# Unit Six

# Analysis of Geographic Data

**Unit objectives**

At the end of this unit, you will be able to:

- ✓ Define geographic data analysis
- ✓ Understand spatial data analytical functions
- ✓ Identify types of spatial analysis of both vector and raster data
- ✓ Outline methods of network and surface analysis

## 6.1. Introduction

Analysis of geographic data can be defined as **computing of new information** to derive new insight from the existing data, sometimes called spatial analysis. Spatial analysis is the crux of GIS, the means of adding value to geographic data, and of turning data into useful. Spatial analysis is a set of methods whose results change when locations of objects being analyzed change.

The manipulation, and analysis capabilities of vector and raster GIS are quite different. In vector, points, lines, and polygons are combined during overlays. The software determines the mathematical intersection among these feature types to produce the result. In raster data, new layers are created through the selection of particular pixel (cell) values, or application of mathematical models such as distance functions. Mathematically, multiple layers are combined using algebraic notations. There are hundreds of vector and raster data analytical functions and tools shipped in ArcGIS software. This unit presents vector and raster data analysis separately.

## 6.2 Vector data analysis in GIS

The following are vector analysis categories:

- ➢ **Measurement functions:** allow the calculation of **distances**, **lengths** and **areas**.
- ➢ **Retrieval functions:** allow for selective searches of data
- ➢ **Classifications:** allow the assignment of features to a class based on attribute values.
- ➢ **Overlay functions:** combination of two or more spatial data layers.
- ➢ **Neighborhood or proximity functions:** Evaluate the characteristics of an area surrounding feature location. E.g. buffering

### 6.2.1. Measurement functions

Measurements are numerical values that describe aspects of geographic data. Geometric measurements include length, area, or shape, and of the relationships between pairs of objects, like distance or direction or identification of locations. Distance is calculated between points, lines, and polygons using distance formula. Location are always stored by GIS or can be calculated on the fly. For example, location of one coordinate pair for a point or the centroid of a polygon. Area can be computed, but usually stored by GIS. All these are limited to geometric measurements, but not attribute data measurement, which is performed using query language.

### 6.2.2. Retrieval function

Retrieval functions involve selective search, manipulation, and output of data without modification of the spatial location. These functions include **interactive selection, spatial selection by attribute,** and **spatial selection using topological relationships**. **Spatial selection by attribute condition** uses relational operators or logical operators or combining of attribute conditions. It involves defining of a selection condition on feature attributes using a query language, such as SQL. It displays the results both on the map and in the attribute table. SQL uses atomic conditions or predicate symbols such as (Area < 4000). GIS uses the following **comparison operators** for attribute queries**:** < (less than); **=** (equals)**; <=** (less than or equal)**;** > (greater than)**; >=** (greater than or equal)**; and <>** (does not equal). In addition, GIS uses also **composite** of comparison operators with logical operators. The logical operators are **AND** (returns true if both expressions of A and B are true); **OR** (returns true if one or both of the expressions A and B is true); or **NOT** (returns true if the expression is false).

Spatial selection using topological relationships between two regions derived from the topological invariants (property) of boundary and interior intersections.

- Various forms of topological relationships used in spatial selections:
- Select features that are inside selection objects (containment relationship)
- Select features that intersect
- Select features adjacent to selection objects

- Select features base on their distance

### 6.2.3 Classification

Classification is a technique of purposefully removing detail from an input dataset in order to reveal patterns. In vector, classification will produce a new attribute. In raster, they produce new raster dataset. If the input dataset itself is a classification, the output is called reclassification. A user selects the attributes to be classified and defines the classification method that involves deciding of the number of classes, indicates value ranges to be mapped to the same category, and the correspondences between the old and new attribute values via a classification table.

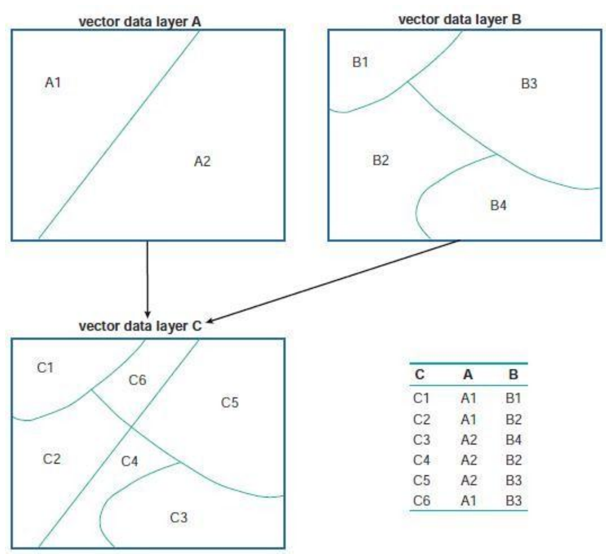
In the right table figure, for example, the original values are ranges, but classified with new nominal values

| Old value | New value |
|---|---|
| 391 - 2474 | 1 |
| 2475 - 6030 | 2 |
| 6031 - 8164 | 3 |

In vector datasets, after classification, adjacent features will have the same class, called spatial merging, aggregation or dissolve. It involves elimination of boundaries by extracting a single attribute from multiple attribute polygons.

### 6.2.4. Overlay functions

Vector overlays are based on the combination of layers and the application of Boolean, logic (union, intersection, erase, clip, etc.). Overlay operation can be arithmetical or logical. Arithmetical overlay involves addition, subtraction, division, and multiplication operations of data in each layer with values in the corresponding location of another layer. Logical overlay involves area selection where set of conditions are satisfied. Overlay operation is used to answer questions like:




| C | A | B |
|---|---|---|
| C1 | A1 | B1 |
| C2 | A1 | B2 |
| C3 | A2 | B4 |
| C4 | A2 | B2 |
| C5 | A2 | B3 |
| C6 | A1 | B3 |

What are the areas where landuse is agricultural and soil type is “A”?

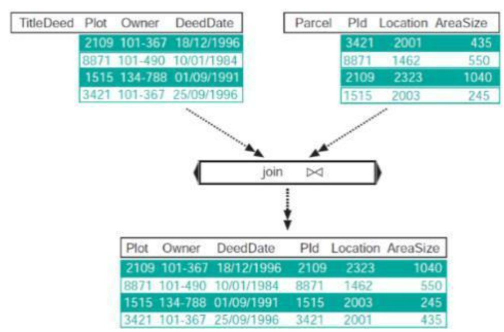
- Overlay operations have three components (output results)
  - ✓ Intersection of the geometry
  - ✓ Spatial join (join of the attribute tables)
  - ✓ Output map extent

## I. Intersection of the geometry

Intersection of the geometry generate the intersection points between two pairs of line segments. These lines are the boundaries of regions that ensure topological correctness of the dataset and identify the number of tuples in the attribute table.

## II. Spatial join

**Attribute join is performed b**ased on a common attribute: On one spatial and one non- spatial layer (Parcel is spatial and TitleDeed is none spatial). In this case, the number of output tuples is equal to the number of input tuples.

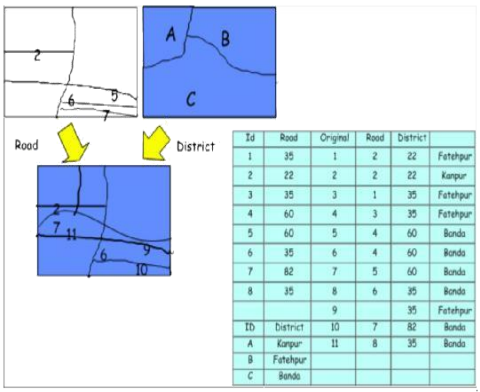



```
SELECT    *
FROM      TitleDeed, Parcel
WHERE     TitleDeed.Plot = Parcel.Id
```

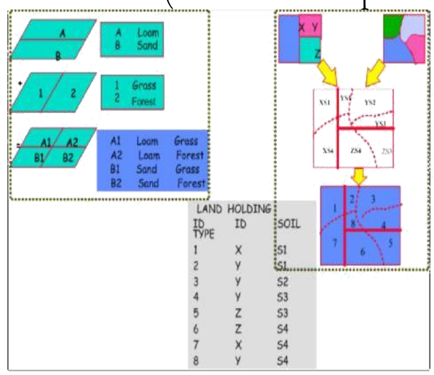
However, in the case of spatial join, a common attribute field is not needed; and the number of output tuples is larger than the number of tuples of the input layers. **Spatial join** transfers the attributes from one feature class to another feature class, based on the spatial relationship of the two feature classes. If two features share the same location, the attributes will be joined.

**The following illustrations show more on spatial join of features.**

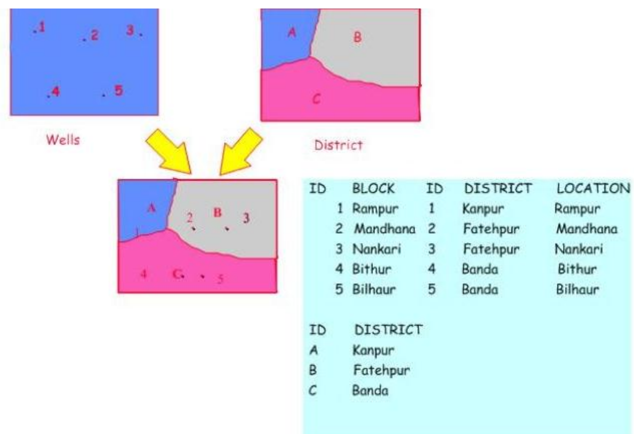
A) Polygon overlay generates new layer by merging/combining two polygons layers.




B) Line on polygon overlay operation finds polygon (of one input theme) in which line features (from second input theme) passes.

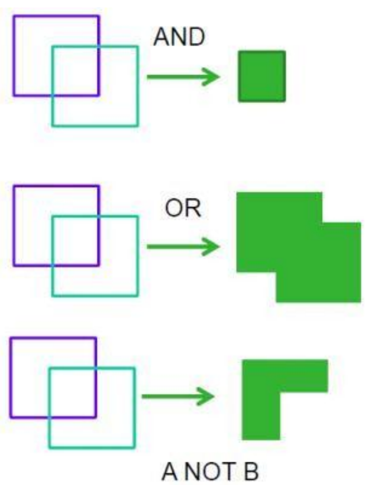



C) Point in polygon overlay finds information on polygons of one theme within which point features of other theme fall.

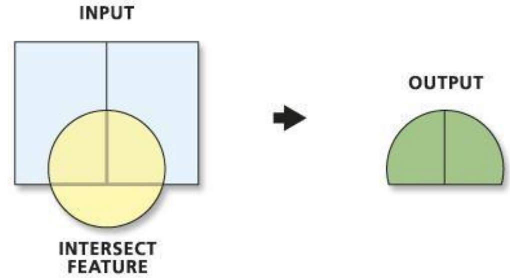
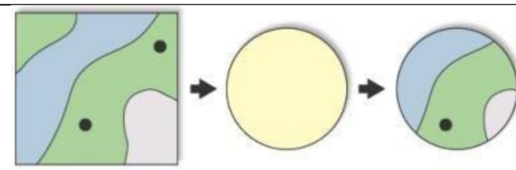
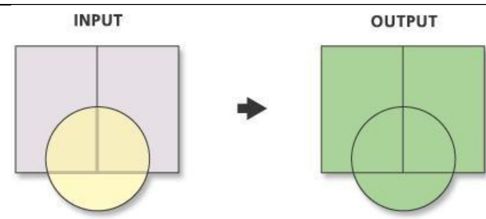
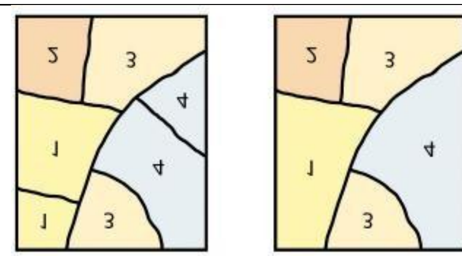
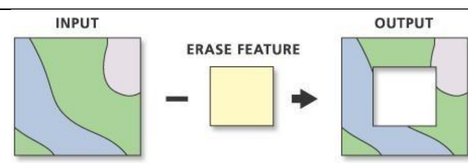



### III. Output map extent

- Two datasets do not cover exactly the same map extent in their size and location.
- Different overlay techniques may conduct the same overlay technique with different output extents.
- For example, Intersect – union, or Clip – Erase. We can relate these overlays to the Boolean operators.



GIS software includes these and more of the following overlay geoprocessing techniques:

| **Tool** | **Description** | **Illustration** |
|---|---|---|
| **Intersect** | Computes a geometric intersection of the input features. The method uses the AND operator. The tool accepts a point, line, or polygon input. Features or portions of features which overlap in all feature classes will be **written to the output feature class.** |   |
| **Clip** | Extract features from an input point, line, or polygon layer that falls within the spatial extent of the clip layer |  |
| **Union** | Computes a geometric intersection of the Input Features. The union overlay method employs the OR operator. It preserves all features, attribute information, and spatial extents from both input layers. |   |
| **Disolve** | The dissolve operation combines adjacent polygon features in a single feature dataset based on a single predetermined attribute. |  |
| **Erase** | The erase tool is opposite to clip tool. The tool preserves only those portions of the input features falling outside the erase features to the output feature class |   |

### 6.2.5. Neighbourhood or proximity functions

These functions use geometric distance to define the neighborhood of one or more target locations. They find out and evaluate the characteristics of an area surrounding feature locations, and create new data of the vicinity location. Buffer zone and Thiessen polygon computations are well-known neighborhood (proximity) operations. To perform proximity analysis, we must:

- State target locations of interest, and their spatial extent.
- Define how to determine the neighborhood for each target.
- Define which characteristic(s) must be computed for each neighborhood.

**Buffer zone generation** creates a spatial envelope (buffer) around a given feature or features. The principle is that we first select one or more target locations, and then determine the area around them within a certain distance. Created buffers may have a fixed or a variable width, depending on the characteristics of the area; and they become polygons. Figure 6-1 (a) shows selected main and minor roads as targets, and computed buffers from them at 75m and 25m respectively. In some cases, zonated buffers are employed, for example, in assessments of street traffic noise effects. Figure 6-1b shows zonated buffers generated at 100, 200 and 300 meters.

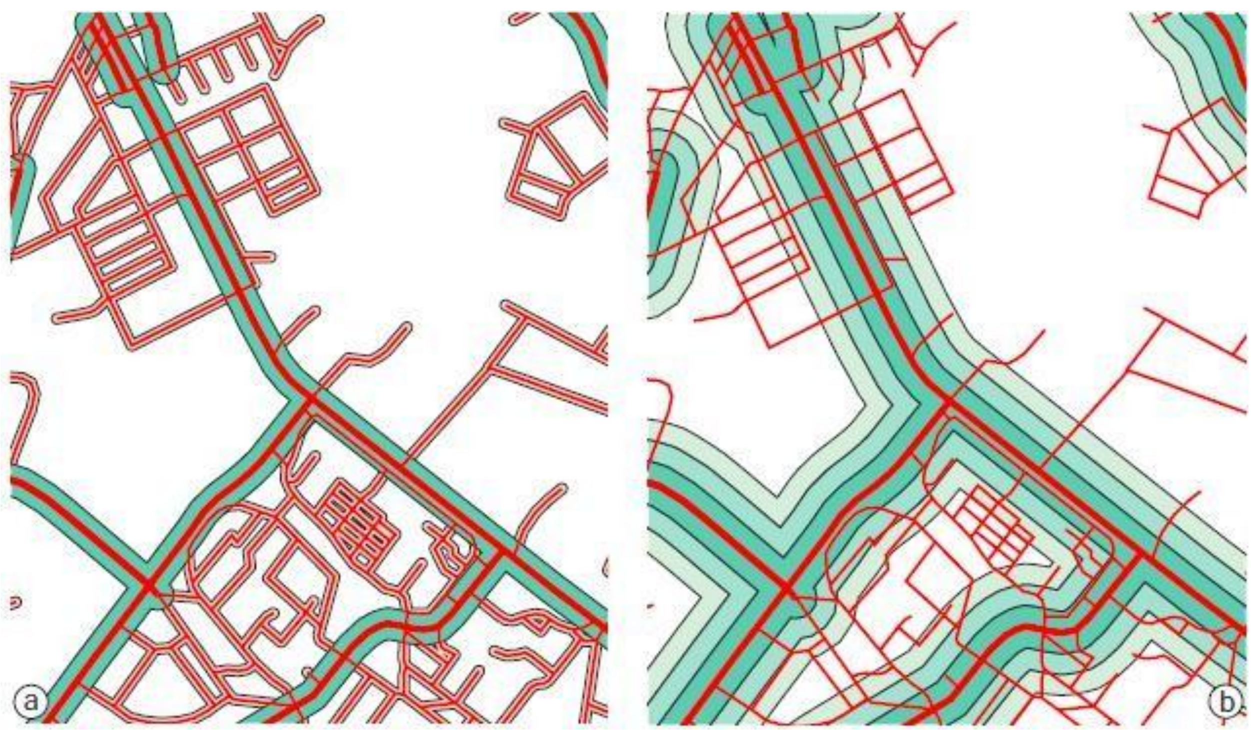
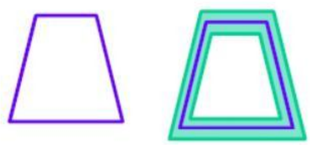


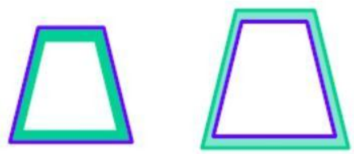

Figure 6-1: Buffer zone generation: (a) around main roads at 75m and around minor roads at 25m; and (b) Zonated buffers generated for three zones: at 100m, at 200m, and at 300 m of the main roads.

Buffering is conceptually simple but computationally difficult. GIS computes buffers for point, line, and polygon features taking in to account of the nods. Polygons border buffer (multiple buffers). There are three types of buffers:

1) **Symmetric buffers**–buffers that go both inside and outside of features
2) **Asymmetric buffers**–buffers only go either inside or outside not both
3) **Inner buffer**–all parts are inside a polygon and buffer distance is away from the boundary.



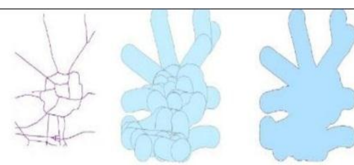
Symmetric buffers



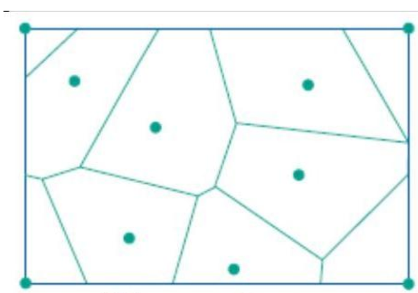
Asymmetric buffers (outside)

Inner buffer

Buffers may be merged or unmerged based on based on attribute values in order to avoid unnecessary features.



Another technique of neighborhood function is Thiessen polygon. Thiessen polygon partitions make use of Euclidian geometric distance to determine neighborhoods. It is useful if we have a spatially distributed set of points as target locations to know the closest target for each location in the study. This technique will generate a polygon around each target location that identifies all those locations that 'belong to' that target. Example is school catchment areas.



Thiessen polygons

**6.3. Raster data analysis in GIS**

**Raster analysis** performs computations on cell-by-cell basis. GIS has a language called **map algebra** that allows computing of new rasters from existing ones, using a range of functions and operators. When producing a new raster, we must provide a name and define how it is computed. In GIS, there are four classes of cell-based raster operations:

- **Local operations** - those that work on single cell locations.
- **Focal operations** - those that work on cell locations within a neighborhood.
- **Zonal operations** - those that work on cell locations within zones.
- **Global operations** - those that work on all cells within the raster.

### 6.3.1. Local operations

Local operations perform calculations **on a single cell** at a time. It can be applied to one or more raster layer with the condition that the surrounding cells do not affect the calculation. An example is multiplying each cell by 25.4 to convert rainfall values from inches to millimeters (Figure 6-2). Another technique of local operations is **reclassification**. It is used to reassign cell values in the input raster to create a new output raster, according to predefined rules (Figure 6-3). Reclassification is widely used to change NoData values (gaps in cell values) to something else.

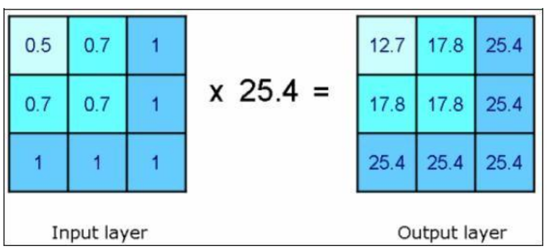
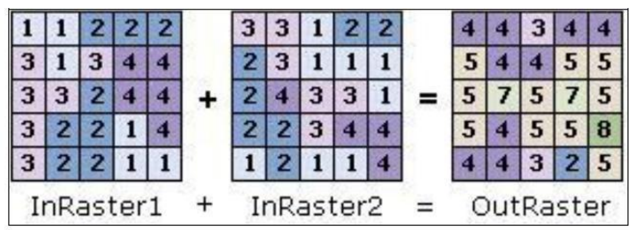



Figure 6-2: Local operations on a single layer




Local Operator on two layers: Sum

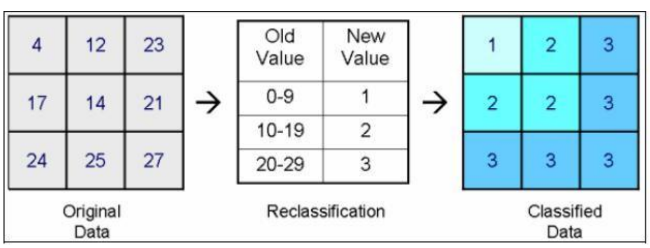



Figure 6-3: Principles of reclassification

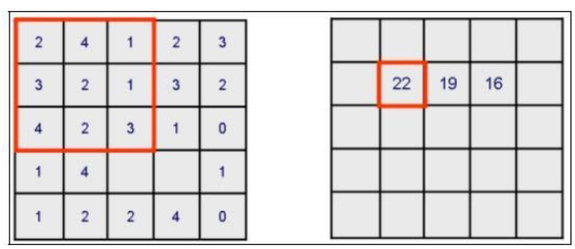


Figure 7-4: Focal operations: Sum

### 6.3.2. Focal operations

Focal operations perform calculations on a **single cell** and **its neighboring cells,** often called **local neighborhood functions.** The principle is to look at the target cell and the surrounding cells and then calculate the total value for the cell in question using a rectangular neighborhood configuration. Usually, the neighborhood size is 3x3 rectangle, **the center = focal cell.** An example is illustrated by the operation of summation (Figure 6-4).

### 6.3.3. Zonal operations

Zonal operations compute new value for each zone with a common value. The specified layers associated with are not just with that location itself but with all locations that occur in a specified zone, and on another specified layer with a common value (most frequent value). The operation works with single or two raster layers. In the Figure shown below, the mean of the input value computed for each zone; and assigned to all cells in that zone.

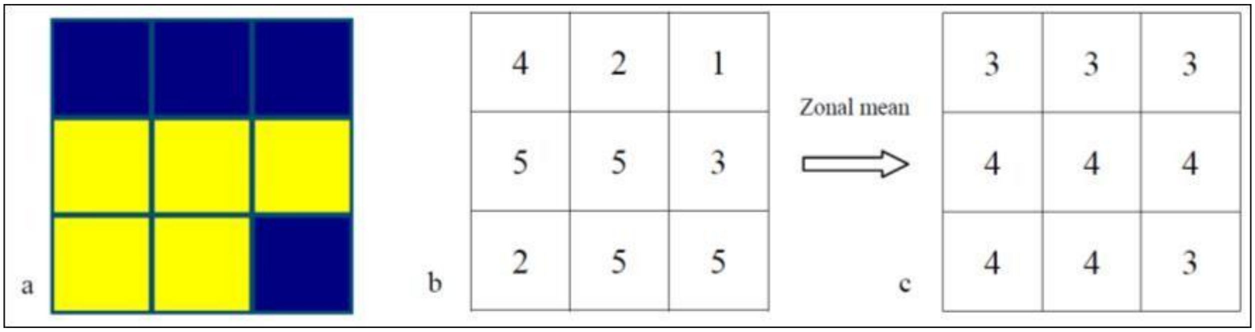



Figure 6-5: Zonal operations: Mean

### 6.3.4. Global operations

Global operations perform calculations on a raster as a whole. The output value is potentially a function of all cells in the input datasets of each location. The most common global operations are Euclidian Distance, Density, Surface Analysis, and Interpolation.

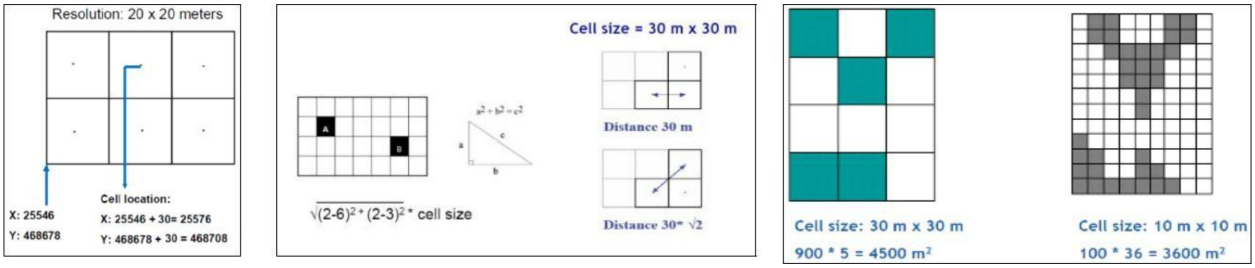
### 6.4. Types of raster data analysis

Types of raster data analysis can be grouped as follows:

1. Measurements, classification, and overlays
2. Neighborhood, and raster based surface analysis

### 6.4.1. Raster measurement

Raster measurements include location, distance, and area. Location of a single cell is derived from anchor point and resolution. A cell's location can be defined by the coordinates of its lower left corner or its midpoint (Fig.6-6, left). Distance measurements of a raster is based on the Euclidean distance between two cells is the distance between their mid points (Fig. 6-6, middle). Area size of a raster dataset is calculated as the number of cells multiplied by the cell size of the raster. The number of cells is also called the frequency or count (Fig.6-6, right). However, the area depends on the raster size. A small cell size raster gives better area measurement results.




**Location** **Distance** **Area**

Figure 6-6: Raster measurements

### 6.4.2. Classification

Classifications remove details to reveal patterns. In raster, classification will produce a new raster layer. If the input dataset is already a classified dataset, it is called reclassification. There are two methods of classification: user controlled classification and **automatic classification.**

User controlled (Re) Classification

In user-controlled classification we indicate the classification attribute ("old" and "new"). This is normally done via a classification table. The right table shows the original old values classified as new (nominal) values.

| Code | Old value | New value |
|---|---|---|
| 10 | Planned Residential | Residential |
| 20 | Industrial | Commercial |
| 30 | Commercial | Commercial |

**Automatic (Re) classification:** GIS software perform automatic classification, in which a user only specifies the number of classes in the output dataset. The system automatically determines the class break points. The two main techniques of determining break points in use are equal interval and equal frequency.

**Equal interval is c**alculated as (Vmax – Vmin)/n; where,

Vmax is the maximum attribute value, and

Vmin is the minimum attribute value; and n is the number of classes.

In Figure 6-7, it is calculated as: (10-1) / 5 ≈ 2, and each class "contains" two values.

**Equal frequency,** also called quintile, and calculated as:

$$\frac{\textbf{Total number of features}}{\textbf{Number of classes (n)}}$$

The objective is to create categories with roughly equal number of cells. Below, the respective classification tables, with the number of cells involved.

| | | | | |
|---|---|---|---|---|
| 1 | 1 | 1 | 2 | 8 |
| 4 | 4 | 5 | 4 | 9 |
| 4 | 3 | 3 | 2 | 10 |
| 4 | 5 | 6 | 8 | 8 |
| 4 | 2 | 1 | 1 | 1 |

(a) original raster

| | | | | |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 4 |
| 2 | 2 | 3 | 2 | 5 |
| 2 | 2 | 2 | 1 | 5 |
| 2 | 3 | 3 | 4 | 4 |
| 2 | 1 | 1 | 1 | 1 |

(b) equal interval classification

| original value | new value | # cells |
|---|---|---|
| 1,2 | 1 | 9 |
| 3,4 | 2 | 8 |
| 5,6 | 3 | 3 |
| 7,8 | 4 | 3 |
| 9,10 | 5 | 2 |

| | | | | |
|---|---|---|---|---|
| 1 | 1 | 1 | 2 | 5 |
| 3 | 3 | 4 | 3 | 5 |
| 3 | 2 | 2 | 2 | 5 |
| 3 | 4 | 4 | 5 | 5 |
| 3 | 2 | 1 | 1 | 1 |

(c) equal frequency classification

| original value | new value | # cells |
|---|---|---|
| 1 | 1 | 6 |
| 2,3 | 2 | 5 |
| 4 | 3 | 6 |
| 5,6 | 4 | 3 |
| 8,9,10 | 5 | 5 |

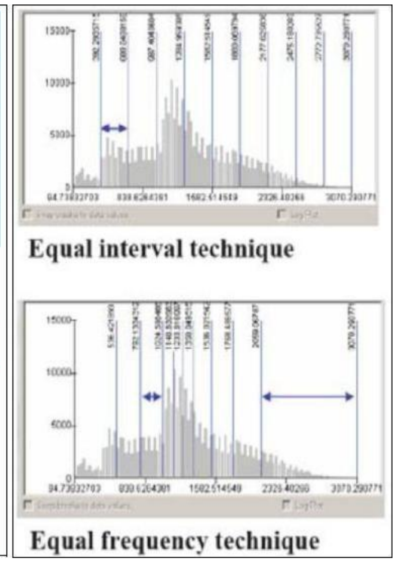



Figure 6-7: Automatic classification techniques

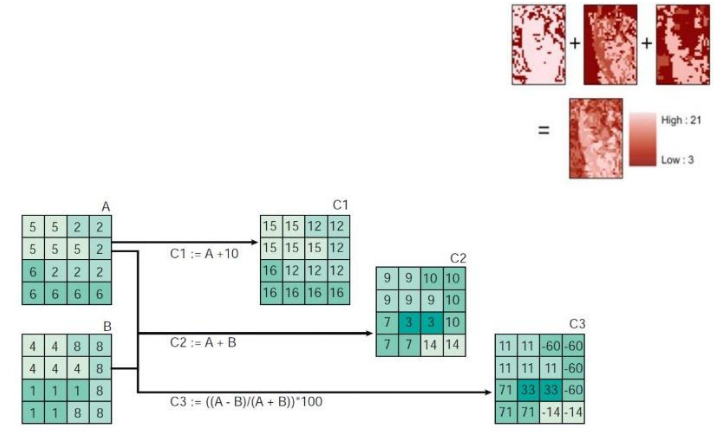
Figure 6-7 shows examples of two automatic classification techniques: (a) the original raster with cell values; (b) classification based on equal intervals; (c) classification based on equal frequencies. In Figure 6-7 (right), the horizontal axis is original value. The vertical axis is frequency. The vertical lines indicate class boundaries, and they are equally spaced, hence "Equal interval". In the lower Figure, the values in each vertical lines are nearly equal in their number, hence "Equal frequency". In this case, when the frequency goes up, the class intervals become narrower.

### 6.4.3. Overlays

In raster overlay, new cell values are calculated using **map algebra/raster calculus**. Raster overlay is performed on cell-by-cell basis; and there is no geometric calculation. **Map algebra/raster calculus** allow a GIS to compute new rasters from existing ones, using a range of functions and operators. Among them are **arithmetic; comparison, logical; conditional expressions;** and **decision tables.**

#### 6.4.3.1. Overlay – Arithmetic operators

The standard Arithmetic operators are multiplication (×), division (/), subtraction (-) and addition (+). Obviously, these arithmetic operators should only be used on appropriate data values, i.e., not on classification values. The above right figure shows three raster layers (steep slopes, soils, and vegetation) are combined using athematic operator, addition, and ranked for suitability modeling on a scale of 1 to 7. In this case, when the layers are added, each cell is ranked on a scale of 3 to 21. The following picture shows another example of arithmetic map algebra expressions.




#### 6.4.3.2. Overlay - comparison operators

Map algebra also allows the comparison of rasters cell by cell. The standard comparison operators are (<, <=, =, >=, > and <>) used to test whether one expression is larger, smaller, equal, etc. than another expression. They can be used in combination with logical operators or conditional functions. Using only a comparison operator in a formula, the formula will be a Boolean statement, e.g. MapC := MapA <> MapB.

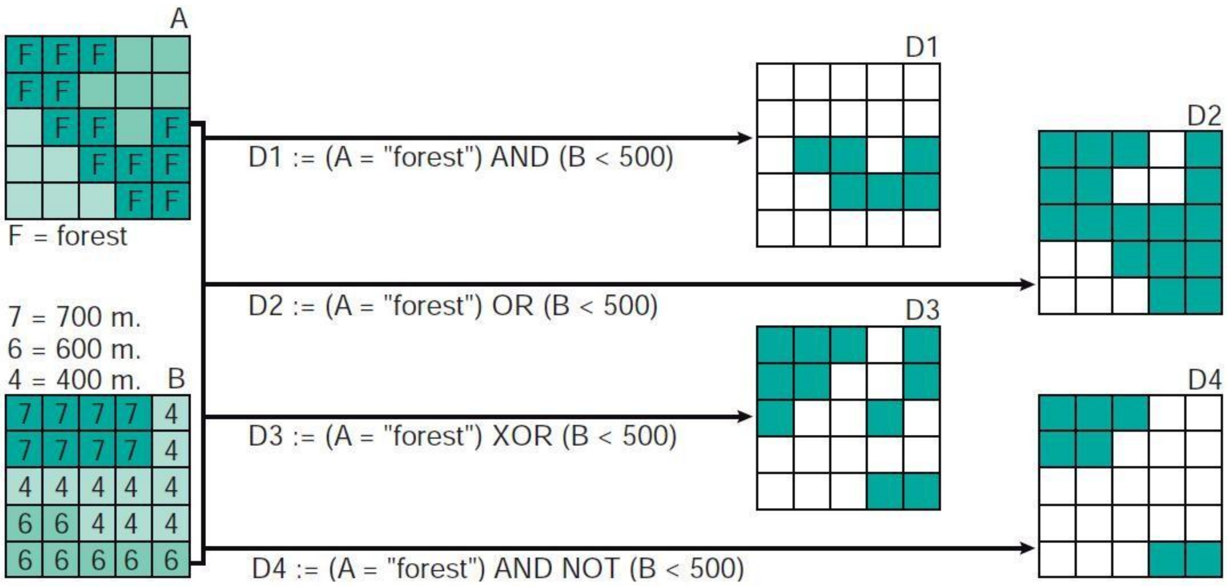

Input raster Test result Output_raster := Input Raster <15

A comparison operator will store truth values—either true (1) or false (0)—in the output raster. A cell value in the output raster will be ‘1’ if the cell’s value of the input raster is true (<15), and it will be ‘0” if the cell’s value of the input raster is false (≥15).

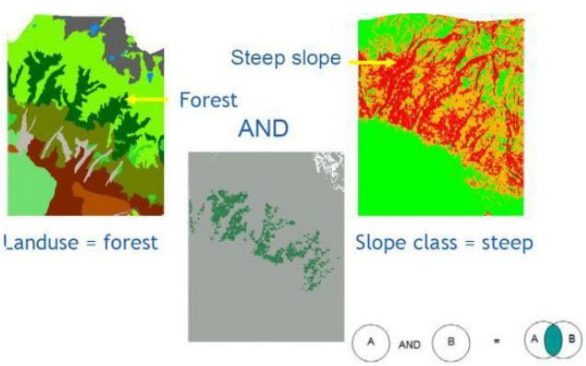
#### 6.4.3.3. Overlay – logical operators and Boolean operators

Logical operators produce rasters with the truth-values of true and false.

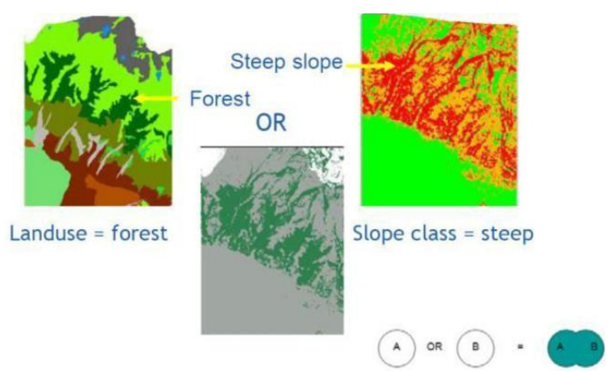
Logical operators use connectives of AND, OR and NOT in many raster overlays. Another common connective offered in map algebra is exclusive OR (XOR). The expression A XOR B is true only if either A or B is true, but not both. The Figures shown below are example of a complex logical and comparison operators combined with Boolean operators. **A** is a classified raster for land use, and **B** holds elevation values. The Figures provides various raster computations searching for forests at specific elevations. For example, rasterD1 indicates forest (F) below 500m of elevation, D2 indicates areas below 500m that are forests, raster D3 indicates areas that are either forest or below 500m (but not at the same time), and raster D4 indicates forests above 500m.



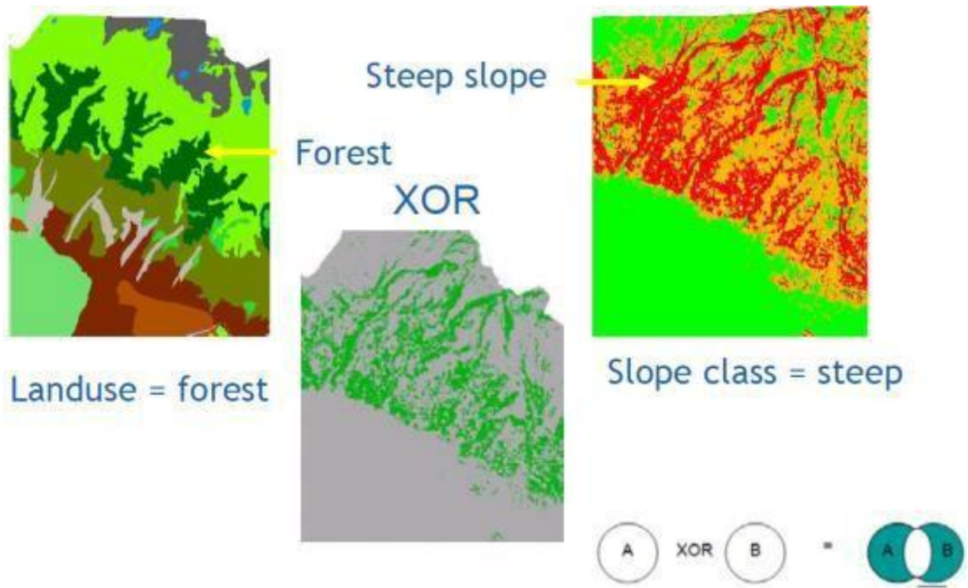

The following Figures shows how these operators work in application of land use and slope.




**Overlay – logical operator AND**



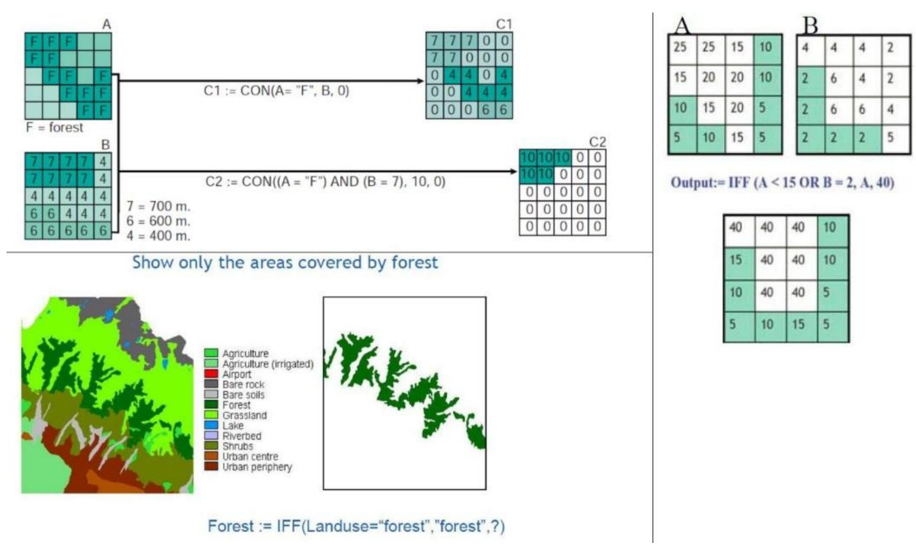

**Overlay – logical operator OR**




Logical operater_**XOR**

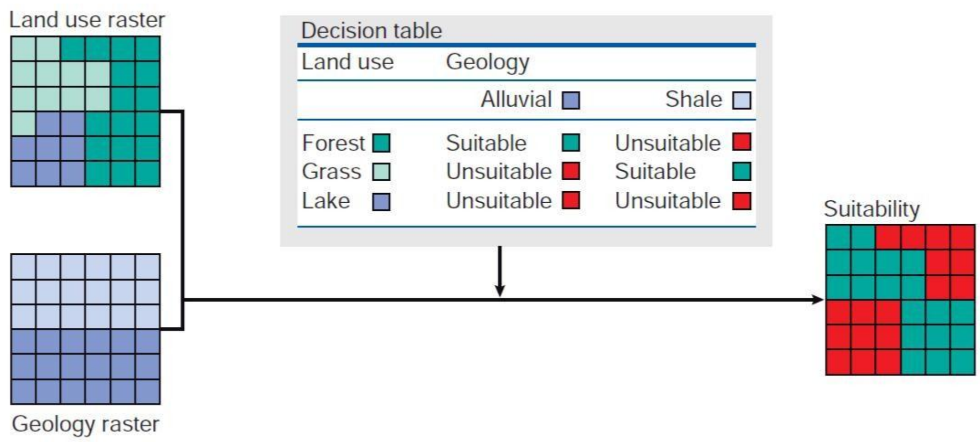
### 6.4.3.4. Overlay - conditional expressions

A conditional expression allows us to test whether a condition is fulfilled. The general format is: **Output raster := CON(condition, then expression, else expression).** If the expression is **true**, the then_expression will be evaluated**.** If the expression is **false,** the else_expression will be executed. A conditional expression can lead to many different types of outputs, numeric value but also true/false. Below is another examples of conditional expressions in map algebra. Here A is a classified raster holding **land-use** data, and B is an **elevation-value** raster.



### 6.4.3.5. Overlays using a decision table

Decision table is the same as Conditional statements, but presented in a different way. They guide the overlay process by list all possible combinations of input values, and of output values. Decision tables are used for complicated conditional expressions using multiple criteria.




The output raster of map algebra expression (CON expression) written as:

**Suitability:= CON((Landuse = “Forest” AND Geology = “Alluvial”) OR (Landuse = “Grass” AND Geology = “Shale”), “Suitable”, “Unsuitable”)**

The above table lists two input layers: Landuse and Geology. Landuse has three values (Forest, Grass, and Lake) and Geology has two values (Alluvia and Shale). For each combination, it lists also the output values. In raster overlays, decision table is used in suitability study.

**Suitability Analysis**

The general purpose to do suitability analysis is **to rank potential sites** according to suitability for a proposed type of activity. The basic requirements are a set of “**factor**” or **criteria maps**, organized to rate sites relative to one or more characteristics. Requirements are also used as a technique for appropriately combining factors

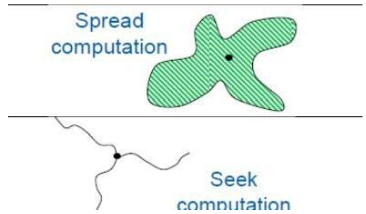
### 6.4.4. Neighbourhood or proximity functions

Neighborhood functions include buffer, spread, and flow computations.

Proximity computation makes use of the geometric distance function

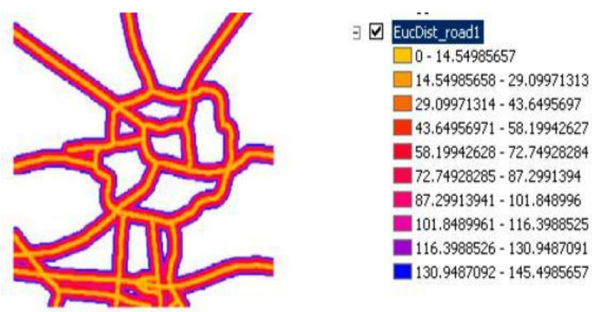
Spread computation assumes that the phenomenon spreads in all directions, but not necessarily equally easily in all direction

In Seek computation the phenomenon will choose a least-resistance path. E.g., Flow computation

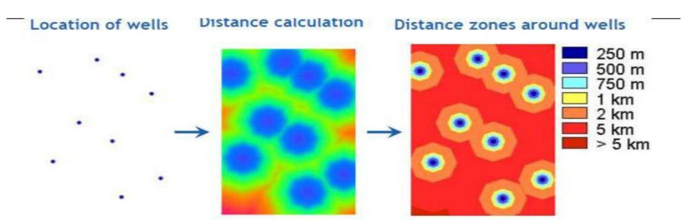



**Proximity computation - buffer**

To perform neighborhood analysis on a raster cell, we need target cell(s), and distance between the cell centers. The distance from a non-target cell to the target is the minimal distance one can find between that non-target cell and any target cell. The right figure shows buffer zone generation in raster data.



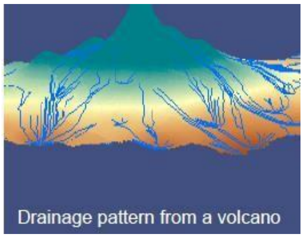

This figure shows also buffer zone generation in raster data to determine the distance at 250m, 500m, 750m, 1 km, 2 km, 5 km, and > 5km from a well. The right figure uses the roads as target.




In spread computation, the neighborhood of a target location not only depends on distance but also on **direction** and **differences** in the terrain in different directions. Target locations contain a "**source material**" that spreads over time such as water and soil contaminants, air (temperature, humidity), and people. Typical examples, where spread computation is useful for:

- Determination of inundated area due to dam construction
- Determination of flooded area due to dike burst
- Spreading of pollution
- Commuters leaving a train station

Flow computation, also called seek computation applies when a phenomenon does not spread in all directions, but chooses a least cost path. Typical example is determination of drainage patterns in a catchment



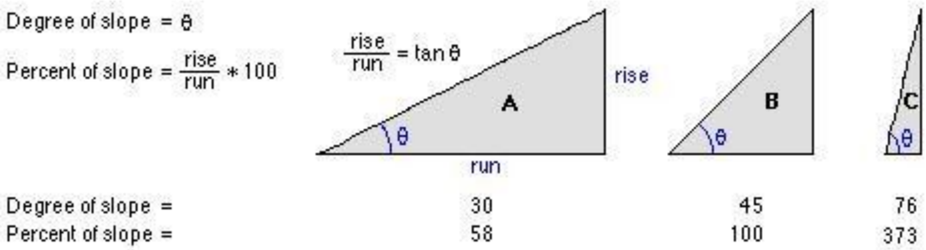

### 6.4.5. Raster based surface analysis

Continuous fields have a number of characteristics since the field values change continuously.

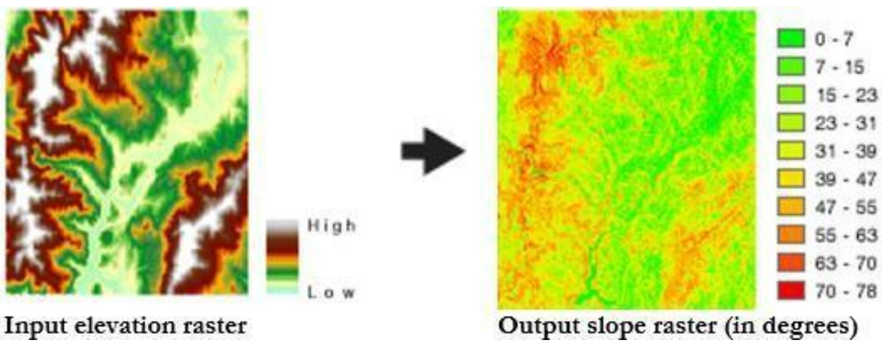
These changes are expressed, for example, in terms of slope angle, slope aspect, curvatures of the slope, viewshed, hillshade, and contour.

#### 6.4.5.1. Slope analysis

Slope is related to changes in values of a continuous field. For each cell, the Slope tool calculates the maximum rate of change in value over the distance from that cell to its neighbors. It identifies the steepest downhill descent from the cell. **This is called slope angle**. Slope angle is expressed in degrees or percent (percent rise), for any or all locations. The percent rise better understood as the rise divided by the run, multiplied by 100. Consider triangle B below.

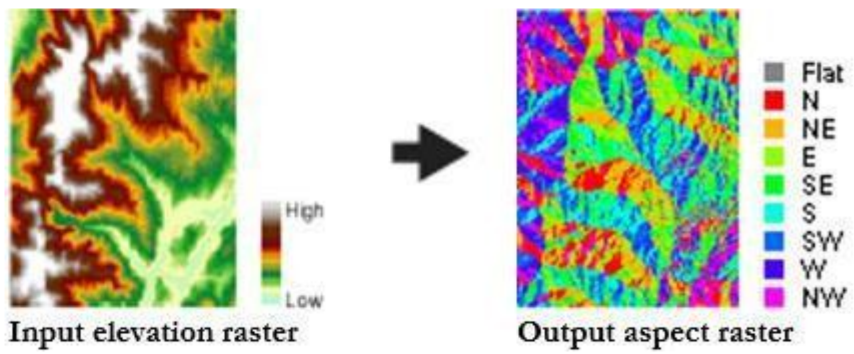



When the angle is 45 degrees, the rise is equal to the run, and the percent rise is 100 percent. As the slope angle approaches vertical ($90^o$), as in triangle *C*, the percent rise begins to approach infinity. In the following Figures, steeper slopes are shaded red on the output slope raster.



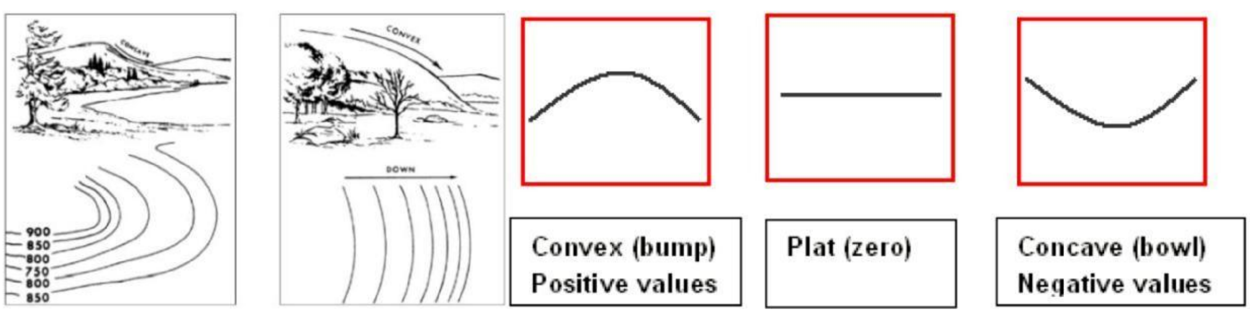
### 6.4.5.2. Slope aspect analysis

Aspect can be considered as the slope direction, which tells us toward which a slope is facing. Aspect identifies the downslope direction of the maximum rate of change in cell value from each cell to its neighbors. **Aspect calculates** the slope in degrees (between 0 and 360o), for any or all locations. The values of the output raster usually measured clockwise direction with reference to the North. The direction, also called azimuth - compass direction of the aspect.



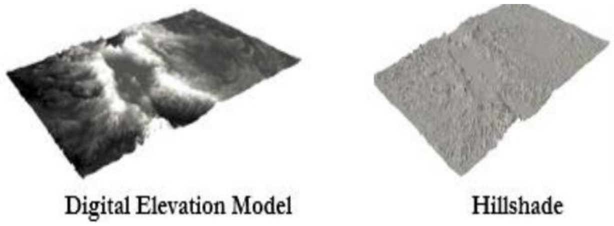

### 6.4.5.3. Curvature- convexity/concavity analysis

Curvature of the terrain shows change of slope angles. The curvature is Negative when the slope is concave and positive when the slope is convex. Curvature is used to model drainage and to understand erosion and runoff processes. When we calculate curvature, our output raster will have positive values (bump), negative values (bowls) and zeros (flat).



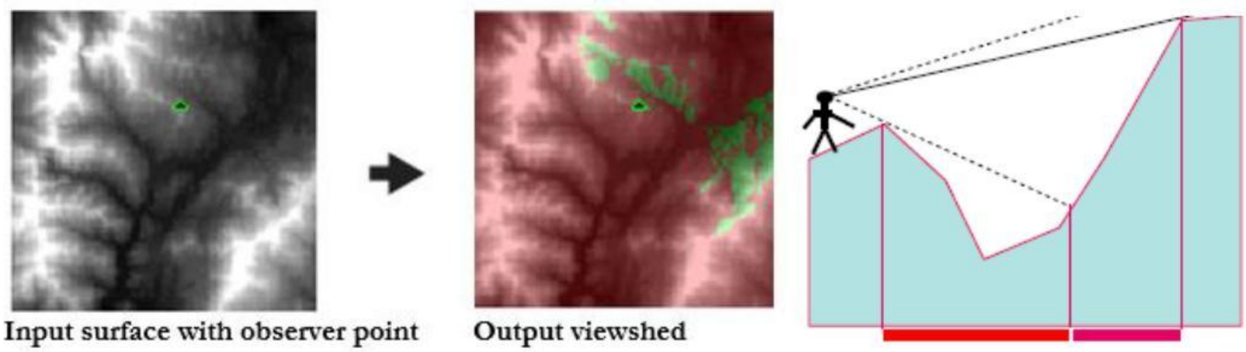

### 6.4.5.4. Hillshade analysis

Hillshading depicts the general pattern of the terrain by calculating the location of shadows and amount of sun incident illumination on the terrain surface. Hillshading is used to portray relief difference and terrain morphology in hilly and mountainous areas using DEM. The amount of reflected light at each location in raster depends on its orientation relative to the illumination source. Illumination source is usually set at $45^0$ northwest above the horizon.



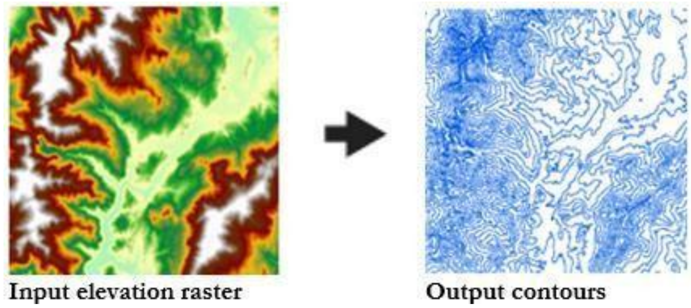

### 6.4.5.5. Viewshed or visibility analysis

A viewshed analysis determines the raster surface locations visible to a set of observer features. The observer points feature class can contain points or lines. The nodes and vertices of lines will be used as observation points. The visibility of each cell center is determined by comparing the altitude angle to the cell center with the altitude angle to the local horizon. If the point lies above the local horizon, it is considered visible. In the example below, the elevation raster displays heights of the land (left figure), and the observation tower is marked as a green triangle (right figure). Cells in green are visible, while cells in red are not from the observation tower.




### 6.4.5.6. Contour analysis

Contours are lines, often referred to as isolines that connect locations of equal value that represents continuous fields, such as elevation, temperature, or pollution. The distribution of the contour lines shows how values change across a surface. Where there is little change in a value, the lines are spaced farther apart. Where the values rise or fall rapidly, the lines are closer together. The example below shows an input elevation dataset and an output contour dataset. The areas where the contours are closer together indicate the steeper locations. They correspond to areas of higher elevation (in white on the input elevation dataset).



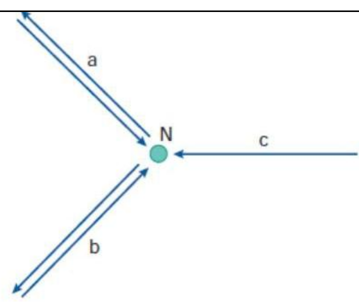

## 6.5. Network analysis

Network analysis is performed on either raster or vector data layers, but commonly done on vector. Many network analysis applications are 2D planar, such as streams or rivers, roads, pipelines, and telecommunication lines networks. Network analysis is about studying how ‘goods’ can be transported along these lines. They can be directed in which transportation is only in one direction, e.g., rivers or one-way roads; or they can be undirected in which goods can be transported in both directions (roads). GIS software supports network analysis functions.

### 6.5.1. Optimal or shortest path finding

Optimal or shortest path finding generates a least-cost path on a network between a pair of predefined locations, using both geometric and attribute data. The aim is to find a sequence of connected lines to traverse from the origin to the destination at the lowest cost. The cost function can be the total length of all lines on the path. The cost function also called Impedance.

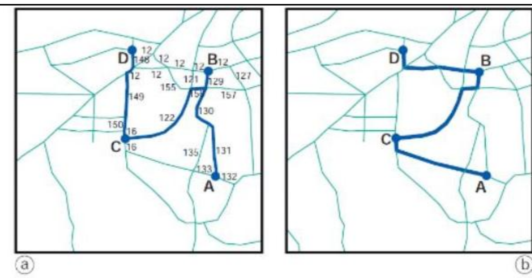
Network neighborhood of node N with associated turning costs at N. Turning at N onto c is prohibited because of direction, so no costs are mentioned for turning onto c.




| from | onto | cost |
|---|---|---|
| a | a | x |
| a | b | 5 |
| b | a | 8 |
| b | b | 15 |
| c | a | 3 |
| c | b | 6 |

Problems related to optimal path finding may require **ordered** or **unordered optimal path finding**. In ordered optimal path finding, the sequence of visiting these extra nodes is required; in unordered optimal path finding, it does not.

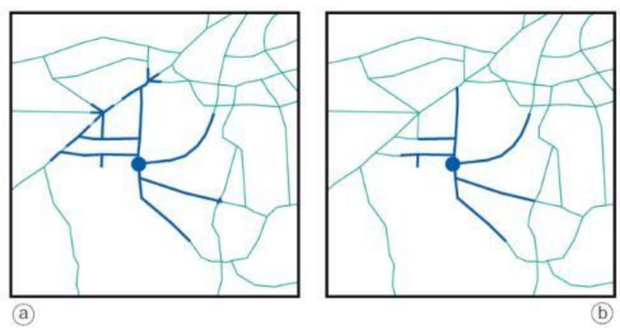
In both cases, extra nodes need to be visited along the path. Ordered (a) and unordered (b) optimal path finding. In both cases, a path had to be found from A to D, in (a) by visiting B and then C, in (b) both nodes also but in arbitrary order.




## 6.5.2. Network partitioning

Network partitioning assigns network elements (nodes or line segments) to different locations using predefined criteria, in a mutually exclusive way. Typically, the target locations play the role of service center for the network such as medical treatment, education, water supply. This type of network partitioning also known as a **network allocation problem.**

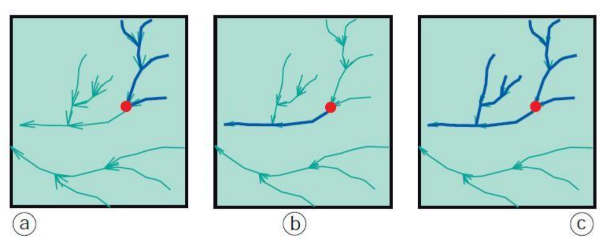
Network allocation for a school assignment problem is shown in the right Figure. In (a), the street segments within 2km of the school are identified; in (b), the selection of (a) is further restricted to accommodate the school's capacity for the new school year.




## 6.5.3. Trace analysis

Trace analysis performed to understand which part of a network is 'conditionally connected' to a chosen node, known as **trace origin** of a given target location. Trace analysis address problems of pollution along river or streamlines, and tracking of failures in energy distribution networks.

Figure (a), tracing conditions were set to trace all the way upstream; (b) traces all the way downstream; and in part
(c) there are no conditions on direction of the tracing path, thereby tracing all connected lines from the trace origin.



## 6.6 Hydrology Analysis

The objectives of hydrology analysis are:

- ✓ To explore the application of GIS in stream and watershed delineation.
- ✓ To understand how water flows in an area is useful for flood management, water use planning and basin development.
- ✓ To model the movement of water across the surface.

The hydrological tool in ArcGIS is used to extract hydrologic information from a Digital Elevation Model (DEM). The major steps to extract streams and delineate watershed includes:

- ➢ Fill DEM
- ➢ Flow direction
- ➢ Flow accumulation
- ➢ Extract streams
- ➢ Take Pour point
- ➢ Extract watershed
- ➢ Clip streams by watershed boundary