2

Key concepts 1
GIS

2.1 Introduction

This chapter is the first of three intended to introduce a set of key concepts that are central to the approaches described in the rest of the book. This chapter focuses on some key principles of GIS and is followed by chapters on some fundamentals of statistics and spatial data analysis. This chapter will briefly discuss ways in which real-world entities can be represented in a GIS. The focus will then move on to methods for spatial data collection, errors, visualization of spatial data, and simple approaches for extracting information from data sets. More specifically, the chapter covers:

- data and data models, including topology—representing reality in a GIS using data models and databases
- spatial referencing systems and projections—how spatial data are spatially referenced and how parts of the Earth's surface can be represented on a map
- spatial scale—how far spatial properties vary over a given area
- data collection—how spatial data are generated
- sources of errors in spatial data
- visualization of spatial data—a key first step in any analysis
- querying spatial data—extracting information from spatial databases.

The following section deals with data models used commonly in GIS.

2.2 Data and data models

A model is simply a means of representing 'reality' and spatial data models provide abstractions of spatially referenced features in the real world. The focus of this book is on analysis of spatial data rather than the ways in which spatial data are structured. However, a very brief introduction to data models was considered useful as knowledge of analysis of spatial data requires at least a basic understanding of data structures. Most books on GIS stress the division of data models into the well-known raster and vector representations. The key properties of the two data models, which will be useful in making sense of the rest of this book, are outlined here. The data model available determines the choice of method for spatial analysis as do the characteristics of the particular data set.

Representations of real-world features are often divided into (1) entities and (2) fields (Burrough and McDonnell, 1998). Entities are conceptually distinct objects like point locations, roads, or administrative boundaries. Fields convey the idea of values of some property at all locations. For example, elevation can be measured or estimated at all places and elevation does not usually have distinct edges, in contrast with, for example, buildings. Objects that are well described as distinct entities are sensibly represented using the vector data model. Properties that tend to vary quite smoothly from place to place (i.e. they are spatially continuous and their values do not tend to change abruptly from place to place) are frequently represented using the raster data model. There are notable exceptions and these include isolines and contours, which are vector-based representations of continuous phenomena such as temperature or elevation (there are, of course, exceptions—a cliff edge represents an abrupt change in elevation and so temperature is perhaps a more conceptually straightforward example). The raster and vector data models are briefly defined in turn. A more in-depth account of the way in which information is stored using the two data models is given by Wise (2002).

2.2.1 Raster data

While it is assumed that readers are familiar with raster grids, some key issues are addressed here. Raster grids are conceptually simple structures, comprising square cells with numeric values or classes attached to each cell. A simple example of a raster grid is given in Figure 2.1; in this case the value represents elevations in metres. Where the cells contain categorical or integer (i.e. whole number) values the number of instances of each class may be stored in a table. In cases where values with decimal places are used, all information is conventionally stored in the raster itself. There are huge amounts of data available in raster grid format—remotely sensed imagery (see Section 2.8.3) comprises a major component of such data sources. The spatial resolution of a raster refers to the area in the real world covered by a cell. For example, a grid with a spatial resolution of 5 m covers an area of 5 by 5 (=25) square metres. Remotely sensed images with very fine spatial resolutions (e.g. 1 m) have been generated for many parts of the world, although ease of access (cost, etc.) varies geographically.

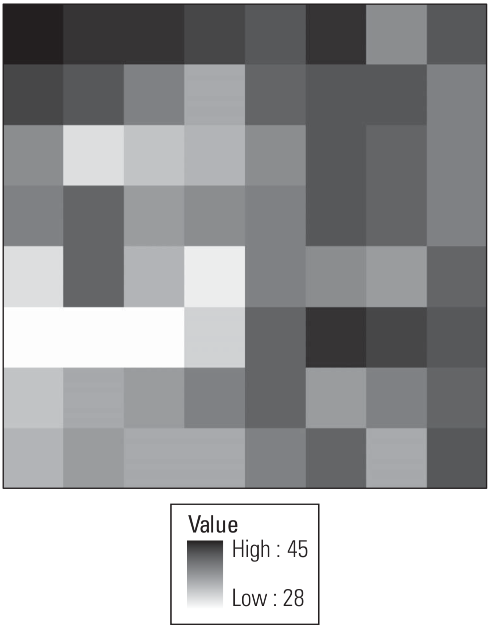


Figure 2.1 Raster image used to illustrate various methods. Units are elevations in metres.

2.2.2 Vector data

As with the raster model, it is assumed that readers are familiar with the vector data model, but some background is given about vector data storage formats and their relevance for spatial data analysis. Whereas features in raster grids are identified simply by the row and column position of cells, vector data comprise explicit spatial coordinates of the features that make up objects. Vector data comprise points (with x- and y-coordinates), lines (line segments (or arcs) connected by points), and area polygons (lines with the same start and end point). An example of some line features is given in Figure 2.2. This example shows that line features comprise two forms of point locations—vertices, which represent change in direction of arcs, and nodes, which represent the start or end of arcs, including locations where different arcs connect. Of course, vectors representing real-world features are usually much more complex than those shown in Figure 2.2 (and may have many vertices). Note that, while x and y are used to represent two-dimensional position, z is often used to indicate the third dimension (elevation) and also, as in this book, values of any property (e.g. precipitation amount) that are associated with particular x- and y-coordinates.

Vector data can be stored as what are sometimes called 'spaghetti' data—that is, strings of unconnected line segments. In this case, relationships between objects (e.g. which line is connected to which other lines) are not encoded. However, explicit information on the relationships between objects reduces the computational demands of subsequent analyses and analysis of vector data is usually preceded by the construction of topology, as discussed in the next section. Conventionally, vector data are divided

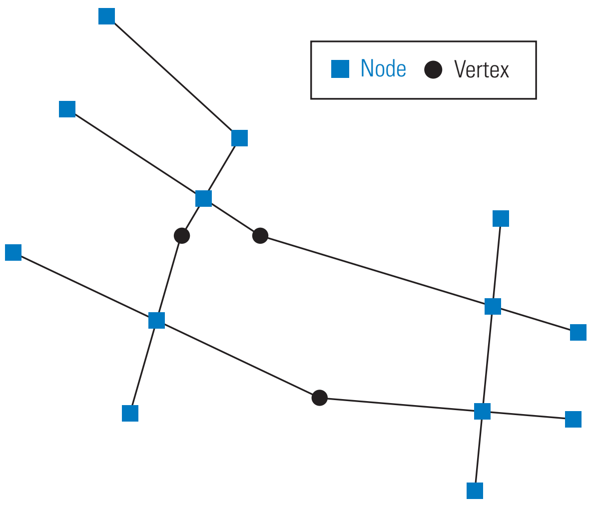


Figure 2.2 Vector line features used to illustrate network analysis methods in Chapter 6.

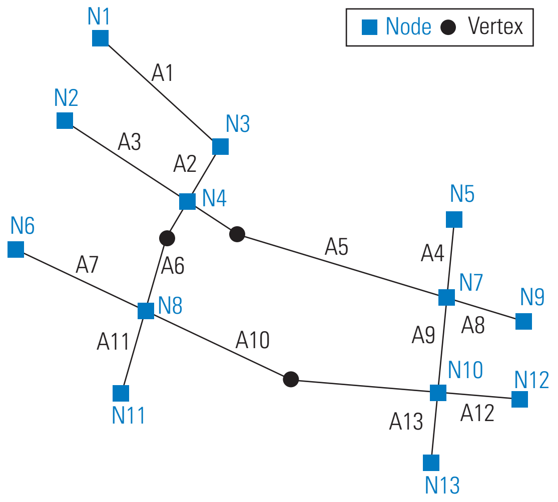
into their spatial component and their attribute component. Attributes linked to each spatial feature are often stored using a relational database system (see Section 2.3 for more on databases) and this is demonstrated in the following subsection.

2.2.3 Topology

This section is concerned with vector representation of objects and connections between features. Topology can be defined as 'the mathematical study of objects which are preserved through deformations, twistings and stretchings. (Tearing, however, is not allowed.)' (Weisstein, 2003, p. 2990). In other words, if a map showing a set of zones is stretched, zones separated by other zones cannot become neighbours. For this to be possible, the map would have to be cut (or torn as the quote above indicates) or folded over on itself (Wise, 2002). In a GIS (as well as a computer-aided design system, etc.), information on the connections between objects and, where appropriate, neighbouring objects, may be stored. Operations concerned with connections between objects (e.g. administrative areas or roads) are dependent on information about topological relationships.

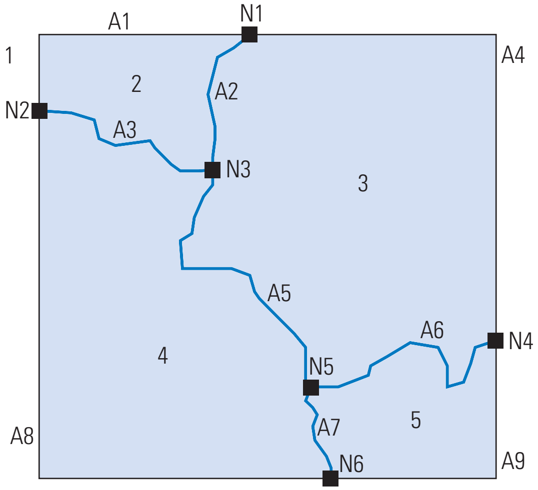
Obviously, to represent a point only the point coordinates are required. In Figure 2.3, an example of line topology representation is given. In this case, arcs are indicated with the prefix 'A' and the nodes with the prefix 'N'. Changes in the direction of arcs between nodes are represented by vertices, as shown previously in Figure 2.2. Note from the table that direction is represented—there is a 'from node' and a 'to node'. N3 may, at first sight, appear to be a vertex but it connects two distinct arcs—A1 and A2.

Figure 2.4 gives an example of vector polygon topology representation. With this representation, information on arcs and the polygons to which they belong is contained in one table while information on the area and perimeter of the polygons is contained in another table. The table at the top has separate codes for each arc, the



Arc	From node	To node	Length
1	1	3	1.358
2	3	4	0.529
3	2	4	1.230
4	5	7	0.665
5	4	7	2.335
6	4	8	0.991
7	6	8	1.204
8	7	9	0.674
9	7	10	0.797
10	8	10	2.588
11	8	11	0.728
12	10	12	0.687
13	10	13	0.598

Figure 2.3 Line topology representation. The prefix 'N' refers to node and the prefix 'A' to arc.



Arc	From node	To node	Left polygon	Right polygon
1	2	1	1	2
2	3	1	2	3
3	3	2	4	2
4	1	4	1	3
5	5	3	4	3
6	4	5	5	3
7	6	5	4	5
8	6	2	1	4
9	4	6	1	5

Record	Area	Perimeter	Polygon
1	−4208522	8206.625	1
2	391022.2	2931.427	2
3	1837280	5843.305	3
4	1552205	5846.086	4
5	428014.2	3045.462	5

Figure 2.4 Polygon topology representation. The polygon numbers are those without a prefix. The prefix 'N' refers to node and the prefix 'A' to arc.

start and end node (indicated with the prefix 'N' on the figure) of each arc (given by the prefix 'A' on the figure) and the polygon that lies to the left and right of each arc (given by numbers without a prefix). Note that the negative area for polygon 1 corresponds to the area outside the edge of the region and it is the negative of the sum of areas within the boundary (i.e. if you take the absolute value of this first area figure, this gives the total area within the outer edge). Additional tables with the same arc identifiers can be added easily with a relational database structure. With a relational database, each object has a unique identifier (a 'primary key'). If information in other tables is available, this information can be linked together if both sets of data use the same identifiers.

With this representation, information is recorded on connectivity (arcs are connected by nodes), containment (enclosed polygons can be identified, although there are none in the example), and contiguity (arcs shared between polygons, thus determining contiguity, are indicated, e.g. polygons 3 and 4 share arc 5). The name of the well-known software environment ArcInfo™ (now part of ArcGIS™) reflects the relationship between arcs and attribute information.

2.2.4 Rasters and vectors in GIS software

Most GIS environments allow for conversion from rasters to vectors or from vectors to rasters. Clearly, the spatial resolution of the raster will limit the positional accuracy of vector objects generated through raster to vector conversion. For many standard operations (e.g. measuring proximity to objects; see Section 4.5) either vectors or rasters can be used and so the choice of data model may, for many applications, be less important than it was in formally less flexible GIS software environments.

2.3 Databases

GIS are often defined as having at their core a spatial database. In essence, a database is simply a set of structured information. In order to store, manage, and access computer-based data a variety of database structures have been developed. The most frequently encountered is based on the relational database structure introduced by Codd (1970). The topological structures discussed in the previous section are based on tables linked by common identifiers, and this is the basis of a relational database. For example, in the case of the line topology representation in Figure 2.3, the arcs themselves and both tables have a unique arc number—there is only one arc with the label '1'. The information in the two tables can easily be joined together using this common identifier or key. In the case of the polygon topology representation in Figure 2.4, there are unique arc and polygon numbers. The polygons listed in the table at the top correspond to the polygons listed in the table at the bottom and the two tables can be linked using the polygon identifiers. One objective in setting up a relational database is to store only necessary information and to enable efficient and effective access to the data. The setting up of such a database may be non-trivial if there are multiple tables linked in different ways. While the relational database structure is encountered most frequently, other database structures are widely used. Object-oriented databases are quite commonplace; such systems organize data in a structured hierarchy. One example is the nesting of an individual within a house and a house within a census area and a census area within a larger administrative unit. Efficient organization of data is important for spatial data analysis (particularly in the case of large data sets) as speed of processing may, in many cases, be markedly increased if the database is well structured. Burrough and McDonnell (1998) and Worboys and Duckman (2004) provide more detailed descriptions of database systems. Conventionally, attributes of vector spatial data are often stored using a relational database system. Raster data are often stored as self-contained objects. However, as noted before, in cases where a raster grid is composed of integer values, an associated table may be used that records the number of instances of each value.

2.3.1 Database management

Data stored in a database are accessed using a database management system (DBMS). A DBMS offers facilities for updating the database and extracting data in a flexible way. The DBMS includes facilities to import data into the database, to manage user access, to update the database structure and content, and to conduct queries to extract specific information. The querying of spatial databases is the subject of Section 2.11. A good summary of DBMS is provided by Longley *et al.* (2005a).

2.3.2 The Geodatabase

The term 'Geodatabase' is sometimes used to mean a spatial database in general. More specifically, the Geodatabase is the core means of storing and managing spatial data in the ArcGIS™ environment. The Geodatabase offers various benefits over conventional GIS file formats such as shapefiles or coverages (both widely used and well-established vector file formats). In particular, the Geodatabase combines data into a single integrated database rather than storing each layer in discrete files. The Geodatabase has built-in rules which help to maintain the integrity of the database and reduce database maintenance.[1]

2.4 Referencing systems and projections

The location of spatial objects is usually recorded using some kind of spatial referencing system such as longitudes and latitudes, or eastings and northings using some kind of national grid system. Given that the surface of the Earth (as well as other bodies) is not flat, one of a variety of projection systems can be used to transform locations on a sphere to features on a flat surface. It is essential that data are projected appropriately, and a brief summary of some key issues is provided for context.

A projection can be defined as comprising the ellipsoid, the datum, and the projection. The ellipsoid is the smooth approximate shape of the Earth and an ellipsoid can be selected which best represents the surface of the Earth for the whole planet or for a given area. The surface which represents deviations from the ellipsoid is termed the 'geoid' (Clarke, 1999). The datum is the origin or centre and rotation of the ellipsoid.

1 http://www.esri.com/news/arcuser/0701/migrating.html

The datum is defined such that it has the best fit to the surface of the Earth over the area of interest.

Any projection entails making compromises, for example some projections distort areas while others distort distances. Conformal projections preserve shape while equal area projections preserve areas (as their name suggests) and not shapes—clearly a projection cannot preserve both properties. A common example of a conformal projection is the Mercator projection and an example of an equal area projection is the Albers equal area projection. Conformal projections like the Mercator projection preserve angles locally and thus the Mercator projection is well-suited to the purpose of navigation, for which it was developed. Projections also exist which preserve distances along one or several lines (Clarke, 1999). An awareness of projection systems is important in working with spatial data. If, for example, the concern is with areas of countries then an equal area projection must be used. The impact of choice of projection is a function of the spatial scale of the map. In short, the larger the area of concern, the greater the impact of selection of an appropriate projection will be (Clarke, 1999) and the impact of poor choice of projection when using a map with a representative fraction (i.e. ratio of distances on a map to distances in the same units in the real world) of 1:1,000,000 will be greater than when using a 1:10,000 map. A brief introduction to projections is provided by Longley *et al.* (2005a) while a more detailed account is given by Seeger (2005). Scale is discussed in Section 2.7.

2.5 Georeferencing

To make spatial data useable, they must be linked to some kind of spatial referencing system, as detailed above. The process of attaching spatial information to data is called georeferencing. An example of georeferencing is the use of coordinates obtained using a global positioning system (GPS) receiver to link positions on a remotely sensed image to positions that have been surveyed, and the linked survey points are called ground control points (GCPs). The process of transforming an image, whereby the transformed image fits well to the GCPs and the image is then in coordinate space, is termed 'georectification'. Georectified images can be overlaid or combined with other data that are georeferenced using the same system. The term 'geocoding' is generally used in relation to the determination of geographic coordinates from an address or related data (McDonnell and Kemp, 1995) and this is outlined next.

2.6 Geocoding

An example of geocoding is the conversion of addresses into geographic coordinates. A variety of databases and software environments exist to facilitate links between names, addresses, postcodes or zip codes, and geographic coordinates. Such data sources and tools may be part of an essential prior step to spatial analysis. In some contexts, information may be available on road networks and segments of properties that are located on the roads, but not on the specific location of properties. In such cases, given this knowledge, the locations of individual properties can be predicted through address interpolation (or address matching). The concept is illustrated in Figure 2.5, where the coordinates of the road junctions are known and the positions of properties in the segment are interpolated given that odd-numbered houses are located on the eastern side of the road while even-numbered houses are located to the west. The interpolated position of number 16 is given. Of course, such an approach may give misleading results where the size of properties, or the spacing between them, varies. An introduction to geocoding and related issues is provided by Longley *et al.* (2005a).

2.7 Spatial scale

Spatial data analysis is dependent on the sample size, density, and, where relevant, the level and type of aggregation (i.e. ways of spatially grouping values; an example is counting the population within different administrative zones). The level and type of aggregation are the subject of Section 4.9. This section discusses the issue of spatial scale. In this context spatial scale is defined as the scale at which the property of interest varies (but there are many definitions and Lloyd (2006) discusses some of these). For example, in a mountainous landscape, elevation values may differ a great deal between one location and another over very short distances. On a river flood plain elevation values may, in contrast, differ very little over even very large distances. In the former case, the spatial variation may be described as being of a fine scale or a high frequency. In the latter case, the spatial variation may be considered coarse scale or

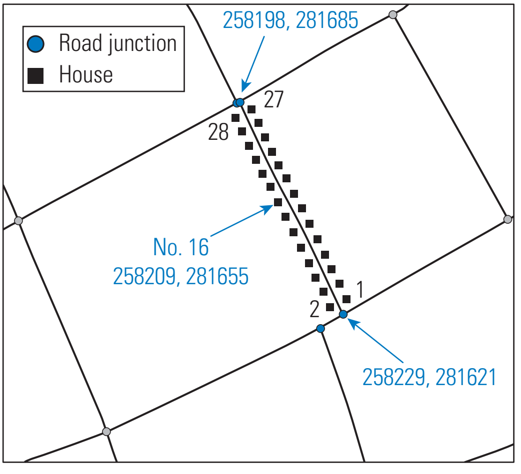


Figure 2.5 Geocoding: address interpolation.

low frequency. To capture spatial variation in the case of a mountainous terrain, a finer sampling grid would be needed than would be the case for the river flood plain. If the sample spacing is too large, then important information may be lost. If it is too small, then some of then effort expended in sampling will have been wasted—some of the data will be redundant and add little information. It is therefore important to consider how the available sample meets the requirements of the analysis and Section 2.8.1 discusses this issue further.

2.8 Spatial data collection

This section introduces the basic principles of some widely used means of spatial data collection. Spatial data may derive from secondary sources (e.g. paper maps) or data may be collected by, or for, a particular user with a particular purpose in mind (termed 'primary data'). The first subsection deals with the topic of spatial sampling. Paper-based secondary data sources are the focus of Section 2.6.2. Spatial data can be divided into those collected by remote sensing and those collected by ground survey, and Sections 2.6.3 and 2.6.4 introduce these two modes of data collection.

2.8.1 Spatial sampling

Any spatial data set is a sample. A remotely sensed image may cover the entire area of interest, but there are limits to the spatial resolution of such imagery. In any application making use of spatial data, it is necessary to find a balance between the amount of information required and the number and spatial positioning of observations. In terms of measurements on the ground, perhaps using GPS or some more traditional survey technology, making observations on a fine grid over the entire study area, will allow detailed characterization of the particular property of interest. However, such a strategy will be wasteful of effort and money if many neighbouring observations have similar characteristics, and thus contain very similar information (see the previous section for a related discussion). The objective of sampling design is, therefore, to design a sampling scheme whereby the maximum possible information is acquired for the minimum effort.

Various commonly used strategies for sampling exist. These may be based on making observations at locations with particular characteristics. Following the example of mapping elevations, making measurements at locations where there is a break of slope would be sensible. Other strategies are based on random selection of sampling locations, with a key objective being to minimize bias. Note that many routines exist for randomly selecting locations or values from a list. Using such approaches, a particular number of locations might be randomly selected from across the whole study area and measurements could then be made at these locations. Other strategies are based on, for example, random selection of locations within areas (a stratified sampling scheme), thus ensuring that particular areas are represented, but that the locations within them are determined randomly. One way of matching spatial variation to sample density is offered by the body of approaches known as geostatistics. With such an approach, the variogram (see Section 9.7) is used to characterize the spatial variation given some provisional survey and this information can be used to ascertain an optimal sample spacing.

Whether a user obtains the sample themselves or is reliant on data provided to them, it is necessary to be aware of the nature of the sample and to consider any potential issues that might arise in the use of the data as a function of the sample design. An introduction to spatial sampling is provided by Delmelle (2009).

2.8.2 Secondary data sources

In studies utilizing historic maps, or paper-based sources of which digital versions are not readily available, conversion from these paper-based sources to digital versions is necessary. Scanning provides a rapid means of deriving a raster image from paper maps, although it is rarely used now for spatial data generation. Where acquisition of information on particular features is desired then digitization of mapped features is often conducted. The most common approach is probably to scan the map and trace the edges of the features of interest to generate a new vector data layer; this is called 'heads-up' digitizing. Increasingly, spatial analyses are based exclusively on data that were collected directly in digital format and the often tedious process of tracing features from maps through digitizing is unnecessary for many users of GIS. An introduction to historical GIS, which includes discussion of conversion of historic paper maps into digital format, is the book by Gregory and Ell (2007).

2.8.3 Remote sensing

Huge amounts of spatial data are generated through various technologies that fall within the umbrella term 'remote sensing'. This may refer to airborne or spaceborne sensors (or simply cameras) or technologies for measuring subsurface characteristics such as electrical resistivity. Conventional aerial photography has been an important source of data historically. Figure 2.6 shows an example of an orthophoto in an area of Maryland in the USA. The data were obtained from http://terraserver-usa.com/ and more information about digital orthophotos is available at: http://online.wr.usgs.gov/ngpo/doq/. Orthophotos are remotely sensed images that have been georectified (see Section 2.5).

There are now many different kinds of airborne and spaceborne sensors that detect radiation from different parts of the electromagnetic spectrum. Sensors may be divided into two groups:

- passive sensors, which sense naturally available energy
- active sensors, which supply their own source of energy to illuminate selected features. Examples are radar and light detection and ranging (LiDAR). Radar emits pulses of microwave energy and LiDAR emits pulses of laser light.

Figure 2.6 Orthophoto of an area approximately 4 km west of Airedele, Maryland, USA. Area: 2.4 km by 1.6 km. Spatial resolution: 4 m. Image courtesy of the United States Geological Survey.

Sensors vary in terms of the spatial resolution (defined in Section 2.2.1) and the spectral resolution of the output imagery. The term 'spectral resolution' refers to which parts of the electromagnetic spectrum are measured. The number of available bands (i.e. parts of the spectrum) and the specific parts of the spectrum represented determine the uses to which the imagery can be put and these are likely to be selected with particular purposes in mind since different parts of the spectrum will highlight different characteristics of the Earth's surface. The majority of remote sensing systems collect information in one or more visible, infrared, or microwave parts of the spectrum (Lillesand *et al.*, 2007). The infrared parts of the spectrum are large compared to the visible part of the spectrum. The far infrared part of the spectrum, for example, is sensed in the acquisition of thermal imagery. Lillesand *et al.* (2007) provide a detailed account of remote sensing principles and practice.

Remote sensing is often used to derive topographic models, i.e. digital models of the surface of the Earth (or, indeed, elsewhere). The term 'digital elevation model' (DEM) refers to such a model and the most common form is a raster grid with cell values representing elevations above some arbitrary datum such as mean sea level. Some key technologies for constructing DEMs (including airborne LiDAR and radar-based systems) are summarized by Lloyd (2004). In the case of airborne LiDAR, outputs are 'point clouds' from which surfaces can be generated.

2.8.4 Ground survey

There are several widely used means of obtaining information on the spatial position of features through ground survey. Traditional survey techniques include tape and offset surveying and levelling. Ground survey techniques make use of the fact that we can measure the position of a point in three dimensions through the measurement of angles and distances from other positions. Angles may be measured using tools such as theodolites, while the advent of electronic distance measurers allows rapid and highly accurate measurement of distances. The measurement of angles and distances may be combined using a high-precision system called a total station. Total stations are capable of obtaining positional measurements accurate to 1 mm or so. Global positioning system (GPS) technology has a major role to play in ground survey. GPS receivers vary markedly in their characteristics and costs. They range from small handheld systems capable of obtaining positional measurements accurate to within a few metres to differential systems costing tens of thousands of pounds (or dollars) but capable of obtaining positional measurements accurate to within 1 cm. With differential GPS, a stationary base station receiver is used to refine measurements made by one or more roving receivers. Another important technology for generation of spatial data is terrestrial LiDAR. This approach can be used to generate very accurate models of topographic surfaces and other objects on the surface of the Earth (see Pietro *et al.* (2008) for an example). Introductions to ground survey are provided by Lloyd (2004) and Longley *et al.* (2005a) and a very detailed account is the book by Bannister *et al.* (1998).

2.9 Sources of data error

All data are only representations of reality and are subject to a variety of factors which may affect their quality. This book outlines a wide range of methods for extracting information from data, but for these analyses to be worthwhile the data must be of sufficient quality. Sources of error in data include errors made during data collection, data input errors, and inappropriate data model choices. Identifying obvious errors caused by factors like equipment failure may often be straightforward, but many major sources of error may go unnoticed. Modelling and analysis of spatial data may introduce further errors. Manual conversion from paper-based to digital formats is a major potential source of error. In addition, any data collection technology has a limited precision even if that technology is employed properly and there are no external factors (such as atmospheric conditions) that have an impact on the accuracy of measurements. In terms of data model choice, if the spatial resolution of a raster image is coarse in relation to the area of objects of interest then those objects will not be well represented by the raster. A particular example is the representation of a linear feature like a river, which may be blocky in appearance if represented by a raster.

The accuracy of a data product can be conceived of as having two parts—bias and precision. A biased set of measurements may consistently over-estimate or under-estimate the ‘true’ value while precision refers to the repeatability of values. In other words, if there is apparently random variation in measurements and repeated measurements differ in some inconsistent way then the measurements are of low precision. In contrast, if repeated measurements are similar then the measurements are of high precision. The possibility of errors being introduced at any stage of data processing or analysis should be taken into account and the generation of high-quality graphic outputs should not disguise the fact that any output is only equal in quality to the lowest quality input. This topic is discussed further below. All spatial data should be associated with metadata—that is, data about the data sources which indicate key information on how and when the data were collected, as well as detailing any conversions or modifications undertaken. If detailed metadata are kept, these act as an invaluable resource for future users of the data in that they provide a means of assessing factors that may have an effect on applications which make use of these data.

2.9.1 Uncertainty in spatial data analysis

The use of alternative procedures, or selecting different options in the application of one method, will often lead to different results. It is essential in any use of spatial data to take into account such potential problems. The modelling of propagation of errors from one processing stage to another, and of the degree of uncertainty in representations of features and their attributes, are significant areas of research. The quality of outputs from a spatial analysis is a function of (1) the quality of the data, (2) the quality of the model, and (3) interactions between the data and the model (Burrough and McDonnell, 1998). When data from different sources are combined, the effects of many different kinds of uncertainties (e.g. measurement errors, scale differences, temporal differences, and other factors) may also combine. Spatial data quality and uncertainties in spatial data are among the subjects of the book chapter by Brown and Heuvelink (2008).

2.10 Visualizing spatial data

Visualization is the first stage of any spatial analysis. Simple viewing of a spatial data set may seem conceptually straightforward. However, there may be a multitude of decisions that have to be taken into account when visualizing data which may impact strongly on interpretations of those data and on the ways in which any analysis might proceed. Simple point patterns (i.e. point event locations with no attributes attached) are often presented using points to represent each event location. In the case of objects with categorical attributes (e.g. urban area or rural area), depiction may be based on the selection of different colours or shades to represent each category. In such cases, selection of colours or shades that enable differentiation between categories is important; a map with two classes depicted using similar shades or colours may be very difficult to use. Continuous variables (e.g. measurements of an airborne pollutant) are usually represented using a range of colours or shades (e.g. white for small values, shades of grey for intermediate values, and black for large values). Where the data model is a grid, a continuous grey scale or colour scale may be used, as shown in Figure 2.1.

A common means of displaying areal data (e.g. population densities in administrative zones) or values attached to other discrete objects such as points, in particular,

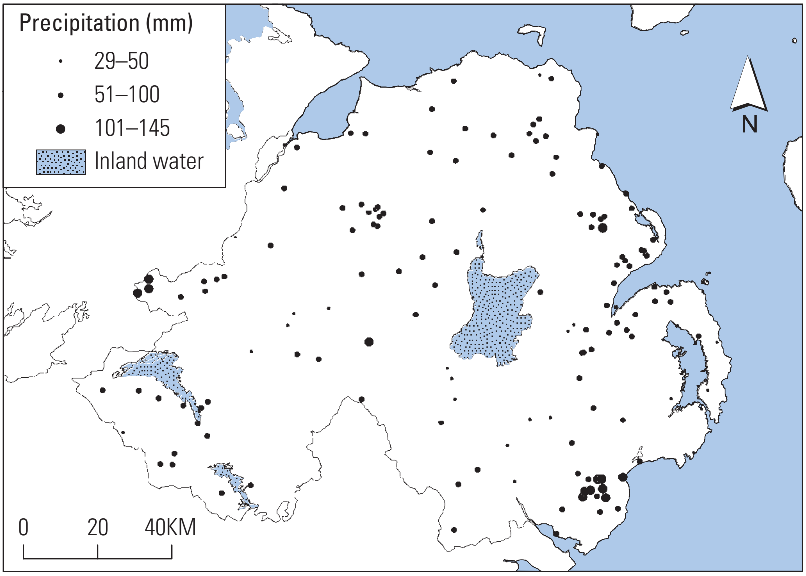


Figure 2.7 Precipitation amounts in Northern Ireland in July 2006.

is a choropleth map. With such maps, ranges of values are assigned a particular shading or colour. In other words, the possible values are divided into classes (typically five or less). Choropleth maps often show areas of uniformity separated from one another by abrupt edges (Tate *et al.*, 2008). The map is then a function of the classes used to display values and the form of the zones for which the values are provided—that is, in the same way that the zones are merely one possible way of spatially dividing a continuously varying phenomenon so are the classes used to represent values one way of subdividing the full range of values. It has been argued that one way of reducing such problems is to convert areal data into surfaces (Tate *et al.*, 2008; see Section 9.9). Figure 2.7 shows a map of point values (precipitation amounts) represented using symbols of different sizes; this is a common means of displaying point values. Even simple approaches open up complex issues—if a range of values is divided into five classes, the different class thresholds used may result in visually very different maps than those based on, say, six or seven classes. Figure 2.8 shows two different groupings of the same set of area values into three sets of different classes. While both maps are based on the same data, the patterns in the maps appear, in many respects, quite different.

There are many more sophisticated means of visualizing spatial data, including three dimensional visualizations (e.g. see Figure 9.5), familiar to users of Google Earth™ (http://earth.google.co.uk/), and cartograms. Cartograms distort the form of features to highlight particular characteristics, for example zones with large populations may be made proportionately larger than zones with small populations such that the modified zones better reflect the attributes that they contain. Throughout this book, maps and other visual outputs are presented as central components of the analyses of which they are part. Introductions to various aspects of spatial data visualization are provided

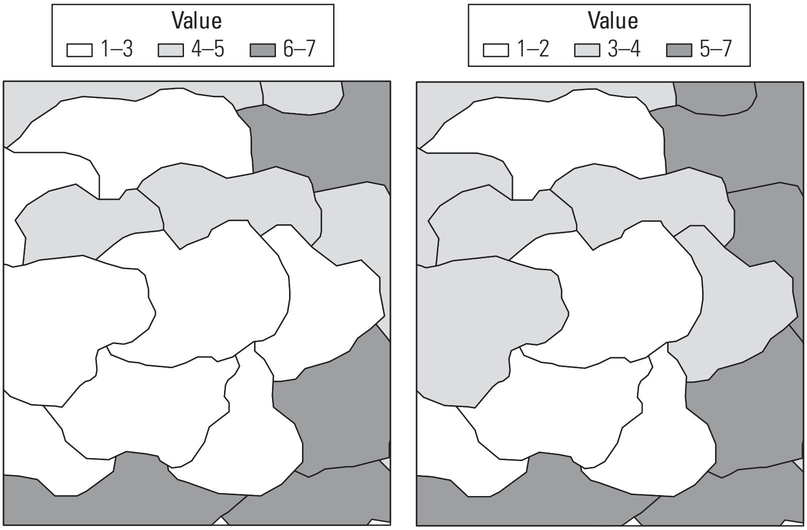


Figure 2.8 Alternative categorizations for vector polygon data.

in the book edited by Wilson and Fotheringham (2008). A good summary account of methods for visualizing spatial data (geovisualization) is given by Longley *et al.* (2005a). The data values may be visualized in other indirect ways, an example being the use of plots or graphs to summarize data values. This topic is amongst those explored in the following chapter.

2.11 Querying data

At its simplest level, the analysis of spatial data could involve selecting and mapping areas or features that have particular properties. For the example of measurement of some pollutant at point locations, all points with a pollution level above some critical threshold could be highlighted. In a standard database system a querying language (such as structured query language) is likely to be used to select entries in the tables that make up the database. Likewise, in a GIS such a query language can be used to select a subset of the data set. Such queries take logical forms such as Nitrogen-DioxidePPB > 21, selecting all locations where the nitrogen dioxide amount measured is greater than 21 parts per billion. The concept of Boolean logic and its application for querying spatial data is outlined next.

2.11.1 Boolean logic

Selection or combination of spatial features is often conducted using logical (or Boolean) operators (like the nitrogen dioxide example above). For example, population zones in

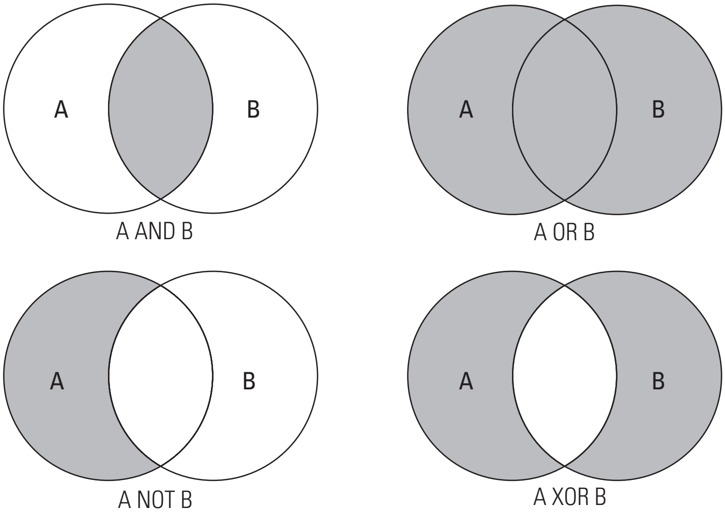
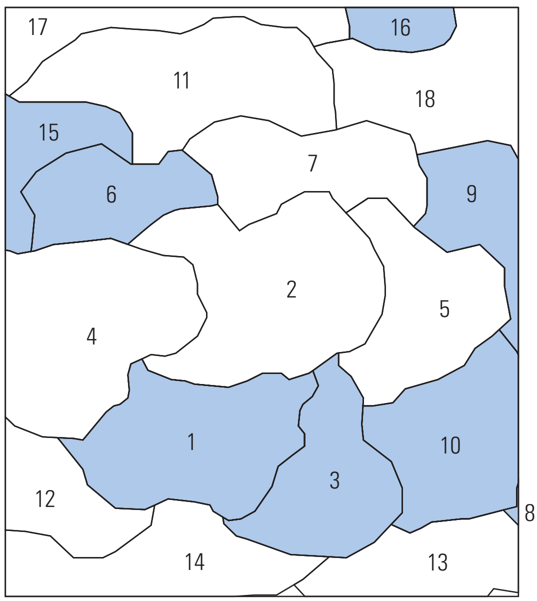


Figure 2.9 Boolean logic.

ID	Population
1	52
2	111
3	43
4	123
5	215
6	26
7	139
8	75
9	45
10	73
11	432
12	321
13	121
14	175
15	64
16	32
17	234
18	107

Selection for query:
Population < 100



Labels are ID numbers

Figure 2.10 Selected polygons and corresponding table entries.

a layer may be selected that have more than a specific number of people above a particular age, but less than a certain percentage (%) who have long-term illnesses. Such an argument would be structured like this: No.People75plus > 2500 AND LongTermIll% < 10. Figure 2.9 shows Venn diagrams (used to display possible combinations between groups) illustrating the outcome of the application of Boolean logic to two sets. In these cases, one, either, both, or mutually exclusive sets are selected. The Boolean AND selects objects fulfilling two criteria, OR selects objects fulfilling either of the criteria, NOT selects objects fulfilling one criterion but not the other and XOR (exclusive OR) selects objects fulfilling one criterion or the other, but not both. Figure 2.10 shows a simple example of a query in practice.

Queries constructed using Boolean logic can be used to select features in any desired combination. In the case of two or more criteria, query statements are easily extended. The application of Boolean logic for the overlay of multiple data layers (the identification of common areas of two or more sets of polygons) is discussed in Chapter 5. With Boolean logic, membership of a class is definite. An alternative approach is fuzzy logic, which recognizes uncertainty in assigning features to classes (see Longley *et al.* (2005a) for a summary). For example, boundaries between two soil types are not likely to be clearly defined and instead some form of classification that accounts for the probability of there being one soil type or another at a particular location is likely to be more appropriate than a 'hard' classification of the type described above (see Section 3.4 for a discussion about probabilities).

Summary

This chapter covers a wide variety of concepts that are important in the analysis of spatial data. The focus was on key GIS concepts, including data models, databases, projections, georeferencing and geocoding, spatial scale, spatial data collection, errors, visualization, and querying spatial data. Such issues are central to understanding the material covered in the rest of the book. Knowledge of data models is important as the data provide the basis of any analysis. Some understanding of database principles and data extraction (querying) is also central to a large proportion of analyses. Understanding of how data are collected, and the limitations of particular approaches, is essential background to the application of spatial data. No data are perfect representations of reality and so an awareness of potential sources of error is crucial. Finally, visualization and querying of spatial databases are common first steps in any spatial analysis.

Further reading

The further reading section of the previous chapter cited some useful introductions to GIS. Some of the books listed in that section provide in-depth material on some of the topics outlined in this chapter. In particular, issues such as databases, query of spatial data, and errors are dealt with by **Burrough and McDonnell (1998)**, **Longley *et al.* (2005a)**, and **Heywood *et al.* (2006)**. Descriptions of spatial data formats and storage are given by **Wise (2002)**. Spatial data collection is a vast topic; useful introductions to survey and remote sensing are provided by **Bannister *et al.* (1998)** and **Lillesand *et al.* (2007)**, respectively.