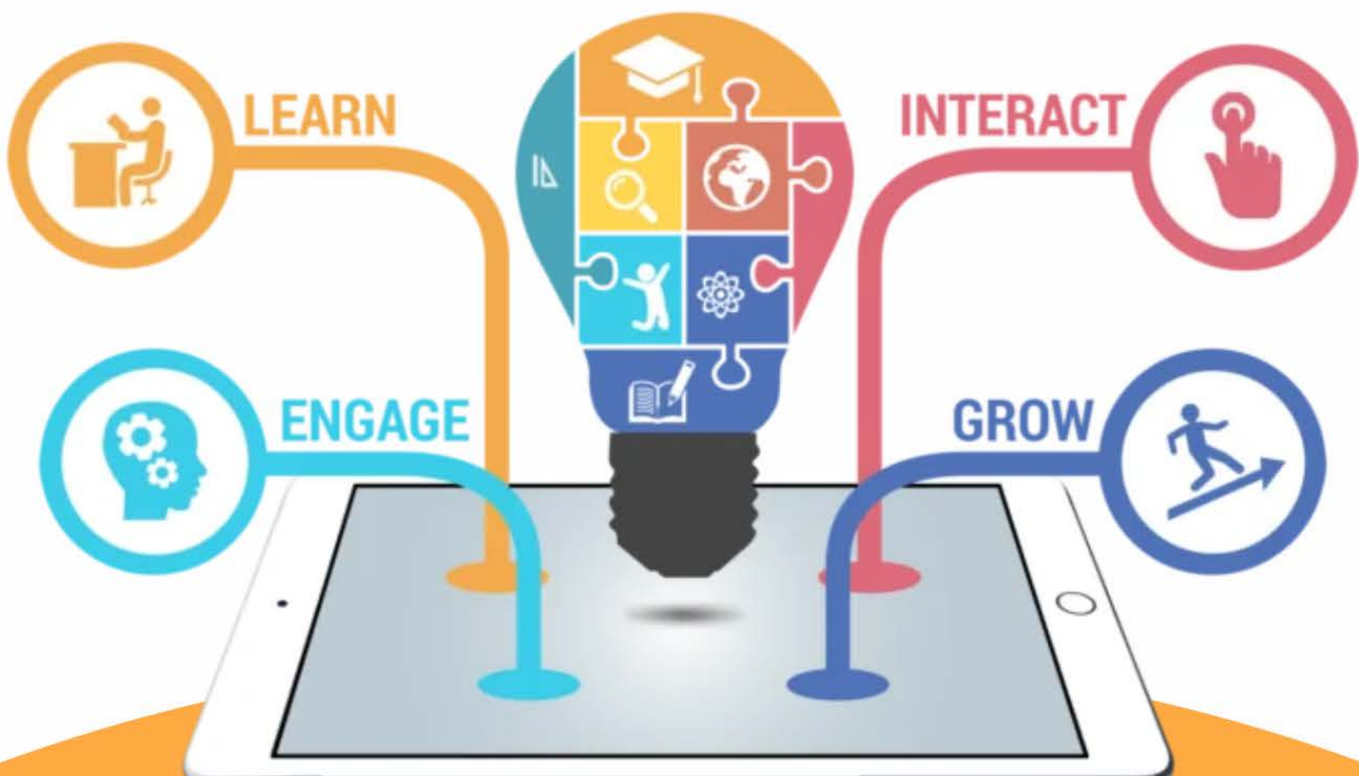
BSC-IT SEM 6

PRINCIPLES OF GEOGRAPHIC INFORMATION



With lots of efforts, research, reviews we have launched the prerecorded series of academics for multiple universities, bundled in userfriendly application "The Shikshak"

The Shikshak EdTech

Students should do **topic-wise study** rather than question-wise study for several reasons:

1. **Comprehensive understanding**: Topic-wise study allows students to have a thorough understanding of a particular topic. It helps in building a strong foundation of knowledge on a subject. Once they have a good understanding of a particular topic, they can answer any question related to it.
2. **Efficient use of time**: When students study topic-wise, they can cover a range of questions related to a particular topic in one go. This way, they can utilize their time more efficiently instead of jumping from one question to another and losing focus.
3. **Better retention**: Studying a topic in-depth helps students retain the information for a longer time. It is because they learn the concepts in a logical sequence, making it easier for them to remember.
4. **Effective exam preparation**: Most exams are organized based on topics or units, so studying topic-wise will enable students to be well-prepared for the exam. They will have a good grasp of all the topics that will appear on the exam.
5. **Build analytical skills**: When students study topic-wise, they develop their analytical skills by understanding how various concepts in a subject connect with each other. This helps them develop a deeper understanding of the subject, making them better problem solvers.

In conclusion, studying topic-wise is more beneficial for students as it enables them to develop a better understanding of a subject, retain information better, utilize their time more efficiently, and be well-prepared for exams.

TheShikshak Edu App is an online learning platform that offers a range of resources and tools to help students pursuing BScIT and BScCS programs. Here are some ways in which TheShikshak Edu App can benefit BScIT and BScCS students:

1. **Comprehensive course material**: TheShikshak Edu App offers comprehensive course material for BScIT and BScCS students, covering all topics and concepts required in these programs.
2. **Track their progress** : analytics program helps student to know which topics are remaining and which are lowest watched lectures
3. **Expert guidance**: TheShikshak Edu App has a team of experienced instructors who provide expert guidance and support to students. Students can get their doubts clarified and receive personalized feedback on their performance.

UNIT 1

Chapter 1: A Gentle Introduction to GIS

The nature of GIS

Fundamental Observations:

- Many aspects of our daily lives and our environment are constantly changing, and not always for the better. Some of these changes appear to have natural causes (e.g. volcanic eruptions, meteorite impacts), while others are the result of human modification of the environment (**e.g.** land use changes or land reclamation from the sea).
- There are also a large number of global changes for which the cause remains unclear: these include global warming, landslides and soil erosion.

Defining GIS:

A GIS is a computer-based system that provides the following four sets of capabilities to handle georeferenced data:

1. *Data capture and preparation*
2. *Data management*, including storage and maintenance
3. *Data manipulation and analysis*
4. *Data presentation*

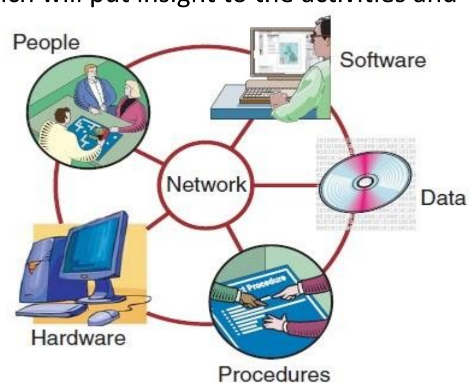
- This implies that a GIS user can expect support from the system to enter (georeferenced) data, to analyse it in various ways, and to produce presentations (including maps and other types) from the data.
- This would include support for various kinds of coordinate systems and transformations between them, options for analysis of the georeferenced data.

GISystem

- Geographic Information System (GISystem) is the most used concept of GIS.
- GISystem as a computerized system designed to dealing with the collection, storage, manipulation, analysis, visualization and displaying geographic information.
- GISystem is a tool to perform the spatial analysis which will put insight to the activities and phenomena carrying out everyday.

- GISystem include different components:-

1. Hardware
2. Software
3. Data/Information
4. Users/People
5. Procedures/Methods and Network



The major components of GISystem

GIScience

- Geographic Information Science (GIScience) is advocated to address a set of intellectual and scientific questions which go well beyond the technical capabilities of GISystem
- GIScience: Talks about GIS as a scientific discipline of study in the academia. This is the science behind the technology aimed at enhancing knowledge of Geospatial concepts and their computational implementations. The major contributing disciplines are:-
  1. Computer science
  2. Mathematics/Statistics
  3. Geomatics (Land Surveying, Photogrammetry, Remote Sensing, Geodesy, GPS, Drone mapping)
  4. Geography and
  5. Cartography

GIS Application

- Geographic Information Application is the kind of services dealing with the geographic information, such as the design and development of the GIS, geographic information retrieval, analysis, etc. For example, MapQuest (www.mapquest.com) provides a routing service for people to find the best driving route between two points.
- GIService allows GIS users to access specific functions that are provided by remote sites through the internet.
- Some examples are: MapQuest, Google maps, Bing Maps, Yahoo Maps, Apple Maps, Yandex Maps, OpenStreetMap and WikiMapia Maps.

Spatial data and geoinformation

- By data, we mean representations that can be operated upon by a computer.
- spatial data mean data that contains positional values, such as (x, y) co-ordinates. Sometimes the more precise phrase geospatial data is used as a further refinement, which refers to spatial data that is georeferenced.
- By information, we mean data that has been interpreted by a human being. Humans work with and act upon information, not data.
- Human perception and mental processing leads to information, and hopefully understanding and knowledge.
- Geoinformation is a specific type of information resulting from the interpretation of spatial data.
- As this information is intended to reduce uncertainty in decision-making, any errors and uncertainties in spatial information products may have practical, fi- nancial and even legal implications for the user.
- Traditionally, most spatial data were collected and held by individual, special- ized organizations
- In recent years, increasing availability and decreasing cost of data capture equipment has resulted in many users collecting their own data. However, the collection and maintenance of ‘base’ data remain the responsibility of the various governmental
- agencies, such as National Mapping Agen- cies (NMAs), which are responsible for collecting topographic data for the entire country following pre-set standards.

- Other agencies such as geological survey companies, energy supply companies, local government departments, and many others, all collect and maintain spatial data for their own particular pur-poses.

The real world and representations of it

- One of the main uses of GIS is as a tool to help us make decisions. Specifically, we often want to know the best location for a new facility, the most likely sites for mosquito habitat, or perhaps identify areas with a high risk of flooding so that we can formulate the best policy for prevention.
- In using GIS to help make these decisions, we need to represent some part of the real world as it is, as it was, or perhaps as we think it will be.
- We need to restrict ourselves to 'some part' of the real world simply because it cannot be represented completely.

Model and modelling

- '**Modelling'** is a term used in many different ways and which has many different meanings. A representation of some part of the real world can be considered a model because the representation will have certain characteristics in common with the real world.
- Specifically, those which we have identified in our model design. This then allows us to study and operate on the model itself instead of the real world in order to test what happens under various conditions, and help us answer 'what if' questions.
- We can change the data or alter the parameters of the model, and investigate the effects of the changes.
- **Models**—as representations—come in many different flavors.
- In the GIS environment, the most familiar model is that of a map. A map is a miniature repre-sentation of some part of the real world.
- Paper maps are the most common, but digital maps also exist.
- Databases are another important class of models. A database can store a considerable amount of data, and also provides various functions to operate on the stored data. The collection of stored data represents some real world phenomena, so it too is a model.

Maps

- **maps** are perhaps the best known (conventional) models of the real world.
- Maps have been used for thousands of years to represent information about the real world, and continue to be extremely useful for many applications in various domains.
- A disadvantage of the traditional paper map is that it is generally restricted to two- dimensional static representations, and that it is always displayed in a fixed scale. The map scale determines the Map spatial resolution of the graphic feature representation.
- A map is always a graphic representation at a certain level of detail, which is determined by the scale.
- Map sheets have physical boundaries, and features spanning two map sheets have to be cut into pieces.
- Cartography, as the science and art of map making, functions as an interpreter, translating real world phenomena (primary data) into correct, clear and understandable representations for our

use.

- Maps also become a data source for other applications, including the development of other maps.

Databases

- A database is a repository for storing large amounts of data. It comes with a number of useful functions:
- A database can be used by multiple users at the same time—i.e. it allows concurrent use
- A database offers a number of techniques for storing data and allows the use of the most efficient one—i.e. it supports storage optimization
- A database allows the imposition of rules on the stored data; rules that will be automatically checked after each update to the data—i.e. it supports data integrity
- A database offers an easy to use data manipulation language, which allows the execution of all sorts of data extraction and data updates—i.e. it has a query facility,
- A database will try to execute each query in the data manipulation lan- guage in the most efficient way—i.e. it offers query optimization.

Databases can store almost any kind of data. Modern database systems, as we shall see below

DAYMEASUREMENTS

BuoyDate		*S*	*STWSHumidTemp10*			. ..
B0749	1997/12/03	28.2 °C	NNW 4.2	72%	22.2 °C	. ..
B9204	1997/12/03	26.5 °C	NW 4.6	63%	20.8 °C	. ..
B1686	1997/12/03	27.8 °C	NNW 3.8	78%	22.8 °C	. ..
B0988	1997/12/03	27.4 °C	N 1.6	82%	23.8 °C	. ..
B3821	1997/12/03	27.5 °C	W 3.2	51%	20.8 °C	. ..
B6202	1997/12/03	26.5 °C	SW 4.3	67%	20.5 °C	. ..
B1536	1997/12/03	27.7 °C	SSW 4.8	58%	21.4 °C	. ..
B0138	1997/12/03	26.2 °C	W 1.9	62%	21.8 °C	. ..
B6823	1997/12/03	23.2 °C	S 3.6	61%	22.2 °C	. ..
. ..	. ..	. ..	. ..	. ..	. ..	. ..

Table 1.2: A stored table (in part) of daily buoy measurements. Illustrated are only measurements for December 3rd, 1997, though measurements for other dates are in the table as well. *Humid* is the air humidity just above the sea, *Temp10* is the measured water temperature at 10 metres depth. Other measurements are not shown.

Spatial databases and spatial analysis

- A GIS must store its data in some way. For this purpose the previous generation of software was equipped with relatively rudimentary facilities.
- Since the 1990's there has been an increasing trend in GIS applications that used a GIS for spatial

analysis, and used a database for storage.

- In more recent years, spatial databases (also known as geodatabases) have emerged. Besides traditional administrative data, they can store representations of real world geographic phenomena for use in a GIS.
- databases are special because they use additional techniques different from tables to store these spatial representations.
- Spatial analysis is the generic term for all manipulations of spatial data carried out to improve one's understanding of the geographic phenomena that the data represents.
- It involves questions about how the data in various layers might re- late to each other, and how it varies over space.
- For example, in the El Nin˜o case, we may want to identify the the steepest gradient in water temperature.
- The aim of spatial analysis is usually to gain a better understanding of geographic phenomena through discovering patterns that were previously unknown to us, or to build arguments on which to base important decisions.
- It should be noted that some GIS functions for spatial analysis are simple and easy-to- use, others are much more sophisticated, and demand higher levels of analytical and operating skills.
- Successful spatial analysis requires appropriate software, hardware, and perhaps most importantly, a competent user.

Chapter 2: Geographicinformation andSpatial database

Models and representations of the real world

- As discussed in the previous chapter, we use GISs to help analyse and under- stand more about processes and phenomena in the real world.
- Modelling is the process of producing an abstraction of the 'real world' so that some part of it can be more easily handled.
- the process of modelling, or building a representation which has certain characteristics in common with the real world.
- In practical terms, this refers to the process of representing key aspects of the real world digitally (inside a computer). These representations are made up of spatial data, stored in memory in the form of bits and bytes, on media such as the hard drive of a computer.
- This digital representation can then be subjected to various analytical functions (computations) in the GIS, and the output can be visualized in various ways.

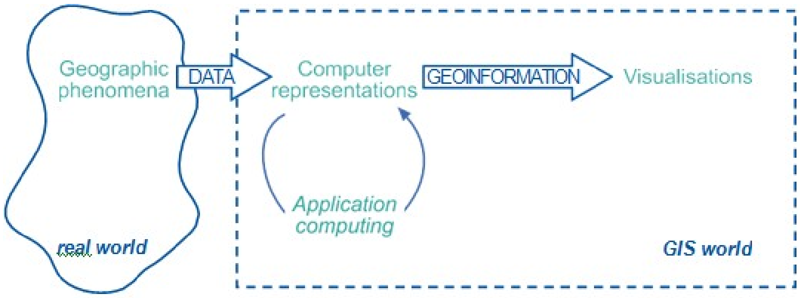


Figure 2.1: Representing relevant aspects of real-world phenomena inside a GIS to build models or simulations.

Defining geographic phenomena

- A GIS operates under the assumption that the relevant spatial phenomena occur in a two- or three-dimensional Euclidean space, unless otherwise specified.
- Euclidean space can be informally defined as a model of space in which locations are represented by coordinates—(x, y) in 2D; (x, y, z) in 3D—and distance and direction can defined with geometric formulas.In the 2D case, this is known as the Euclidean plane, which is the most common Euclidean space in GIS use.
- In order to be able to represent relevant aspects real world phenomena inside a GIS, we first need to define what it is we are referring to.
    - Can be named or described,
    - Can be georeferenced, and
    - Can be assigned a time (interval) at which it is/was present

---

Types of geographic phenomena

- The geographic phenomena come in so many different ‘flavours’, which we will try to categorize below. Before doing so, we must make two further observations.
- Firstly, In order to be able to represent a phenomenon in a GIS requires us to state what it is, and where it is. We must provide a description—or at least a name—on the one hand, and a georeference on the other hand.
- Secondly, some phenomena manifest themselves essentially everywhere in the study area, while others only do so in certain localities. If we define our study area as the equatorial Pacific Ocean, we can say that Sea Surface Temperature can be measured anywhere in the study area. Therefore, it is a typical example of a (geographic) field.
- A (geographic) field is a geographic phenomenon for which, for every point in the study area, a value can be determined.
- Some common examples of geographic fields are air temperature, barometric pressure and elevation.

Elevation in the Falset study area, Tarragona province, Spain. The area is approximately 25 20 km. The illustration has been aesthetically improved by a technique known as ‘hillshading’. In this case, it is as if the sun shines from the north-west, giving a shadow effect towards the south-east. Thus, colour alone is not a good indicator of elevation; observe that elevation is a continuous function over the space.

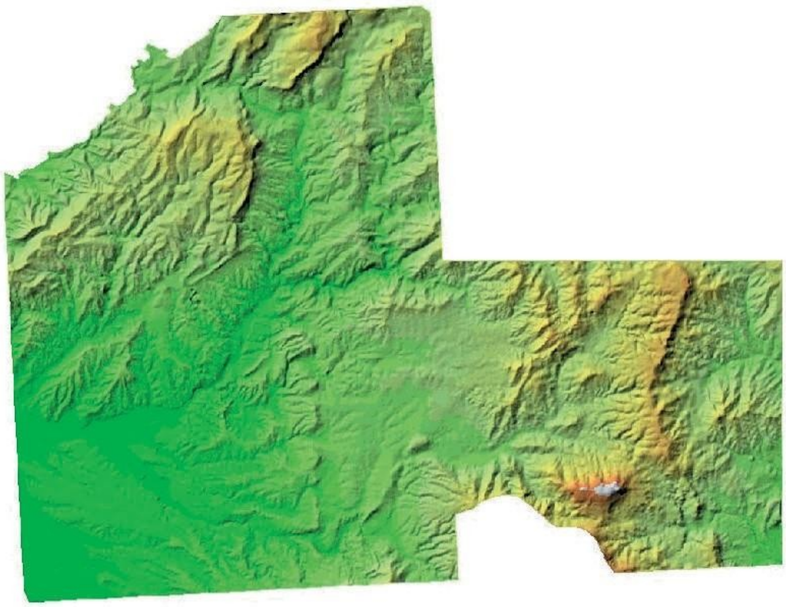
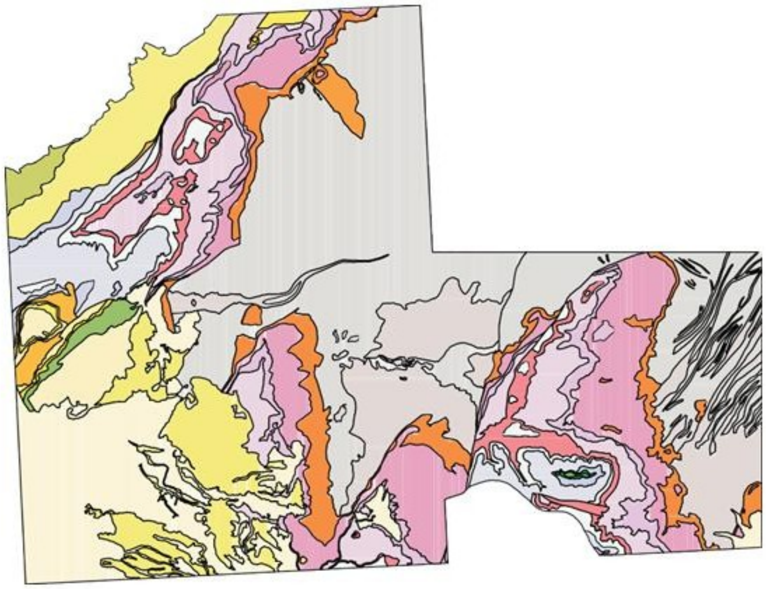


Figure 2.2: A continuous field example, namely the *elevation* in the study area of Falset, Spain. Data source: Department of Earth Systems Analysis (ESA, ITC)

Geographic fields

- A field is a geographic phenomenon that has a value 'everywhere' in the study area. We can therefore think of a field as a mathematical function f that associates a specific value with any position in the study area.
- Hence if (x, y) is a position in the study area, then f (x, y) stands for the value of the field of at locality (x, y).
- Fields can be discrete or continuous. In a continuous field, the underlying function is assumed to be 'mathematically smooth', meaning that the field values along any path through the study area do not change abruptly, but only gradually.
- Good examples of continuous fields are air temperature, barometric pressure, soil salinity and elevation. Continuity means that all changes in field values are gradual.
- Discrete fields divide the study space in mutually exclusive, bounded parts, with all locations in one part having the same field value.
- Typical examples are land classifications, for instance, using either geological classes, soil type, land use type, crop type or natural vegetation type.



Observe that—typical for fields—with any location only a single geological unit is associated. As this is a *discrete* field, value changes are discontinuous, and therefore locations on the boundary between two units are *not* associated with a particular value (i.e. with a geological unit).

Figure 2.3: A discrete field indicating geological units, used in a foundation engineering study for constructing buildings. The same study area as in Figure2.2.
Data source: Department of Earth Systems Analysis (ESA, ITC)

Data types and values

- Since we have now differentiated between continuous and discrete fields, we may also look at different kinds of data values which we can use to represent our 'phenomena'. It is important to note that some of these data types limit the types of analyses that we can do on the data itself:
- Nominal data values are values that provide a name or identifier so that we can discriminate between different values, but that is about all we can do. Specifically, we cannot do true computations with these values. An example are the names of geological units.
- Ordinal data values are data values that can be put in some natural sequence but that do not allow any other type of computation. Household income, for instance, could be classified as being either 'low', 'average' or 'high'.
- Interval data values are quantitative, in that they allow simple forms of computation like addition and subtraction. However, interval data has no arithmetic zero value, and does not support multiplication or division
- Ratio data values allow most, if not all, forms of arithmetic computation. Rational data have a natural zero value, and multiplication and division of values are possible operators (distances measured in metres are an example).

Geographic objects

- When a geographic phenomenon is not present everywhere in the study area, but somehow 'sparsely' populates it, we look at it as a collection of geographic objects. Such objects are usually easily distinguished and named, and their position in space is determined by a combination of one or more of the following parameters:

- *Location* (where is it?),
- *Shape* (what form is it?),
- *Size* (how big is it?), and
- *Orientation* (in which direction is it facing?).

- Collections of geographic objects can be interesting phenomena at a higher aggregation level: forest plots form forests, groups of parcels form suburbs, streams, brooks and rivers form a river drainage system, roads form a road network, and SST buoys form an SST sensor network. It is sometimes useful to view geographic phenomena at this more aggregated level and look at characteristics like coverage, connectedness ,and capacity. For example:
  - Which part of the road network is within 5 km of a petrol station? (A coverage question)
  - What is the shortest route between two cities via the road network? (A connectedness question)
  - How many cars can optimally travel from one city to another in an hour? (A capacity question)

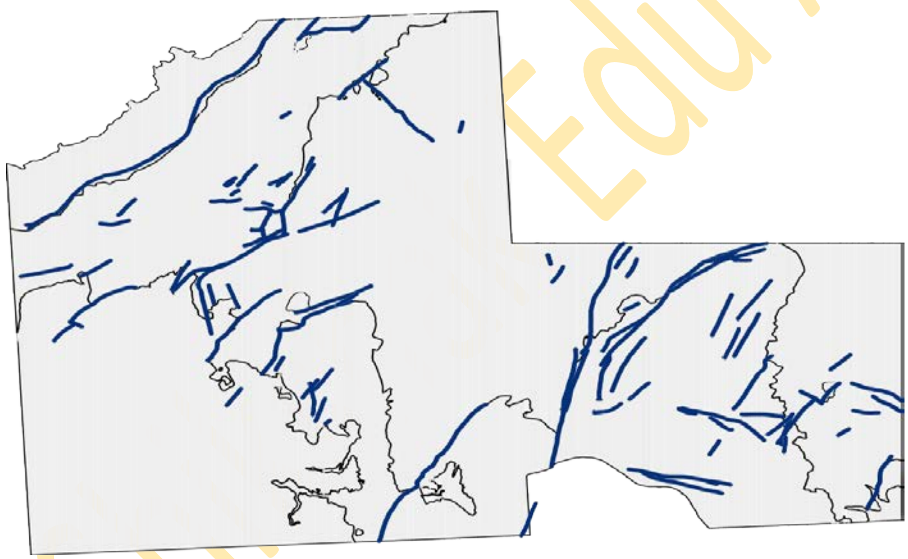


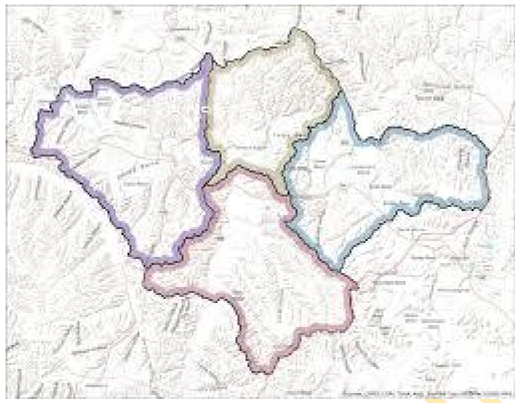
Fig: A number of geological faults in the same study area as in Figure2.2.Faults are indicated in blue; the study area, with the main geo- logical era's is set in grey in the background only as a reference. Data source: Department of Earth Systems Analysis (ITC).

Boundaries

- Where shape and/or size of contiguous areas matter, the notion of boundary comes into play. This is true for geographic objects but also for the constituents of a discrete geographic field Location, shape and size are fully determined if we know an area's boundary, so the boundary is a good candidate for representing it.
- This is especially true for areas that have naturally crisp boundaries.
- Fuzzy boundaries contrast with crisp boundaries in that the boundary is not a precise line, but rather itself an area of transition.

Computer representations of geographic information

- Up to this point, we have not looked at how geoinformation, like fields and objects, is represented in a computer. After the discussion of the main characteristics of geographic phenomena above, let us now examine representation in more detail.
- We have seen that various geographic phenomena have the characteristics of continuous functions over space.



- Elevation, for instance, can be measured at many locations, even within one's own backyard, and each location may give a different value.

In order to represent such a phenomenon faithfully in computer memory, we could either:

- Try to store as many (*location*,*elevation*) observation pairs as possible, or
- Try to find a symbolic representation of the elevation field function, as a formula in x and y—like ($3.0678x^2 + 20.08x\ 7.34y$) or so—which can be evaluated to give us the elevation at any given (x, y) location.
- Both of these approaches have their drawbacks. The first suffers from the fact that we will never be able to store all elevation values for all locations; after all, there are infinitely many locations.
- The second approach suffers from the fact that we do not know just what this function should look like, and that it would be extremely difficult to derive such a function for larger areas.

Regular tessellations

- A tessellation (or tiling) is a partitioning of space into mutually exclusive cells that together make up the complete study space. With each cell, some (thematic) value is associated to characterize that part of space. Three regular tessellation types are illustrated in Figure2.5.
- In a regular tessellation, the cells are the same shape and size. The simplest example is a rectangular raster of unit squares, represented in a computer in the 2D case as an array of n m elements (see Figure2.5–left).

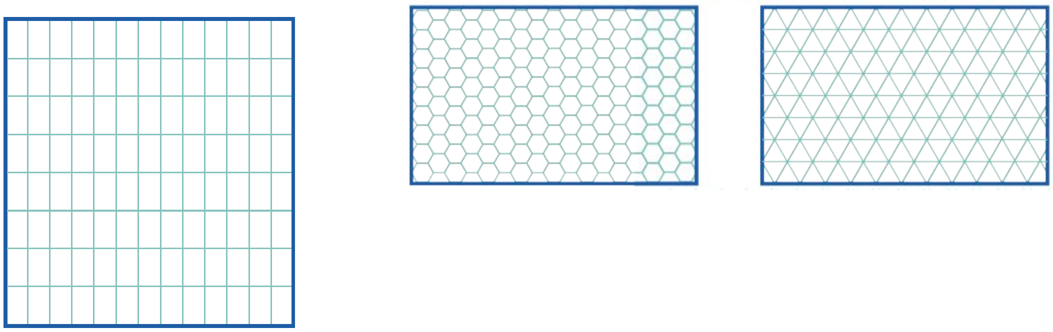


Figure 2.5: The three most common regular tessellation types: square cells, hexagonal cells, and triangular cells.

- In all regular tessellations, the cells are of the same shape and size, and the field attribute value assigned to a cell is associated with the entire area occu- pied by the cell. The square cell tessellation is by far the most commonly used, mainly because georeferencing a cell is so straightforward. These tessellations are known under various names in different GIS packages, but most frequently as rasters.
- A raster is a set of regularly spaced (and contiguous) cells with associated (field) values. The associated values represent cell values, not point values. This means that the value for a cell is assumed to be valid for all locations within the cell.

Irregular tessellations

- Irregular tessellations are more complex than the regular ones, but they are also more adaptive, which typically leads to a reduction in the amount of memory used to store the data.
- A well-known data structure in this family—upon which many more variations have been based—is the region quadtree. It is based on a regular tessellation of square cells, but takes advantage of cases where neighbouring cells have the same field value, so that they can together be represented as one bigger cell.

A simple illustration is provided in Figure2.7.

- It shows a small 8 8 raster with three possible field values: white, green and blue. The quadtree that represents this raster is constructed by repeatedly splitting up the area into four quadrants, which are called NW, NE, SE, SW for obvious rea- sons. This procedure stops when all the cells in a quadrant have the same field value.
- The procedure produces an upside-down, tree-like structure, known as a quadtree. In main memory, the nodes of a quadtree (both circles and squares in the figure below) are represented as records. The links between them are pointers, a programming technique to address (i.e. to point to) other records.

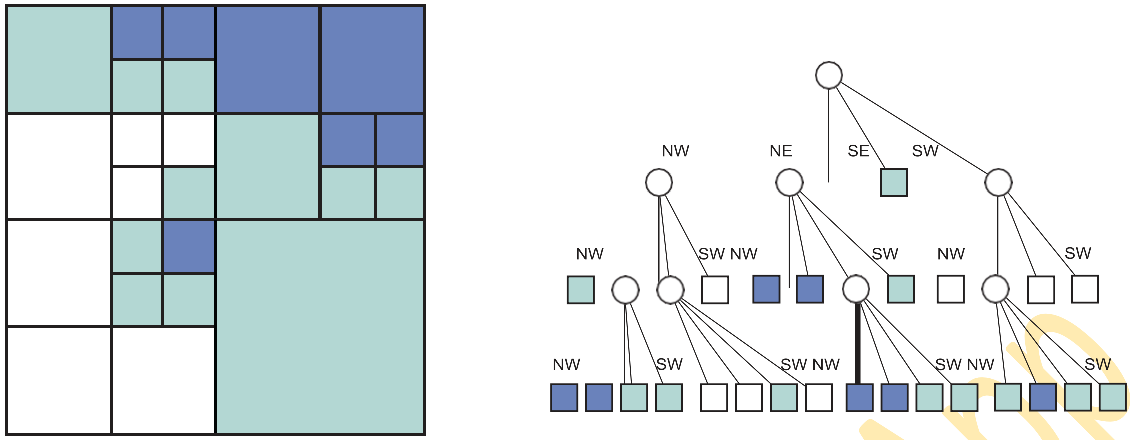
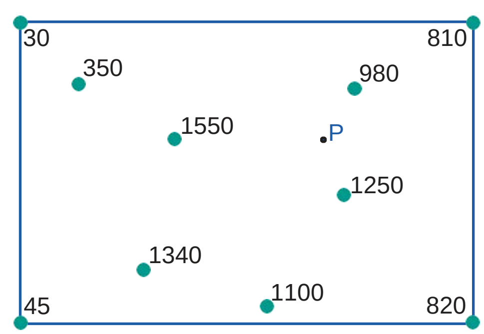


Figure 2.7: An 8 8, three-valued raster (here: colours) and its repre- sentation as a region quadtree. To construct the quadtree, the field is successively split into four quadrants until parts have only a single field value. After the first split, the southeast quadrant is entirely green, and this is indicated by a green square at level two of the tree. Other quadrants had to be split further.

Vector representations

- Tessellations do not explicitly store georeferences of the phenomena they repre- sent. Instead, they provide a georeference of the lower left corner of the raster, for instance, plus an indicator of the raster's resolution, thereby implicitly pro- viding georeferences for all cells in the raster. In vector representations, an attempt is made to explicitly associate georeferences with the geographic phenomena. georeference is a coordinate pair from some geographic space, and is also known as a vector. This explains the name.
- Below, we discuss various vector representa- tions. We start with our discussion with the TIN, a representation for geographic fields that can be considered a hybrid between tessellations and vector representations.

Figure 2.8: Input locations and their (elevation) values for a TIN construction. The location *P* is an arbitrary location that has no associated elevation measurement.



Triangulated Irregular Networks

- A commonly used data structure in GIS software is the triangulated irregular net- work, or TIN.

- It is one of the standard implementation techniques for digital terrain models, but it can be used to represent any continuous field. The principles behind a TIN are simple.
- It is built from a set of locations for which we have a measurement, for instance an elevation. The locations can be arbitrarily scattered in space, and are usually not on a nice regular grid. Any location together with its elevation value can be viewed as a point in three-dimensional space.

Two tessellations are illustrated in Figure2.9.

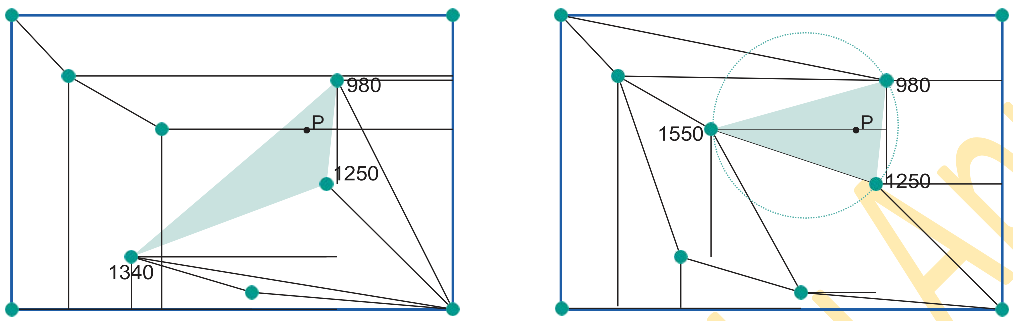


Figure 2.9: Two triangulations based on the input locations of Figure2.8.
(a) one with many 'stretched' triangles; (b) the triangles are more equilateral; this is a *Delaunay triangulation*

- In three-dimensional space, three points uniquely determine a plane, as long as they are not collinear, i.e. they must not be positioned on the same line. A plane fitted through these points has a fixed aspect and gradient, and can be used to compute an approximation of elevation of other locations.3 Since we can pick many triples of points, we can construct many such planes, and therefore we can have many elevation approximations for a single location, such as P (Figure2.8). So, it is wise to restrict the use of a plane to the triangular area 'between' the three points.

Point representations

- Points are defined as single coordinate pairs (x, y) when we work in 2D, or co- ordinate triplets (x, y, z) when we work in 3D.
- Points are used to represent objects that are best described as shape- and size- less, one-dimensional features. Whether this is the case really depends on the purposes of the spatial application and also on the spatial extent of the objects compared to the scale applied in the application. For a tourist city map, a park will not usually be considered a point feature, but perhaps a museum will, and certainly a public phone booth might be represented as a point.

Line representations

- Line data are used to represent one-dimensional objects such as roads, railroads, canals, rivers and power lines. Again, there is an issue of relevance for the application and the scale that the application requires. For the example application of mapping tourist information, bus, subway and streetcar routes are likely to be relevant line features. Some cadastral systems, on the other hand, may consider roads to be two-dimensional features,
- i.e. having a width as well.
- Above, we discussed the notion that arbitrary, continuous curvilinear features are as equally difficult to represent as continuous fields. GISs therefore approximate such features (finitely!) as lists of

nodes. The two end nodes and zero or more internal nodes or vertices define a line. Other terms for 'line' that are commonly used in some GISs are polyline, arc or edge. A node or vertex is like a point (as discussed above) but it only serves to define the line, and provide shape in order to obtain a better approximation of the actual feature.

- The straight parts of a line between two consecutive vertices or end nodes are called line segments.

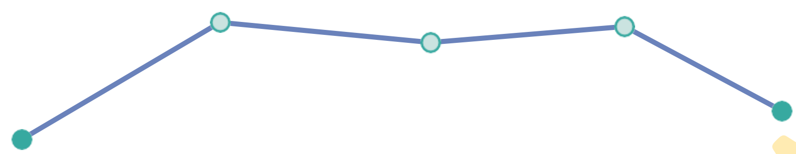


Figure 2.10: A line is de- fined by its two end nodes and zero or more internal nodes, also known as vertices. This line representation has three vertices, and therefore four line segments.

Area representations

- When area objects are stored using a vector approach, the usual technique is to apply a boundary model. This means that each area feature is represented by some arc/node structure that determines a polygon as the area's boundary. Common sense dictates that area features of the same kind are best stored in a single data layer, represented by mutually non-overlapping polygons. In essence, what we then get is an application- determined (i.e. adaptive) partition of space.

In the case that the object can be perceived as having a fuzzy boundary, a polygon is an even worse approximation, though potentially the only one possible. An example is provided in Figure2.11.

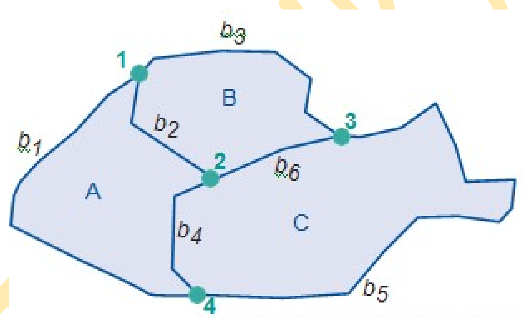


Figure 2.11: Areas as they are represented by their boundaries. Each boundary is a cyclic se- quence of line features; each line—as before—is a sequence of two end nodes, with in between, zero or more vertices.

General spatial topology

- Topology deals with spatial properties that do not change under certain transformations. For example, features drawn on a sheet of rubber (as in Figure2.13) can be made to change in shape and size by stretching and pulling the sheet. However, some properties of these features do not change:
- Area E is still inside area D
- The neighbourhood relationships between A, B, C, D, and E stay intact, and their boundaries have the same start and end nodes, and

- The areas are still bounded by the same boundaries, only the shapes and lengths of their perimeters have changed.

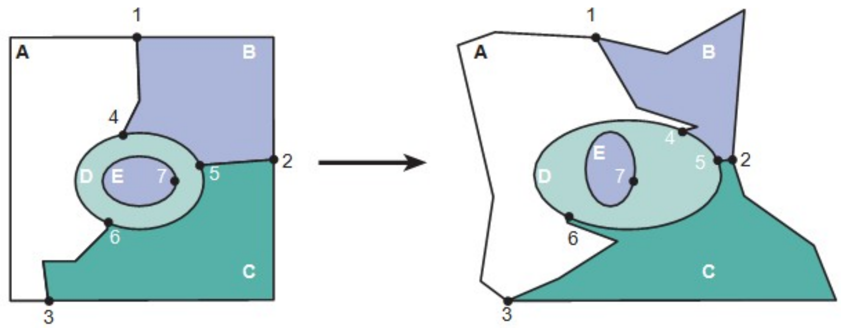


Figure 2.13: Rubber sheet transformation: The space is transformed, yet many relationships between the constituents remain unchanged.

> Topology refers to the spatial relationships between geographical elements in a data set that do not change under a continuous transformation.

- Topological relationships are built from simple elements into more complex elements: nodes define line segments, and line segments connect to define lines, which in turn define polygons.

Topological relationships

The mathematical properties of the geometric space used for spatial data can be described as follows:

- **The space is a three-dimensional Euclidean space** where for every point we can determine its three-dimensional coordinates as a triple (x, y, z) of real numbers. In this space, we can define features like points, lines, polygons, and volumes as geometric primitives of the respective dimension. A point is zero-dimensional, a line one- dimensional, a polygon two-dimensional, and a volume is a three-dimensional primitive.
- **The space is a metric space**, which means that we can always compute the distance between two points according to a given distance function. Such a function is also known as a metric.
- **The space is a topological space**, of which the definition is a bit complicated. In essence, for every point in the space we can find a neighbourhood around it that fully belongs to that space as well.
- **Interior and boundary** are properties of spatial features that remain invariant under topological mappings. This means, that under any topological mapping, the interior and the boundary of a feature remains unbroken and intact.

The topology of two dimensions

- We can use the topological properties of interior and boundary to define relationships between spatial features.

- Suppose we consider a spatial region A. It has a boundary and an interior, both seen as (infinite) sets of points, and which are denoted by *boundary*(A) and *interior* (A), respectively. We consider all possible combinations of intersections ( ) between the boundary and the interior of A with those of another region B, and test whether they are the empty set ( ) or not. From these intersection patterns, we can derive eight (mutually exclusive) spatial relationships between two regions. If, for instance, the interiors of A and B do not intersect, but their boundaries do, yet a boundary of one does not intersect the interior of the other, we say that A and B *meet*. In mathematics, we can therefore define the *meets* relationship using set theory, as

$$
\begin{aligned}
A\ meets B \overset{\text{def}}{=}\ & interior(A) \cap interior(B) = \emptyset\ \wedge \\
& boundary(A) \cap boundary(B)\ f= \emptyset\ \wedge \\
& interior(A) \cap boundary(B) = \emptyset\ \wedge \\
& boundary(A) \cap interior(B) = \emptyset.
\end{aligned}
$$

In the above formula, the symbol $\wedge$ expresses the logical connective 'and'. Thus, the formula states four properties that must all hold for the formula to be true.

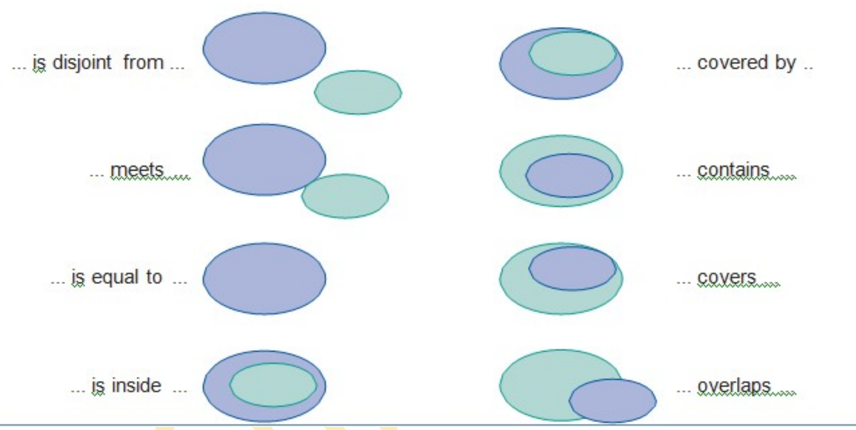


Figure 2.15: Spatial relationships between two regions derived from the topological invariants of intersections of boundary and interior. The relationships can be read with the green region on the left . . . and the blue region on the right . . .

Figure2.15 shows all eight spatial relationships: *disjoint*, *meets*, *equals*, *inside*, *covered by*, *contains*, *covers*, and *overlaps*. These relationships can be used in queries against a spatial database, and represent the 'building blocks' of more complex spatial queries.

The three-dimensional case

- It is not without reason that our discussion of vector representations and spatial topology has focused mostly on objects in two-dimensional space. The history of spatial data handling is almost purely 2D, and this is remains the case for the majority of present-day GIS applications. Many application domains make use of elevational, but these are usually accommodated by so-called $2\frac{1}{2}$D data structures. These $2\frac{1}{2}$D data structures are similar to the (above discussed) 2D data structures using points, lines and areas.
- There is, on the other hand, one important aspect in which $2\frac{1}{2}$D data does differ from standard 2D data, and that is in their association of an additional z- value with each 0- simplex ('node').

Thus, nodes also have an elevation value associated with them. Essentially, this allows the GIS user to represent 1- and 2-simplices that are non- horizontal, and therefore, a piecewise planar, 'wrinkled surface' can be constructed as well, much like a TIN. Note however, that one cannot have two different nodes with identical x- and y-coordinates, but different z-values. Such nodes would constitute a perfectly vertical feature, and this is not allowed. Consequently, true solids cannot be represented in a $2\frac{1}{2}$D GIS.

Scale and resolution

- Map scale can be defined as the ratio between the distance on a paper map and the distance of the same stretch in the terrain. A 1:50,000 scale map means that1 cm on the map represents 50,000 cm, i.e. 500 m, in the terrain. 'Large-scale' means that the ratio is large, so typically it means there is much detail, as in a 1:1,000 paper map. 'Small-scale' in contrast means a small ratio, hence less detail, as in a 1:2,500,000 paper map. When applied to spatial data, the term resolution is commonly associated with the cell width of the tessellation applied.
- Digital spatial data, as stored in a GIS, is essentially without scale: scale is a ratio notion associated with visual output, like a map or on-screen display, not with the data that was used to produce the map.
- When digital spatial data sets have been collected with a specific map-making purpose in mind, and these maps were designed to be of a single map scale, like 1:25,000, we might suppose that the data carries the characteristics of "a 1:25,000 digital data set."

Representations of geographic fields

- A geographic field can be represented through a tessellation, through a TIN or through a vector representation. The choice between them is determined by the requirements of the application at hand.
- It is more common to use tessellations, notably rasters, for field representation, but vector representations are in use too. We have already looked at TINs. We provide an example of the other two below

1. **Raster representation of a field**

- In Figure 2.17, we illustrate how a raster represents a continuous field like elevation. Different shades of blue indicate different elevation values, with darker blues indicating higher elevations. The choice of a blue colour spectrum is only to make the illustration aesthetically pleasing; real elevation values are stored in the raster, so instead we could have printed a real number value in each cell. This would not have made the figure very legible, however

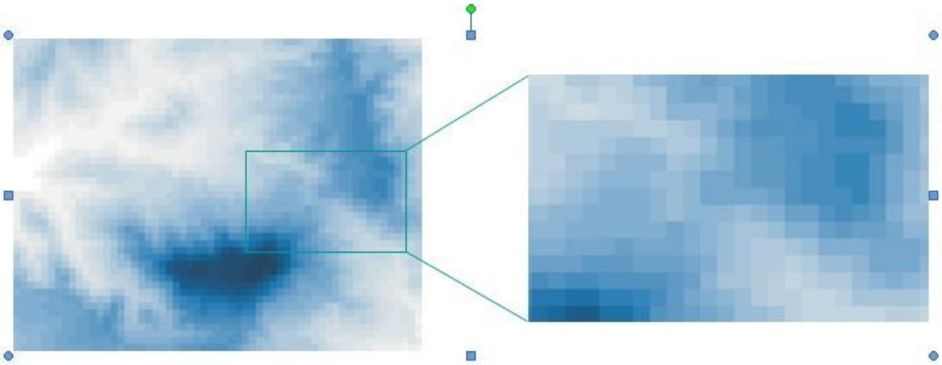


Figure 2.17: A raster representation (in part) of the elevation of the study area of Figure2.2. Actual elevation values are indicated as shades of blue. The depicted area is the north-east flank of the mountain in the south-east of the study area. The right-hand side of the figure is a zoomed-in part of that of the left.

- A raster can be thought of as a long list of field values: actually, there should be m n such values. The list is preceded with some extra information, like a single georeference as the origin of the whole raster, a cell size indicator, the integer values for m and n, and a data type indicator for interpreting cell values. Rasters and quadtrees do not store the georeference of each cell, but infer it from the above information *about* the raster.

2. Vector representation of a field

- We briefly mention a final representation for fields like elevation, but using a vector representation. This technique uses isolines of the field. An *isoline* is a linear feature that connects the points with equal field value. When the field is elevation, we also speak of *contour lines*. The elevation of the Falset study area is represented with contour lines in Figure2.18. Both TINs and isoline representations use vectors.

Figure 2.18: A vector-based elevation field representation for the study area of Figure2.2. Indicated are elevation isolines at a resolution of 25 metres.
Data source: Department of Earth Systems Analysis (ESA, ITC)

- Isolines as a *representation mechanism* are not very common, however. They are in use as a *geoinformation visualization technique* (in mapping, for instance), but commonly using a TIN for representing this type of field is the better choice. Many GIS packages provide functions to generate an isoline visualization from a TIN.

Representation of geographic objects

- The representation of geographic objects is most naturally supported with vectors. After all, objects are identified by the parameters of location, shape, size and orientation and many of these parameters can be expressed in terms of vectors. However, tessellations are still commonly used for representing geographic objects as well, and we discuss why below.

1. Tessellations to represent geographic objects

- Remotely sensed images are an important data source for GIS applications. Un- processed digital images contain many pixels, with each pixel carrying a reflectance value. Various techniques exist to process digital images into *classified images* that can be stored in a GIS as a raster.
- Image classification attempts to characterize each pixel into one of a finite list of classes, thereby obtaining an interpretation of the contents of the image. The classes recognized can be crop types as in the case of Figure 2.19 or urban land use classes as in the case of Figure 2.20.

These figures illustrate the unprocessed images (a) as well as a classified version of the image (b)

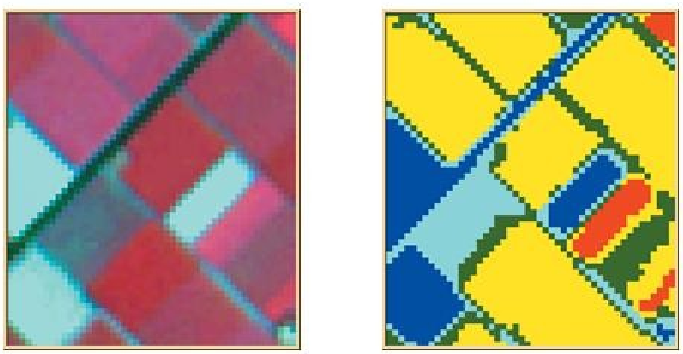
(a)

(b)

Figure 2.19: An unprocessed digital image (a) and a classified raster (b) of an agricultural area.

2. Vector representations for geographic objects

- The somehow more natural way to represent geographic objects is by vector representations. We have discussed most issues already in Section2.3.3, and a small example suffices at this stage.

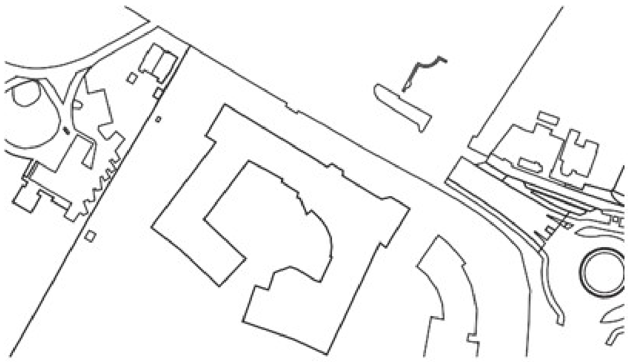


Figure 2.22: Various objects (buildings, bike and road lanes, railroad tracks) represented as area objects in a vector representation.

- In Figure 2.22, a number of geographic objects in the vicinity of the ITC building have been depicted. These objects are represented as area representations in a boundary model. Nodes and vertices of the polylines that make up the object's boundaries are not illustrated, though they obviously are stored.

Organizing and managing spatialdata

- In the previous sections, we have discussed various types of geographic information and ways of representing them. We have looked at case-by-case examples, however, we have purposefully avoided looking at how various sorts of spatial data are combined in a single system.

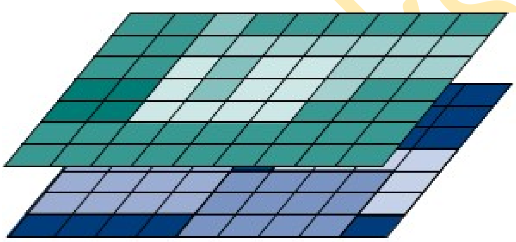


Figure 2.23: Different rasters can be overlaid to look for spatial correlations.

- The main principle of *data organization* applied in GIS systems is that of a spatial data layer. A *spatial data layer* is either a representation of a continuous or discrete field, or a collection of objects of the same kind. Usually, the data is organized so that similar elements are in a single data layer. For example, all telephone booth point objects would be in one layer, and all road line objects in another. A data layer contains spatial data—of any of the types discussed above— as well as attribute (or: thematic) data, which further describes the field or objects in the layer.
- Attribute data is quite often arranged in tabular form, maintained in some kind of geodatabase, as we will see in Chapter3. An example of two field data layers is provided in

Figure2.23.

- Data layers can be overlaid with each other, inside the GIS package, so as to study combinations of geographic phenomena. We shall see later that a GIS can be used to study the *spatial relationships* between different phenomena, requiring computations which overlay one data layer with another. This is schematically depicted in Figure 2.24 for two different object layers.

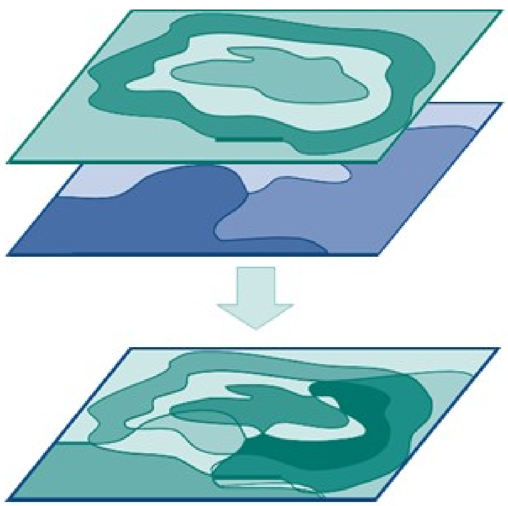


Figure 2.24: Two different object layers can be overlaid to look for spatial correlations, and the result can be used as a separate (object) layer.

The temporal dimension

- Besides having geometric, thematic and topological properties, geographic phenomena are also *dynamic*; they change over time. For an increasing number of applications, these changes themselves are the key aspect of the phenomenon to study. Examples include identifying the owners of a land parcel in 1972, or how land cover in a certain area changed from native forest to pastures over a specific time period. We can note that some features or phenomena change slowly, such as geological features, or as in the example of land cover given above. Other phenomena change very rapidly, such as the movement of people or atmospheric conditions. For different applications, different scales of measurement will apply.
- **Examples of the kinds of questions involving time include:**
  - o Where and when did something happen?
  - o How fast did this change occur?
  - o In which order did the changes happen?
- The way we represent relevant components of the real world in our models can influence the kinds of questions we can or cannot answer. This chapter has al- ready discussed representation issues for spatial features, but has so far ignored the problematic issues for incorporating time. The main reason lies in the fact that GISs still offer limited support for the representation of time. As a result, most studies require substantial efforts from the GIS user in data preparation and data manipulation. Also, besides representing an object or field

in 2D or 3D space, the temporal dimension is of a continuous nature. Therefore in order to represent it in a computer, we have to ‘discretize’ the time dimension.

UNIT 2

Data management and processing systems

Chapter 1 : Hardware and software trends

- Computers are also becoming increasingly affordable. Hand-held computers are now commonplace in business and personal use, equipping field surveyors with powerful tools, complete with GPS capabilities for instantaneous georeferencing.
- To support these hardware trends, software providers continue to produce application programs and operating systems that, while providing a lot more functionality, also consume significantly more memory.
- In general, software technology has developed somewhat slower and often cannot fully utilise the possibilities offered by the exponentially growing hardware capabilities.
- Existing software obviously performs better when run on faster computers.
- Alongside these trends, there have also been significant developments in computer networks. In essence, today almost any computer on Earth can connect to some network, and contact computers virtually anywhere else, allowing fast and reliable exchange of (spatial) data.
- Mobile phones are more and more frequently being used to connect to computers on the Internet. The UMTS protocol (Universal Mobile Telecommunications System), allows digital communication of text, audio, and video at a rate of approximately 2 Mbps.
- Bluetooth version 2.0 is a standard that offers up to 3 Mbps connections, especially between palm- and laptop computers and their peripheral devices, such as a mobile phone, GPS or printer at short range.
- Wireless LANs (Local Area Networks), under the so-called WiFi standard, nowadays offer a bandwidth of up to 108 Mbps on a single connection point, *to be shared*between computers. They are more and more used for constructing a computer network in office buildings and in private homes.

Chapter 2 : Geographic information systems

- It was identified in Chapter 1 that a GIS provides a range of capabilities to handle georeferenced data, including:
  1. Data capture and preparation,
  2. Data management (storage and maintenance)
  3. Data manipulation and analysis, and
  4. Data presentation
  - For many years, analogue data sources were used, processing was done manually, and paper maps were produced. The introduction of modern techniques has led to an increased use of computers and digital information in all aspects of spatial data handling. The software technology used in this domain is centered around geographic information systems.
  - Typical planning projects require data sources, both spatial and non-spatial, from different national institutes, like national mapping agencies, geological, soil, and forest survey institutes, and national census bureaus.

GIS software

- GIS can be considered to be a data store (i.e. a system that stores spatial data), a toolbox, a technology, an information source or a field of science. The main characteristics of a GIS software package are its analytical functions that provide means for deriving new geoinformation from existing spatial and attribute data.
- The use of tools for problem solving is one thing, but the production of these tools is something quite different. Not all tools are equally well-suited for a particular application, and they can be improved and perfected to better serve a particular need or application.
- The discipline of *geographic information science* is driven by the use of our GIS tools, and these are in turn improved by new insights and information gained through their application in various scientific fields.
- All GIS packages available on the market have their strengths and weaknesses, typically resulting from the development history and/or intended application domain(s) of the package.
- Well-known, full-fledged GIS packages include ILWIS, Intergraph's GeoMedia, ESRI's ArcGIS, and MapInfo from Map-Info Corp.

GIS architecture and functionality

- A geographic information system in the wider sense consists of software, data, people, and an organization in which it is used.
- Before moving on, we should also note that organizational factors will define the context and rules for the capture, processing and sharing of geoinformation, as well as the role which GIS plays in the organization as a whole.
- A GIS consists of several *functional components*—components which support key GIS functions. These are data capture and preparation, data storage, data analysis, and presentation of spatial data.
- Figure3.1shows a diagram of these components, with arrows indicating the data flow in the system. For a particular GIS, each of these components may provide many or only a few functions.

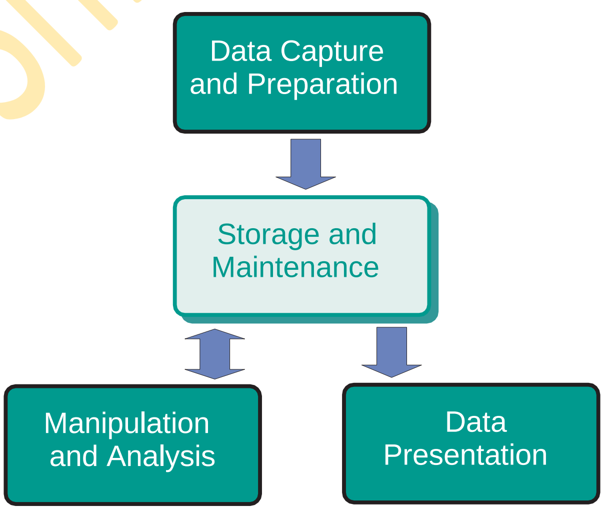


Figure 3.1: Functional components of a GIS

Spatial Data Infrastructure (SDI)

- An SDI is defined as "the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data".
- Fundamental to those arrangements are—in a wider sense— the agreements between organizations and in the narrow sense, the agreements between software systems on *how* to share the geographic information. In SDI, standards are often the starting point for those agreements. Standards exist for
- All facets of GIS, ranging from data capture to data presentation.
- They are developed by different organizations, of which the most prominent are the International Organization for Standardisation (ISO) and the Open Geospatial Consortium (OGC).
- Typically, an SDI provides its users with different facilities for finding, viewing, downloading and processing data. Because the organizations in an SDI are normally widely distributed over space, computer networks are used as the means of communication.
- With the development of the internet, the functional components of GIS have been gradually become available as web-based applications.

Chapter 3 : Stages of spatial data handling

Spatial data capture and preparation

- The functions for capturing data are closely related to the disciplines of surveying engineering, photogrammetry, remote sensing, and the processes of digitizing, i.e. the conversion of analogue data into digital representations.
- Remote sensing, in particular, is the field that provides photographs and images as the raw base data from which spatial data sets are derived.
- Surveys of the study area often need to be conducted for data that cannot be obtained with remote sensing techniques, or to validate data thus obtained.
- Traditional techniques for obtaining spatial data, typically from paper sources, included *manual digitizing* and *scanning.*
- Table3.2lists the main methods and devices used for data capture. In recent years there has been a significant increase in the availability and sharing of digital (geospatial) data.
- The data, once obtained in some digital format, may not be quite ready for use in the system. This may be because the format obtained from the capturing process is not quite the format required for storage and further use, which means that some type of data conversion is required.
- In part, this problem may also arise when the captured data represents only raw base data, out of which the real data objects of interest to the system still need to be constructed.

Method	*Devices*
Manual digitizing	• coordinate entry via keyboard • digitizing tablet with cursor • mouse cursor on the computer monitor (heads-up digitizing) • (digital) photogrammetry
Automatic digitizing	• scanner
Semi-automatic digitizing	• line-following software
Input of available digital data	• CD-ROM or DVD-ROM • via computer network or internet (including geo-webservices)

Table 3.2: Spatial data input methods and devices used

Spatial data storage and maintenance

- The way that data is stored plays a central role in the processing and the eventual understanding of that data. In most of the available systems, spatial data is organized in layers by theme and/or scale
- For instance, the data may be organized in thematic categories, such as land use, topography and administrative subdivisions, or according to map scale.
- In a GIS, features are represented geometry of features is represented with primitives of the respective dimension: a windmill probably as a point, an agricultural field as a polygon. The primitives follow either the vector, as in the example, or the raster approach.
- vector data types describe an object through its boundary, thus dividing the space into parts that are occupied by the respective objects. The raster approach subdivides space into (regular) cells, mostly as a square tessellation of dimension two or three.
- These cells are called either *cells* or *pixels* in2D, and *voxels* in 3D. The data indicates for every cell which real world feature it covers, in case it represents a discrete field.
- In case of a continuous field, the cell holds a representative value for that field. Table3.3lists advantages and disadvantages of raster and vector representations.

Raster representation	*Vector representation*
advantages	
• simple data structure • simple implementation of overlays • efficient for image processing	• efficient representation of topology • adapts well to scale changes • allows representing networks • allows easy association with attribute data
disadvantages	
• less compact data structure • difficulties in representing topology • cell boundaries independent of feature boundaries	• complex data structure • overlay more difficult to implement • inefficient for image processing • more update-intensive

Table 3.3: Raster and vector representations compared

- GIS software packages provide support for both spatial and attribute data, i.e. they accommodate spatial data storage using a vector approach, and attribute data using tables. Historically, however, *database management systems* (DBMSs) have been based on the notion of tables for data storage.
- For some time, substantial GIS applications have been able to link to an external database to store attribute data and make use of its superior data management functions.
- currently, All major GIS packages provide facilities to link with a DBMS and ex- change attribute data with it.

Spatial query and analysis

- The most distinguishing parts of a GIS are its functions for spatial analysis, i.e. operators that use spatial data to derive new geoinformation.
- *Spatial queries* and *process models* play an important role in this functionality. One of the key uses of GISs has been to support spatial decisions.
- *Spatial decision support systems* (SDSS) are a category of information systems composed of a database, GIS software, models, and a so-called knowledge engine which allow users to deal specifically with locational problems.
- The analysis functions of a GIS use the spatial and non-spatial attributes of the data in a spatial database to provide answers to user questions. GIS functions are used for maintenance of the data, and for analysing the data in order to infer information from it.
- *Analysis* of spatial datacan be defined as computing new information that provides new insight from the existing, stored spatial data.

Spatial data presentation

- The presentation of spatial data, whether in print or on-screen, in maps or in tabular displays, or as 'raw data', is closely related to the disciplines of cartography, printing and publishing. The presentation may either be an end-product, for ex- ample as a printed atlas, or an intermediate product, as in spatial data made available through the internet.

Table 3.4: Spatial data presentation

Method	*Devices*
Hard copy	• printer • plotter (pen plotter, ink-jet printer, thermal transfer printer, electrostatic plotter) • film writer
Soft copy	• computer screen
Output of digital data sets	• magnetic tape • CD-ROM or DVD • the Internet

Table3.4lists several different methods and devices used for the presentation of spatial data. Cartography and scientific visualization make use of these methods and devices to produce their products.

Chapter 4 : Database management systems

- A *database* is a large, computerized collection of structured data.
- Designing a database is not an easy task.
- Firstly, one has to consider carefully what the database purpose is, and who its users will be.
- Secondly, one needs to identify the available data sources and define the format in which the data will be organized within the database. This format is usually called the *database structure*. Lastly, data can be entered into the database.

Reasons for using a DBMS

There are various reasons why one would want to use a DBMS for data storage and processing.

- A DBMS supports the storage and manipulation of *very large data sets*.
- A DBMS can be instructed to guard over *data correctness*.
- A DBMS supports the *concurrent use* of the same data set by many users.
- A DBMS provides a high-level, *declarative query language*
- A DBMS supports the use of a *data model*. A data model is a language with which one can define a database structure and manipulate the data stored in it.
- A DBMS includes *data backup* and *recovery* functions to ensure data avail- ability at all times.
- A DBMS allows the control of *data redundancy*.

Alternatives for data management

- The decision whether or not to use a DBMS will depend, among other things, on how much data there is or will be, what type of use will be made of it, and how many users might be involved.
- On the small-scale side of the spectrum—when the data set is small, its use relatively simple, and with just one user—we might use simple text files, and a text processor. Think of a personal address book as an example, or a small set of simple field observations. Text files offer no support for data analysis whatsoever, except perhaps in alphabetical sorting.
- If our data set is still small and numeric by nature, and we have a single type of use in mind, a spreadsheet program will suffice. This might be the case if we have a number of field observations with measurements that we want to prepare for statistical analysis, for example. However, if we carry out region- or nation- wide censuses, with many observation stations and/or field observers and all sorts of different measurements, one quickly needs a database to keep track of all the data. It should also be noted that spreadsheets do not accommodate con- current use of the data set well, although they do support some data analysis, especially when it comes to calculations over a single table, like averages, sums, minimum and maximum values.
- All such computations are usually restricted to just a single table of data. When one wants to relate the values in the table with values of another nature in some other table, some expertise and significant amounts of time are usually required to make this happen.

The relational data model

A data model is a language that allows the definition of:

- The structures that will be used to store the base data,
- The integrity constraints that the stored data has to obey at all moments in time, and
- The computer programs used to manipulate the data
  - For the relational data model, the structures used to define the database are attributes, tuples and relations. Computer programs either perform data extraction from the database without altering it, in which case we call them queries, or they change the database contents, and we speak of updates or transactions.
  - Let us look at a tiny database example from a cadastral setting. It is illustrated in Figure3.2. This database consists of three tables, one for storing people's details, one for storing parcel details and a third one for storing details concerning title deeds.
  - Various sources of information are kept in the database such as a taxation identifier (TaxId) for people, a parcel identifier (PId) for parcels and the date of a title deed (DeedDate).

PrivatePerson TaxId	Surname	BirthDate
101-367	Garcia	10/05/1952
134-788	Chen	26/01/1964
101-490	Fakolo	14/09/1931

Parcel PId	Location	AreaSize
3421	2001	435
8871	1462	550
2109	2323	1040
1515	2003	245

TitleDeed Plot	Owner	DeedDate
2109	101-367	18/12/1996
8871	101-490	10/01/1984
1515	134-788	01/09/1991
3421	101-367	25/09/1996

Figure 3.2: A small example database consisting of three relations (tables), all with three attributes, and resp. three, four and four tuples. PrivatePerson / Parcel / TitleDeed are the names of the three tables. Surname is an attribute of the PrivatePerson table; the Surname attribute value for person with TaxId '101-367' is 'Garcia.'

Relations, tuples and attributes

- In the relational data model, a database is viewed as a collection of *relations*, commonly also known as *tables*.
- A table or relation is itself a collection of *tuples* (or records). In fact, each table is a collection of tuples *that are similarly shaped.*
- By this, we mean that a tuple has a fixed number of named fields, also known as attributes. All tuples in the same relation have the same named fields. In a diagram, as in Figure3.2, relations can be displayed as tabular form data.
- An *attribute* is a named field of a tuple, with which each tuple associates a value, the tuple's *attribute value.*
- The example relations provided in the figure should clarify this. The Private- Person table has three tuples; the Surname attribute value for the first tuple illustrated is 'Garcia.'
-

PrivatePerson	(TaxId : string, Surname : string, Birthdate : date)
Parcel	(Pid : number, Location : polygon, AreaSize : number)
TitleDeed	(Plot : number, Owner : string, DeedDate : date)

Table 3.5: The relation schemas for the three tables of the database in Figure3.2.

- When a relation is created, we need to indicate what type of tuples it will store. This means that we must

1. Provide a *name* for the relation,
2. Indicate which *attributes* it will have, and
3. Set the *domain* of each attribute

Finding tuples and building links between them

- Database systems are particularly good at storing large quantities of data. The DBMS must support quick searches amongst many tuples. This is why the relational data model uses the notion of a key.
- A key of a relation comprises one or more attributes. A value for these attributes uniquely identifies a tuple. If we have a value for each of the key attributes we are guaranteed to find no more than one tuple in the table with that combination of values, such that there is no tuple for the given combination. Every relation has a key.
- A tuple can refer to another tuple by storing that other tuple's key value. This attribute is called a foreign key because it refers to the primary key of another relation. Two tuples of the same relation instance can have identical foreign key values.

Querying a relational database

- A query is a computer program that extracts data from the database that meet the conditions indicated in the query. The first query operator is called tuple selection; Tuple selection works like a filter: it allows tuples that meet the selection condition to pass, and disallows tuples that do not meet the condition.
- The operator is given some input relation, as well as a selection condition about tuples in the input relation. A selection condition is a truth statement about a tuple's attribute values such as: Distance <1000.
- The second operator is called attribute projection. Besides an input relation, this operator requires a list of attributes, all of which should be attributes of the schema of the input relation. Attribute projection works like a tuple formatter: it passes through all tuples of the input, but reshapes each of them in the same way.
- The output relation of this operator has as its schema only the list of attributes given, and we say that the operator projects onto these attributes. The most common way of defining queries in a relational database is through the SQL language. SQL stands for Structured Query Language.
- SELECT * FROM PARCEL WHERE AreaSize>1000.
- SELECT PId, Location FROM Parcel.
- SELECT queries do not create stored tables in the database. This is why the result tables have no name: they are virtual tables. The result of a query is a table that is shown to the user who executed the query. Whenever the user closes her/his view on the query result, that result is lost. The SQL code for the query can be stored for future use.
- The user can re-execute the query again to obtain a view on the result once more. Our third query operator differs from the two above in that it requires two input relations. The operator is called the join.
- It takes two input relations and produces one output relation, by gluing two tuples together, if they meet a specified condition The number of attributes therefore increases.

Chapter 5 : GIS and spatial databases

Linking GIS and DBMS

- GIS software provides support for spatial data and thematic or attribute data. GIS stores spatial data and attribute data separately. This required the GIS to provide a link between the spatial data (represented with rasters or vectors), and their non-spatial attribute data. The strength of GIS technology lies in its built-in 'understanding' of geographic space and all functions that derive from this, for purposes such as storage, analysis, and map production.
- GIS packages themselves can store tabular data; however, they do not always provide a full-fledged query language to operate on the tables. DBMSs have a long tradition in handling attribute (i.e. administrative, non-spatial, tabular, thematic) data in a secure way, for multiple users at the same time.
- DBMS offer much better table functionality, since they are specifically designed for this purpose. A iot of the data in GIS applications is attribute data, so it made sense to use a DBMS for it. For this reason, many GIS applications have made use of external DBMS for data support.
- With raster representations, each raster cell stores a characteristic value. This value can be used to look up attribute data in an accompanying database table.
- For instance, the land use raster of Figure3.7indicates the land use class for each of its cells, while an accompanying table provides full descriptions for all classes, including perhaps some statistical information for each of the types. Observe the similarity with the key/foreign key concept in relational databases.

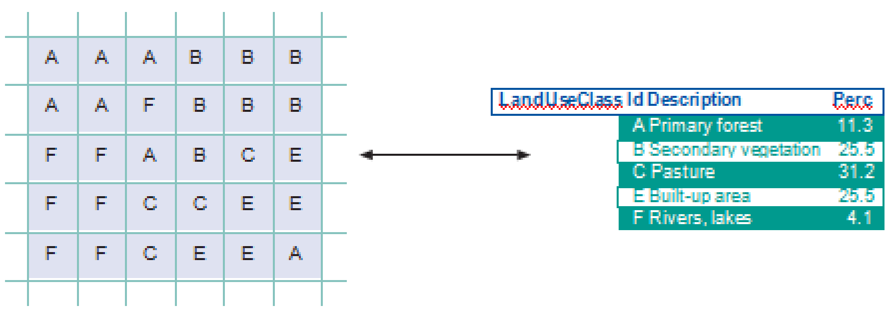


Figure 3.7: A raster representing land use and a related table providing full text descriptions (amongst others) of each land use class.

Spatial database functionality

- DBMS vendors have recognized the need for storing more complex data, like spatial data. The main problem was that there is additional functionality needed by DBMS in order to process and manage spatial data. Object-oriented and object-relational data models were developed for just this purpose. These extend standard relational models with support for objects, including 'spatial' objects.
- GIS software packages are able to store spatial data using a range of commercial and open source DBMSs such as Oracle, Informix, IBM DB2, Sybase, and PostgreSQL, with the help of spatial extensions. Some GIS software have integrated database 'engines', and therefore do not need these extensions.
- ESRI's ArcGIS and QGIS for example, have data base software built-in. This means that the

designer of a GIS application can choose whether to store the application data in the GIS or in the DBMS. Spatial databases, also known as geodatabases, are implemented directly on existing DBMS, using extension software to allow them to handle spatial objects.

- A spatial database allows users to store query and manipulate collections of spatial data.
- There are several advantages in doing this, spatial data can be stored in a special database column, known as the geometry column, (or feature or shape, depending on the specific software package),. This means GISs can rely fully on DBMS support for spatial data, making use of a DBMS for data query and storage (and multi-user support), and GIS for spatial functionality. Small-scale GIS applications may not require a multi-user capability, and can be supported by spatial data support from a personal database.
- A geodatabase allows a wide variety of users to access large data sets (both geographic and alphanumeric), and the management of their relations, guaranteeing their integrity. The Open Geospatial Consortium (OGC) has released a series of standards relating to geodatabases that (amongst other things), define :
    - Which tables must be present in a spatial database (i.e. geometry columns table and spatial reference system table)
    - The data formats, called 'Simple Features' (i.e. point, line, polygon, etc.)
    - A set of SQL-like instructions for geographic analysis.

Querying a spatial database

- A Spatial DBMS provides support for geographic co-ordinate systems and transformations. It also provides storage of the relationships between features, including the creation and storage of topological relationships.
- As a result one is able to use functions for 'spatial query' (exploring spatial relationships). To illustrate, a spatial query using SQL to find all the Metro City within 20 km of a River GANGA would look like this:
    - SELECT C. Name FROM River AS R, City as C WHERE C. Type = "METRO" AND R. name = "GANGA" AND ST Intersects (C. Geometry, CT_Buffer(R. Geometry, 20000)
- In this case the WHERE clause uses the ST_Intersects function to perform a spatial join between a 20000 m buffer of the selected River and the selected subset of Cities. The Geometry column carries the spatial data.

UNIT 3

SPATIAL REFERENCING AND POSITIONING

Chapter 1 : SPATIAL REFERENCING

One of the defining features of GIS is their ability to combine spatially referenced data. A frequently occurring issue is the need to combine spatial data from different sources that use different spatial reference systems. This section provides a broad background of relevant concepts relating to the nature of spatial reference systems and the translation of data from one spatial referencing system into another.

Reference surfaces for mapping

- The surface of the Earth is anything but uniform. The oceans can be treated as reasonably uniform, but the surface or topography of the land masses exhibits large vertical variations between mountains and valleys.
- These variations make it impossible to approximate the shape of the Earth with any reasonably simple mathematical model. Two main reference surfaces have been established to approximate the shape of the Earth. One reference surface is called the Geoid, the other reference surface is the ellipsoid as shown in the figure below.

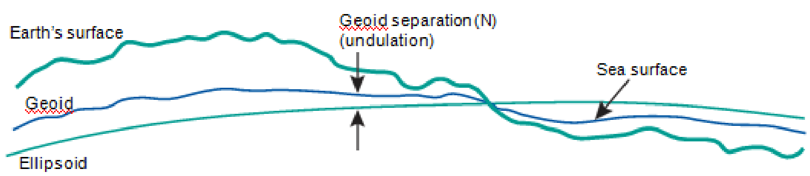


Figure 4.1: The Earth's surface, and two reference surfaces used to approximate it: the Geoid, and a reference ellipsoid. The Geoid separation (N) is the deviation between the Geoid and a reference ellipsoid.

The Geoid and the vertical datum

- Imagine that the entire Earth's surface is covered by water. If ignored tidal and current effects on this 'global ocean', the resultant water surface is affected only by gravity. This has an effect on the shape of this surface because the direction of gravity- more commonly known as plumb line-is dependent on the mass distribution inside the Earth.
- Due to irregularities or mass anomalies in this distribution the 'global ocean' results in an undulated surface. This surface is called the Geoid. The plumb line through any surface point is always perpendicular to it.

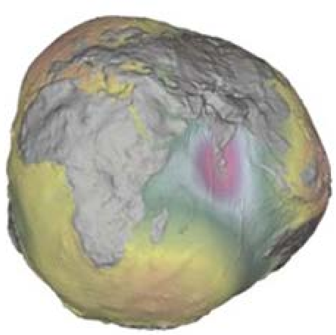


Figure 4.2: The Geoid, exaggerated to illustrate the complexity of its surface.

- The Geoid is used to describe heights. In order to establish the Geoid as reference for heights, the ocean's water level is registered at coastal places over several years using tide gauges (mareographs).
- Averaging the registrations largely eliminates variations of the sea level with time. The resulting water level represents an approximation to the Geoid and is called the mean sea level.

The ellipsoid

- The physical surface, called Geoid, is used as a reference surface for heights. Also a reference surface for the description of the horizontal coordinates of points of interest is required.
- This will later used to project these horizontal coordinates onto a mapping plane, the reference surface for horizontal coordinates requires a mathematical definition and description. The most convenient geometric reference is the oblate ellipsoid.
- It provides a relatively simple figure which fits the Geoid to a first order approximation, though for small scale mapping purposes a sphere may be used. An ellipsoid is formed when an ellipse is rotated about its minor axis. This ellipse which defines an ellipsoid or spheroid is called a meridian ellipse.

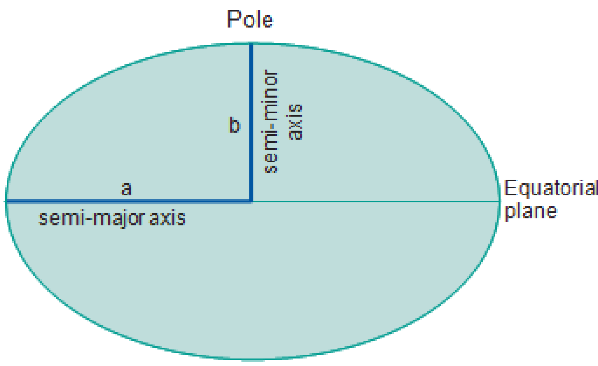


Figure 4.4: An oblate ellipse, defined by its semi-major axis a and semi-minor axis b.

- The shape of an ellipsoid may be defined in a number of ways, but in geodetic practice the definition is usually by its semi-major axis and flattening, Flattening f is dependent on both the semi-major axis a and the semi-minor axis b.

The local horizontal datum

- Ellipsoids have varying position and orientations. An ellipsoid is positioned and oriented with respect to the local mean sea level by adopting a latitude (ϕ) and longitude (λ) and ellipsoidal height (h) of a so-called fundamental point and an azimuth to an additional point.
- We say that this defines a *local horizontal datum*. Notice that the term horizontal datum and geodetic datum are being treated as equivalent and interchangeable words.
- A local horizontal datum is realized through a triangulation network. Such a network consists of monumented points forming a network of triangular mesh elements (Figure4.6). The angles in each triangle are measured in addition to at least one side of a triangle; the fundamental point is also a point in the triangulation network.
- The angle measurements and the adopted coordinates of the fundamental point are then used to derive geographic coordinates (ϕ, λ) for all monumented points of the triangulation network.

Figure 4.6: The old primary triangulation network in the Netherlands made up of 77 points (mostly church towers). The extension and re-measurement of the network is nowadays done through satellite measurements. Adapted from original figure by 'Dutch Cadastre and Land Registers' now called *het Kadaster*.

The global horizontal datum

- Local horizontal datums have been established to fit the Geoid well over the area of local interest, which in the past was never larger than a continent. With increasing demands for global surveying activities are underway to establish global reference surfaces.
- The objective is to make geodetic results mutually comparable and to provide coherent results also to other disciplines like astronomy and geophysics.
- The most important global (geocentric) spatial reference system for the GIS community is the International Terrestrial Reference System (ITRS).
- It is a three dimensional coordinate system with a well-defined origin (the centre of mass of the Earth) and three orthogonal coordinate axes (X, Y, Z).
- The Z-axis points towards a mean Earth north pole. The X-axis is oriented towards a mean Greenwich meridian and is orthogonal to the Z-axis. The Y -axis completes the right-handed reference coordinate system.

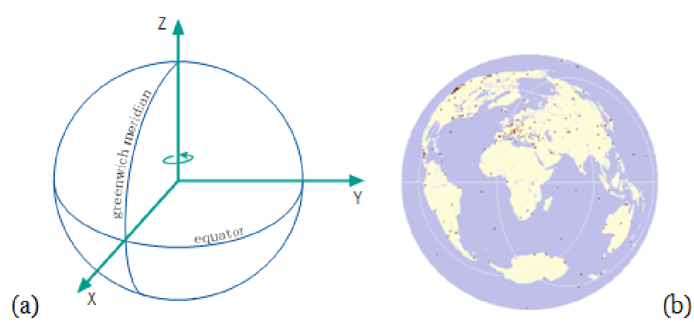


Figure 4.7: (a) The International Terrestrial Reference System (ITRS), and; (b) the International Terrestrial Reference Frame (ITRF) visualized as a distributed set of ground control stations (represented by red points).

- We can easily transform ITRF coordinates (X, Y and Z in metres) into geo- graphic coordinates (ϕ, λ,h) with respect to the GRS80 ellipsoid without the loss of accuracy. However, the ellipsoidal height h, obtained through this straight- forward transformation, has no physical meaning and does not correspond to intuitive human perception of height. We therefore use the height H, above the Geoid (see Figure4.8).

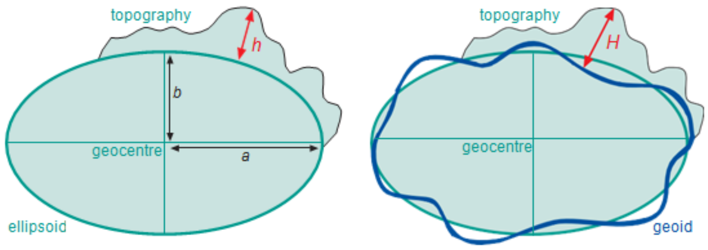


Figure 4.8: Height h above the geocentric ellipsoid, and height H above the Geoid. The first is measured orthogonal to the ellipsoid, the second orthogonal to the Geoid.

Coordinate systems

- Different kinds of coordinate systems are used to position data in space. Spatial (or global) coordinate systems are used to locate data either on the Earth's surface in a 3D space, or on the Earth's reference surface in a 2D space. The geographic coordinate system in 2D and 3D space and the geocentric coordinate system, also known as the 3D Cartesian coordinate system. Planar coordinate systems on the other hand are used to locate data on the flat surface of the map in a 2D space.

1. 2D Geographic coordinates (ϕ, λ)

- The most widely used global coordinate system consists of lines of geographic latitude (phi or cf) or <p) and longitude (lambda or A). Lines of equal latitude are called parallels. They form circles on the surface of the ellipsoid4. Lines of equal longitude are called meridians and they form ellipses (meridian ellipses) on the ellipsoid.

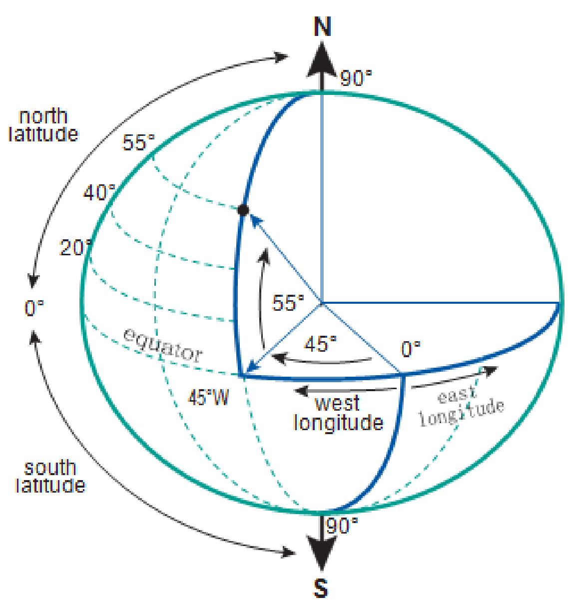


Figure 4.9: The latitude (φ) and longitude (λ) angles represent the 2D geographic coordinate system.

- The latitude (cf>) of a point P is the angle between the ellipsoidal normal through P ' and the equatorial plane. Latitude is zero on the equator (cf» = 0°), and increases towards the two poles to maximum values of 4> = +90° (N 90°) at the North Pole and cj) = -90° (S 90°) at the South Pole.
- The longitude (A) is the angle between the meridian ellipse which passes through Greenwich and the meridian ellipse containing the point in question. It is measured in the equatorial plane from the meridian of Greenwich (A = 0°) either eastwards through A = + 180° (E 180°) or westwards through A = -180° (W 180°).

2. **3D Geographic coordinates** (ϕ, λ, h)

- 3D geographic coordinates (ϕ, λ, h) are obtained by introducing the ellipsoidal height h to the system. The ellipsoidal height (h) of a point is the vertical distance of the point in question above the ellipsoid.
- It is measured in distance units along the ellipsoidal normal from the point to the ellipsoid surface. 3D geographic coordinates can be used to define a position on the surface of the Earth (point P in Figure4.10).

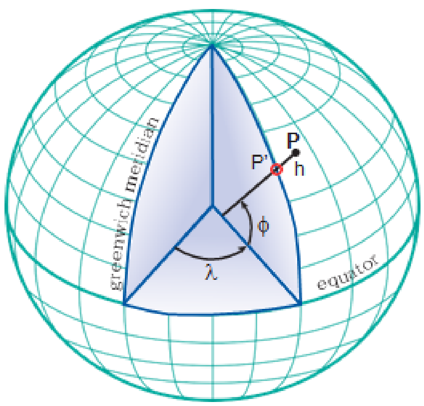


Figure 4.10: The latitude (φ) and longitude (λ) angles and the ellipsoidal height (h) represent the 3D geographic coordinate system.

3. **3D Geocentric coordinates** (X, Y, Z)

- An alternative method of defining a 3D position on the surface of the Earth is by means of geocentric coordinates (X,Y,Z), also known as *3D Cartesian coordi- nates*.
- The system has its origin at the mass-centre of the Earth with the X and Y axes in the plane of the equator. The X-axis passes through the meridian of Greenwich, and the Z-axis coincides with the Earth's axis of rotation.
- The three axes are mutually orthogonal and form a right-handed system. Geocentric coordinates can be used to define a position on the surface of the Earth (point P in Figure4.11).

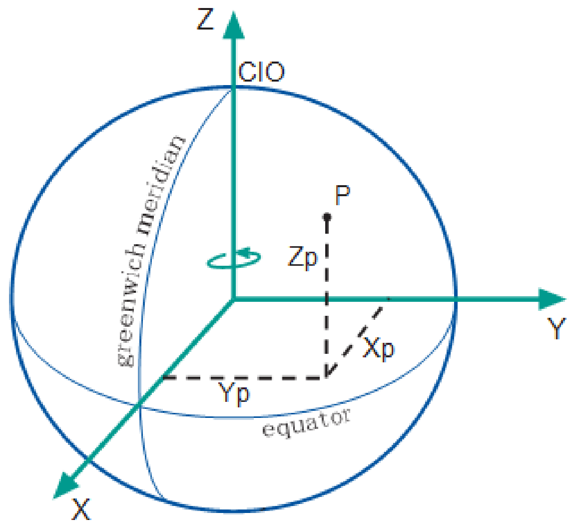


Figure 4.11: An illustration of the 3D geocentric coordinate system (see text for further explanation).

- It should be noted that the rotational axis of the earth changes its position over time (referred to as polar motion). To compensate for this, the mean position of the pole in the year 1903 (based on observations between 1900 and 1905) has been used to define the so-called Conventional International Origin (CIO).

4. **2D Cartesian coordinates** (X, Y)

- A flat map has only two dimensions: width (left to right) and length (bottom to top). Transforming the three dimensional Earth into a two-dimensional map is subject of map projections and coordinate transformation. Like in several other cartographic applications, two-dimensional Cartesian coordinates (x, y), also known as planar rectangular coordinates, are used to describe the location of any point unambiguously. The two coordinates x and y for point P, specify any location P on the map.

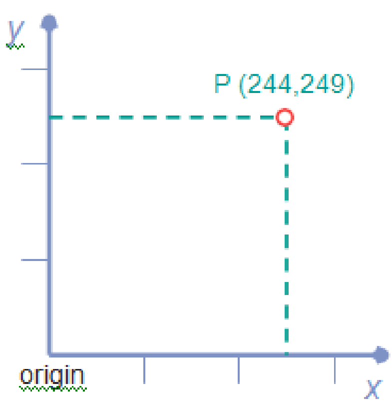


Figure 4.12: An illustration of the 2D Cartesian coordinate system (see text for further explanation).

5. **2D Polar coordinates** (α,d)

- Polar coordinate is the distance "d" from the origin to the point concerned and the angle a between a fixed (or zero) direction and the direction to the point. The angle a is called azimuth or bearing and is measured in a clockwise direction.
- It is given in angular units while the distance d is expressed in length units. Bearings are always related to a fixed direction (initial bearing) or a datum line.

- In principle, this reference line can be chosen freely. However, in practice three different directions are widely used: True North, Grid North and Magnetic North. The corresponding bearings are called: true (or geodetic) bearing, grid bearing and magnetic (or compass) bearing.

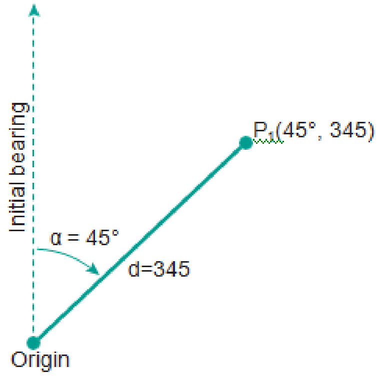


Figure 4.14: An illustration of the 2D Polar coordinate system (see text for further explanation).

Map projections

- A map projection is a mathematically described technique of how to represent the Earth's curved surface on a flat map. To represent parts of the surface of the Earth on a flat paper map or on a computer screen, the curved horizontal reference surface must be mapped onto the 2D mapping plane. The reference surface for large-scale mapping is usually an oblate ellipsoid, and for small-scale mapping, a sphere.
- Mapping onto a 2D mapping plane means transforming each point on the reference surface with geographic coordinates (<f>, A) to a set of Cartesian coordinates (x, y) representing positions on the map plane.

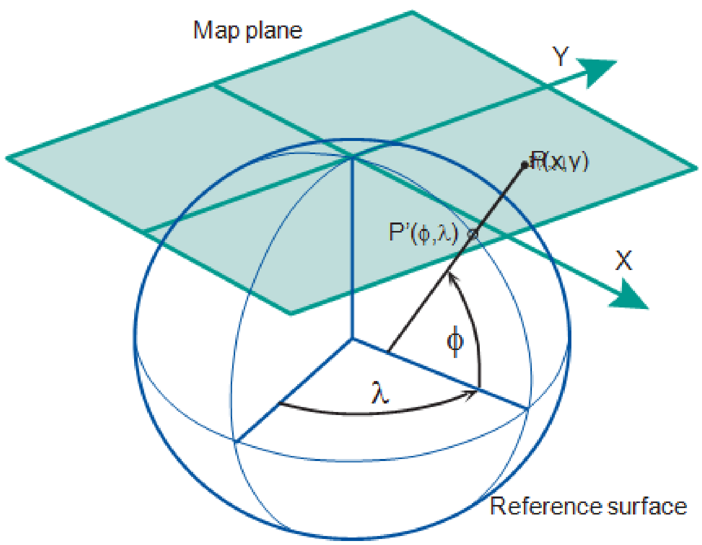


Figure 4.15: Example of a map projection where the reference surface with geographic coordinates (φ, λ) is projected onto the 2D mapping plane with 2D Cartesian coordinates (x, y).

- The actual mapping cannot usually be visualized as a true geometric projection, directly onto the mapping plane. This is achieved through mapping equations.
- A forward mapping equation transforms the geographic coordinates (4>, A) of a point on the curved reference surface to a set of planar Cartesian coordinates (x, y), representing the position of the same

point on the map plane : (x, y) = f((f>, A) The corresponding inverse mapping equation transforms mathematically the planar Cartesian coordinates (x, y) of a point on the map plane to a set of geographic coordinates (cf>, A) on the curved reference surface: ($, A) = f(x, y).

Classification of map projections

- A large number of map projections have been developed, each with its own specific qualities. These qualities in turn make resulting maps useful for certain purposes. By definition, any map projection is associated with scale distortions.
- There is simply no way to flatten out a piece of ellipsoidal or spherical surface without stretching some parts of the surface more than others. Some map projections can be visualized as true geometric projections directly onto the mapping plane, in which case we call it an azimuthal projection, or onto an intermediate surface, which is then rolled out into the mapping plane.
- Typical choices for such intermediate surfaces are cones and cylinders. Such map projections are then called conical, and cylindrical, respectively.

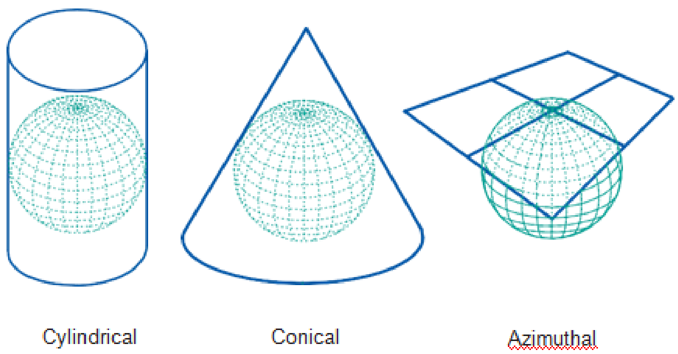


Figure 4.16: Classes of map projections

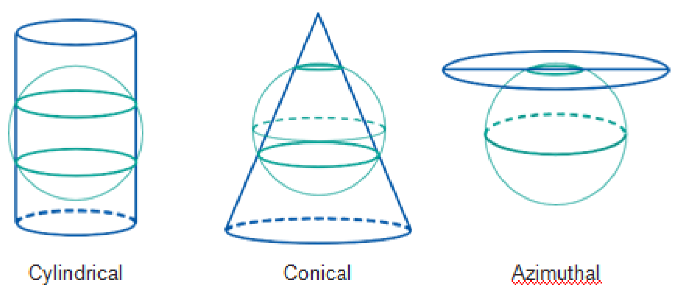


Figure 4.17: Three secant projection classes

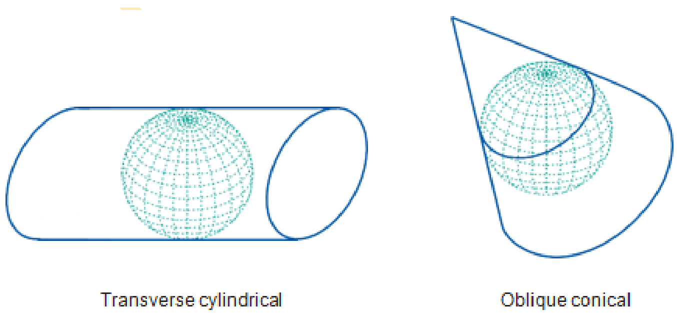


Figure 4.18: A transverse and an oblique projection

Coordinate transformations

- Map and GIS users are mostly confronted in their work with transformations from one two-dimensional coordinate system to another. This includes the trans- formation of polar coordinates delivered by the surveyor into Cartesian map coordinates or the transformation from one 2D Cartesian (x,y) system of a spe- cific map projection into another 2D Cartesian (x^j, y^j) system of a defined map projection.
- Datum transformations are transformations from a 3D coordinate system (i.e. horizontal datum) into another 3D coordinate system. These kinds of transfor- mations are also important for map and GIS users. They are usually collecting spatial data in the field using satellite navigation technology and need to repre- sent this data on published map on a local horizontal datum.

1. 2D Polar to 2D Cartesian transformations

- The transformation of polar coordinates (a, d), into Cartesian map coordinates (x, y) is done when field measurements, angular and distance measurements are transformed into map coordinates. The equation for this transformation is:

  x = d(sin(a))
  y = d(cos(a))
  The inverse equation is: a = tan'1 (x/y)
  d2 = x2 + y2

- A more realistic case makes use of a translation and a rotation to transform one system to the other.

Changing map projection

- Forward and inverse mapping equations are normally used to transform data from one map projection to another. The inverse equation of the source projection is used first to transform source projection coordinates (x,y) to geographic coordinates (ϕ, λ).
- Next, the forward equation of the target projection is used to transform the geographic coordinates (ϕ,λ) into target projection coordinates (x^j, y^j).
- The first equation takes us from a projection A into geographic coordinates. The second takes us from geographic coordinates (ϕ, λ) to another map projection B. These principles are illustrated in Figure4.22.

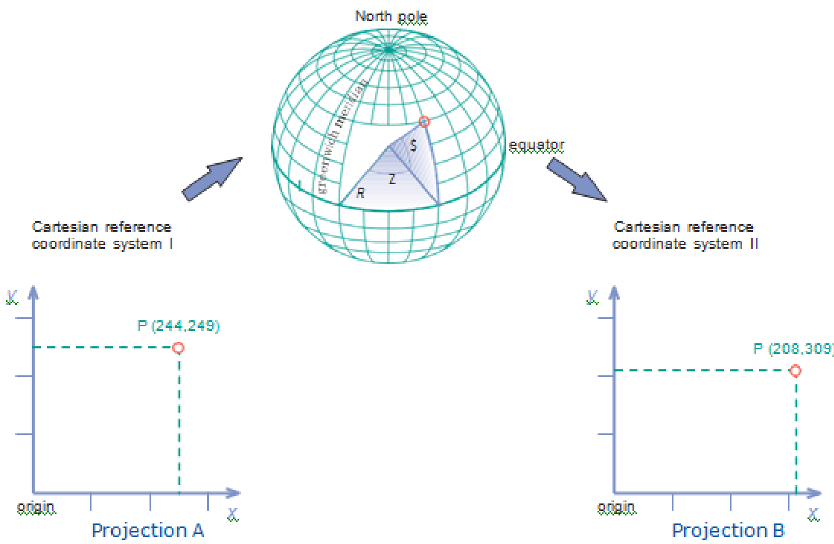


Figure 4.22: The principle of changing from one map projection into another.

Datum transformations

- A change of map projection may also include a change of the horizontal datum. This is the case when the source projection is based upon a different horizontal datum than the target projection. If the difference in horizontal datums is ignored, there will not be a perfect match between adjacent maps of neighboring countries or between overlaid maps originating from different projections.
- It may result in up to several hundred meters difference in the resulting coordinates. Therefore, spatial data with different underlying horizontal datums may need a so-called datum transformation.
- Suppose we wish to transform spatial data from the UTM projection to the Dutch RD system, and that the data in the UTM system are related to the European Datum 1950 (ED50), while the Dutch RD system is based on the Amersfoort datum.
- In this example the change of map projection should be combined with a datum transformation step for a perfect match. This is illustrated in Figure4.23.

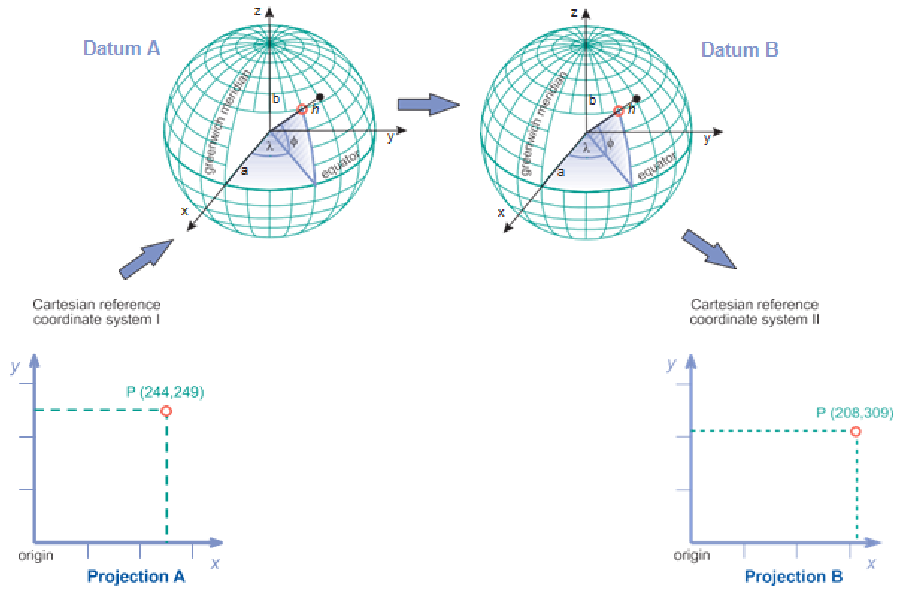


Figure 4.23: The principle of changing from one projection into another, combined with a datum transformation from datum A to datum B.

Chapter 2 : Satellite-based positioning

- Satellites are used in geocentric reference systems, and increase the level of spatial accuracy substantially. They are critical tools in geodetic engineering for the maintenance of the ITRF. They also play a key role in mapping, surveying, and in a growing number of applications requiring positioning techniques.
- Nowadays, for fieldwork that includes spatial data acquisition, the use of satellite-based positioning is considered indispensable. Satellite-based positioning was developed and implemented to address military needs, somewhat analogously to the early development of the internet.
- The technology is now widely available for civilians use. The requirements for the development of the positioning system were:
  - ✓ Suitability for all kinds of military use: ground troops and vehicles, aircraft and missiles, ships;
  - ✓ Requiring only low-cost equipment with low energy consumption at the receiver end;
  - ✓ Provision of results in real time for an unlimited number of users concurrently;
  - ✓ Support for different levels of accuracy (military versus civilian);
  - ✓ Around-the-clock and weather-proof availability;
  - ✓ Use of a single geodetic datum;
  - ✓ Protection against intentional and unintentional disturbance, for instance, through a design allowing for redundancy.
- A satellite-based positioning system set-up involves implementation of three hardware segments:

1. The ***space segment***, i.e. the satellites that orbit the Earth, and the radio signals that they emit,
2. The ***control segment***, i.e. the ground stations that monitor and maintain the space segment components, and
3. The ***user segment***, i.e. the users with their hard- and software to conduct positioning

Absolute positioning

The working principles of absolute, satellite-based positioning are fairly simple:

1. A satellite, equipped with a clock, at a specific moment sends a radio message that includes:
   a) The satellite identifier,
   b) Its position in orbit, and
   c) Its clock reading.
2. A receiver on or above the planet, also equipped with a clock, receives the message slightly later, and reads its own clock.
3. From the time delay observed between the two clock readings, and knowing the speed of radio transmission through the medium between (satelite) sender and receiver, the receiver can compute the distance to the sender, also known as the satellite's *pseudorange*.

- The *pseudorange* of a satellite with respect to a receiver, is its apparent distance to the receiver, computed from the time delay with which its radio signal is received.
- Such a computation determines the position of the receiver to be on a sphere of radius equal to the computed pseudorange (refer to Figure4.24(a)).

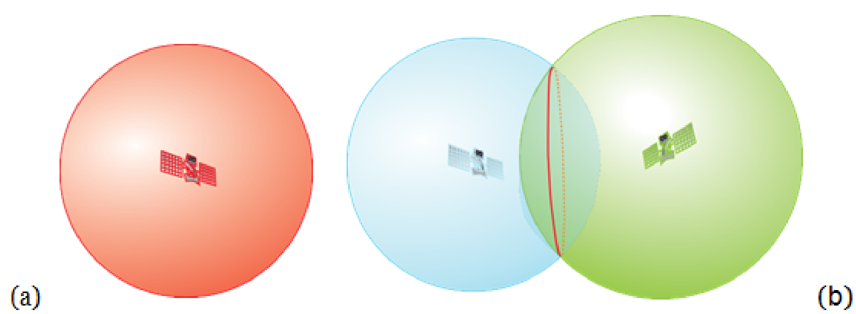


Figure 4.24: Pseudorange positioning: (a) With just one satellite the position is determined by a sphere, (b) With two satellites, it is determined by the intersection of two spheres, a circle. Not shown: with three satellites, it is the intersection of three spheres.

Time, clocks and world time

- While latitude was determined with a sextant from the position of the Sun in the sky, they carried clocks with them to determine the longitude of their position. Early ship clocks were unreliable, having a drift of multiple seconds a day, which could result in positional error of a few kilometers.
- Before any notion of standard time existed, villages and cities simply kept track of their local time determined from position of the Sun in the sky. When trains became an important means of transportation, these local time systems became problematic as the schedules required a single time system.
- Such a time system needed the definition of time zones: typically as 24 geographic strips between certain longitudes that are multiples of 15°. This all gave rise to Greenwich Mean Time (GMT). GMT was the world time standard of choice. It was a system based on the mean solar time at the meridian of Greenwich, United Kingdom, which is the conventional O-meridian in geography.

Errors in absolute positioning

Errors related to the space segment

- As a first source of error, the operators of the control segment may intentionally deteriorate radio signals of the satellites to the general public, to avoid optimal use of the system by the enemy, for instance in times of global political tension and war.
- This selective availability—meaning that the military forces allied with the control segment will still have access to undisturbed signals—may cause error that is an order of magnitude larger than all other error sources combined.
- Secondly, the satellite message may contain incorrect information. Assuming that it will always know its own identifier, the satellite may make two kinds of error:

1. **Incorrect clock reading**

- Even atomic clocks can be off by a small margin, and since Einstein, we know that travelling clocks are slower than resident clocks, due to a so-called relativistic effect. If one understands that a clock that is off by 0.000001 sec causes a computation error in the satellite's pseudorange of approximately 300 m, it is clear that these satellite clocks require very strict monitoring.

2. **Incorrect orbit position**

- The orbit of a satellite around our planet is easy to describe mathematically if both bodies are considered point masses, but in real life they are not.
- For the same reasons that the Geoid is not a simply shaped surface, the Earth's gravitation field that a satellite experiences in orbit is not simple either. Moreover, it is disturbed by solar and lunar gravitation, making its flight path slightly erratic and difficult to forecast exactly.

Errors related to the medium

- The medium between sender and receiver may be of influence to the radio signals. The middle atmospheric layers of stratosphere and mesosphere are relatively harmless and of little hindrance to radio waves, but this is not true of the lower and upper layer. They are, respectively:
  - The **troposphere** : the approximate 14 km high airspace just above the Earth's surface, which holds much of the atmosphere's oxygen and which envelopes all phenomena that we call the weather. It is an obstacle that delays radio waves in a rather variable way.
  - The **ionosphere** : the most outward part of the atmosphere that starts at an altitude of 90 km, holding many electrically charged atoms, thereby forming a protection against various forms of radiation from space, including to some extent radio waves. The degree of ionization shows a distinct night and day rhythm, and also depends on solar activity.

Errors related to the receiver's environment

- The error occurring when a radio signal is received via two or more paths between sender and receiver, some of which typically via a bounce off of some nearby surface, like a building or rock face. The term applied to this phenomenon is multi-path; when it occurs the multiple receptions of the same signal may interfere with each other. Multipath is a difficult to avoid error source.

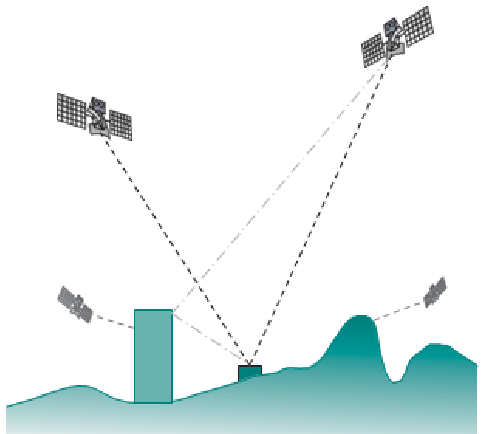


Figure 4.26: At any point in time, a number of satellites will be above the receiver's horizon. But not all of them will be 'in view' (like the left and right satellites), and for others multi-path signal reception may occur.

Errors related to the relative geometry of satellites and receiver

- There is one more source of error that is unrelated to individual radio signal characteristics, but that rather depends on the combination of the satellite signals used for positioning. Of importance is their constellation in the sky from the receiver perspective.
- Referring to Figure4.27, one will understand that the sphere intersection technique of positioning will provide more precise results when the four satellites are nicely spread over the sky, and thus that the satellite constellation of Figure4.27(b) is preferred over the one of4.27(a).

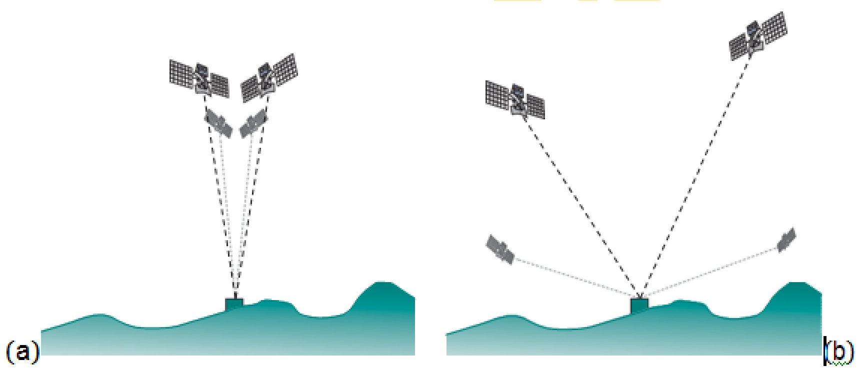


Figure 4.27: Geometric dilution of precision. The four satellites used for positioning can be in a bad constellation (a) or in a better constellation (b).

Relative positioning

- One technique to remove errors from positioning computations is to perform many position computations, and to determine the average over the solutions. Many receivers allow the user to do so.
- It should however be clear from the above that averaging may address random errors like signal noise, selective availability (SA) and multi-path to some extent, but not systematic sources of error, like incorrect satellite data, atmospheric delays, and GDOP effects. These sources should be removed before averaging is applied. It has been shown that averaging over 60 minutes in absolute, single-point positioning based on code measurements, before systematic error removal, leads only to a 10-20% improvement of accuracy.

- In such cases, receiver averaging is therefore of limited value, and requires long periods under near-optimal conditions.
- In relative positioning, also known as differential positioning, one tries to remove some of the systematic error sources by taking into account measurements of these errors in a nearby stationary reference receiver with an accurately known position.
- By using these systematic error findings at the reference, the position of the target receiver of interest will become known much more precisely.

Network positioning

- Network positioning is an integrated, systematic network of reference receivers covering a large area like a continent or even the whole globe. The organization of such a network can take different shapes, augmenting an already existing satellite-based system.
- A general architecture consists of a network of reference stations, strategically positioned in the area to be covered, each of which is constantly monitoring signals and their errors for all positioning satellites in view. One or more control centres receive the reference station data, verify this for correctness, and relay (uplink) this information to a geostationary satellite.
- The satellite will retransmit the correctional data to the area that it covers, so that target receivers, using their own approximate position, can determine how to correct for satellite signal error, and consequently obtain much more accurate position fixes.

Code versus phase measurements

- So far, we have assumed that the receiver determines the range of a satellite by measuring time delay on the received ranging code. There exists a more advanced range determination technique known as carrier phase measurement. This typically requires more advanced receiver technology, and longer observation sessions.
- Carrier phase measurement can currently only be used with relative positioning, as absolute positioning using this method is not yet well developed. The technique aims to determine the number of cycles of the (sine-shaped) radio signal between sender and receiver.
- Each cycle corresponds to one wavelength of the signal, which in the applied L-band frequencies is 19-24 cm. Since this number of cycles cannot be directly measured, it is determined, in a long observation session, from the change in carrier phase with time. This happens because the satellite is orbiting itself. From its orbit parameters and the change in phase over time, the number of cycles can be derived.

Positioning technology

- At present, two satellite-based positioning systems are operational (GPS and GLONASS), and a third is in the implementation phase (Galileo). Respectively, these are American, Russian and European systems. Any of these, but especially GPS and Galileo, will be improved over time, and will be augmented with new techniques.

GPS

- The NAVSTAR Global Positioning System (GPS) was declared operational in 1994, providing Precise Positioning Services (PPS) to US and allied military forces as well as US government agencies, and Standard Positioning Services (SPS) to civilians throughout the world.
- Its space segment nominally consists of 24 satellites, each of which orbits the Earth in llh58m at an altitude of 20,200 km. There can be any number of satellites active, typically between 21 and 27.
- The satellites are organized in six orbital planes, somewhat irregularly spaced, with an angle of inclination of 55-63° with the equatorial plane, nominally having four satellites each (see Figure 4.28).
- This means that a receiver on Earth will have between five and eight (sometimes up to twelve) satellites in view at any point in time. Software packages exist to help in planning GPS surveys, identifying expected satellite set-up for any location and time.

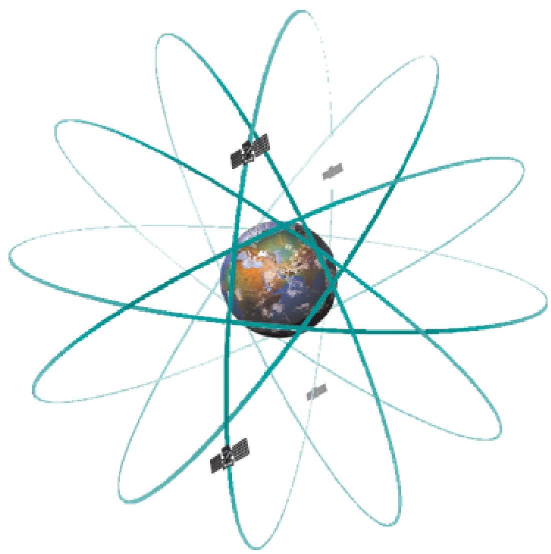


Figure 4.28: Constellation of satellites, four shown in only one orbit plane, in the GPS system.

GLONASS

- What GPS is to the US military, is GLONASS to the Russian military, specifically the Russian Space Forces. Both systems were primarily designed on the basis of military requirements.
- The big difference between the two is that GPS generated a major interest in civil applications, thus having an important economic impact. This cannot be said of GLONASS.
- The GLONASS space segment consists of nominally 24 satellites, organized in three orbital planes, with an inclination of 64.8° with the equator. Orbiting altitude is 19,130 km, with a period of revolution of 11 hours 16 min. GLONASS uses the PZ–90 as its reference system, and like GPS uses UTC as time reference, though with an offset for Russian daylight.

Galileo

- In the 1990's, the European Union (EU) judged that it needed to have its own satellite-based positioning system, to become independent of the GPS monopoly and to support its own economic growth by providing services of high reliability under civilian control.
- Galileo is the name of this EU system. The vision is that satellite-based positioning will become even bigger due to the emergence of mobile phones equipped with receivers, perhaps with some 400 million users by the year 2015.
- Development of the system has experienced substantial delays, and at the time of writing European ministers insist that Galileo should be up and running by the end of 2013. The completed system will have 27 satellites, with three in reserve, orbiting in one of three, equally spaced, circular orbits at an elevation of 23,222 km, inclined 56° with the equator. This higher inclination, when compared to that of GPS, has been chosen to provide better positioning coverage at high latitudes, such as northern Scandinavia where GPS performs rather poorly.

DATA ENTRY AND PREPARATION

Chapter 3 : Spatial data input

Spatial data can be obtained from various sources. It can be collected from scratch, using direct spatial data acquisition techniques, or indirectly, by making use of existing spatial data collected by others.

Direct spatial data capture

- One way to obtain spatial data is by direct observation of the relevant geographic phenomena. This can be done through ground-based field surveys, or by using remote sensors in satellites or airplanes.
- Many Earth sciences have developed their own survey techniques, as ground-based techniques remain the most important source for reliable data in many cases.
- Data which is captured directly from the environment is known as *primary data*
- Remotely sensed imagery is usually not fit for immediate use, as various sources of error and distortion may have been present, and the imagery should first be freed from these. This is the domain of remote sensing, and these issues are discussed further in ***Principles of Remote Sensing***.
- An *image* refers to raw data produced by an electronic sensor, which are not pictorial, but arrays of digital numbers related to some property of an object or scene, such as the amount of reflected light.
- *Factors of cost and available time may be a hindrance in using existing remotely sensed images because previous projects sometimes have acquired data that may not fit the current project's purpose.*

Indirect spatial data capture

- In contrast to direct methods of data capture described above, spatial data can also be sourced indirectly. This includes data derived from existing paper maps through scanning, data digitized from a satellite image, processed data purchased from data capture firms or international agencies, and so on. This type of data is known as secondary data.
- Any data which is not captured directly from the environment is known as secondary data.

Digitizing

- A traditional method of obtaining spatial data is through digitizing existing paper maps. This can be done using various techniques. Before adopting this approach, one must be aware that positional errors already in the paper map will further accumulate, and one must be willing to accept these errors.
- There are two forms of digitizing: *on-tablet* and *on-screen* manual digitizing. In on-tablet digitizing, the original map is fitted on a special surface (the tablet), while in on-screen digitizing, a scanned image of the map (or some other image) is shown on the computer screen.
- In both of these forms, an operator follows the map's features with a mouse device, thereby tracing the lines, and storing location coordinates relative to a number of previously defined control points.
- The function of these points is to 'lock' a coordinate system onto the digitized data: the control points on the map have known coordinates, and by digitizing them we tell the system implicitly where all other digitized locations are. At least three control points are needed, but preferably more should be digitized to allow a check on the positional errors made.

Scanning

- A scanner is an input device that illuminates a document and measures the intensity of the reflected light with a CCD array. The result is an image as a matrix of pixels, each of which holds an intensity value.
- Office scanners have a fixed maximum resolution, expressed as the highest number of pixels they can identify per inch; the unit is dots-per- inch (dpi). For manual on-screen digitizing of a paper map, a resolution of 200-300 dpi is usually sufficient, depending on the thickness of the thinnest lines. For manual on-screen digitizing of aerial photographs, higher resolutions are recommended — typically, at least 800 dpi.
- After scanning, the resulting image can be improved with various image processing techniques. It is important to understand that scanning does *not* result in a structured data set of classified and coded objects. Additional work is required to recognize features and to associate categories and other thematic at- tributes with them.

Vectorization

- The process of distilling points, lines and polygons from a scanned image is called vectorization. As scanned lines may be several pixels wide, they are often first thinned to retain only the centreline. The remaining centreline pixels are converted to series of (x, y) coordinate pairs, defining a polyline.
- Subsequently, features are formed and attributes are attached to them. This process may be entirely automated or performed semi- automatically, with the assistance of an operator. Pattern recognition methods—like Optical Character Recognition (OCR) for text—can be used for the automatic detection of graphic symbols and text.

- Vectorization causes errors such as small spikes along lines, rounded comers, errors in T- & X-junctions, displaced lines or jagged curves. These errors are corrected in an automatic or interactive post-processing phase. The phases of the vectorization process are illustrated in Figure below.

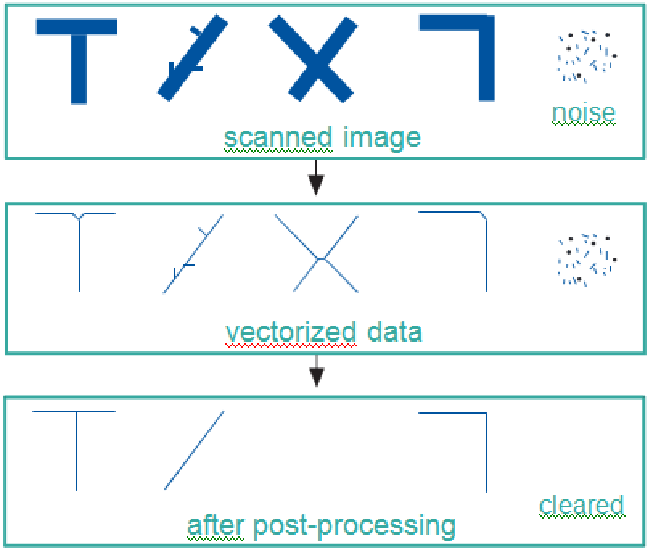


Figure 5.1: The phases of the vectorization process and the various sorts of small error caused by it. The post-processing phase makes the final repairs.

Selecting a digitizing technique

- The choice of digitizing technique depends on the quality, complexity and con- tents of the input document. Complex images are better manually digitized; simple images are better automatically digitized. Images that are full of detail and symbols—like topographic maps and aerial photographs—are therefore better manually digitized.
- The optimal choice may be a combination of methods. For example, contour line film separations can be automatically digitized and used to produce a DEM. Existing topographic maps must be digitized manually, but new, geometrically corrected aerial photographs, with vector data from the topographic maps displayed directly over it, can be used for updating existing data files by means of manual on-screen digitizing.

Obtaining Spatial Data Elsewhere

- Spatial data has been collected in digital form at increasing rate, stored in various databases by the individual producers for their own use and for commercial purposes. More and more of this data is being shared among GIS users. This is for several reasons.
- Some of this data is freely available, although other data is only available commercially, as is the case for most satellite imagery. High quality data remain both costly and time- consuming to collect and verify, as well as the fact that more and more GIS applications are looking at not just local, but national or even global processes.

Clearinghouses and web portals

- Spatial data can also be acquired from centralized repositories. More often those repositories are embedded in Spatial Data Infrastructures, which make the data available through what is sometimes called a spatial data clearinghouse.
- This is essentially a marketplace where data users can 'shop'. It will be no surprise that such markets for digital data have an entrance through the internet. The first entrance is typically

formed by a web portal which categorizes all available data and provides a local search engine and links to data documentation also called metadata.

Metadata

- Metadata is defined as background information that describes all necessary information about the data itself. More generally, it is known as 'data about data'.
- This includes: • Identification information : Data source(s), time of acquisition, etc. • Data quality information : Positional, attribute and temporal accuracy, lineage, etc. • Entity and attribute information: Related attributes, units of measure, etc.

Data formats and standards

- An important problem in any environment involved in digital data exchange is that of data formats and data standards. Different formats were implemented by different GIS vendors; different standards came about with different standardization committees. The phrase 'data standard' refers to an agreed upon way of representing data in a system in terms of content, type and format.
- The good news about both formats and standards is that there are many to choose from; the bad news is that this can lead to a range of conversion problems. Several metadata standards for digital spatial data exist, including the International Organization for Standardization (ISO) and the Open Geospatial Consortium (OGC) standards.

Chapter 4 : DATA QUALITY

- GIS is being increasingly used for geospatial decision support applications, with increasing reliance on secondary data sourced through data providers or via the internet, through geo-webservices.
- The implications of using low-quality data in important decisions are potentially severe. There is also a danger that uninformed GIS users introduce errors by incorrectly applying geometric and other transformations to the spatial data held in their database.
- The main issues related to data quality in spatial data are positional, temporal and attribute accuracy, lineage, completeness, and logical consistency.

Accuracy and Precision

- Accuracy should not be confused with precision, which is a statement of the smallest unit of measurement to which data can be recorded. In conventional surveying and mapping practice, accuracy and precision are closely related. Instruments with an appropriate precision are employed and surveying methods chosen, to meet specified accuracy tolerances. In GIS, however, the numerical precision of computer processing and storage usually exceeds the accuracy of the data.
- Using graphs that display the probability distribution (for which see below) of a measurement against the true value T, the relationship between accuracy and precision can be clarified. In Figure5.2, we depict the cases of good/bad accuracy against good/bad precision. An *accurate* measurement has a mean close to the true value; a *precise* measurement has a sufficiently small variance.

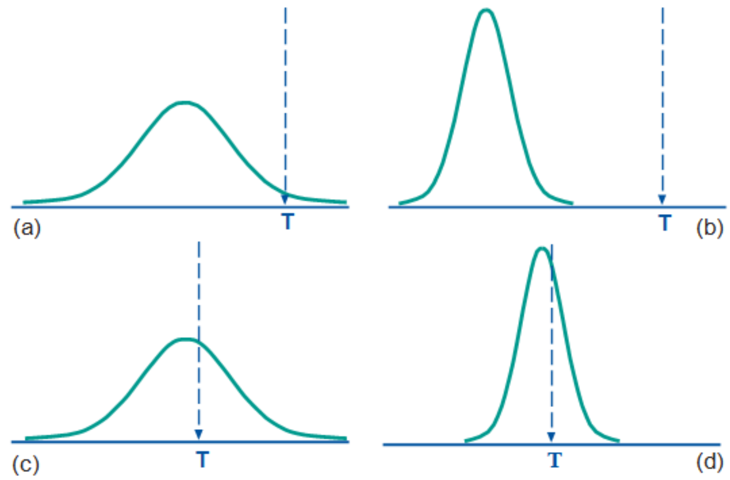


Figure 5.2: A measurement probability function and the underlying true value T: (a) bad accuracy and precision, (b) bad accuracy/good precision, (c) good accuracy/bad precision, and (d) good accuracy and precision.

Positional accuracy

- The surveying and mapping profession has a long tradition of determining and minimizing errors. This applies particularly to land surveying and photogrammetry, both of which tend to regard positional and height errors as undesirable.
- Cartographers also strive to reduce geometric and attribute errors in their products, and, in addition, define quality in specifically cartographic terms, for example quality of linework, layout, and clarity of text. It must be stressed that all measurements made with surveying and photogrammetric instruments are subject to error.

- **These include:**
  1. Human errors in measurement (e.g. reading errors) generally referred to as gross errors or blunders. These are usually large errors resulting from care lessness which could be avoided through careful observation, although it is never absolutely certain that all blunders have been avoided or eleminated.
  2. Instrumental or systematic errors (e.g. due to misadjustment of instruments). This leads to errors that vary systematically in sign and/or magnitude, but can go undetected by repeating the measurement with the same instrument. Systematic errors are paticularly dangerous because they tend to accumulate.
  3. Random errors caused by natural variations in the quantity being measured. These are effectively the errors that remain after blunders and systematic errors have been removed. They are usually small, and dealt with in least-squares adjustment, more general ways of quantifying positional accuracy using root mean square error (RMSE).

- Measurement errors are generally described in terms of accuracy. In the case of spatial data, accuracy may relate not only to the determination of coordinates (positional error) but also to the measurement of quantitative attribute data. The accuracy of a single measurement can be defined as:
- "The closeness of observations, computations or estimates to the true values or the values perceived to be true".

Accuracy tolerances

- Many kinds of measurement can be naturally represented by a bell-shaped probability density function p. This function is known as the normal (or Gaussian) distribution of a continuous, random variable, in the figure indicated as Y. It shape is determined by two parameters: ja, which is the mean expected value for Y, and o which is the standard deviation of Y . A small a leads to a more attenuated bell shape.

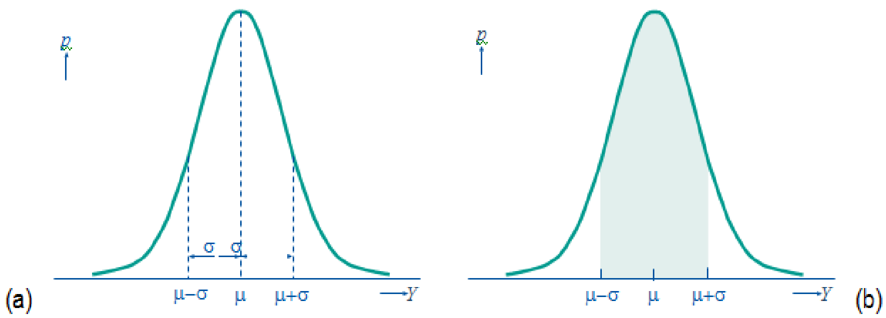


Figure 5.4: (a) Probability density function p of a variable Y, with its mean μ and standard deviation σ. *(b)* The probability that Y is in the range $[\mu-\sigma, \mu+\sigma]$.

- Any probability density function p has the characteristic that the area between its curve and the horizontal axis has size 1. Probabilities P can be inferred from p as the size of an area under p's curve. Figure above, for instance, depicts P (x - a < Y < x - a), i.e. the probability that the value for Y is within distance a from |a. In a normal distribution this specific probability for Y is always 0.6826.

Attribute Accuracy

- Two types of attribute accuracies, related to the type of data it is dealing with:
  - For nominal or categorical data, the accuracy of labeling (for example the type of land cover, road surface, etc).
  - For numerical data, numerical accuracy (such as the concentration of pollutants in the soil, height of trees in forests, etc).
- It follows that depending on the data type, assessment of attribute accuracy may range from a simple check on the labelling of features—for example, is a road classified as a metalled road actually surfaced or not?—to complex statistical procedures for assessing the accuracy of numerical data, such as the percentage of pollutants present in the soil.

Temporal Accuracy

- Spatial data sets captured through remotely sensed data has increased enormously over the last decade. These data can provide useful temporal information such as changes in land ownership and the monitoring of environmental processes such as deforestation.
- Analogous to its positional and attribute components, the quality of spatial data may also be assessed in terms of its temporal accuracy. For a static feature this refers to the difference in the values of its coordinates at two different times.

- This includes not only the accuracy and precision of time measurements but also the temporal consistency of different data sets. Because the positional and attribute components of spatial data may change together or independently, it is also necessary to consider their temporal validity. For example, the boundaries of a land parcel may remain fixed over a period of many years whereas the ownership attribute may change more frequently.

Lineage

- Lineage describes the history of a data set. In the case of published maps, some lineage information may be provided as part of the metadata, in the form of a note on the data sources and procedures used in the compilation of the data.
- Examples include the date and scale of aerial photography, and the date of field verification. For digital data sets, however, lineage may be defined as: "that part of the data quality statement that contains information that describes the source of observations or materials, data acquisition and compilation methods, conversions, transformations, analyses and derivations that the data has been subjected to, and the assumptions and criteria applied at any stage of its life."

Completeness

- Completeness refers to whether there are data lacking in the database compared to what exists in the real world. Essentially, it is important to be able to assess what does and what does not belong to a complete dataset as intended by its producer.
- It might be incomplete (i.e. it is 'missing' features which exist in the real world), or overcomplete (i.e. it contains 'extra' features which do not belong'within the scope of the data set as it is defined). Completeness can relate to spatial, temporal, or thematic aspects of a data set.
- For example, a data set of property boundaries might be spatially incomplete because it contains only 10 out of 12 suburbs; it might be temporally incomplete because it does not include recently subdivided properties; and it might be thematically over complete because it also includes building footprints.

Logical consistency

- For any particular application, (predefined) logical rules concern:
  - The compatibility of data with other data in a data set (e.g. in terms of data format),
  - The absence of any contradictions within a data set,
  - The topological consistency of the data set, and
  - The allowed attribute value ranges, as well as combinations of attributes. For example, attribute values for population, area, and population density must agree for all entities in the database. The absence of any inconsistencies does not necessarily imply that the data are accurate.

Chapter 5 : DATA PREPARATION

- Spatial data preparation aims to make the acquired spatial data fit for use. Im- ages may require enhancements and corrections of the classification scheme of the data.
- Vector data also may require editing, such as the trimming of over- shoots of lines at intersections, deleting duplicate lines, closing gaps in lines, and generating polygons.
- Data may require conversion to either vector format or raster format to match other data sets which will be used in the analysis. Additionally, the data preparation process includes associating attribute data with the spatial features through either manual input or reading digital attribute files into the GIS/DBMS.

Data checks and repairs

- Acquired data sets must be checked for quality in terms of the accuracy, consistency and completeness parameters discussed above. Often, errors can be identified automatically, after which manual editing methods can be applied to correct the errors. Alternatively, some software may identify and automatically correct certain types of errors.
- Below, we focus on the geometric, topological, and attribute components of spatial data.
- 'Clean-up' operations are often performed in a standard sequence. For example, crossing lines are split before dangling lines are erased, and nodes are created at intersections before polygons are generated. Thefce are illustrated in Table below.

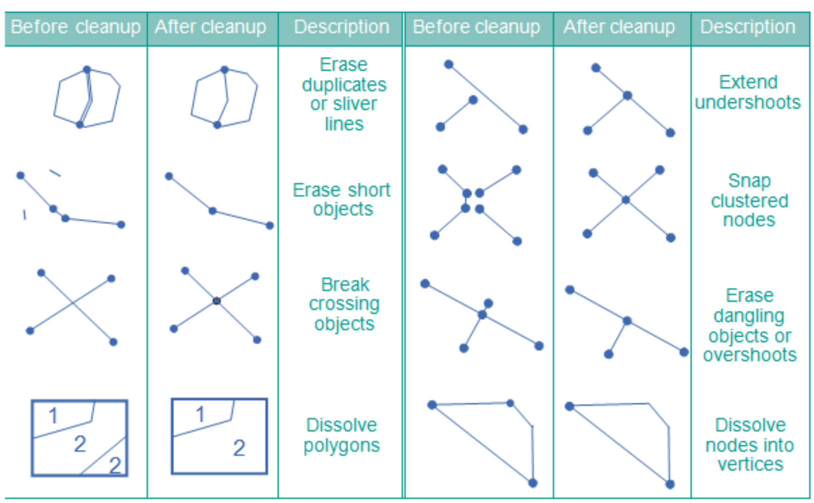


Table 5.2: Clean-up oper- ations for vector data

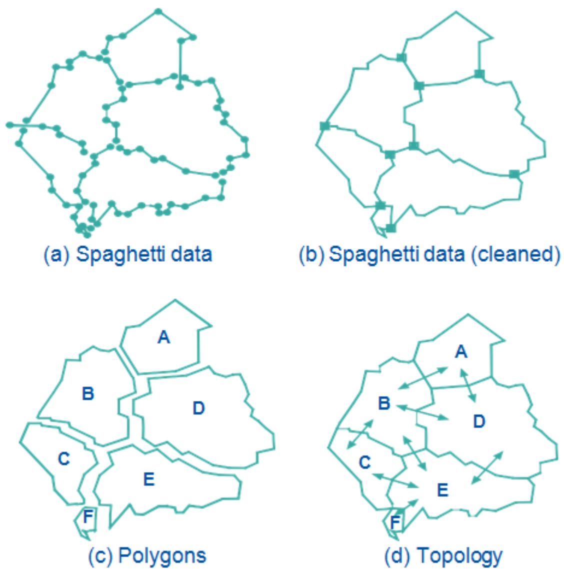


Figure 5.9: Successive clean-up operations for vector data, turning spaghetti data into topological structure.

Rasterization or vectorization

- ***Vectorization*** produces a vector data set from a raster. We have looked at this in some sense already: namely in the production of a vector set from a scanned image. Another form of vectorization takes place when we want to identify features or patterns in remotely sensed imagery. The keywords here are *feature extraction* and *pattern recognition*, which are dealt with in *Principles of Remote Sensing.*
- If much or all of the subsequent spatial data analysis is to be carried out on raster data, one may want to convert vector data sets to raster data. This process is known as rasterization.
- It involves assigning point, line and polygon attribute values to raster cells that overlap with the respective point, line or polygon. To avoid information loss, the raster resolution should be carefully chosen on the basis of the geometric resolution.
- A cell size which is too large may result in cells that cover parts of multiple vector features, and then ambiguity arises as to what value to assign to the cell. If, on the other hand, the cell size is too small, the file size of the raster may increase significantly.

Topology generation

- Topological relations may sometimes be needed, for instance in networks, e.g. the questions of line connectivity, flow direction, and which lines have over- and underpasses. For polygons, questions that may arise involve polygon inclusion: Is a polygon inside another one, or is the outer polygon simply around the inner polygon? Many of these questions are mostly questions of data semantics, and can therefore usually only be answered by a human operator.

Combining data from multiple sources

- A GIS project usually involves multiple data sets, so the next step addresses the issue of how these multiple sets relate to each other. There are four fundamental cases to be considered in the combination of data from different sources:
  1. They may be about the same area, but differ in accuracy,

2. They may be about the same area, but differ in choice of representation,
3. They may be about adjacent areas, and have to be merged into a single data set.
4. They may be about the same or adjacent areas, but referenced in different coordinate systems.

- The following may be the situation :
  - Differences in accuracy
  - Differences in representation
  - Merging data sets of adjacent areas
  - Differences in coordinate systems

Differences in accuracy

- These are clearly relevant in any combination of data sets which may themselves have varying levels of accuracy. Images come at a certain resolution, and paper maps at a certain scale. This typically results in differences of resolution of acquired data sets, all the more since map features are sometimes intentionally displaced to improve readability of the map.
- For instance, the course of a river will only be approximated roughly on a small- scale map, and a village on its northern bank should be depicted north of the river, even if this means it has to be displaced on the map a little bit.
- The small scale causes an accuracy error. If we want to combine a digitized version of that map, with a digitized version of a large-scale map, we must be aware that features may not be where they seem to be. Analogous examples can be given for images at different resolutions.
- There can be good reasons for having data sets at different scales. A good example is found in mapping organizations; European organizations maintain a single source database that contains the base data.

Differences in representation

- Some advanced GIS applications require the possibility of representing the same geographic phenomenon in different ways. These are called multi representation systems. The production of maps at various scales is an example, but there are numerous others.
- The commonality is that phenomena must sometimes be viewed as points, and at other times as polygons. For example, a small-scale national road network analysis may represent villages as point objects, but a nation-wide urban population density study should regard all municipalities as represented by polygons.
- The links between various representations for the same object maintained by the system allows switching between them, and many fancy applications of their use seem possible. A comparison is illustrated in **Figure5.11.**

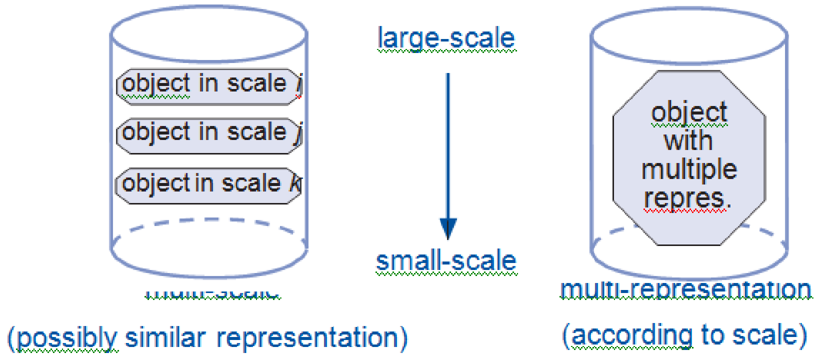


Figure 5.11: Multi-scale and multi-representation systems compared: the main difference is that multi-representation systems have a built-in 'understanding' that different representations belong together.

Merging data sets of adjacent areas

- When individual data sets have been prepared as described above, they some- times have to be matched into a single 'seamless' data set, whilst ensuring that the appearance of the integrated geometry is as homogeneous as possible. Edge matching is the process of joining two or more map sheets, for instance, after they have separately been digitized.

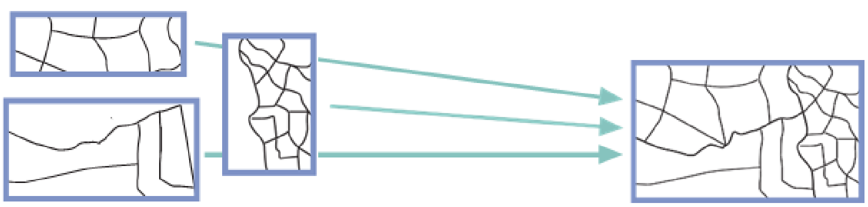


Figure 5.12: Multiple adjacent data sets, after cleaning, can be matched and merged into a single one.

Differences in coordinate systems

- Map projections provide means to map geographic coordinates onto a flat surface (for map production), and vice versa. It may be the case that data layers which are to be combined or merged in some way are referenced in different coordinate systems, or are based upon different datums.
- As a result, data may need coordinate transformation, or both a coordinate transformation and datum transformation. It may also be the case that data has been digitized from an existing map or data layer. In this case, geometric transformations help to transform device coordinates (coordinates from digitizing tablets or screen coordinates) into world coordinates (geographic coordinates, meters, etc.).

Chapter 6 : POINT DATA TRANSFORMATION

- We may have captured a sample of points (or acquired a dataset of such points), but wish to derive a value for the phenomenon at another location or for the whole extent of our study area. We may want to transform our points into other representations in order to facilitate interpretation and/or integration with other data.

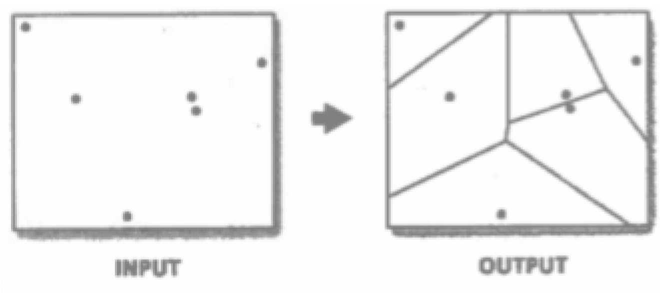
- Examples include defining homogeneous areas (polygons) from our point data, or deriving contour lines. This is generally referred to as interpolation, i.e. the calculation of a value from 'surrounding' observations. The principle of spatial autocorrelation plays a central part in the process of interpolation.
- In order to predict the value of a point for a given (x, y) location, we could simply find the 'nearest' known value to the point, and assign that value. This is the simplest form of interpolation, known as nearest-neighbour interpolation. We might instead choose to use the distance that points are away from (x, y) to weight their importance in our calculation.

Figure 5.13: A geographic field representation obtained from two point measurements: (a) for qualitative (categorical), and (b) for quantitative (continuous) point measurements. The value measured at P is represented as dark green, that at Q as light green.

- A simple example is given in Figure5.13. Our field survey has taken only two measurements, one at P and one at Q. The values obtained in these two locations are represented by a dark and light green tint, respectively. If we are dealing with qualitative data, and we have no further knowledge, the only assumption we can make for other locations is that those nearer to P probably have P 's value, whereas those nearer to Q have Q's value. This is illustrated in part (a)

Interpolating Discrete Data

- If we are dealing with discrete (nominal, categorical or ordinal) data, we are effectively restricted to using nearest-neighbour interpolation. This is the situation shown in Figure below, though usually we would have many more points.
- In a nearest- neighbour interpolation, each location is assigned the value of the closest measured point. Effectively, this technique will construct 'zones' around the points of measurement, with each point belonging to a zone assigned the same value. Effectively, this represents an assignment of an existing value (or category) to a location.
- If the desired output was a polygon layer, we could construct Thiessen polygons around the points of measurement. The boundaries of such polygons, by definition, are the locations for which more than one point of measurement is the closest point. If the desired output was in the form of a raster layer, we could rasterize the Thiessen polygons.



Interpolating Continuous Data

- Interpolation of values from continuous measurements is significantly more complex. Since the data are continuous, we can make use of measured values for interpolation. There are many continuous geographic fields—elevation, temperature and ground water salinity are just a few examples. Continuous fields are represented as rasters, and we will almost by default assume that they are.
- The main alternative for continuous field representation is a polyline vector layer, in which the lines are isolines. We will also address these issues of representation below.
- The aim is to use measurements to obtain a representation of the entire field using point samples. In this section we outline four techniques to do so:

1. Trend surface fitting using regression,
2. Triangulation,
3. Spatial moving averages using inverse distance weighting,
4. Kriging.

Trend surface fitting:

- In trend surface fitting, the assumption is that the entire study area can be represented by a formula f(x, y) that for a given location with coordinates (x, y) will give us the approximated value of the field in that location. The key objective in trend surface fitting is to derive a formula that best describes the field. Various classes of formulae exist, with the simplest being the one that describes a flat, but tilted plane: f(x, y) = ci • x + c2 • y + c3.
- The field under consideration can be best approximated by a tilted plane, then the problem of finding the best plane is the problem of determining best values for the coefficients c1, c2 and C3.

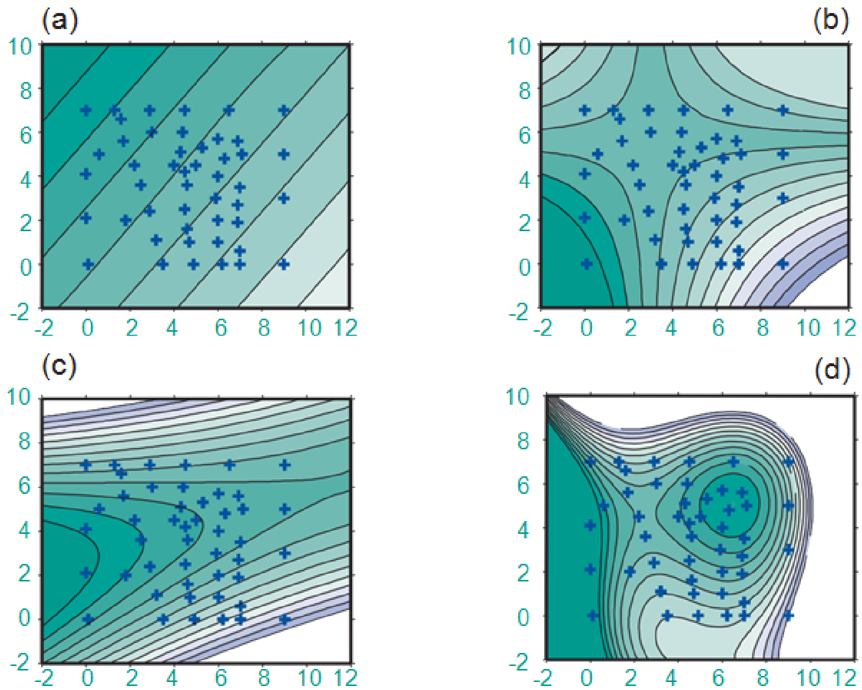


Figure 5.15: Various global trend surfaces obtained from regression techniques: (a) simple tilted plane; (b) bilinear saddle; (c) quadratic surface; (d) cubic surface. Values range from white (low), via blue, and light green to dark green (high).

- In figure 5.15, We have used the same set of point measurements, with four different approximation functions. Part (a) has been determined under the assumption that the field can be approximated by a tilted plane, in this case with a downward slope to the southeast. The values found by regression techniques were: ci = -1.83934, c2 = 1.61645 and c3 = 70.8782, giving f(x, y) = -1.83934 • x + 1.61645 • y + 70.8782.

Triangulation

- Another way of interpolating point measurements is by triangulation. Triangulated Irregular Networks (TINs) technique constructs a triangulation of the study area from the known measurement points. Preferably, the triangulation should be a Delaunay triangulation.
- After having obtained it, we may define for which values of the field we want to construct isolines. For instance, for elevation, we might want to have the 100 m- isoline, the 200 m-isoline, and so on.
- For each edge of a triangle, a geometric computation can be performed that indicates which isolines intersect it, and at what positions they do so. A list of computed locations, all at the same field value, is used by the GIS to construct the isoline. This is illustrated in Figure below

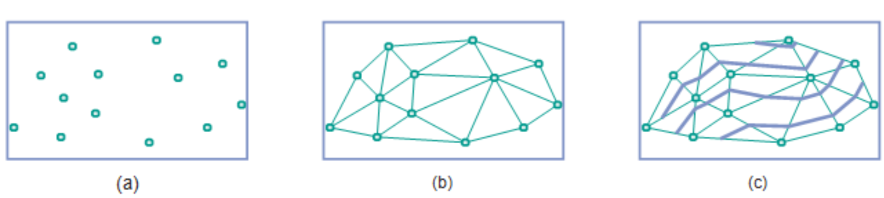


Figure 5.16: Triangulation as a means of interpolation. (a) known point measurements; (b) constructed triangulation on known points; (c) isolines constructed from the triangulation.

Spatial moving averages using inverse distance weighting

- Moving window averaging attempts to directly derive a raster dataset from a set of sample points. This is why it is sometimes also called 'gridding'. The principle behind this technique is illustrated in Figure below.
- The cell values for the output raster are computed one by one. To achieve this, a 'window' (also known as a kernel) is defined, and initially placed over the top left raster cell. Measurement points falling inside the window contribute to the averaging computation, those outside the window do not.

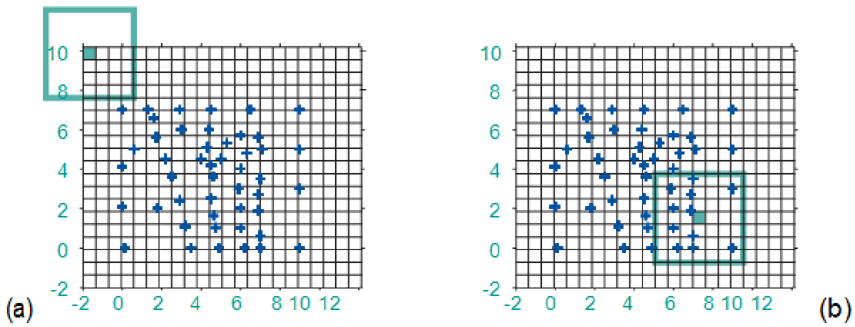


Figure 5.17: The principle of moving window averaging. In blue, the measurement points. A virtual window is moved over the raster cells one by one, and some averaging function computes a field value for the cell, using measurements within the window.

- In part (b) of the figure, the 295th cell value out of the 418 in total, is being computed. This computation is based on eleven measurements, while that of the first cell had no measurements available. Where this is the case, the cell should be assigned a value that signals this 'non-availability of measurements'.
- The principle of spatial autocorrelation suggests that measurements closer to the cell centre should have greater influence on the predicted value than those further away. In order to account for this, a distance factor can be brought into the averaging function. Functions that do this are called inverse distance weighting functions (IDW). This is one of the most commonly used functions in interpolating spatial data.

Kriging

- Kriging was originally developed my mining geologists attempting to derive accurate estimates of mineral deposits in a given area from limited sample measurements. It is an advanced interpolation technique belonging to the field of geostatistics, which can deliver good results if applied properly and with enough sample points.
- Kriging is usually used when the variation of an attribute and/or the density of sample points is such that simple methods of interpolation may give unreliable predictions.
- The first step in the kriging procedure is to compare successive pairs of point measurements to generate a semi-variogram.
- In the second step, the semi-variogram is used to calculate the weights used in interpolation. Although kriging is a powerful technique, it should not be applied without a good understanding of geostatistics, including the principle of spatial autocorrelation. It should be noted that there is no single best interpolation method, since each method has advantages and disadvantages in particular contexts.
- As a general guide, the following questions should be considered in selecting an appropriate method of interpolation:
    - o For what type of application will the results be used?

- What data type is being interpolated (e.g. categorical or continuous)?
- What is the nature of the surface (for example, is it a 'simple' or complex surface)?
- What is the scale and resolution of the data (for example, the distance between sample points)?

UNIT 4

Spatial data analysis

Chapter 1 : Classification of analytical GIS capabilities

There are many ways to classify the analytical functions of a GIS.

1) Classification, retrieval, and measurement functions

All functions in this category are performed on a single (vector or raster) data layer, often using the associated attribute data.

- Classification allows the assignment of features to a class on the basisof attribute values or attributes ranges (i.e. definition of data patterns). On the basis of reflectance characteristics found in a raster, pixels may be classified as representing different crops, such as cotton and jute.
- Retrieval functions allow the selective search of data. Example: retrieve all agricultural fields where cotton is grown.
- Generalization is a function that joins different classes of objects with common characteristics to a higher level (generalized) class. For example, we might generalize fields where potato or maize, and possibly other crops, are grown as 'kharif crop fields'.
- Measurement functions allow the calculation of distances, lengths, or areas.

2) Overlay functions

These belong to the most frequently used functions in a GIS application. They allow the combination of two (or more) spatial data layers comparing them position by position, and treating areas of overlap—and of non-overlap —in distinct ways. In this way, we can find

- The cotton fields on black soils (select the 'cotton' cover in the crop data layer and the 'black' cover in the soil data layer and perform an intersection),
- The fields where cotton or jowar is the crop (select both areas of 'cotton' and 'jowar' cover in the crop data layer and take their union),
- The cotton fields not on red soils (perform a difference operator of areas with 'cotton' cover with the areas having red soil),
- The fields that do not have wheat as crop (take the complement of the wheat areas).

3) Neighborhood functions

Whereas overlays combine features at the same location, neighborhood functions evaluate the characteristics of an area surrounding a feature's location. A neighborhood function 'scans' the neighborhood of the given feature(s), and performs a computation on it.

- Search functions allow the retrieval of features that fall within a given search window. This window may be a rectangle, circle, or polygon.
- Buffer zone generation (or buffering) is one of the best known neighborhood functions. It determines a spatial envelope (buffer) around (a) given feature(s). The created buffer may have a fixed width, or a variable width that depends on characteristics of the area.
- Interpolation functions predict unknown values using the known values at nearby locations. This typically occurs for continuous fields, like elevation, when the data actually stored does not provide the direct answer for the location(s) of interest.

- Topographic functions determine characteristics of an area by looking at the immediate neighborhood as well. Typical examples are slope computations on digital terrain models (i.e. continuous spatial fields). The slope in a location is defined as the plane tangent to the topography in that location. Various computations can be performed, such as determination of slope angle, slope aspect, slope length, contour lines.
- These are lines that connect points with the same value (for elevation, depth, temperature, barometric pressure, water salinity etc).

4) Connectivity functions

These functions work on the basis of networks, including road networks, water courses in coastal zones, and communication lines in mobile telephony. These networks represent spatial linkages between features. Main functions of this type include:

- Contiguity functions evaluate a characteristic of a set of connected spatial units. One can think of the search for a contiguous area of forest of certain size and shape in a satellite image.
- Network analytic functions are used to compute over connected line features that make up a network. The network may consist of roads, public transport routes, high voltage lines or other forms of transportation infrastructure. Analysis of such networks may entail shortest path computations (in terms of distance or travel time) between two points in a network for routing purposes. Other forms are to find all points reachable within a given distance or duration from a start point for allocation purposes, or determination of the capacity of the network for transportation between an indicated source location and sink location.
- Visibility functions also fit in this list as they are used to compute the points visible from a given location (viewshed modelling or viewshed mapping) using a digital terrain model.

Chapter 2 : RETRIEVAL, CLASSIFICATION AND MEASUREMENT

Measurement

- Geometric measurement on spatial features includes counting, distance and area size computations. For the sake of simplicity, this section discusses such measurements in a planar spatial reference system.
- We limit ourselves to geometric measurements, and do not include attribute data measurement. Measurements on vector data are more advanced, thus, also more complex, than those on raster data.

Measurements on vector data

- The primitives of vector data sets are point, (poly)line and polygon. Related geometric measurements are location, length, distance and area size. Some of these are geometric properties of a feature in isolation (location, length, area size); others (distance) require two features to be identified.
- The location property of a vector feature is always stored by the GIS: a single coordinate pair for a point, or a list of pairs for a polyline or polygon boundary. Occasionally, there is a need to obtain the location of the centroid of a polygon; some GISs store these also, others compute them 'on-the-fly'
- **Length** is a geometric property associated with polylines, by themselves, or in their function as polygon boundary.

- **Area size** is associated with polygon features. Again, it can be computed, but usually is stored with the polygon as an extra attribute value. This speeds up the computation of other functions that require area size values.
- Another geometric measurement used by the GIS is the minimal bounding box computation. It applies to polylines and polygons, and determines the minimal rectangle-with sides parallel to the axes of the spatial reference system-that covers the feature. This is illustrated in Figure6.1.

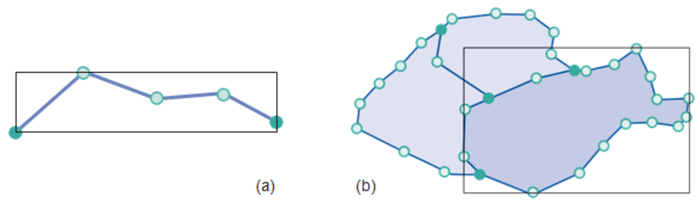


Figure 6.1: The minimal bounding box of (a) a polyline, and (b) a polygon

Measurements on raster data

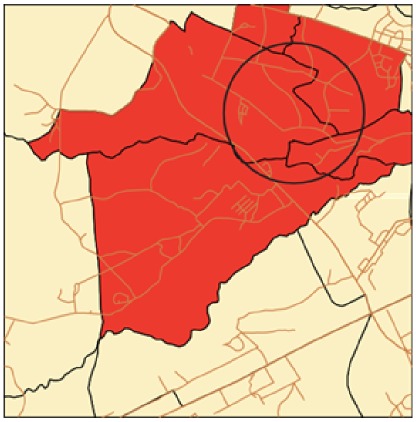
- Measurements on raster data layers are simpler because of the regularity of the cells. The area size of a cell is constant, and is determined by the cell resolution.
- Horizontal and vertical resolution may differ, but typically do not. Together with the location of a so- called anchor point, this is the only geometric information stored with the raster data, so all other measurements by the GIS are computed. The anchor point is fixed by convention to be the lower left (or sometimes upper left) location of the raster.
- Location of an individual cell derives from the raster's anchor point, the cell resolution, and the position of the cell in the raster.
- Again, there are two conventions: the cell's location can be its lower left comer, or the cell's midpoint. These conventions are set by the software in use, and in case of low resolution data they become more important to be aware of. The area size of a selected part of the raster (a group of cells) is calculated as the number of cells multiplied by the cell area size.

Spatial Selection Queries

- When exploring a spatial data set, the first thing one usually wants is to select certain features, to (temporarily) restrict the exploration. Such selections can be made on geometric/spatial grounds, or on the basis of attribute data associated with the spatial features.

Interactive spatial selection

- In interactive spatial selection, one defines the selection condition by pointing at or drawing spatial objects on the screen display, after having indicated the spatial data layer(s) from which to select features. The interactively defined objects are called the selection objects; they can be points, lines, or polygons.
- Interactive spatial selection answers questions like "What is at . . . ?" In Figure6.2, the selection object is a circle and the selected objects are the red polygons; they overlap with the selection object.

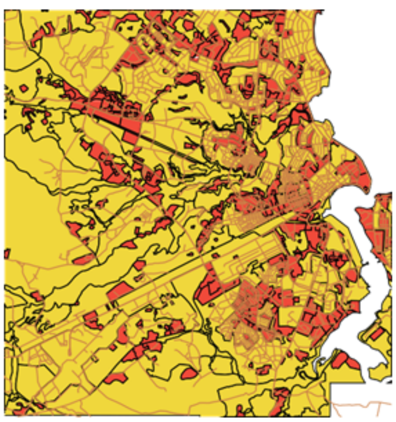


Area	Perimeter	Ward_id	Ward_nam	District	Pop88	Pop92
65420360.0000	41654.940000	1	KUNDUCHI	Kinondoni	22106	27212.00
24813620.0000	30755.620000	2	KAWE	Kinondoni	32854	40443.00
16698500.0000	26403.560000	3	MSASANI	Kinondoni	51225	63058.00
61848610.0000	49645.160000	4	UBUNGO	Kinondoni	47281	58203.00
4468546.00000	12480.130000	5	MANZESE	Kinondoni	59467	73204.00
4999599.00000	10356.850000	6	TANDALE	Kinondoni	58357	71837.00
4102216.00000	8951.096000	7	MWANANYAMALA	Kinondoni	72986	89809.00
3749640.00000	9447.420000	8	KINONDONI	Kinondoni	42301	52073.00
2087509.00000	7502.250000	9	UPANGA WEST	Ilala	9552	11425.00
2265313.00000	9026.755000	10	KIVUKONI	Ilala	5391	6254.00
1400024.00000	6553.255000	11	NDUGUMBI	Kinondoni	32548	40067.00
559966.900000	4369.110000	12	MAGOMENI	Kinondoni	16935	20851.00
1445370.00000	5651.258000	13	UPANGA EAST	Ilala	11019	12782.00
6214376.00000	14552.060000	14	MABIBO	Kinondoni	43381	53403.00
2496622.00000	7121.255000	15	MAKURUMILA	Kinondoni	54141	66646.00
1262026.00000	4555.793000	16	MZIMUNI	Kinondoni	23959	29530.00
35362240.0000	25976.090000	17	KINYEREZI	Ilala	3044	3531.00
1010613.00000	5292.771000	18	JANGWANI	Ilala	15297	17745.00
473743.500000	3042.066000	19	KISUTU	Ilala	8399	9743.00
1754043.00000	7742.157000	20	KIGOGO	Kinondoni	21267	26150.00
22964950.0000	36964.000000	21	KIGAMBONI	Temeke	23203	27655.00
1291479.00000	5187.690000	22	MCHIKICHINI	Ilala	14852	17228.00
720322.100000	4342.732000	23	MCHAFUKOGE	Ilala	8439	9789.00
9296131.00000	16221.520000	24	TABATA	Ilala	18454	21407.00
463620.700000	3304.072000	25	KARIAKOO	Ilala	12506	14507.00
3364653.00000	9566.751000	26	BUGURUNI	Ilala	45258	56012.00
2632573.00000	6970.156000	27	ILALA	Ilala	35372	41032.00
912452.800000	4021.937000	28	GEREZANI	Ilala	7490	8688.00
6725125.00000	12579.590000	29	KURASINI	Temeke	26727	31871.00

Figure 6.2: All city wards that overlap with the selection object—here a circle—are selected (left), and their corresponding attribute records are highlighted (right, only part of the table is shown). Data from an urban application in Dar es Salaam, Tanzania. Data source: Dept. of Urban & Regional Planning and Geo-information Management, ITC.

Spatial selection by attribute conditions

- It is also possible to select features by using selection conditions on feature attributes. These conditions are formulated in SQL if the attribute data reside in a geodatabase. This type of selection answers questions like "where are the features with...?"



Area	IDs	LandUse
174308.70	2	30
2066475.00	3	70
214582.50	4	80
29313.86	5	80
73328.08	6	80
53303.30	7	80
614530.10	8	20
1637161.00	9	80
156357.40	10	70
59202.20	11	20
83289.59	12	80
225642.20	13	20
28377.33	14	40
228930.30	15	30
986242.30	16	70

Figure 6.3: Spatial selection using the attribute condition *Area* < 400000 on land use areas in Dar es Salaam. Spatial features on left, associated attribute data (in part) on right. Data source: Dept. of Urban & Regional Planning and Geo-information Management, ITC.

- Figure6.3shows an example of selection by attribute condition. The query ex- pression is *Area* < 400000, which can be interpreted as "select all the land use areas of which the size is less than 400, 000." The polygons in red are the selected areas; their associated records are also highlighted in red.

Combining attribute conditions

- When multiple criteria have to be used for selection, we need to carefully express all of these in a single composite condition. The tools for this come from a field of mathematical logic, known as *propositional calculus.*
- Atomic conditions such as Area < 400000, and LandUse = 80. Atomic conditions use apredicate symbol, such as < (less than) or = (equals). Other possibilities are <= (less than or equal), > (greater than), >= (greater than or equal) and o (does not equal). Any of these symbols is combined with an expression on the left and one on the right.

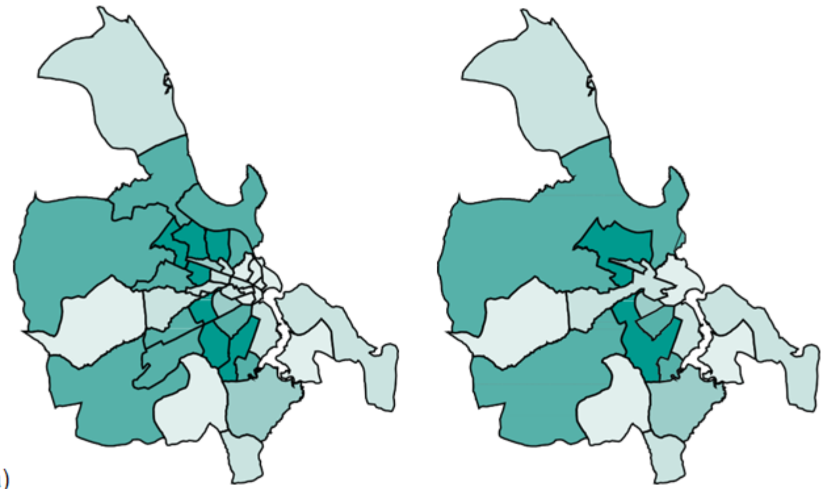
- Atomic conditions can be combined into composite conditions using logical connectives. The most important ones are AND, OR, NOT and the bracket pair (• • •). If we write a composite condition like Area < 400000 AND LandUse = 80,

Spatial selection using topological relationships

- Various forms of topological relationship can be useful to select features as well. The steps carried out are:
    1. To select one or more features as the selection objects, and
    2. To apply a chosen spatial relationship function to determine the selected features that have that relationship with the selection objects.
- **Selecting features that are inside selection objects** This type of query uses the containment relationship between spatial objects. Obviously, polygons can contain polygons, lines or points, and lines can contain lines or points, but no other containment relationships are possible.
- **Selecting features that intersect** The intersect operator identifies features that are not disjoint to include points and lines.
- **Selecting features adjacent to selection objects** Adjacency is the meet relationship. It expresses that features share boundaries, and therefore it applies only to line and polygon features.
- **Selecting features based on their distance** One may also want to use the distance function of the GIS as a tool in selecting features.
- **Afterthought on selecting features** The selection conditions on attribute values can be combined using logical connectives like AND,OR and NOT. A fact is that the other techniques of selecting features can usually also be combined.

Classification

- *Classification* is a technique of purposefully removing detail from an input data set, in the hope of revealing important patterns (of spatial distribution). In the process, we produce an output data set, so that the input set can be left intact.
- We do so by assigning a characteristic value to each element in the input set, which is usually a collection of spatial features that can be raster cells or points, lines or polygons. If the number of characteristic values is small in comparison to the size of the input set, we have *classified* the input set.
- The pattern that we look for may be the distribution of household income in a city. Temperature Shift is called the classification parameter. If we know for each ward in the city the associated average recorded temperature, will have many different values.
- It can be defined in three different categories (or: classes): 'low', 'Moderate' and 'high', and provide value ranges for each category. If these three categories are mapped in a sensible color scheme, this may reveal interesting information. This has been done for Dares Salaam in Figure6.9in two ways.



(a) (b)

Figure 6.9: Two classifications of average annual household income per ward in Dar es Salaam, Tanzania. Higher income areas in darker greens. Five categories were identified. (a) with original polygons left intact; (b) with original polygons merged when in same category. The data used for this illustration are *not* factual.

User-controlled classification

- In user-controlled classification, a user selects the attribute(s) that will be used as the classification parameter(s) and defines the classification method. The latter involves declaring the number of classes as well as the correspondence between the old attribute values and the new classes. This is usually done via a classification table.
- Another case exists when the classification parameter is nominal or at least dis- crete. Such an example is given in Figure6.10.

Code	*Old category*	*New category*
10	Planned residential	Residential
20	Industry	Commercial
30	Commercial	Commercial
40	Institutional	Public
50	Transport	Public
60	Recreational	Public
70	Non built-up	Non built-up
80	Unplanned residential	Residential

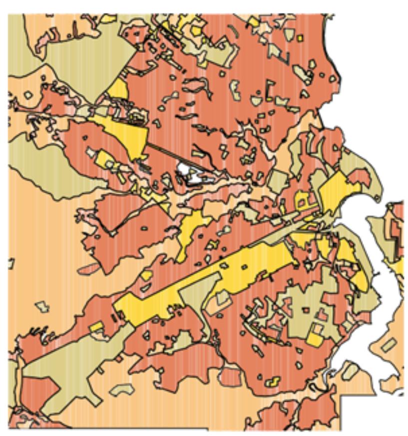


Figure 6.10: An example of a classification on a discrete parameter, namely land use unit in the city of Dar es Salaam, Tanzania. Colour scheme: Residential (brown), Commercial (yellow), Public (Olive), Non built-up (orange). Data source: Dept. of Urban & Regional Planning and Geo-information Management, ITC.

Automatic classification

User-controlled classifications require a classification table or user interaction. GIS software can also perform automatic classification, in which a user only specifies the number of classes in the output data set. The system automati- cally determines the class break points. Two main techniques of determining break points are in use.

1. **Equal interval technique**

- minimum and maximum values vmin and vmax of the classification parameter are determined and the (constant) interval size for each category is calculated as (vmax - vmin)/n, where n is the number of classes chosen by the user. This classification is useful in revealing the distribution patterns as it determines the number of features in each category.

2. **Equal frequency technique**
   - This technique is also known as quantile classification. The objective is to create categories with roughly equal numbers of features per category. The total number of features is determined first and by the required number of categories, the number of features per category is calculated. The class break points are then determined by counting off the features in order of classification parameter value.
   - Both techniques are illustrated on a small 5 × 5 raster in Figure 6.11.

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 6.11: Example of two automatic classification techniques: (a) the original raster with cell values; (b) classification based on equal intervals; (c) classification based on equal frequencies. Below, the respective classification tables, with a tally of the number of cells involved.

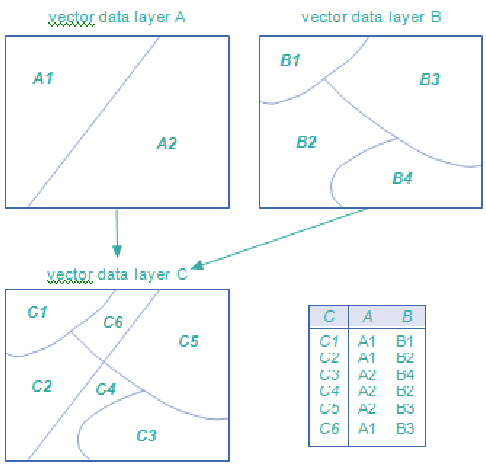
Chapter 3 : OVERLAY FUNCTIONS

- Overlay is a technique of combining two spatial data layers and producing a third from them. The binary operators that we discuss are known as spatial overlay operators. We will firstly discuss vector overlay operators, and then focus on the raster case.
- Standard overlay operators take two input data layers, and assume they are georeferenced in the same system, and overlap in study area. If either of these requirements is not met, the use of an overlay operator is senseless.
- The principle of spatial overlay is to compare the characteristics of the same location in both data layers, and to produce a result for each location in the output data layer. The specific result to produce is determined by the user. It might involve a calculation, or some other logical function to be applied to every area or location.

Vector Overlay Operators

- In the vector domain, overlay is computationally more demanding than in the raster domain. Here we will only discuss overlays from polygon data layers, but we note that most of the ideas also apply to overlay operations with point or line data layers.

- The standard overlay operator for two layers of polygons is the polygon intersection operator. It is fundamental, as many other overlay operators proposed in the literature or implemented in systems can be defined in terms of it.
- The result of this operator is the collection of all possible polygon intersections; the attribute table result is a join—in the relational database of the two input attribute tables. This output attribute table only contains one table for each intersection polygon found, and this explains why we call this operator a spatial join.

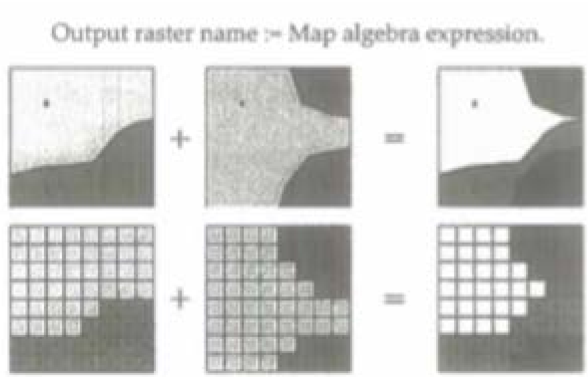


C	A	B
C1	A1	B1
C2	A1	B2
C3	A2	B4
C4	A2	B2
C5	A2	B3
C6	A1	B3

Figure 6.12: The polygon intersect (overlay) operator. Two polygon layers A and B produce a new polygon layer (with associated attribute table) that contains all intersections of polygons from A and B. Figure after [8].

Raster Overlay Operators

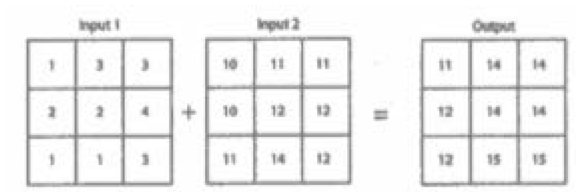
- Vector overlay operators are useful, but geometrically complicated, and this sometimes results in poor operator performance. Raster overlays do not suffer from this disadvantage, as most of them perform their computations cell by cell, and thus they are fast.
- GIS that support raster processing have a language to express operations on raster referred to as map algebra, or raster calculus, allowing a GIS to compute new raster from existing ones, using a range of functions and operators.
- The key operations using a logical structured language differs for different GIS software packages. When producing a new raster we must provide a name for it, and define how it is computed. This is done in an assignment statement of the following format:



- The expression on the right is evaluated by the GIS, and the raster in which it results is then stored under the name on the left. The expression may contain references to existing rasters, operators and functions; the format is made clear below. The raster names and constants that are used in the expression are called its operands.

Arithmetic operators

- Various arithmetic operators are supported. The standard ones are multiplication (*), division (/), subtraction (-) and addition (+). Other arithmetic operators may include modulo division (MOD) and integer division (DIV). Modulo division returns the remainder of division: for example, 11 MOD 5 will return 1 as 10 - 5 * 2 = 1. Similarly, 10 DIV 2 will return 5.



Comparison and logical operators

- Map algebra also allows the comparison of rasters cell by cell. To this end, we may use the standard comparison operators (<, <=, =, >=, > and o) that we introduced before. A simple raster comparison assignment is: C: = A o B, will store truth value either true orfalse in the output raster C. Logical connectives like AND, OR, XOR, NOT are also supported in map algebra.

Conditional expressions

- The above comparison and logical operators produce rasters with the truth value true and false. In practice, we often need a conditional expression with them that allows us to test whether a condition is fulfilled. The general format is:

  Output raster: = CON (condition, then expression, else expression).

- Here, condition is the tested condition, then expression is evaluated if condition holds, and else expression is evaluated if it does not hold.
- For example an expression like CON (Gridln>3,1, 0) will evaluate to 1 for each cell in the output raster where the same cell in Gridin is classified as greater than 3. In each cell where this is not true, the else expression is evaluated, resulting in 0.

Overlays Using a Decision Table

- Conditional expressions are powerful tools in cases where multiple criteria must be taken into account. A small size example may illustrate this. Consider a suitability study in which a land use classification and a geological classification must be used.
- The respective rasters. Do main expertise dictates that some combinations of land use and are a Type result in suitable areas, whereas other combinations do not. In our example, NA Land on CITY and RURAL areas are considered suitable combinations, while the others are not.

A map algebra expression

Suitability : = CON ((Landuse = "Non-Agriculture" AND areaType = "CITY") OR (Landuse = "Non-Agriculture" AND areaType = "RURAL"), "Suitable", "Unsuitable")

- The above type of computation becomes simpler by setting up a separate decision table that will guide the raster overlay process. This extra table carries domain expertise, and dictates which combinations of input raster cell values will produce which output raster cell value. This gives us a raster overlay operator using a decision table, as illustrated in Figure below.

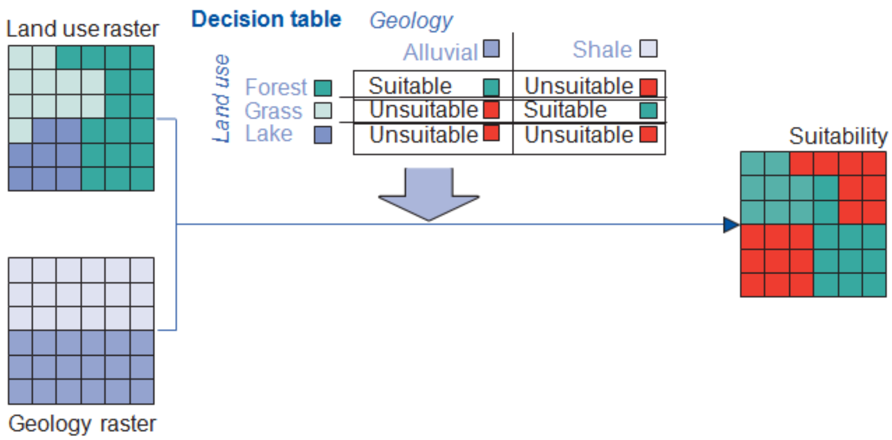


Figure 6.19: The use of a decision table in raster overlay. The overlay is computed in a suitability study, in which land use and geology are important factors. The meaning of values in both input rasters, as well as the output raster can be understood from the decisiontable.

Chapter 5 : NEIGHBORHOOD FUNCTIONS

- The principle in Neighborhood function is to find out the characteristics of the vicinity, here called neighborhood, of a location. After all, many suitability questions, for instance, depend not only on what is at the location, but also on what is near the location. Thus, the GIS must allow us 'to look around locally'.

To perform neighborhood analysis, we must:

1. State which target locations are of interest to us, and define their spatial extent,
2. Define how to determine the neighborhood for each target,
3. Define which characteristic(s) must be computed for each neighborhood.

For instance, our target might be a nearby ATM. Its neighborhood could be defined as:

- o An area within 100m walking distance of an State Bank ATM, or
- o An area within 2 km travel distance, or
- o All roads within 500 m travel distance, or
- o All other Bank ATM within 5 minutes travel time, or
- o All Banks, for which the ATM is the closest.

To discover about the phenomena that exist or occur in the neighborhood. E. g. spatial extent, also require statistical information like:

- o The total population of the area,
- o Average household income, or

- o The distribution of high-risk industries located in the neighborhood.

Proximity Computations

- In proximity computations, we use geometric distance to define the neighborhood of one or more target locations. The most common and useful technique is buffer zone generation.

Buffer zone generation

- The principle of buffer zone generation is simple : we select one or more target locations, and then determine the area around them, within a certain distance. In Figure below, the main roads were selected as targets, and a 75 meter buffer was computed from them.

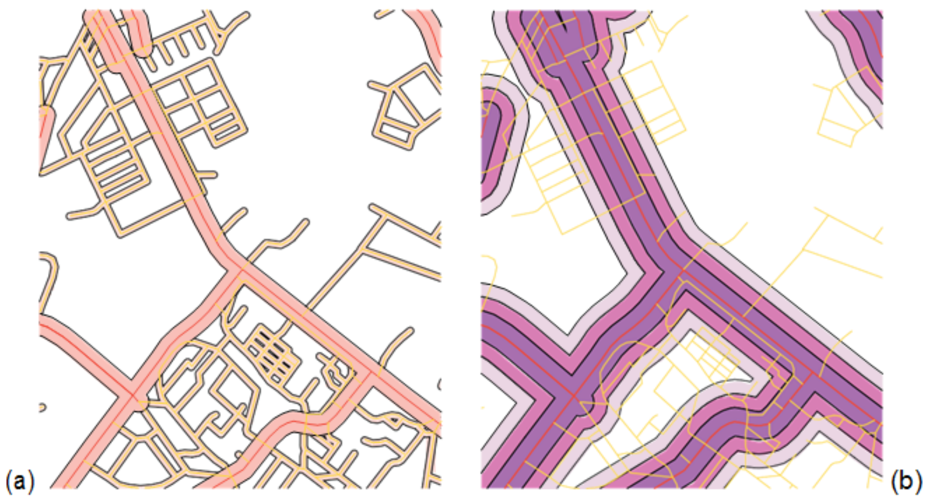


Figure 6.20: Buffer zone generation: (a) around main and minor roads. Different distances were applied: 25 metres for minor roads, 75 metres for main roads. (b) Zonated buffer zones around main roads. Three different zones were obtained: at 100 metres from main road, at 200, and at 300 metres.

- In vector-based buffer generation, the buffers themselves become polygon features, usually in a separate data layer, that can be used in further spatial analysis.
- Buffer generation on rasters is a fairly simple function. The target location or locations are always represented by a selection of the raster's cells, and geometric distance is defined, using cell resolution as the unit.

Thiessen polygon generation

- Thiessen polygon partitions make use of geometric distance for determining neighbourhoods. This is useful if we have a spatially distributed set of points as target locations, and we want to know for each location in the study to which target it is closest.
- This technique will generate a polygon around each target location that identifies all those locations that 'belong to' that target. We have already seen the use of Thiessen polygons in the context of interpolation of point data.
- Given an input point set that will be the polygon's midpoints, it is not difficult to construct such a partition. It is even much easier to construct if we already have a Delaunay triangulation for the same input point set.

- Figure below repeats the Delaunay triangulation of the Thiesse polygon partition constructed from it is on the right.

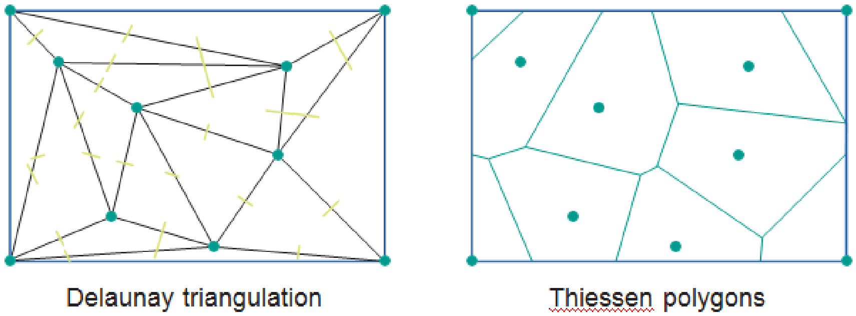


Figure 6.21: Thiessen polygon construction (right) from a Delaunay triangulation (left): perpendiculars of the triangles form the boundaries of the polygons.

Computation of Diffusion

- The determination of neighborhood of one or more target locations may depend not only on distance—cases which we discussed above—but also on direction and differences in the terrain in different directions. This typically is the case when the target location contains a 'source material' that spreads over time, referred to as diffusion.
- This 'source material' may be air, water or soil pollution, commuters exiting a train station, people from an opened-up refugee camp, a water spring uphill, or the radio waves emitted from a radio relay station. In all these cases, one will not expect the spread to occur evenly in all directions. There will be local terrain factors that influence the spread, making it easier or more difficult.
- Diffusion computation involves one or more target locations, which are better called source locations in this context. They are the locations of the source of whatever spreads. The computation also involves a local resistance raster, which for each cell provides a value that indicates how difficult it is for the 'source material' to pass by that cell.
- The value in the cell must be normalized: i.e. valid for a standardized length (usually the cell's width) of spread path. From the source location(s) and the local resistance raster, the GIS will be able to compute a new raster that indicates how much minimal total resistance the spread has witnessed for reaching a raster cell. This process is illustrated in Figure below.

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)

14.50	14.95	15.95	17.45	22.45
12.00	12.45	14.61	16.66	21.44
8.00	8.95	11.95	13.66	19.66
4.00	6.36	8.00	10.00	11.00
0.00	3.00	4.50	5.50	6.50

(b)

Figure 6.22: Computation of diffusion on a raster. The lower left green cell is the source location, indicated in the local resistance raster (a). The raster in (b) is the minimal total resistance raster computed by the GIS. (The GIS will work in higher precision real arithmetic than what is illustrated here.)

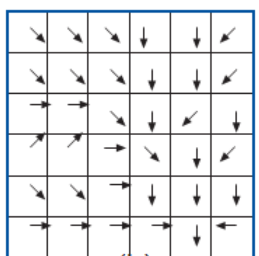
- While computing total resistances, the GIS take proper care of the path lengths. Obviously, the diffusion from a cell csrc to its neighbor cell to the east ce is shorter than to the cell that is its northeast neighbor cne.

Flow Computation

- Flow computations determine how a phenomenon spreads over the area, in principle in all directions, though with varying difficulty or resistance. There are also cases where a phenomenon does not spread in all directions, but moves or 'flows' along a given, least- cost path, determined again by local terrain characteristics.
- The typical case arises when we want to determine the drainage patterns in a catchment: the rainfall water 'chooses' a way to leave the area.
- Cells with a high accumulated flow count represent areas of concentrated flow, and thus may belong to a stream. By using some appropriately chosen threshold value in a map algebra expression, we may decide whether they do. Cells with an accumulated flow count of zero are local topographic highs, and can be used to identify ridges.

156	144	138	142	116	98
148	134	112	98	92	100
138	106	88	74	76	96
128	116	110	44	62	48
136	122	94	42	32	38
148	106	68	24	22	24

(a)



(b)

0	0	0	0	0	0
0	1	1	2	2	0
0	3	7	5	4	0
0	0	0	20	0	1
0	0	0	1	24	0
0	2	4	7	35	1

(c)

Figure 6.23: Flow computations on a raster: (a) the original elevation raster, (b) the flow direction raster computed from it, (c) accumulated flow count raster.

Raster Based Surface Analysis

- Continuous fields have a number of characteristics not shared by discrete fields. Since the field changes continuously, we can talk about slope angle, slope aspect and concavity/convexity of the slope. These notions also not applicable to discrete fields.
- The discussions here use terrain elevation as the prototypical example of a continuous field, but all issues discussed are equally applicable to other types of continuous fields.
- Nonetheless, we regularly refer to the continuous field representation as a DEM, to conform with the most common situation.

Applications

There are numerous examples where more advanced computations on continuous field representations are needed. A short list is provided below.

- **Slope angle calculation**
  - The calculation of the slope steepness, expressed as an angle in degrees or percentages, for any or all locations.
- **Slope aspect calculation**

- The calculation of the aspect (or orientation) of the slope in degrees (between 0 and 360 degrees), for any or all locations.

- **Slope length calculation**
  - With the use of neighborhood operations, it is possible to calculate for each cell the nearest distance to a watershed boundary (the upslope length) and to the nearest stream (the downslope length).
- **Three-dimensional map display**
  - With GIS software, three-dimensional views of a DEM can be constructed, in which the location of the viewer, the angle under which s/he is looking, the zoom angle, and the amplification factor of relief exaggeration can be specified.
- **Determination of change in elevation through time**
  - The cut-and-fill volume of soil to be removed or to be brought in to make a site ready for construction can be computed by overlaying the DEM of the site before the work begins with the DEM of the expected modified topography.
- **Automatic catchment delineation**
  - Catchment boundaries or drainage lines can be automatically generated from a good quality DEM with the use of neighborhood functions
- **Dynamic modeling**
  - DEMs are increasingly used in GIS-based dynamic modeling, such as the computation of surface run-off and erosion, groundwater flow, the delineation of areas affected by pollution, the computation of areas that will be covered by processes such as debris flows and lava flows.
- **Visibility analysis**
  - A viewshed is the area that can be 'seen', i.e. in the direct line-of-sight from a specified target location.

Filtering

- The principle of filtering is quite similar to that of moving window averaging. We define a window and let the GIS move,it over the raster cell-by-cell. For each cell, the system performs some computation, and assigns the result of this computation to the cell in the output raster.
- The. difference withimoving window averaging is that the moving window in filtering is itself a little raster, which contains cell values that are used in the computation for the output cell value.
- This little raster is a filter, also known as a kernel which may be square (such as a 3x3 kernel), but it does not have to be. The values in the filter are used as weight factors.

Chapter 6 : Analysis

NETWORK ANALYSIS

- A completely different set of analytical functions in GIS consists of computations on networks. A network is a connected set of lines, representing some geographic phenomenon, typically of the transportation type.
- The 'goods' transported can be almost anything: people, cars and other vehicles along a road network, commercial goods along a logistic network, phone calls along a telephone network, or water pollution along a stream/river network.
- Network analysis can be performed on either raster or vector data layers, but they are more commonly done in the latter, as line features can be associated with a network, and hence can be

assigned typical transportation characteristics such as capacity and cost per unit. A fundamental characteristic of any network is whether the network lines are considered directed or not.

- Additional application-specific rules are usually required to define what can and cannot happen in the network. Most GISs provide rule-based tools that allow the definition of these extra application rules. Various classical spatial analysis functions on networks are supported by GIS software packages. The most important ones are:

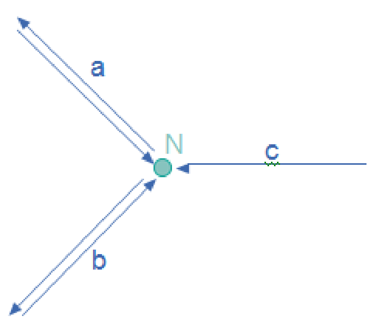
1. Optimal path finding which generates a least cost-path on a network between a pair of predefined locations using both geometric and attribute data.
2. Network partitioning which assigns network elements (nodes or line segments) to different locations using predefined criteria.

The two typical functions discussed here are.

- Optimal path finding
- Network partitioning

Optimal path finding

- Optimal path finding techniques are used when a least-cost path between two nodes in a network must be found. The two nodes are called origin and destination, respectively. The aim is to find a sequence of connected lines to traverse from the origin to the destination at the lowest possible cost.
- The cost function can be simple: for instance, it can be defined as the total length of all lines on the path.
- The cost function can also be more elaborate and take into account not only length of the lines, but also their capacity, maximum transmission (travel) rate and other line characteristics, for instance to obtain a reasonable approximation of travel time.
- There can even be cases in which the nodes visited add to the cost of the path as well. These may be called turning costs, which are defined in a separate turning cost table for each node, indicating the cost of turning at the node when entering from one line and continuing on another. This is illustrated in Figure below.



from	onto	cost
a	a	∞
a	b	5
b	a	8
b	b	15
c	a	3
c	b	6

Figure 6.27: Network neighbourhood 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. A turning cost of infinity (∞) means that it is also prohibited.

- Notice that it is possible to travel on line b in Figure above, then take a U-turn at node N, and return along a to where one came from. The question is whether doing this makes sense in optimal path finding.

Network partitioning

- In network partitioning, the purpose is to assign lines and/or nodes of the network, in a mutually exclusive way, to a number of target locations. Typically, the target locations play the role of service centre for the network. This may be any type of service: medical treatment, education, water supply. This type of network partitioning is known as a network allocation problem.
- Another problem is trace analysis. Here, one wants to determine that part of the network that is upstream (or downstream) from a given target location. Such problems exist in pollution tracing along river/stream systems, but also in network failure chasing in energy distribution networks.

Network allocation

- In network allocation, we have a number of target locations that function as resource centres, and the problem is which part of the network to exclusively assign to which service centre.
- This may sound like a simple allocation problem, in which a service centre is assigned those lines (segments) to which it is nearest, but usually the problem statement is more complicated. These further complications stem from the requirements to take into account
- The capacity with which a centre can produce the resources (whether they are medical operations, school pupil positions, kilowatts, or bottles of milk), and
- The consumption of the resources, which may vary amongst lines or line segments. After all, some streets have more accidents, more children who live there, more industry in high demand of electricity or just more thirsty workers.

Trace analysis

- Trace analysis is performed when we want to understand which part of a network is 'conditionally connected' to a chosen node on the network, known as the trace origin. For a node or line to be conditionally connected, it means that a path exists from the node/line to the trace origin, and that the connecting path fulfills the conditions set.
- What these conditions are depends on the application, and they may involve direction of the path, capacity, length, or resource consumption along it. The condition typically is a logical expression, as we have seen before, for example:
- The path must be directed from the node/line to the trace origin,
- Its capacity (defined as the minimum capacity of the lines that constitute the path) must be above a given threshold, and
- The path's length must not exceed a given maximum length.

Chapter 7 : GIS AND APPLICATION MODELS

- Models are simplified abstractions of reality representing or describing its most important elements and their interactions. Modelling and GIS are more or less inseparable, as GIS is itself a tool for modelling 'the real world'.

- The solution to a (spatial) problem usually depends on a large number of parameters. Since these parameters are often interrelated, their interaction is made more precise in an application model.
- The nature of application models varies enormously. GIS applications for famine relief programs, for instance, are very different from earthquake risk assessment applications, though both can make use of GIS to derive a solution. Many kinds of application models exist, and they can be classified in many different ways.

Here we identify five characteristics of GIS-based application models :

1. The purpose of the model,
2. The methodology underlying the model,
3. The scale at which the model works,
4. Its dimensionality - i.e. whether the model includes spatial, temporal or spatial and temporal dimensions, and
5. Its implementation logic - i.e. the extent to which the model uses existing knowledge about the implementation context.

- It is important to note that the categories above are merely different characteristics of any given application model. Any model can be described according to these characteristics. Each is briefly discussed below.

- **Purpose of the model** refers to whether the model is descriptive, prescriptive or predictive in nature. Descriptive models attempt to answer the "what is" question. Prescriptive models usually answer the "what should be" question by determining the best solution from a given set of conditions.
- **Methodology** refers to the operational components of the model. Stochastic models use statistical or probability functions to represent random or semi-random behaviour of phenomena. In contrast, deterministic models are based upon a well-defined cause and effect relationship. Examples of deterministic models include hydrological flow and pollution models, where the 'effect' can often be described by numerical methods and differential equations.
- **Scale** refers to whether the components of the model are individual or aggregate in nature. Essentially this refers to the 'level' at which the model operates. Individual-based models are based on individual entities, such as the agent-based models described above, whereas aggregate models deal with 'grouped' data, such as population census data.
- **Dimensionality** is the term chosen to refer to whether a model is static or dynamic, and spatial or aspatial. Some models are explicitly spatial, meaning they operate in some geographically defined space. Some models are aspatial, meaning they have no direct spatial reference.
- **Implementation logic** refers to how the model uses existing theory or knowledge to create new knowledge. Deductive approaches use knowledge of the overall situation in order to predict outcome conditions. This includes models that have some kind of formalized set of criteria, often with known weightings for the inputs, and existing algorithms are used to derive outcomes.

Chapter 8 : ERROR PROPAGATION IN SPATIAL DATA PROCESSING

How Errors Propagate

- Error may be present in source data. It is important to note that the acquisition of base data to a high standard of quality still does not guarantee that the results of further, complex processing can be treated with certainty. As the number of processing steps increases, it becomes difficult to predict the behavior of error propagation.
- These various errors may affect the outcome of spatial data manipulations. In addition, further errors may be introduced during the various processing steps as illustrated in Figure below.

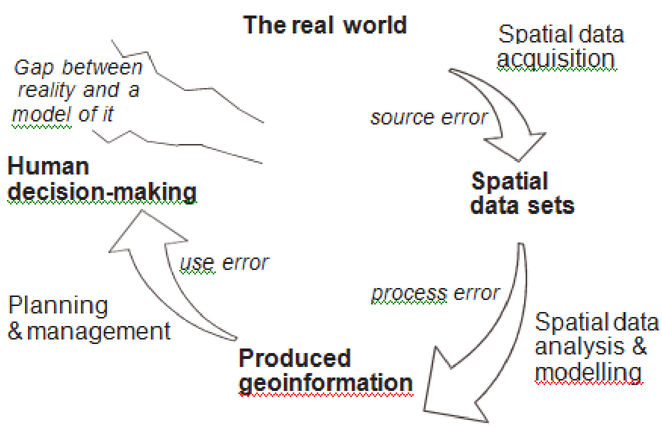


Figure 6.31: Error propagation in spatial data handling

- One of the most commonly applied operations in geographic information systems is analysis by overlaying two or more spatial data layers. Each such layer will contain errors, due to both inherent inaccuracies in the source data and errors arising from some form of computer processing, for example, rasterization. During the process of spatial overlay, all the errors in the individual data layers contribute to the final error of the output.

Quantifying error propagation

- Chrisman noted that "the ultimate arbiter of cartographic error is the real world, not a mathematical formulation". It is an unavoidable fact that we will never be able to capture and represent everything that happens in the real world perfectly in a GIS. Hence there is much to recommend the use of testing procedures for accuracy assessment.
- Various perspectives, motives and approaches to dealing with uncertainty have given rise to a wide range of conceptual models and indices for the description and measurement of error in spatial data.
- All these approaches have their origins in academic research and have strong theoretical bases in mathematics and statistics. Here we identify two main approaches for assessing the nature and amount of error propagation:
    1. Testing the accuracy of each state by measurement against the real world, and
    2. Modelling error propagation, either analytically or by means of simulation techniques.

- Modeling of error propagation has been defined by Veregin as: "the application of formal mathematical models that describe the mechanisms whereby errors in source data layers are modified by particular data transformation operations."

UNIT 5

Data visualization

Chapter 1 : GIS AND MAPS

- A map is "a representation or abstraction of geographic reality. A tool for presenting geographic information in a way that is visual, digital or tactile."
- The definition holds three key words. The "geographic reality" represents the object of study, our world. "Representation" and "abstraction" refer to models of these geographic phenomena. The second sentence reflects the appearance of the map. A map is a reduced and simplified representation of the Earth's surface on a plane.
- Maps and GIS are closely related to each other. Maps can be used as input for a GIS. Also play a key role in relation to all the functional components of a GIS.
- A map can often be the most suitable tool to solve the question contains "where", and provide the answer. "Where do I find GPO?" and "Where do B. Sc. IT colleges are located?". The answers could be in non-map form like "in the FORT Region" or "in all over Mumbai." These answers could be satisfying; however, they do not give the full picture.
- A map would put these answers in a spatial context. It could show where in the Netherlands Enschede is to be found and where it is located with respect to Schiphol–Amsterdam airport, where most students arrive. A world map would refine the answer "from all over the world," since it reveals that most students arrive from Africa and Asia, and only a few come from the Americas, Australia and Europe as can be seen in Figure7.1.

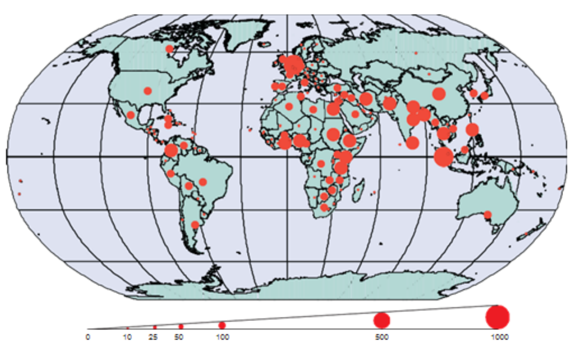


Figure 7.1: Maps and location—"Where did ITC cartography students come from?" Map scale is 1 : 200, 000, 000.

THE VISUALIZATION PROCESS

- The characteristic of maps and their function in relation to the spatial data handling process was explained in the previous section. In this context the cartographic visualization process is considered to be the translation or conversion of spatial data from a database into graphics. These are predominantly map like products.

- During the visualization process, cartographic methods and techniques are applied. These can be considered to form a kind of grammar that allows for the optimal design and production for the use of maps, depending on the application(see Figure7.8).

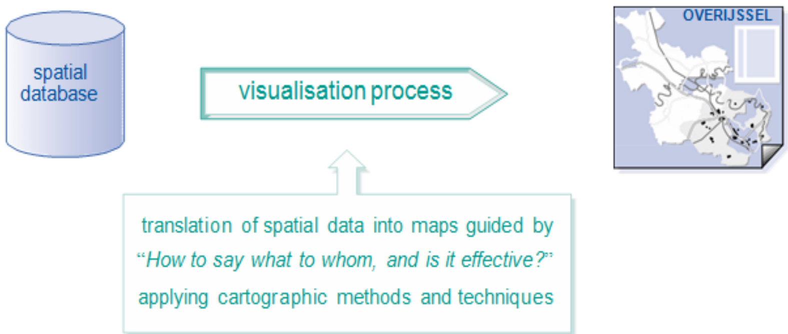


Figure 7.8: The carto-graphic visualization pro-cess. Source: Figure 2–1 in [30].

- The producer of these visual products may be a professional cartographer, but may also be a discipline expert, for instance, mapping vegetation stands using remote sensing images, or health statistics in the slums of a city. To enable the translation from spatial data into graphics; we assume that the data are available and that the spatial database is well structured.
- The visualization process can vary greatly depending on where in the spatial data handling process it takes place and the purpose for which it is needed. Visualizations can be, created during any phase of the spatial data handling process. They can be simple or complex, while the production time can be short or long.
- Some examples are the creation of a full, traditional topographic map sheet, a newspaper map, a sketch map, a map from an electronic atlas, an animation showing the growth of a city, a three-dimensional view of a building or a mountain, or even a real-time map display of traffic conditions.
- The visualization process is always influenced by several factors. Some of these questions can be answered by just looking at the content of the spatial database:
    - What will be the scale of the map: large, small, other? This introduces the problem of generalization. Generalization addresses the meaningful reduction of the map content during scale reduction.
    - Are we dealing with topographic or thematic data? These two categories traditionally resulted in different design approaches as was explained in the previous section.
    - More important for the design is the question of whether the data to be represented are of a quantitative or qualitative nature.

Chapter 2 : VISUALIZATION STRATEGIES : PRESENT OR EXPLORE?

- The cartographer's main task was the creation of good cartographic products. The main function of maps is to communicate geographic information, i.e. to inform the map user about location and nature of geographic phenomena and spatial patterns.
- This has been the map's function throughout history. Well-trained cartographers are designing and producing maps, supported by a whole set of cartographic tools and theory. The widespread use of GIS has increased the number of maps tremendously.
- Many of these maps are not produced as final products, but rather as intermediaries to support the user in her/his work dealing with spatial data. The map has started to play a completely new

role: it is not only a communication tool, but also has become an aid in the user's visual thinking process.

- This thinking process is accelerated by the continued developments in hardware and software. Media like DVD-ROMs and the WWW allow dynamic presentation and also user interaction. These went along with changing scientific and societal needs for georeferenced data and maps. Users now expect immediate and real-time access to the data; data that have become abundant in many sectors of the geoinformation world. This abundance of data, seen as a 'paradise' by some sectors, is a major problem in other sectors.
- We lack the tools for user-friendly queries and retrieval when studying the massive amount of spatial data produced by sensors, which is now available via the WWW. A new branch of science is currently evolving to deal with this problem of abundance. In the geo-disciplines, it is called visual data mining.
- These developments have given the term *visualization* an enhanced meaning. According to the dictionary, it means 'to make visible' or 'to represent in graphical form'.
- Developments in scientific visualization stimulated DiBiase [18] to define a model for map-based scientific visualization, also known as *geovisualization*. It covers both the presentation and exploration functions of the map (see Figure7.9). Presentation is described as 'public visual communication' since it concerns maps aimed at a wide audience. Exploration is defined as 'private visual thinking' because it is often an individual playing with the spatial data to determine its significance.

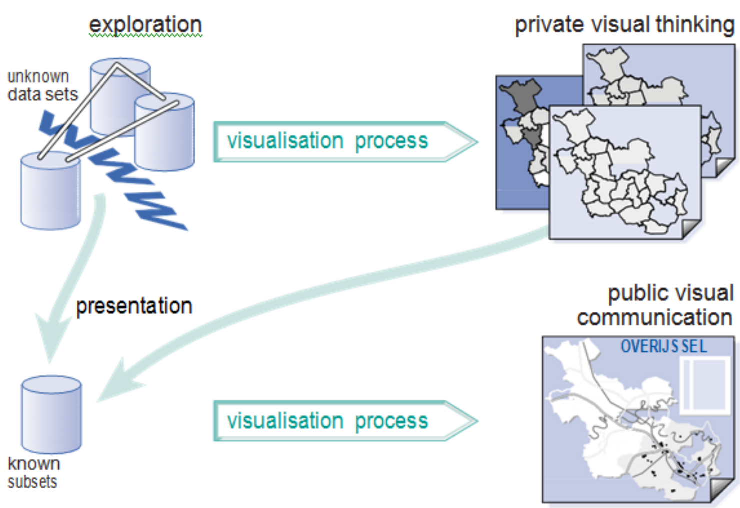


Figure 7.9: Private visual thinking and public visual communication. Source: Modified from Figure 2–2 in [30].

THE CARTOGRAPHIC TOOLBOX

What kind of data do I have?

- To derive the proper symbology for a map one has to execute a cartographic data analysis. The core of this analysis process is to access the characteristics of the data to find out how they can be visualized, so that the map user properly interprets them. The first step in the analysis process is to find a common - denominator for all the data. This common denominator will then be used as the title of the map.

- For instance, if all data are related to land use, collected in 2015, the title could be Landuse of . . . 2015. Secondly, the individual component(s), such as landuse, and probably relief, should be analyzed and their nature described. Later, these components should be visible in the map legend.
- Different types of data in relation to how it might map or display them.

Data will be of a qualitative or quantitative nature.

- **Qualitative** data is also called nominal or categorical data. This data exists as discrete, named values without a natural order amongst the values. Examples are the different languages (e.g. English, Hindi, Marathi, Tamil), the different soil types (e.g. sand, clay, peat) or the different land use categories (e.g. arable land, pasture). In the map, qualitative data are classified according to disciplinary insights such as a soil classification system represented as basic geographic units: homogeneous areas associated with a single soil type, recognized by the soil classification.
- **Quantitative** data can be measured, either along an interval or ratio scale. For data measured on an interval scale, the exact distance between values is known, but there is no absolute zero on the scale. Temperature is an example: 40° C is not twice as warm as 20° C, and 0° C is not an absolute zero. Quantitative data with a ratio scale does have a known absolute zero. An example is income: someone earning ? 1000 earns twice as much as someone with an income of ? 500. In order to generate maps, quantitative data are often classified into categories according to some mathematical method.

How Can I Map My Data?

- Basic elements of a map, irrespective of the medium on which it is displayed, are point symbols, line symbols, area symbols, and text. The appearance of point, line, and area symbols can vary depending on thfeir nature.
- Most maps in this book show symbols in different size, shape and color. Points can vary in form or color to represent the location of shops or they can vary in size to represent aggregated values (like number of inhabitants) for an administrative area. Lines can vary in color to distinguish between administrative boundaries and rivers, or vary in shape to show the difference between railroads and roads.
- Areas follow the same principles: difference in color distinguishes between different vegetation stands. Although the variations in symbol appearance are only limited by the imagination they can be grouped together in a few categories. Bertin distinguished six categories, which he called the visual variables and which may be applied to point, line and area symbols.

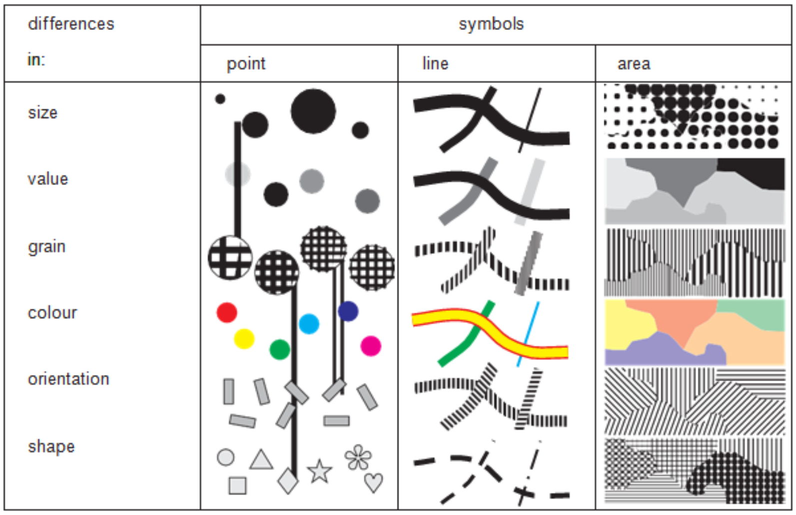


Figure 7.11: Bertin's six visual variables illustrated. Source: Plate 1 in [31].

- These visual variables can be used to make one symbol different from another. In doing this, map makers in principle have free choice, provided they do not violate the rules of cartographic grammar. They do not have that choice when deciding where to locate the symbol in the map. The symbol should be located where features belong. Visual variables influence the map user's perception in different ways. What is perceived depends on the human capacity to see or perceive:
  - What is of equal importance (e.g. all red symbols represent danger),
  - Order (e.g. the population density varies from low to high—represented by light and dark color tints, respectively),
  - Quantities (e.g. symbols changing in size with small symbols for small amounts), or An instant overview of the mapped theme.

Chapter 3 : HOW TO MAP...?

How to Map Qualitative Data

- Qualitative data is also called nominal or categorical data. If, after a long fieldwork, finally delineated the boundaries of a soil type in India, cartographer likely will be interested in a map showing these areas. The geographic units in the map will have to represent the individual watersheds. In such a map, each of the watersheds should get equal attention, and none should stand out above the others.

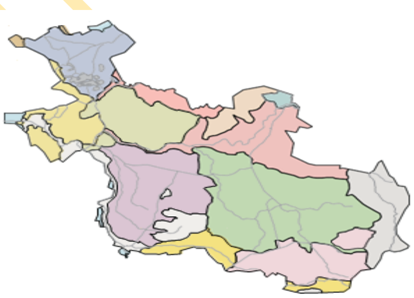


Figure 7.12: A good example of mapping qualitative data

- The application of colour would be the best solution since is has characteristics that allow one to quickly differentiate between different geographic units. How- ever, since none of the watersheds is more important than the others, the colours used have to be of equal visual weight or brightness. Figure7.12gives an example of a correct map.

How to Map Quantitative Data

- When, after executing a census, one would for instance like to create a map with the number of people living in each municipality, one deals with absolute quantitative data. The geographic units will logically be the municipalities.
- The final map should allow the user to determine the amount per municipality and also offer an overview of the geographic distribution of the phenomenon. To reach this objective, the symbols used should have quantitative perception properties. Symbols varying in size fulfil this demand. Figure7.14shows the final map for the province of Overijssel.

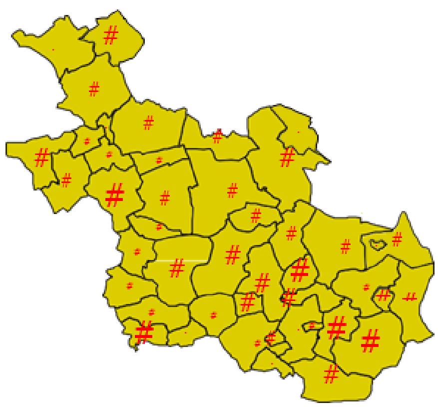


Figure 7.14: Mapping ab- solute quantitative data

- The fact that it is easy to make errors can be seen in Figure7.15. In7.15(a), differ- ent tints of green (the visual variable 'value') have been used to represent *absolute* population numbers. The reader might get a reasonable impression of the indi- vidual amounts but not of the actual geographic distribution of the population, as the size of the geographic units will influence the perceptional properties too much. Imagine a small and a large unit having the same number of inhabitants.
- The large unit would visually attract more attention, giving the impression there are more people than in the small unit. Another issue is that the population is not necessarily homogeneously distributed within the geographic units. Colour has also been misused in Figure7.15(b).

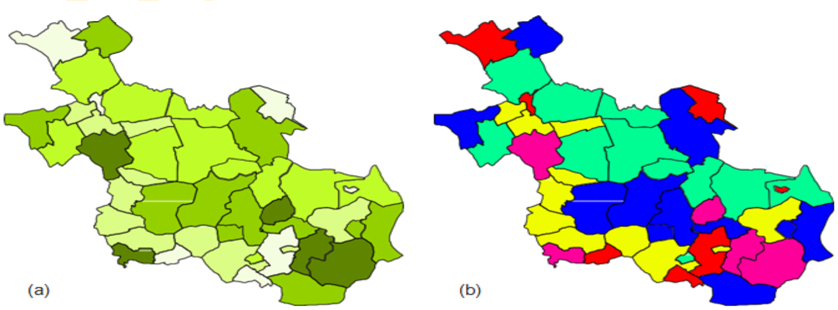


Figure 7.15: Poorly de- signed maps displaying absolute quantitative data: (a) wrong use of green tints for absolute popula- tion figures; (b) incorrect use of colour

How To Map The Terrain Elevation

- Terrain elevation can be mapped using different methods. Often, one will have collected an elevation data set for individual points like peaks, or other characteristic points in the terrain. Obviously, one can map the individual points and add the height information as text. However, a contour map, in which the lines connect points of equal elevation, is generally used.
- To visually improve the information content of such a map the space between the contour lines can be filled with color and value information following a convention, e.g. green for low elevation and brown for high elevation areas. This technique is known as hypsometric or layer tinting. Even more advanced is the addition of shaded relief. This will improve the impression of the three-dimensional relief.
- The shaded relief map uses the full three-dimensional information to create shading effects. This map, represented on a two-dimensional surface, can also be floated in three- dimensional space to give it a teal three-dimensional appearance of a 'virtual world', as shown in Figure (d). Looking at such a representation one can immediately imagine that it will not always be effective. Certain (low) objects in the map will easily disappear behind other (higher) objects.

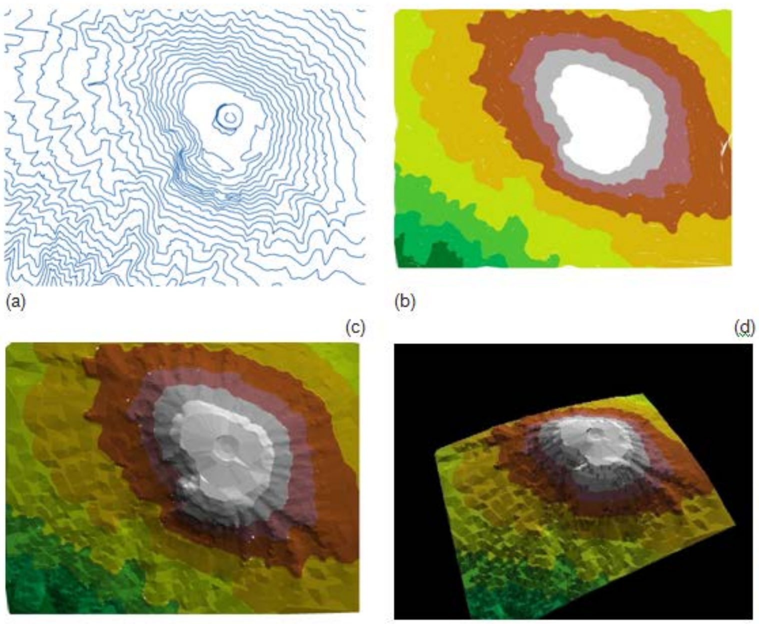


Figure 7.18: visualization of terrain elevation: (a) contour map; (b) map with layer tints; (c) shaded relief map; (d) 3D view of the terrain

- Socio-economic data can also be viewed in three dimensions. This may result in dramatic images, which will be long remembered by the map user. Figure7.19 shows the absolute population figures of Overijssel in three dimensions.

Figure 7.19: Quantitative data visualized in three di- mensions

How To Map Time Series

- Advances in spatial data handling have not only made the third dimension part of GIS routines. Nowadays, the handling of time-dependent data is also part of these routines. This has been caused by the increasing availability of data captured at different periods in time.
- Next to this data abundance, the GIS community wants to analyse changes caused by real world processes. To that end, single time slice data are no longer sufficient, and the visualization of these processes cannot be supported with only static paper maps.
- Mapping time means mapping change. This may be change in a feature's geometry, in its attributes or both. Examples of changing geometry are the evolving coastline of the Mumbai, the location of India's national boundaries, or the position of weather fronts.
- The changes of a land parcel's owner, landuse, or changes in road traffic intensity are examples of changing attributes. Urban growth is a combination of both. The urban boundaries expand and simultaneously the land use shifts from rural to urban. If maps are to represent events like these, they should be suggestive of such change.

It is possible to distinguish between three temporal cartographic techniques (see Figure7.20):

- ✓ ***Single static map**:* Specific graphic variables and symbols are used to indi- cate change or represent an event. Figure7.20(a) applies the visual variable value to represent the age of the built-up areas;
- ✓ ***Series of static maps**:* A single map in the series represents a 'snapshot' in time. Together, the maps depict a process of change. Change is perceived by the succession of individual maps depicting the situation in successive snapshots. It could be said that the temporal sequence is represented by a spatial sequence, which the user has to follow, to perceive the temporal variation. The number of images should be limited since it is difficult for the human eye to follow long series of maps (Figure7.20(b));
- ✓ ***Animated map**:* Change is perceived to happen in a single image by display- ing several snapshots after each other just like a video cut with successive frames. The difference with the series of maps is that the variation can be deduced from real 'change' in the image itself, not from a spatial sequence (Figure7.20(c)).

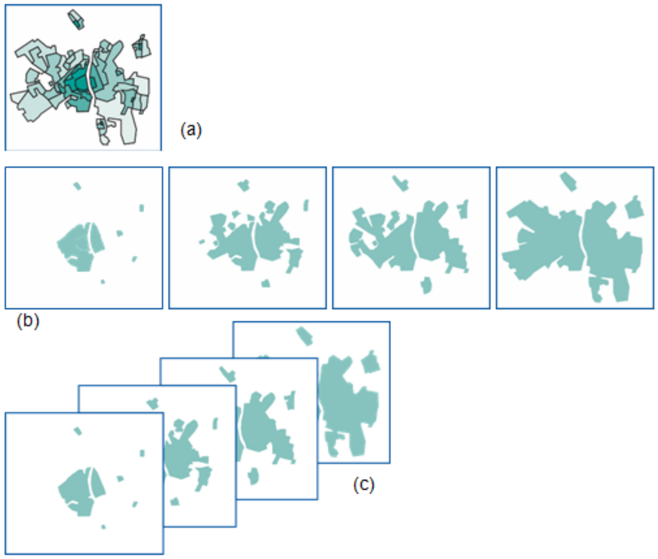


Figure 7.20: Mapping change; example of the urban growth of the city of Maastricht, The Netherlands: (a) single map, in which tints represent age of the built-up area; (b) series of maps; (c) (simulation of an) animation.

Chapter 4: MAP COSMETICS

- Most maps in this chapter are correct from a cartographic grammar perspective. However, many of them lack the additional information needed to be fully understood that is usually placed in the margin of printed maps. Each map should have, next to the map image, a title, informing the user about the topic visualized. A legend is necessary to understand how the topic is depicted.
- Additional marginal information to be found on a map is a scale indicator, a north arrow for orientation, the map datum and map projection used, and some lineage information, (such as data sources, dates of data collection, methods used, etc.). Further information can be added that indicates when the map was issued, and by whom (author / publisher). All this information allows the user to obtain an impression of the quality of the map, and is comparable with metadata describing the contents of a database or data layer.
- Figure below illustrates these map elements. On paper maps, these elements (if all relevant) have to appear next to the map face itself. Maps presented on screen often go without marginal information, partly because of space constraints. However, on-screen maps are often interactive, and clicking on a map element may reveal additional information from the database. Legends and titles are often available on demand as well.
- Text is used to transfer information in addition to the symbols used. This can be done by the application of the visual variables to the text as well. Italics—cf. the visual variable of orientation—have been used for building names to distinguish them from road names. Another common example is the use of colour to differentiate (at nominal level) between hydrographic names (in blue) and other names (in black). The text should also be placed in a proper position with respect to the object to which it refers.

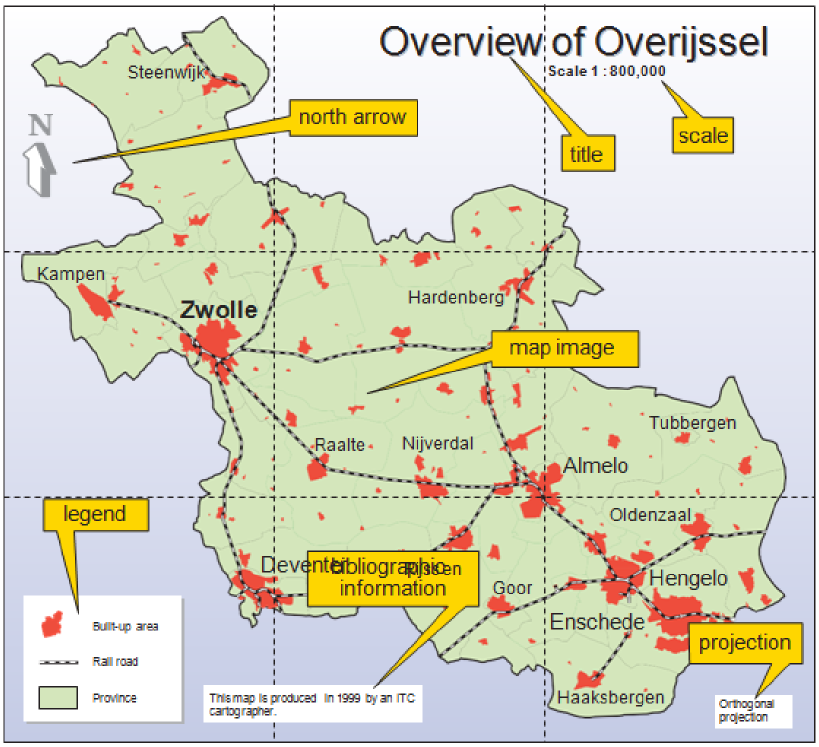


Figure 7.21: The paper map and its (marginal) information. Source: Figure 5–10 in [30].

MAP DISSEMINATION

- The map design will not only be influenced by the nature of the data to be mapped or the intended audience (the 'what' and 'whom' from "How do I say What to Whom, and is it Effective"), the output medium also plays a role. Traditionally, maps were produced on paper, and many still are. Currently, most maps are presented on screen, for a quick view, for an internal presentation or for presentation on the WWW.
- Compared to maps on paper, on-screen maps have to be smaller, and therefore their contents should be carefully selected. This might seem a disadvantage, but presenting maps on-screen offers very interesting alternatives. A mouse click could also open the link to a database, and reveal much more information than a paper map could ever offer. Links to other than tabular or map data could also be made available.
- Maps and multimedia (photography, sound, video or animation) can be integrated. Some of today's electronic atlases, such as the Encarta World Atlas are good examples of how multimedia elements can be integrated with the map. Pointing to a country on a world map starts the national anthem of the country or shows its flag. It can be used to explore a country's language; moving the mouse would start a short sentence in the region's dialects.
- The World Wide Web is a popular medium used to present and disseminate spatial data. Here, maps can play their traditional role, for instance to show the location of objects, or provide insight into spatial patterns, but because of the nature of the internet, the map can also function as an interface to additional information. Geographic locations on the map can be linked to photographs, text, sound or other maps, perhaps even functions such as on-line booking services. Maps can also be used as 'previews' of spatial data products to be acquired through a spatial data clearinghouse that is part of a Spatial Data Infrastructure. For that purpose we can make use of geo-webservices which can provide interactive map views as intermediate between data and web browser.

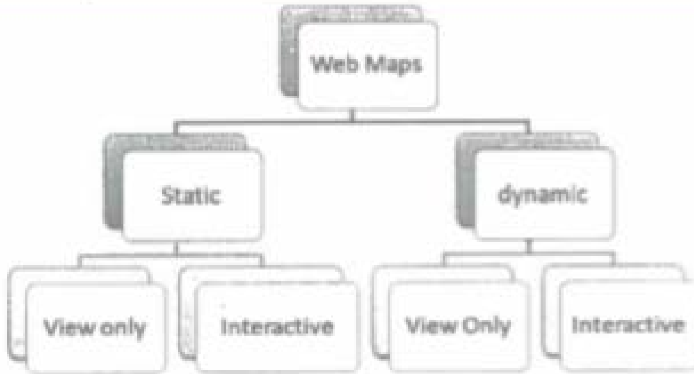


Figure : Types of Web Maps