map into classes with intervals of 50 m.

### 7. Conclusions

GIS software provides extensive functionality that allows a user to approach his or her dataset in a way that combines the spatial and attribute components of their data. This functionality leads to added value being extracted from an existing dataset. All datasets have limitations, and the extra functionality provided by GIS software allows us to use the data in ways that their creator would never have envisaged. As a result, it is important to consider the limitations of all layers when manipulating them with the GIS. It is also important to consider the limitations of the techniques used on the data, particularly those that integrate data together. As long as the results of spatial operations are understood within these limitations, GIS software provides new functionality that should allow new understanding to be derived from spatially referenced data. This shows the usefulness of the combined spatial and attribute data model used by GIS. This allows data to be queried and integrated in ways that no other approach can manage. The key advantage of this is that it allows the complexity of the data to be handled without undesirable simplification of the data.
References

