

Spatial Analysis

Unit: V

Semester: II

Paper Code: GIS 07

Name of Paper: Geographic Information System

PG Diploma in RS & GIS

Dr. SHYAMA PRASAD MUKHERJEE UNIVERSITY, RANCHI

Spatial Analysis

4.1 What is Spatial Analysis?

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

Spatial analysis is done to answer questions about the real world including the present situation of specific areas and features, the change in situation, the trends, the evaluation of capability or possibility using overlay technique and/or modelling and prediction. Therefore, spatial analysis ranges from simple arithmetic and logical operation to complicated model analysis.

Spatial analysis is categorized as follows.

Query: retrieval of attribute data without altering the existing data by means of arithmetic and logical operations.

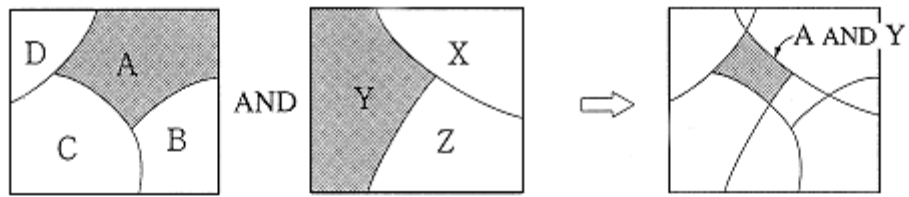
Reclassification: reclassification of attribute data by dissolving a part of the boundaries and merging into new reclassified polygons.

Coverage Rebuilding: rebuilding of the spatial data and the topology by "update", "erase", "clip", "split", "join" or "append".

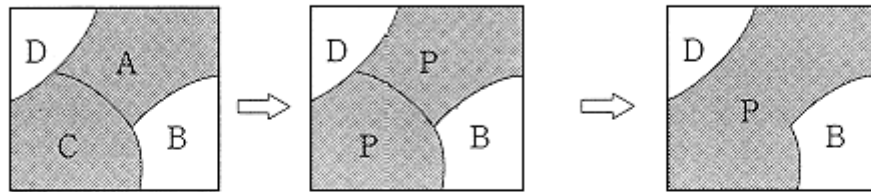
Overlay: Overlaying of more than two layers, including rebuilding topology of the merged points, lines and polygons and operations on the merged attributes for suitability study, risk management and potential evaluation.

Connectivity Analysis: analysis of connectivity between points, lines and polygon in terms of distance, area, travel time, optimum paths etc. Proximity analysis by buffering, seek analysis of optimum paths, network analysis, etc. are included.

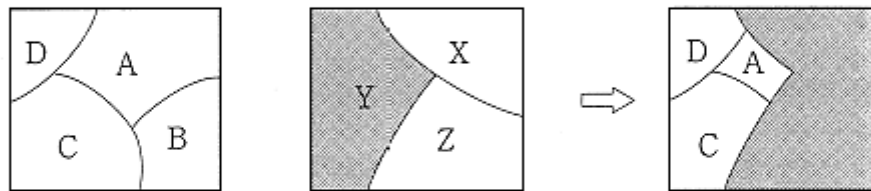
[Figure 4.1](#) shows examples of spatial analysis



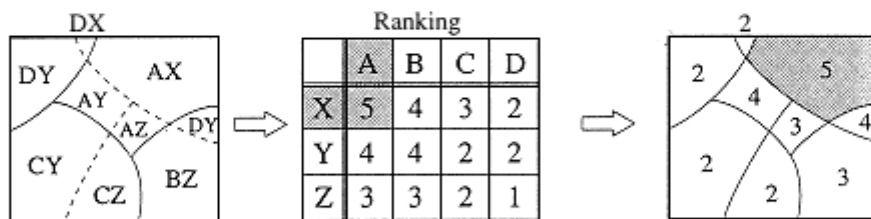
(a) Query: What Area area both A and Y?



(b) Reclassification: Reclassify A and C into P.



(c) Coverage rebuilding: Clip onto Y zone and erase others.



(d) Overlay: What are ranks of overlaid areas?



(e) Connectivity: Where are the buffer zone from boundaries?

Figure 4.1 Examples of Spatial Analysis

4.2 Query

Query is to retrieve the attribute data without altering the existing data according to specifications given by the operator.

The specifications include the following three items, given usually in Standard Query Language (SQL).

SELECT: attribute name (s)

FROM: table

WHERE: condition statement

The conditional statement is represented by the following three types of operator.

relational: $>$, $<$, $=$, \geq , \leq

Arithmetic: $+$, $-$, \times , \div

Boolean (logical): AND, OR, NOT, XOR (exclusive OR)

- ✓ The Boolean operators are used to combine more than two conditions as shown in [Figure 4.2](#).
- ✓ The Boolean operators are based on 0 and 1; 0 if the attributes do not meet the condition and 1 if they do as shown in [Figure 4.2](#).

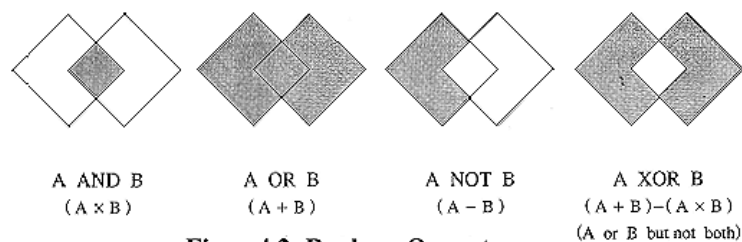


Figure 4.2 Boolean Operators

- ✓ The Boolean operators are based on 0 and 1; 0 if the attribute do not meet the condition and 1 if they do as shown in [Figure 4.3](#).

AND: multiply

OR: add (2 is reclassified to 1)

NOT: subtract (-1 is reclassified to 0)

XOR: (add)- (multiply)

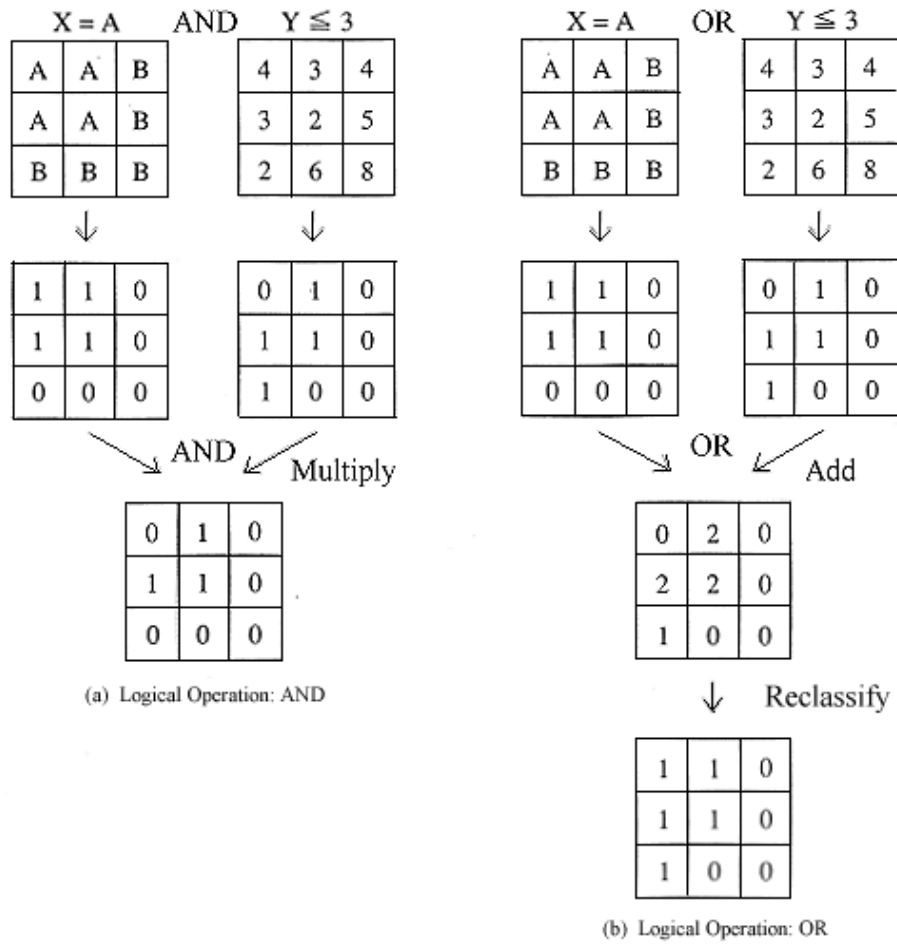


Figure 4.3 Logical Operators using Boolean Operators

4.3 Reclassification

Reclassification is to reassign new thematic values or codes to units of spatial feature, which will result in merging polygons.

A set of "reclassify attributes", "dissolve the boundaries" and "merge the polygons" are used frequently in aggregating area objects, as already shown in [Figure 4.1 \(b\)](#).

Reclassification is executed in the following cases.

Generalization: reassignment of existing data into smaller number of classes. Generalization will result in a reduction of the level of detail.

Ranking: valuation of attributes based on an evaluation model or table specified by

Reselection :selection of features to be kept and removal of unselected features.

[Figure 4.4](#) shows examples of the above three cases.

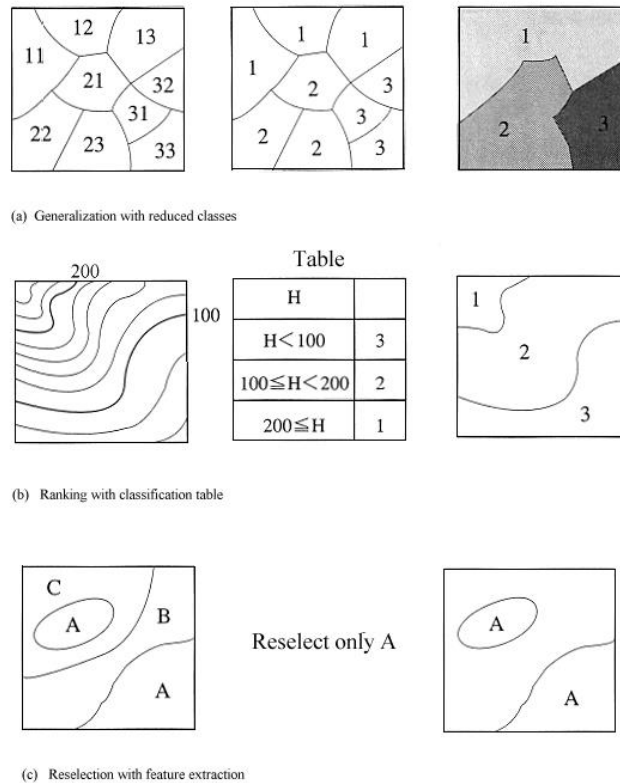


Figure 4.4 Examples of Reclassification

4.4 Coverage Rebuilding

Coverage rebuilding is a boundary operation to create new coverages that are identified and selected by users.

Boundary operations include the following six commands.

- **Clip:** to identify and preserve features within the boundary of interest specified by users. It is called a "cookie cutter".
- **Erase:** to erase features inside the boundary while preserving features outside the boundary.
- **Update:** to replace features within the boundary by cutting out the current polygons and pasting in the updated polygons.

- **Split:** to create new coverages by clipping geographic features with divided borders.
- **Append:** to merge the same feature classes of points and lines from the adjacent coverages.
- **Map Join:** to join the adjacent polygon features into a single coverage and to rebuild to topology. It is called mosaicking.

[Figure 4.5](#) shows the concept of coverage rebuilding.

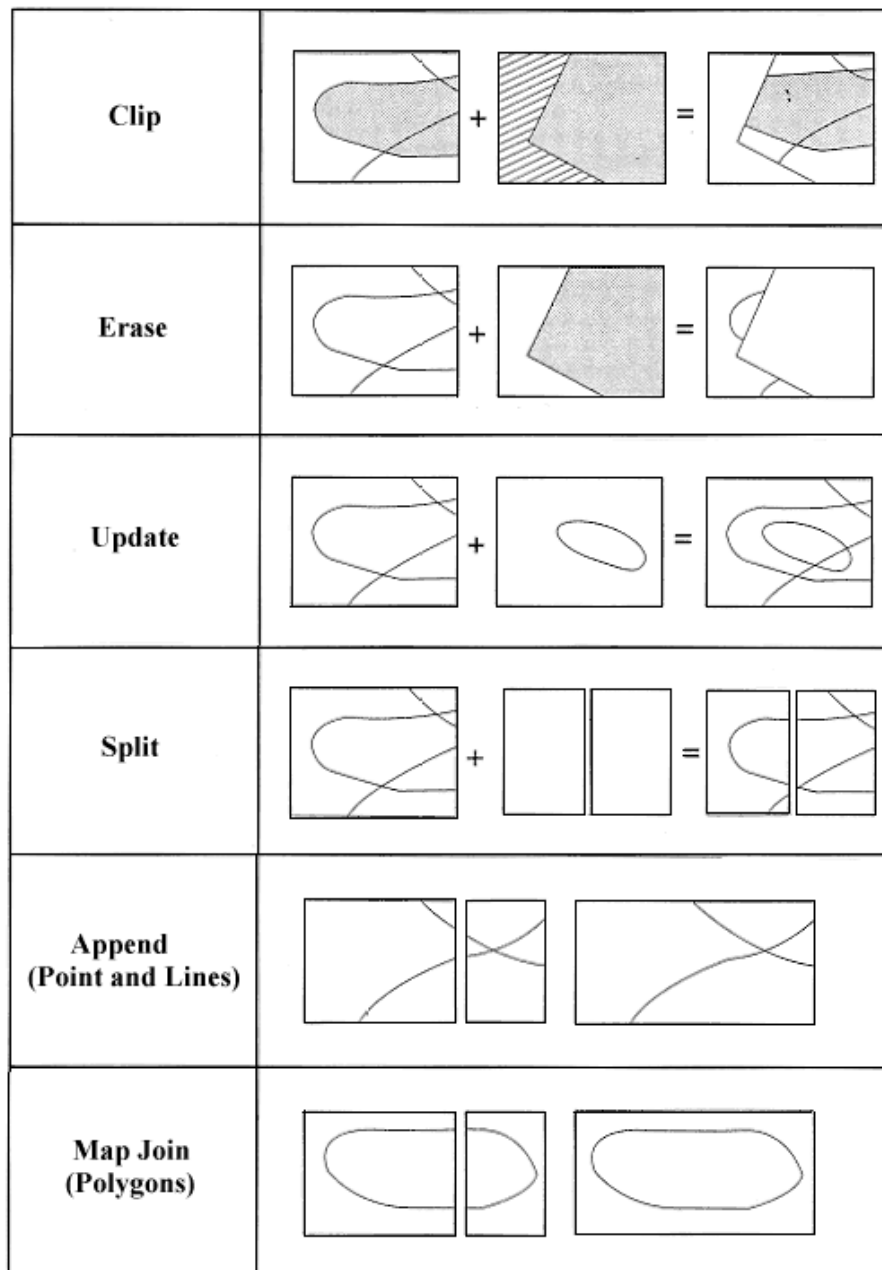


Figure 4.5 Concept of Coverage Rebuilding

4.5 Overlay of Raster Data

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.

Generally, there are two methods of raster-based overlay.

Weighting point method: basically, two layers with the values of $P1$ and $P2$ respectively are overlaid with the weight of $w1$ and $w2$ respectively as follows.

$$P = w1 P1 + w2 P2$$

where $w1 + w2 = 1.0$

The weighting point method is only available when the attributes have numerical values which can be operated arithmetically.

Ranking method: at first the attributes of the two layers are categorized into five ranks as excellent (5), better (4), good (3), poor (2), and bad (1) before a specific purpose of overlay. Then the two different layers of A and B are overlaid by following one of the three ranking tables as shown in [Table 4.1](#).

Table 4.1 Ranking Tables

(a) Minimum Ranking						(b) Multiplication Ranking						(c) Selective Ranking					
B \ A	5	4	3	2	1	B \ A	5	4	3	2	1	B \ A	5	4	3	2	1
5	5	4	3	2	1	5	25	20	15	10	5	5	5	4	3	2	
4	4	4	3	2	1	4	20	16	12	8	4	4	5	4	3	3	2
3	3	3	3	2	1	3	15	12	9	6	3	3	4	3	3	2	1
2	2	2	2	2	1	2	10	8	6	4	2	2	2	3	2	2	1
1	1	1	1	1	1	1	5	4	3	2	1	1	1	2	1	1	1

Minimum Ranking

Lower rank is taken as the new rank of the overlaid pixel as the safety rule.

Multiplication Ranking

Two ranks are multiplied because of more influential effect rather than additional effect.

Selective Ranking

Experts can set up combined ranks depending on professional experience.

For practical purposes, a model of overlay with many layers and the hierarchical structure should be built by users as shown in [Figure 4.6](#).

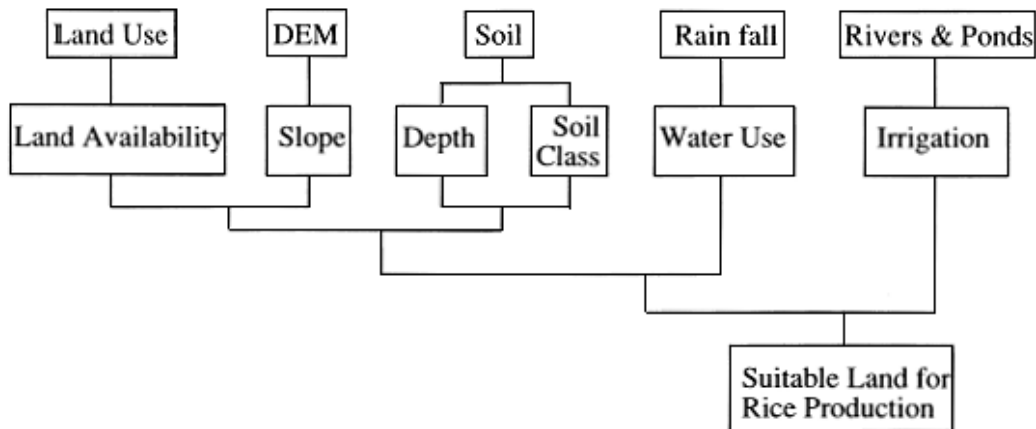


Figure 4.6 Example of a Model of Overlay for Land Suitability Study of Rice Production

4.6 Overlay of Vector Data

Overlay of vector data is a little bit complicated because it must update the topological tables of spatial relationships between points, lines and polygons.

Overlay of vector data results in the creation of new line and area objects with additional intersections or nodes, that need topological overlay.

There are three types of vector overlay.

point in polygon overlay: points are overlaid on polygon map as shown in [Figure 4.7 \(a\)](#). Topology of point in polygon is "is contained in" relationship. Point topology is a new attribute of polygon for each point.

line on polygon overlay: lines are overlaid on polygon map with broken line objects as shown in [Figure 4.7 \(b\)](#). Topology of line on polygon is "is contained in" relationship. Line topology is the attribute of old line ID and containing area ID.

polygon on polygon overlay: two layers of area objectives are overlaid resulting in new polygons and intersections as shown in [Figure 4.7 \(c\)](#). The number of new polygons are usually larger than that of the original polygons. Polygon topology is a list of original polygon IDs.

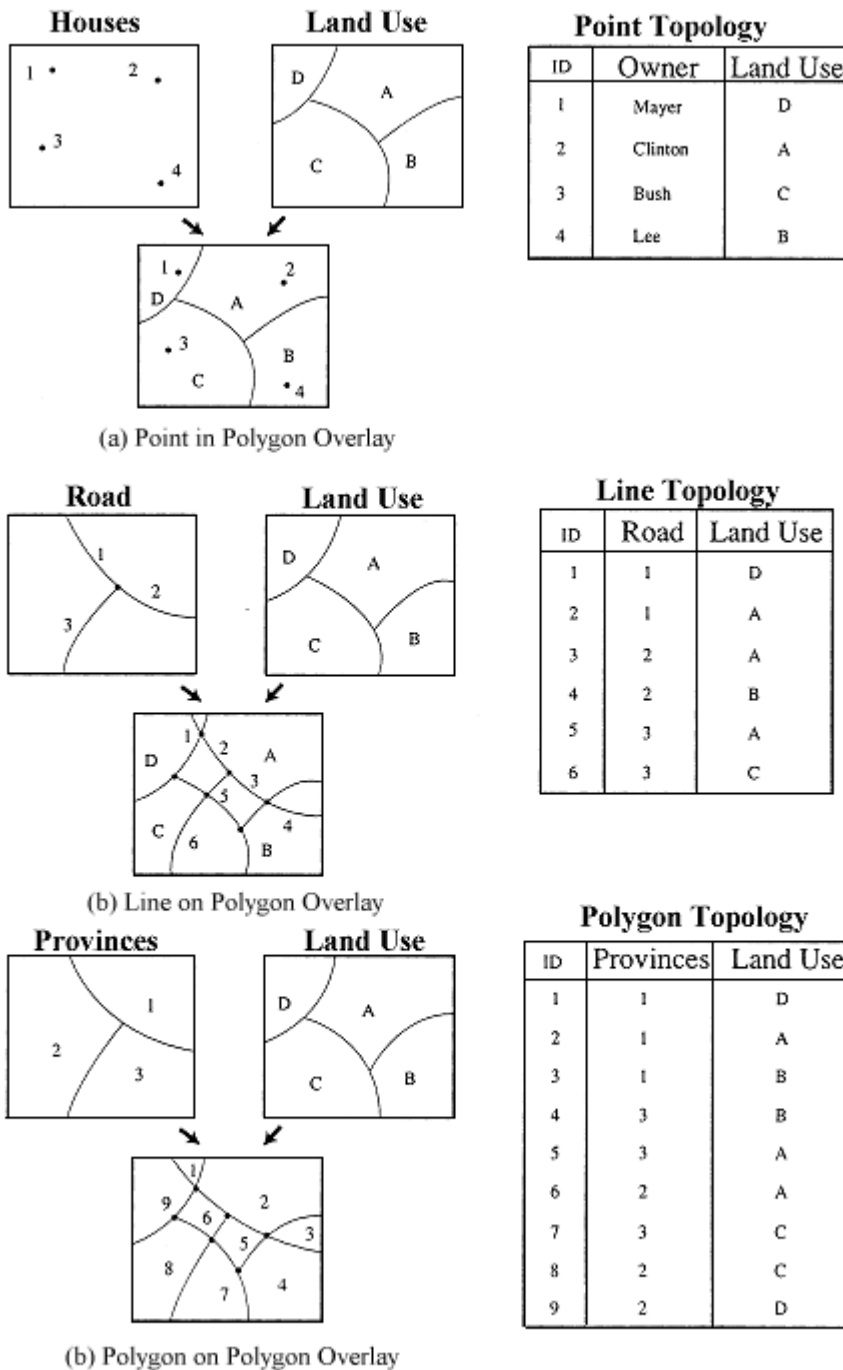


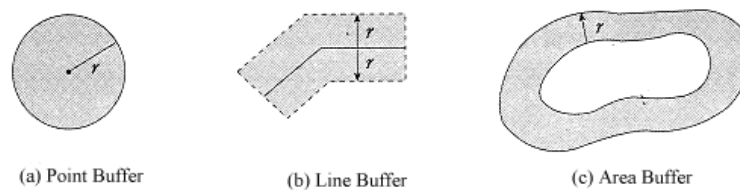
Figure 4.7 Overlay of Vector Data

4.7 Connectivity Analysis

Connectivity analysis is to analyse the connectivity between points, lines and areas in terms of distance, area, travel time, optimum path etc.

Connectivity analysis consists of the following analyses.

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 as shown in [Figure 4.8](#). Buffering is easier to generate for raster data than for vector data.



Figuring 4.8 Buffering

Proximity analysis is not always based on distance but also time. 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.

[Figure 4.9](#) shows walking distance in time (contour lines of every 10 minutes) to the railway station.

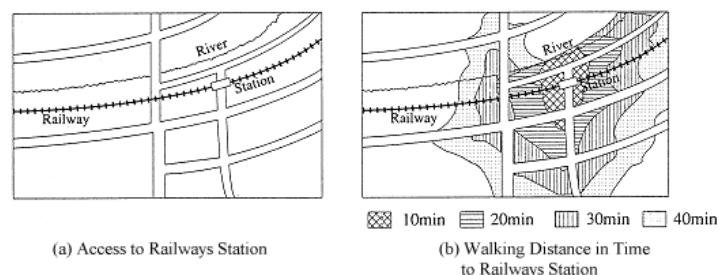


Figure 4.9 Proximity Analysis by Time

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

[Figure 4.10](#) shows two examples of optimum paths based on minimum distance and time respectively.

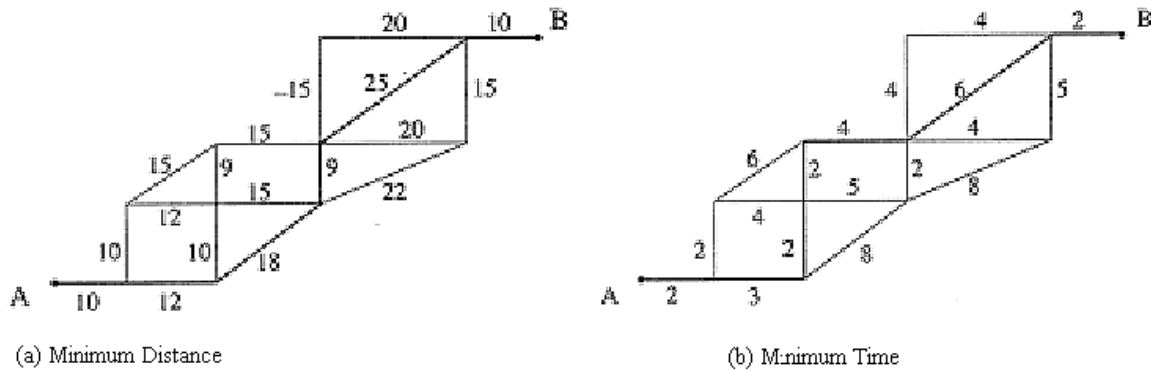


Figure 4.10 Optimum paths by Network Analysis

4.8 Shape Analysis and Measurement

Shape analysis and measurement are very important to analyses the shape of area objects in GIS.

The following parameters are computed from vector data.

Area:
$$A = \frac{1}{2} \sum_{i=1}^n (x_{i+1} - x_i)(y_{i+1} + y_i)$$

where $x_{n+1} = x_1$, $y_{n+1} = y_1$

Perimeter:
$$L = \sum_{i=1}^n \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

Centroid (Centre of Gravity) (see [Figure 4.11](#))

$$x_c = \frac{1}{6A} \sum_{i=1}^n (y_i - y_{i+1})(x_i^2 + x_i x_{i+1} + x_{i+1}^2)$$

$$y_c = \frac{1}{6A} \sum_{i=1}^n (x_i - x_{i+1})(y_i^2 + y_i y_{i+1} + y_{i+1}^2)$$



(a) Convex Polygon (b) Concave Polygon (c) Thin Polygon

Figure 4.11 Examples of Centroid Location

where A : the area of the polygon

The above parameters of area and the centroid can be computed easily from raster data by simply summing the number of pixels of the polygon and averaging the x and y coordinates respectively.

The perimeters can be computed by counting the chain codes around the boundary of the polygon.

The following parameters indicating the shape of the polygon are also used widely in the shape analysis as shown in [Figure 4.12](#).

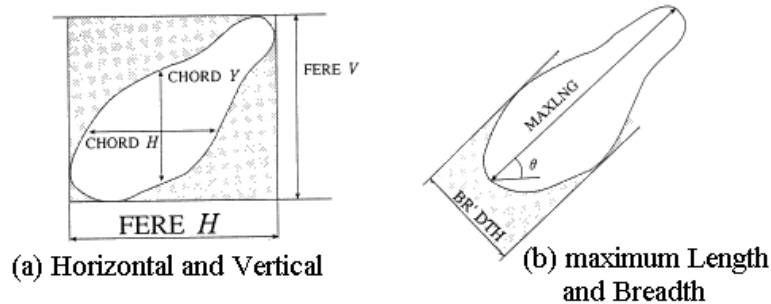


Figure 4.12 Measurement of Shape

Horizontal maximum chord: CHORD H

Vertical maximum chord: CHORD V

Horizontal Feret Diameter: FERE H

Vertical Feret Diameter: FERE V

Maximum Length: MAXLING

Breadth: BRTH

Orientation: q

The following shape factors are also used for shape analysis as shown in [Figure 4.13](#).

Roundness Factor:
$$SF_1 = \frac{(MAXLNG)^2}{A} \times \frac{\pi}{4}$$

Unevenness:
$$SF_2 = \frac{L^2}{A} \times \frac{\pi}{4}$$

Flatness:
$$SF_3 = \frac{MAXLNG}{BR'DTH}$$

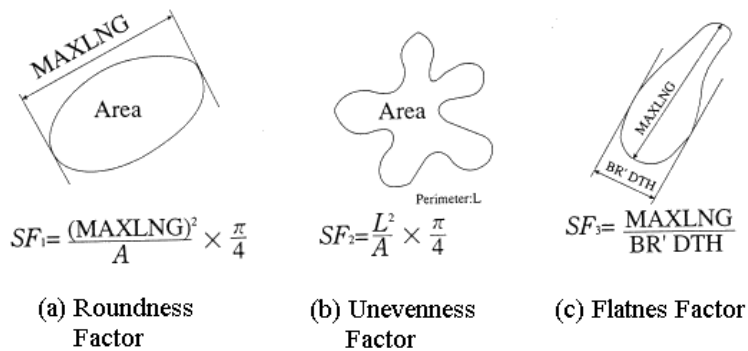


Figure 4.13 Shape Factors